Essays on Univariate and Multivariate Modeling of Financial Market Risks

INAUGURALDISSERTATION

zur Erlangung des akademischen Grades eines
Doctor rerum politicarum
(Dr. rer. pol.)

an der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Technischen Universität Dortmund

Lehrstuhl für Investition und Finanzierung (Finance)
Prof. Dr. Jack Wahl

vorgelegt von

Dipl.-Wirt.-Math. Marcus Scheffer

aus Ratingen

2015

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Risk management emerged during the 1950s as an alternative to insurance markets, which has proven to be insufficient and cost-intensive for protection against pure risk. Since then, risk management has been theorized in a systematic fashion and has received particular attention at distinct levels of its development. The use of derivatives as new tools for managing risks starts in 1970s and introduce a new era of risk management. Due to the increased volatility in interest rates, stock market returns, foreign exchange rates and the prices of commodities companies intensified their financial risk management during the 1980s. As a consequence, the demand for risk management instruments as well as the size of the derivatives market expanded rapidly. The 1990s were characterized by the beginning of international risk regulation, and financial institutions establish internal risk management models to hedge against unanticipated risks and reduce regulatory capital. To cope with new challenges of the market, companies introduce the integrated risk management and the monitoring and control of risks became management objectives. As of 2000, risk management had begun to expand even further, but despite all significant advancements it could not prevent the recent financial crisis.

This gives a strong impetus to explore risk management by applying the full range of scientific research methods: the deductive development of a model, the implementation of models in experimental design as well as the inductive analysis of simulation results. This dissertation deals with all three major landmarks identifying the process of scientific research with regard to the univariate and multivariate modeling of financial market risks. Thus, it contributes to further develop risk assessment and management strategies under different conditions of risk in the form of four self-contained chapters.

The second chapter of this dissertation presents a deductive, theory-based approach in a risk management framework. The objective is to investigate implications of mean reversion in asset prices with respect to the optimal hedging strategy. In the area of forward pricing decisions and optimal hedge ratios, previous studies mainly focus on price evolutions expressed by standard geometric Brownian motion. In this context Benninga et al. (1983) uses the mean-variance paradigm to specify the optimal hedge. Briys and Schlesinger (1993) and Broll et al. (2001), on the other hand, employ the less restrictive expected utility approach. However, it has now become common knowledge that in many cases price changes reveal strong evidence of mean reversion. More precisely, neglecting the dependence of price changes critically adds to incorrect specifications of the optimal hedging decision. As hedge ratios derived with a mis-specified price process reduce only limited risk exposures, this chapter aims to fill the gap of a rigorous hedge analysis originating from mean reversion effects in asset prices. To be precise, the optimal hedging strategy is derived in an expected utility framework for assets following a mean-reverting square root process. The magnitude and the direction of hedging crucially depends on the degree of risk aversion, risk premium, and market conditions for derivatives. Furthermore, the optimal hedge ratio can be decomposed in a preference free component, a speculative component and two price-elastic components, depending on futures market conditions. The study reveals a comprehensive analysis on the interrelation of the so-called mean-reverting backwardation and mean-reverting contango situations, respectively, and the optimal hedge ratio. The

main finding is that investors sensitively react on market fluctuations regarding the consumption-wealth-ratio.

Chapter three and four deal with all three major landmarks identifying the process of risk management by means of simulation. The process starts with a deductive development of a theory, continues with its implementation in an experimental framework, and finally concludes with an inductive analysis of the simulation results. Both chapters focus on further developments of dynamic vine copula models and are motivated by the growing criticism of elliptical models in risk management. Elliptical models such as the multivariate Gaussian distribution cannot fully capture the dependence structures often found in financial asset returns. Starting with the work by Embrechts et al. (2002), several studies have criticized the inadequacy of correlation-based models for modeling the non-linear dependence in financial returns advocating the use of copulas instead. At the same time, elliptical copula models, which have become an industry standard in credit risk modeling following the influential study by Li (2000), have been found to perform just as poorly as their correlation-based counterparts due to their symmetric tail independence (see, e. g., Cherubini et al., 2004, Fischer et al., 2009). Especially in times of financial market turmoil, neglecting the tail dependence between financial time series can have disastrous effects on both banks and insurers as evidenced during the recent financial crisis. To take into account more complex dependence structures, recent works by Joe (1997), Bedford and Cooke (2002) and Whelan (2004) have proposed copula models which are highly flexible but at the same time still tractable even in higher dimensions. Most notably, vine copulas, also called pair-copula constructions, have emerged as the most promising tool for modeling dependence structures in high dimensions. Vine copulas consist of a cascade of conditional bivariate copulas that are combined hierarchically to yield a copula that can be used for building a joint distributional model for a data set. As a result, vine copulas are extremely flexible yet still tractable even in high dimensions as all computations necessary in statistical inference are performed on bivariate data sets. However, the increase in flexibility comes at the price of a greatly increased model risk. To fully specify a

d-dimensional vine copula, $d(d-1)/2$ different pair-copulas have to be selected and estimated from a set of candidate bivariate parametric copula families.

The paper in chapter three is the first to use the recently proposed nonparametric Bernstein copulas (see, e. g., Sancetta and Satchell, 2004, Pfeifer et al., 2009, Diers et al., 2012) as pair-copulas to yield smooth nonparametric vine copula models. The contributions of this approach are twofold: First, the use of Bernstein copulas completely obviates the need for the error-prone selection of pair-copulas from prespecified sets of parametric copulas. The resulting smooth and nonparametric vine copulas do not only constitute extremely flexible tools for modeling high-dimensional dependence structures, they are also characterized by a smaller model risk than their parametric counterparts. Second, Bernstein copulas have been shown to improve on the estimation of the underlying dependence structure by competing nonparametric empirical copulas. For example, Bernstein copulas provide a higher rate of consistency than other common nonparametric estimators and do not suffer from boundary bias. Similarly, other approximations as, e. g., linear B-spline copulas have also been shown to yield lower average squared approximation errors than competing discrete approximations. In addition, the Bernstein copula has recently attracted attention in insurance modeling. According to that, the modeling of a vine model's pair-copulas by the use of smooth approximating functions is thus a crucial extension in terms of reducing model risk. Within the scope of this chapter a comprehensive simulation study illustrates that the Bernstein vine model is well suited for the task of approximating the true dependence structure of a multivariate data sets. In comparison with the current state of the art benchmark modeling Bernstein vines outperform in higher dimensions with respect to the accuracy and numerical stability of the approximation to the true underlying dependence structure. To be precise, the Bernstein vine model works reliably without running into the numerical problems of the heuristic benchmark, i. e. non-convergence of the maximization of the log-likelihood function. By means of a risk management application it is documented that the proposed Bernstein vine model accurately forecasts the Value-at-Risk for multivariate portfolios consisting of com-

modities, stocks, and bonds. This finding is also supported by performing a range of formal statistical backtests.

The follow-up study presented in chapter four tackles the problem of constructing high-dimensional vine copulas in risk management from a different perspective. Using convex combinations of parametric copulas as pair-copulas creates the so-called mixture pair-copula-constructions (Mixture-PCCs in short). Employing bivariate mixture copulas to minimize the possibility of misspecifying a vine model is beneficial for two reasons: First, mixture copulas increase the flexibility of a vine model even further. Simple parametric copulas are often not flexible enough to model complex dependence structures and cannot account for diverse tail dependence patterns. Conversely, choosing suitable mixture components produces significantly better model fits, particularly in the tails. As stated above, the recent financial crisis especially suffers due to the neglect of joint extreme movements in financial time series. Second, Mixture-PCCs completely obviate the need for the error-prone selection of pair-copulas from pre-specified sets of parametric copulas. The bivariate building blocks are estimated sequentially by using the Expectation-Maximization (EM) algorithm to take account of the incomplete data structure. Each mixture pair-copula is fitted separately by estimating the vector including the mixing weights as well as the vectors of the unknown copula parameters. The superiority of the proposed model is illustrated by performing both a simulation and an empirical study on the in-sample and out-of-sample Value-at-Risk forecasting accuracy of Mixture-PCCs and heuristically calibrated vine copulas. The results show that Mixture-PCCs produce Value-at-Risk estimates that possess a comparable and especially in higher dimensions a better in-sample fit than the current state of the art benchmark model. In the empirical study of a four-dimensional financial portfolio, the proposed Mixture-PCC is characterized by a significantly better out-of-sample fit than the benchmark which overestimates portfolio risk. Based on these findings, Mixture-PCCs help risk managers to save on regulatory risk capital while at the same time satisfactorily bounding possible portfolio losses.

The final chapter of this dissertation pushes forward into a new area of research relating to financial markets. It has long been recognized that investor attention plays an important role in asset prices, returns, and the efficiency of security markets. But traditionally, measures of attention were restricted to indirect proxies like trading volume, abnormal returns, media news and advertising expense. Recent research in the empirical finance literature illustrate the usefulness of internet data as a direct measure for investor attention. For example, the work of Kim and Kim (2014) is based on internet message postings to assess the influence of investor attention on stock markets, while Siganos et al. (2014) employ Facebook users sentiment data in the same context. However, in financial economics literature Google search queries have emerged as the most promising method of measuring investor attention. Starting with the work by Da et al. (2011), most applications of Google search data in financial applications focused on asset pricing implications and predicting dynamics of stock market volatility (see, e. g., Da et al., 2015, Dimpfl and Jank, forthcoming, Mondria and Wu, 2011, Vozlyublennaia, 2014, Vlastakis and Markellos, 2012). In the course of these studies, the co-movements between investor attention and stock prices (volatility) has been investigated in detail. In this study, a statistical modeling framework for specifying, estimating, and testing time series of investor attention measured by internet search queries is provided. More precisely, this chapter presents both a univariate as well as a multivariate econometric model for Google search data. The dependence structure of high-dimensional Google search data is shown to be significantly non-linear and asymmetric. Furthermore, the existence of extreme dependence in Google search data pairs and, particularly noteworthy, between stock returns and the corresponding Google Search Data is documented. Finally, the main contribution is to show a striking similarity in the joint distributions of a multivariate bank stock portfolio and the corresponding portfolio of Google search queries, respectively. Following these results, it is hypothesized that investor attention as measured by internet search data reflects the most relevant information as contained in stock returns.

Starting point of this chapter is the proper modeling of internet search queries. While modeling of time series such as stock returns, foreign exchange rates, and CDS spreads has become common practice in financial econometrics, the application of statistical techniques to investor attention measured by internet search data is widely unexplored. To this end, a comprehensive time series analysis is provided and meaningful statistics and other characteristics of Google Search Data are extracted. In the underlying data set both autoregressive dynamics and conditional heteroskedasticity structures are found. Additionally, there is statistical evidence for specific distributional characteristics like the presence of distinct levels of skewness and kurtosis within Google Search Data. The analysis shows that the first- and second-moment dependencies are well captured by asymmetric ARMA-CS-GARCH models. Finally, it is shown that the skewed *t* as well as the skewed *ged* distribution provide good fits to the Google Search Data residuals. In the multivariate econometric framework, the objective is to model the joint distribution of high-dimensional search query data in an extremely flexible way. In this regard, the implementation of a vine copula approach is especially appropriate for two reasons: First, it allows to capture both linear dependences as well as potential non-linearities in the dependence structure. In fact, the existence of strong non-linear dependencies and asymmetries in the Google Search Data is documented. Second, due to their pair copula constructions, vine copulas allow to model different structural behaviors of pairs of variables. As a result, the study provides the first empirical evidence of significant tail dependence in Google search data. Beside their usefulness in capturing inherent dependency patterns of high dimensional data sets, vine copulas provide a powerful tool to depict similarities in the joint distribution of stock returns and Google Search Data. In the empirical application it is documented a striking similarity in the joint distributions of a multivariate bank stock portfolio and the corresponding portfolio of Google search queries, respectively. The remarkable similarities in the dependence patterns necessitates a detailed investigation of the co-movements between investor attention and stock returns. Further analyses provide first empirical evidence for the existence of tail dependence between stock returns and the respective

search query pairs. Furthermore, it is documented that stock returns and Google search data evolve concurrently in real time. These results suggest that investor attention as measured by internet search data reflects the most relevant information as contained in stock returns.

Apart from the introduction, this dissertation contains four chapters that can be read in any particular order and are based on various research topics in financial risk management. The following section provides a comprehensive overview of the papers and presents the corresponding publication details.

1.2 Publication details

Paper I (Chapter 2):

Dynamisches Hedging von Wertpapieren mit Mean Reversion Eigenschaften

Author:

Marcus Scheffer

Abstract:

Der vorliegende Beitrag untersucht ein zeitstetiges intertemporales Konsum- und Hedgeproblem eines Anlegers unter Ungewissheit. Die Preisentwicklungen von Kassa- und Termingeschäft unterliegen einer Mean-Reversion Eigenschaft, die durch Wurzel-Diffusionsprozesse modelliert wird. Der stochastische Zusammenhang wird mittels imperfekt korrelierter Wiener Prozesse abgebildet. Unter diesen Rahmenbedingungen wird eine modifizierte Keynes-Ramsey-Regel hergeleitet und anschließend analysiert. Die optimale dynamische Hedgeentscheidung des Anlegers lässt sich formal in eine präferenzfreie, eine spekulative und zwei preiselastische Komponenten zerlegen. Es zeigt sich, dass die präferenzfreie Komponente das reine Absicherungsmotiv des Investors erfasst, während die spekulative Komponente durch die sogenannte Mean-Reversion-Sharpe-Ratio des Futures in Abhängigkeit von der intertemporalen relativen Risikoaversion determiniert wird. Die beiden preiselastischen Hedge-Terme sind essentiell, um das Konsumrisiko hinsichtlich der intertemporalen Konsumstrategie zu minimieren.

Publication details:

Working paper.

Paper II (Chapter 3):

Forecasting Multivariate Portfolio-Value-at-Risk using Smooth Nonparametric Bernstein Vine Copulas

Authors:

Gregor Weiß, Marcus Scheffer

Abstract:

We consider the problem of accurately forecasting the market risk of a multivariate portfolio consisting of a stock index, U.S. bonds, a bond index and gold. We employ a multivariate GARCH model in which the dependence structure between the assets is modeled via a vine copula. We address the problem of how the parametric pair-copulas in a vine copula should be chosen by proposing to use nonparametric Bernstein copulas as bivariate pair-copulas. An extensive simulation study illustrates that our smooth nonparametric vine copula model is superior to competing parametric vine models calibrated via Akaike's Information Criterion. Our empirical analysis of financial market data demonstrates that our proposed model yields Value-at-Risk forecasts that are significantly more accurate than those of a benchmark parametric model.

Publication details:

Published in *Quantitative Finance*.

Paper III (Chapter 4):

Mixture Pair-Copula-Constructions

Authors:

Gregor Weiß, Marcus Scheffer

Abstract:

We propose the use of convex combinations of parametric copulas as pair-copulas in high-dimensional vine copula models. By doing so, we circumvent the error-prone need to choose and estimate a parametric copula for each pair-copula in a vine model. We show in simulations that our proposed model fits the dependence structure in a given data sample significantly better than a competing benchmark. In our empirical study on the models' accuracy for forecasting the Value-at-Risk of financial portfolios, we show that our proposed mixture pair-copula construction yields significantly better results in backtesting while the benchmark overestimates portfolio risk.

Publication details:

Published in the *Journal of Banking & Finance*.

Paper IV (Chapter 5):

Extreme Dependence in Investor Attention to Bank stocks

Author:

Marcus Scheffer

Abstract:

This paper proposes a univariate and multivariate modeling framework for investor attention measured by daily internet search data. We model the joint distribution of high-dimensional search query data and provide first evidence on strong non-linear and asymmetric dependence in the data. Further, we find a striking similarity between the joint distribution of the returns on a multivariate bank stock portfolio and the corresponding time series of search queries. Furthermore, we document the existence of tail dependence between stock returns and the respective search query pairs. Finally, our results show that investor attention as measured by internet search data and stock returns evolve concurrently in real time. Given this evidence, we hypothesize that investor attention as measured by internet search data and stock returns reflect almost the same information.

Publication details:

Under review at the *Journal of Applied Econometrics*.

Chapter 2

Dynamisches Hedging von Wertpapieren mit Mean Reversion Eigenschaften

2.1 Einleitung

Derivate gehören zu den Instrumenten des modernen Risikomanagements, deren Einsatz die Handelbarkeit unternehmerischer Risiken ermöglicht. Vor dem Hintergrund der gestiegenen Volatilität von Wertpapierrenditen, Zinssätzen, Devisenkursen und Rohstoffpreisen ergibt sich ein erhöhter Absicherungsbedarf, was den Handel mit Derivaten begünstigt und das Volumen der Derivatemärkte extrem vergrößert hat. Werden Derivate im Rahmen der Risikosteuerung in Form von Gegengeschäften zur Verminderung oder Ausschaltung von Preisrisiken (Risikopositionen) eingesetzt, wird dies als Hedging bezeichnet. Innerhalb der finanzwirtschaftlichen Literatur zum unternehmerischen Hedging existiert eine Vielzahl von Arbeiten, die Preisentwicklungen durch geometrisch Brownsche Bewegungen modellieren und die ökonomischen Auswirkungen der optimalen Absicherungspolitik untersuchen. Die Annahme einer geometrisch Brownschen Bewegung impliziert dabei unabhängige Zuwächse des unterliegenden Preisprozesses. Wertentwicklungen von Devisen, Anleihen und Rohstof-

fen weisen diese charakteristische Eigenschaft jedoch nicht auf. Vielmehr unterliegen sie einer Mean Reversion Eigenschaft, einer Tendenz, nach einer temporären Abweichung vom Mittelwert langfristig wieder zu diesem zurückzukehren. In empirischen Studien wurde diese Eigenschaft für Zinsen von Cox et al. (1985) und bei einigen Rohstoffnarten von Pindyck and Rubinfeld (1998) nachgewiesen. Für eine adäquate Modellierung dieser Preisentwicklungen sind somit stochastische Prozesse erforderlich, die eine Mean Reversion Eigenschaft abbilden können. Eine fehlerhafte Spezifizierung der unterliegenden stochastischen Prozesse führt zu einer inadäquaten Risikosteuerung. Insofern wird mit dem vorliegenden Arbeitspapier das Ziel verfolgt die optimale Absicherungspolitik bei unterliegenden Mean Reverting Prozessen zu ermitteln und zu analysieren.

Merton (1969) lieferte das Basismodell zur zeitstetigen intertemporalen Konsum- und Investitionsentscheidung. In einem ähnlichen Modellrahmen gelang es Stulz (1984) und Ho (1984) neue Erkenntnisse hinsichtlich der optimalen Konsum- und Absicherungspolitik zu gewinnen. In beiden Studien sowie einer Vielzahl weiterer Arbeiten wird das unterliegende Preisrisiko entweder durch eine geometrisch Brownsche Bewegung oder einen allgemeinen, nicht weiter spezifizierten stochastischen Prozess modelliert. Erst durch Briys et al. (1990), Briys and Solnik (1992) und Briys and Schlesinger (1993) wurde die optimale Hedge-Entscheidung bei unterliegenden Mean Reverting Prozessen untersucht. Basierend auf der Arbeit von Broll et al. (2010) werden in dieser Studie die ökonomischen Auswirkungen eines dynamischen Futures-Hedgings unter der Prämisse analysiert, dass ein rapräsentativer Entscheider den Erwartungsnutzen seines lebenslangen Konsumstroms unter Berücksichtigung einer bestimmten Vermögensentwicklung maximiert.

Die vorliegende Studie gliedert sich in vier Teile: Im Anschluss an diese Einleitung wird in Kapitel 2.2 zunächst das Grundmodell (Abschnitt 2.2.1) erläutert, in dem der Marktwert des Wertpapierbestands eines repräsentativen Investors einem Kursrisiko ausgesetzt ist und dieses mittels Futures-Hedging abgesichert werden soll. Auf Basis dieses Modells werden in Abschnitt 2.2.2 die zeitstetige Konsum- und

Hedgeentscheidung ermittelt und charakterisiert. Kapitel 2.3 behandelt den Einfluss der Absicherungsmöglichkeit des Kursrisikos mittels Futures auf das optimale intertemporale Konsumprofil des Investors. Unter den hier vorliegenden stochastischen Rahmenbedingungen wird eine modifizierte Keynes-Ramsey-Regel hergeleitet und anschließend analysiert. Das abschließende Kapitel 2.4 fasst die wichtigsten Ergebnisse der Studie zusammen und liefert einen kurzen Ausblick auf offene Forschungsfragen.

2.2 Imperfektes dynamisches Hedging

In diesem Kapitel wird zunächst in Abschnitt 2.2.1 das Grundmodell vorgestellt, in dem einem Investor zur Finanzierung seiner Konsumausgaben ein Wertpapierbestand zur Verfügung steht. Die Wertpapierrendite ist unsicher und damit der realisierbare Konsum ebenfalls.[1] Aufgrund seiner Präferenzen will der Investor dem Renditerisiko mittels Futures-Hedging entgegenwirken. Im Rahmen dieses Modells wird in Abschnitt 2.2.2 die optimale intertemporale Konsum- und Hedgeentscheidung ermittelt.

2.2.1 Modell

Den Ausgangspunkt der nachfolgenden Untersuchungen bildet ein repräsentativer Investor[2], der über einen Wertpapier-Anfangsbestand verfügt. Der Marktwert dieses Bestandes beträgt Φ und unterliegt einem Kursrisiko.[3] Der korrespondierende Kassakurs s wird annahmegemäß durch den Wurzel-Diffusionsprozess

$$ds = \kappa_s(\theta_s - s)dt + \sigma_s \sqrt{s} dz_s \tag{2.1}$$

modelliert, mit $\kappa_s \geq 0$ und $\theta_s, \sigma_s > 0$ sowie einem Wiener Prozess z_s.

[1] Vgl. Wahl and Broll (2003), S. 170.
[2] Wird im Folgenden auch als der Investor bezeichnet.
[3] Vgl. Wahl and Broll (2003), S. 171.

Für die Absicherung des Kursrisikos stehen dem Investor Terminfixgeschäfte zur Verfügung. Konkret handelt es sich dabei um Futures-Kontrakte[4], deren Laufzeit mit dem Planungshorizont des Investors übereinstimmt. Der Kurs des Futures f folgt ebenfalls einem Wurzel-Diffusionsprozess. Mit den Parametern $\kappa_f \geq 0$ und $\theta_f, \sigma_f > 0$ sowie dem Wiener Prozess z_f lautet dieser annahmegemäß:

$$df = \kappa_f(\theta_f - f)dt + \sigma_f \sqrt{f} dz_f. \tag{2.2}$$

Die Modellierung der Preisprozesse (2.1) und (2.2) mit zwei unterschiedlichen Brownschen Bewegungen z_s und z_f ermöglicht es, die Korrelation zwischen dem Wertpapier s und dem Futures f zu berücksichtigen. Es gilt $dz_s dz_f = \rho_{sf} dt$, wobei $\rho_{sf} > 0$ den zeitunabhängigen Korrelationskoeffizienten zwischen dem Kassakurs s und dem Terminkurs f bezeichnet. Hierdurch wird implizit unterstellt, dass der Basiswert des Futures nicht perfekt mit dem Kassakurs korreliert ist.

Aufgrund seiner Präferenzen will der Investor dem Marktwertrisiko seines Wertpapierbestandes über einen vorgegebenen Planungshorizont $[0, T]$ entgegenwirken. Durch den Verkauf von Futures im Umfang von $H > 0$ nimmt er eine kurze Position in Terminkontrakten ein und erreicht somit ein Hedging.[5,6] Aufgrund der imperfekten Korrelation stehen jedoch Inkongruenzen einer vollständigen Absicherung entgegen, wodurch die Risikosteuerung als Cross-Hedging[7] verwirklicht wird. Der Investor intentioniert durch die Absicherung die Stabilisierung einer bestimmten Kon-

[4] Die beliebige Teilbarkeit der Futures-Kontrakte ist in diesem Modellrahmen eine unabdingbare Voraussetzung. Nur für diesen Fall ist gewährleistet, dass der Investor seine optimale Absicherungspolitik realisieren kann. Des Weiteren dürfen vor dem Verfallstermin eines Future-Kontraktes keine Zahlungen anfallen, d. h. es existieren keine zwischenzeitlichen Margin-Zahlungen.

[5] Vgl. Wahl and Broll (2003), S. 171.

[6] Sind der Kassa- und Terminkurs negativ korreliert, d. h. gilt $\rho_{sf} < 0$, erreicht der Investor durch den Kauf von Terminkontrakten ein Hedging.

[7] Cross-Hedging bezeichnet eine Technik zur Absicherung einer Kassaposition mit Hilfe eines Terminkontrakts, der eine andere Basisgröße hat. Bei Anwendung eines Cross Hedge wird davon ausgegangen, dass die beiden Basisgrößen dennoch preisverwandt sind.

sumstrategie. Bezeichnet $C\,dt$ die Konsumausgaben über die Zeitperiode dt, lautet die Entwicklung des Vermögens Φ demnach:

$$d\Phi = \Phi ds/s - Cdt - Hdf/f. \tag{2.3}$$

Das Vermögen des Investors zum Zeitpunkt t hängt somit von dem Vermögen des vorherigen Zeitpunkts $\Phi(t-1)$, der (relativen) Veränderung des Kassakurses, der (relativen) Veränderung des Terminkurses f in Abhängigkeit des Hedge-Volumens H und dem über die Zeitperiode dt realisierten Konsum ab.

Unter Berücksichtigung der Kursprozesse (2.1) und (2.2) und der Hedge-Rate $h = \frac{H}{\Phi}$ ergibt sich für die Vermögensentwicklung (2.3)

$$d\Phi = -\kappa_\Phi(\theta_\Phi - \Phi)dt + \Phi\sigma_\Phi d\omega, \tag{2.4}$$

mit

$$\sigma_\Phi = \sqrt{\frac{f\sigma_s^2 - 2h\sqrt{sf}\sigma_s\sigma_f\rho_{sf} + h^2 s\sigma_f^2}{sf}},$$

und

$$d\omega = \frac{\sigma_s\sqrt{f}dz_s - h\sigma_f\sqrt{s}dz_f}{\sqrt{s}\sqrt{f}\sigma_\Phi},$$

wobei $\kappa_\Phi = ((\frac{\theta_s}{s} - 1)\kappa_s - (\frac{\theta_f}{f} - 1)\kappa_f h)$ und $\theta_\Phi = \frac{C}{\kappa_\Phi}$ gilt. Der Term $d\omega$ konzentriert die stochastischen Elemente der Vermögensentwicklung $d\Phi$. ω weist wegen $\mathbb{E}(d\omega) = 0$ und $\mathbb{E}(d\omega)^2 = dt$ charakteristische Eigenschaften eines Wiener Prozesses auf. Die Vermögensentwicklung $d\Phi$ besitzt die Eigenschaften[8]

$$\mathbb{E}[d\Phi] = -\kappa_\Phi(\theta_\Phi - \Phi)dt \tag{2.5}$$

und[9]

$$\mathbb{V}ar[d\Phi] = \Phi^2\sigma_\Phi^2 dt. \tag{2.6}$$

[8] $\mathbb{E}[\cdot]$ kennzeichnet den Erwartungswertoperator.
[9] $\mathbb{V}ar[\cdot]$ kennzeichnet den Varianzoperator.

Der Investor orientiert sich am Bernoulli-Prinzip[10] und sein Verhalten ist im Sinne der Erwartungsnutzentheorie durch Risikoaversion[11] gekennzeichnet. In diesem Zusammenhang bezeichnet $u(C)$ eine von Neumann-Morgenstern Nutzenfunktion, wobei $u'(C) > 0$ und $u''(C) < 0$ für alle C gilt. Ziel des Investors ist es, seine Konsumstrategie und Absicherungspolitik derart zu wählen, dass der intertemporale Erwartungsnutzen über den Planungshorizont $[0, T]$ maximal wird. Unter Berücksichtigung der Zeitpräferenzrate[12] τ lässt sich das Entscheidungsproblem formal durch das Optimierungsproblem

$$V(\Phi(0), s(0), f(0), 0) = \max_{C,h} \mathbb{E} \int_0^T u(C) e^{-\tau t} dt \tag{2.7}$$

unter der Vermögensentwicklung (2.4) darstellen. V bezeichnet in diesem Kontext die sogenannte Wertfunktion[13] und gibt den optimalen Wert der Zielfunktion ausgehend vom Zeitpunkt 0 an. Zur Lösung des Optmimierungsproblems (2.7) werden im Folgenden Methoden der dynamischen Programmierung angewendet.

2.2.2 Optimale Konsum- und Hedgeentscheidung

Die Dynamische Programmierung bildet eine Gruppe von Verfahren zur Optmierung sequentieller Entscheidungsprobleme.[14] Das Konzept dieser Methode geht auf Bellman (1957) zurück, der ihr gleichzeitig den Namen „Dynamische Programmierung“ gegeben hat. Sequentielle Entscheidungsprobleme zeichnen sich dadurch aus, dass

[10] Das Bernoulli-Prinzip bildet eine theoretische Basis zur Beschreibung von subjektiven Präferenzen, um Entscheidungen unter Risiko bzw. die Wahl zwischen Wahrscheinlichkeitsverteilungen von Ergebnissen zu treffen. Die Grundlage des Prinzips bildet eine Menge von Axiomen, die die Basis einer Theorie der Entscheidung unter Risiko darstellt. Vgl. Franke and Hax (2009), S. 299 ff.

[11] In Entscheidungssituationen unter Risiko ist Risikoaversion die klassische Verhaltensannahme der Entscheidungsträger. Vgl. z.B. Pratt (1964) und Arrow (1965).

[12] Die (positive) Zeitpräferenzrate stellt eine Nutzendiskontrate dar, nach der der Nutzen des zukünftigen Konsums im Zeitpunkt t_1 bezogen auf den Zeitpunkt $t_0 = 0$ um den Faktor $e^{-\tau t_1}$ geringer bewertet wird. Demnach lässt sich die Zeitpräferenzrate als ein Indikator für die Ungeduld hinsichtlich des Nutzens auffassen. Vgl. Albers (1981), S. 465 und Frenkel and Hemmer (1999), S. 74.

[13] In der Literatur wird anstelle der Bezeichnung Wertfunktion auch der Terminus indirekte Nutzenfunktion verwendet. Vgl. z. B. Chang (2004), S. 115. und Munk (2013), S. 235.

[14] Für die Lösung dynamischer Optimierungsprobleme sind neben der Dynamischen Programmierung zwei weitere Methoden bekannt: die klassische Variationsrechnung sowie die Theorie der optimalen Kontrolle (Pontryagins Maximumprinzip). Vgl. z. B. Christiaans (2004), S. 66 f.

sie in mehrere Stufen zerlegt werden können. Dabei wird jede Entscheidung durch die vorangehende beeinflusst. Bei zeitlich sequentiellen Entscheidungsproblemen sind die Stufen zeitabhängig angeordnet, so dass zeitlich nachfolgende und vorangehende Entscheidungen und Zustände resultieren.[15] Die auf den einzelnen Stufen (Zeitpunkte) zu treffenden Entscheidungen werden dabei durch Entscheidungsvariablen repräsentiert.[16] Die Entscheidungsvariablen des hier vorliegenden Optimierungsproblems sind der Konsum C und die Absicherungspolitik h.[17] Nicht kontrollierbare Faktoren werden dagegen als Zustandsvariablen deklariert und beschreiben die Zustände in einem bestimmten Zeitpunkt bzw. auf einer einzelnen Stufe.[18,19] Differenziert wird zwischen endogenen und exogenen Zustandsvariablen. Im Rahmen dieses Modells ist die Stochastik des Kassakurses s und des Terminkurses f durch die Prozesse (2.1) und (2.2) exogen vorgegeben. Das Vermögen Φ ist dagegen eine abhängige Variable, die durch die Entscheidungsvariablen Konsum und Hedge-Rate sowie den exogenen Zustandsvariablen bestimmt wird.[20]

Die Anforderungen an eine optimale Lösung werden durch Bellman in dem sogenannten Optimalitätsprinzip charakterisiert.[21,22] Für das hier betrachtete Entscheidungsproblem bedeutet dies, dass es für die zukünftigen optimalen Entscheidungen für einen beliebig betrachteten Zeitpunkt $t \in [0, T]$ nicht relevant ist, auf welche Art die Zustände im Zeitpunkt t erreicht worden sind. Die Zustände im Zeitpunkt t stellen gleichzeitig die Ausgangszustände für die folgenden Entscheidungen dar. Wird, aus-

[15] Vgl. Albers (1981), S. 342f.

[16] Die Bezeichnung „Entscheidungsvariable“ kommt daher, dass sie Elemente darstellen, die direkt durch den Entscheider beeinflusst werden.

[17] Alternativ kann auch das Konsumniveau $c = \frac{C}{\Phi}$ bzw. das Hedge-Volumen H betrachtet werden.

[18] Die Begriffe Zustand und Zustandsvariable werden im Folgenden synonym verwendet.

[19] In der Literatur wird im Rahmen der Dynamischen Programmierung häufig zwischen drei Typen von Variablen differenziert: Entscheidungs-, Zustands- und Zeitvariablen. Vgl. Chiang (1992), S. 18.

[20] Vgl. Gleichung (2.3).

[21] „An optimal policy has the property that, whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.“ Vgl. Bellman (1957), S. 83.

[22] Das Prinzip bezieht sich auf alle Zeitpunkte $t \in [0, T]$, so dass es gleichzeitig hinreichend für die Optimalität einer Lösung ist.

gehend vom Zeitpunkt t, für die verbleibende Zeitperiode $[t, T]$ die optimale Strategie gewählt, so gibt

$$V(\Phi(t), s(t), f(t), t) = \max_{C(\xi),h(\xi)} \mathbb{E} \int_t^T U(C(\xi))e^{-\tau\xi} d\xi \tag{2.8}$$

unter Berücksichtigung der Vermögensentwicklung (2.4) den maximal erreichbaren Wert der Zielfunktion für die Zeitperiode $[t, T]$ an. Anhand der Darstellung (2.8) wird deutlich, dass das Optimalitätsprinzip von Bellman auch für ein Optimum des Entscheidungsproblems (2.7) über den gesamten Planungshorizont $[0, T]$ notwendig ist.[23]

Aus der Dynamischen Programmierung ist bekannt, dass die Wertfunktion V die sogenannte Bellman-Gleichung löst. Nach Dixit and Pindyck (1994) lautet die korrespondierende Bellman-Gleichung für ein zeitstetiges Optimierungsproblem unter Untersicherheit:

$$\tau V(\Phi(t), s(t), f(t), t) = \max_{C(t),h(t)} \{u(C(t)) + \mathbb{E}[dV(\Phi(t), s(t), f(t), t)]/dt\}. \tag{2.9}$$

Bei optimaler Entscheidung beträgt der Wert des intertemporalen Erwartungsnutzens V. Die linke Seite der Gleichung lässt sich somit als verlangter Nutzen des Investors auffassen. Dieser ist gegeben durch den unmittelbaren Konsumnutzen zuzüglich der erwarteten Änderung der Wertfunktion.[24]

Wird Gleichung (2.9) nach den Entscheidungsvariablen C und h differenziert, ergeben sich unter Berücksichtigung der Vermögensentwicklung (2.4) durch die Bedingungen erster Ordnung die optimale Konsum- und Hedgeentscheidung:[25]

[23] Wird für das Zeitintervall $[t, T]$ eine Politik gewählt, die den Erwartungswert auf der rechten Seite von (2.8) nicht maximiert, so kann der Erwartungswert in (2.7) ebenfalls nicht sein Maximum annehmen.

[24] Beides bewertet unter der Prämisse optimaler Entscheidungen.

[25] Aus Gründen der Übersichtlichkeit wird auf eine gesonderte Kennzeichnung des Zeitindexes t bzgl. der Entscheidungs- und Kontrollparameter verzichtet, d. h. für die folgenden Ausführungen gilt stets $\Phi = \Phi(t), s = s(t), f = f(t), C = C(t), h = h(t)$.

Satz 2.2.1. *Auf Grundlage der Bellman-Gleichung* (2.9) *lautet die optimale Konsum- bzw. Hedgeentscheidung für das in diesem Modell vorliegende Entscheidungsproblem:*

$$u'(C^*) = V_\Phi(\Phi, s, f, t), \tag{2.10}$$

$$h^* = \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf} + \frac{\kappa_f(\theta_f - f)}{\sigma_f^2\frac{V_{\Phi\Phi\Phi}}{V_\Phi}} + \frac{f}{\frac{V_{\Phi\Phi\Phi}}{V_{\Phi f}}} + \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{s}{\frac{V_{\Phi\Phi\Phi}}{V_{\Phi s}}}. \tag{2.11}$$

Beweis: Für die Wertfunktion V ergibt sich nach dem Lemma von Itô[26,27] das totale Differential:

$$\begin{aligned} dV(\Phi, s, f, t) &= V_t dt + V_\Phi d\Phi + V_s ds + V_f df \\ &+ \frac{1}{2}V_{\Phi\Phi}(d\Phi)^2 + \frac{1}{2}V_{ss}(ds)^2 + \frac{1}{2}V_{ff}(df)^2 \\ &+ V_{\Phi s}d\Phi ds + V_{\Phi f}d\Phi df + V_{sf}dsdf. \end{aligned} \tag{2.12}$$

[26] Der Mathematiker Itô (1951) entdeckte ein zentrales Resultat auf dem Gebiet der stochastischen Analysis, welches als Lemma von Itô bezeichnet wird. Folgt die Variable x einem sogenannten Itô Prozess $dx = a(x,t)dt + b(x,t)dz$, wobei dz das Inkrement eines Wiener Prozesses darstellt und a und b jeweils Funktionen der Variablen x und der Zeit t sind, so beantwortet das Lemma von Itô die Frage, welchem Prozess die Funktion $F(x,t)$ folgt. Ist die Funktion $F(x,t)$ zweimal differenzierbar in der Variablen x und einmal differenzierbar in der Zeitkomponente t, so gilt für das stochastische Differential von $F(x,t)$:

$$dF = \left(\frac{\partial F}{\partial t} + a(x,t)\frac{\partial F}{\partial x} + \frac{1}{2}b^2(x,t)\frac{\partial^2 F}{\partial x^2}\right)dt + b(x,t)\frac{\partial F}{\partial x}dz,$$

wobei z derselbe Wiener-Prozess wie der des Itô Prozesses dx ist. $F(x,t)$ folgt somit ebenfalls einem Itô Prozess mit der Drift $\frac{\partial F}{\partial t} + a(x,t)\frac{\partial F}{\partial x} + \frac{1}{2}b^2(x,t)\frac{\partial^2 F}{\partial x^2}$ und der Volatilität $b(x,t)\frac{\partial F}{\partial x}$. Itô´s Lemma basiert demnach auf einer Taylor-Entwicklung zweiten Grades der Funktion $F(x,t)$. Vgl. Dixit and Pindyck (1994), S. 79ff.

[27] Es wird angenommen, dass die Wertfunktion hinreichend oft stetig differenzierbar ist.

Für den Term $\mathbb{E}[dV/dt]$ aus der Bellman-Gleichung (2.9) ergibt sich unter Berücksichtigung der Vermögensentwicklung (2.4) und dem Differential (2.12):[28]

$$\begin{aligned}\mathbb{E}[dV(\Phi,s,f,t)/dt] &= V_t + V_\Phi(-\kappa_\Phi)(\theta_\Phi-\Phi) + V_s\kappa_s(\theta_s-s) + V_f\kappa_f(\theta_f-f) \quad (2.13)\\ &+ \frac{1}{2}\left\{V_{\Phi\Phi}\Phi^2\left(\frac{\sigma_s^2}{s}+h^2\frac{\sigma_f^2}{f}-2h\frac{\sigma_s\sigma_f}{\sqrt{sf}}\rho_{sf}\right)+V_{ss}\sigma_s^2 s+V_{ff}\sigma_f^2 f\right\}\\ &+ V_{\Phi s}\Phi\left(\sigma_s^2-h\sigma_s\sigma_f\frac{\sqrt{s}}{\sqrt{f}}\rho_{sf}\right)+V_{\Phi f}\Phi\left(\sigma_s\sigma_f\frac{\sqrt{f}}{\sqrt{s}}\rho_{sf}-h\Phi\sigma_f^2\right)\\ &+ V_{sf}\sigma_s\sigma_f\sqrt{sf}\rho_{sf}.\end{aligned}$$

Einsetzen von (2.13) in die Bellman-Gleichung (2.9) und Differentiation nach der Kontrollvariablen C ergibt unmittelbar die Bedingung (2.10) für die erwartungsnutzenmaximale Konsumentscheidung:

$$u'(C^*) = V_\Phi(\Phi,s,f,t).$$

Die Bedingung (2.11) für die erwartungsnutzenmaximale Hedgeentscheidung ergibt sich nach Einsetzen von (2.13) in die Gleichung (2.9) durch Differentiation nach h, d. h. es gilt:

$$0 = -V_\Phi\Phi\kappa_f\left(\frac{\theta_f}{f}-1\right)+V_{\Phi\Phi}\Phi^2\left(h\frac{\sigma_f^2}{f}-\frac{\sigma_s\sigma_f}{\sqrt{sf}}\rho_{sf}\right)-V_{\Phi s}\Phi\sigma_s\sigma_f\frac{\sqrt{s}}{\sqrt{f}}\rho_{sf}-V_{\Phi f}\Phi\sigma_f^2.$$

Division durch den Faktor $V_{\Phi\Phi}\Phi^2\frac{\sigma_f^2}{f}$ und anschließendes Auflösen nach h ergibt die Bedingung (2.11) für die erwartungsnutzenmaximale Hedgeentscheidung:

$$h^* = \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}+\frac{\kappa_f(\theta_f-f)}{\sigma_f^2\frac{V_{\Phi\Phi}\Phi}{V_\Phi}}+\frac{f}{\frac{V_{\Phi\Phi}\Phi}{V_{\Phi f}}}+\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{s}{\frac{V_{\Phi\Phi}\Phi}{V_{\Phi s}}}.$$

[28] An dieser Stelle muss erwähnt werden, dass für einen Wiener-Prozess z die Terme $(dtdz_t)$ und (dt^2) Terme der Ordnung $O(dt)$ sind. Demnach werden die Terme $(dtdz_s),(dtdz_f)$ und $(dt)^2$ in dem totalen Differential (2.12) bei der folgenden Gleichung nicht weiter berücksichtigt. Außerdem verhalten sich Terme der Ordnung $(dz_t)^2$ für einen Wiener Prozess wie dt, was zusätzlich für eine wesentliche Vereinfachung bei den Berechnungen führt. Vgl. Ingersoll (1987), S. 267f.

Die Lösungen der Bedingungen erster Ordnung, C^* und h^*, bilden ein eindeutiges, globales Maximum, wenn die Wertfunktion V in Abhängigkeit des Vermögens strikt konkav verläuft. D. h. die notwendigen Bedingungen (2.10) und (2.11) sind in diesem Fall für ein Maximum auch hinreichend. □

Die Optimalitätsbedingung (2.10) lässt sich folgendermaßen interpretieren: Während die linke Seite den Nutzenzuwachs angibt, den eine zusätzliche marginale Konsumeinheit stiftet, gibt die rechte Seite den erwarteten Grenznutzen des momentanen Vermögens an. Dieses aktuelle Vermögen ist die Kapitalakkumulation für den optimalen zukünftigen Konsum, welches sich in der Wertfunktion V widerspiegelt.[29] In der Literatur ist dieser Sachverhalt als „envelope condition" der intertemporalen Optimierung bekannt.[30] Es umhüllt die optimale Hedging-Entscheidung, welche separat von der optimalen Konsumentscheidung getroffen wird.

Die zweite Bedingung des obigen Satzes gibt die optimale Hedge-Rate des Investors an, welche aus vier Summanden besteht.[31] Auffallend ist dabei insbesondere der signifikante Einfluss der Wertfunktion V über verschiedene partielle Ableitungen in den Hedge-Komponenten h_2, h_3 und h_4. Der Term $-\frac{V_{\Phi\Phi\Phi}}{V_\Phi}$ aus dem Nenner des zweiten Summanden von (2.11) weist eine enge Verwandtschaft zu dem Arrow-Pratt-Maß[32] der relativen Risikoaversion auf, das durch $-\frac{U''(C)}{U'(C)}C$ definiert ist. Dies motiviert die nachfolgende Definition:

Definition 1. *Der Term*

$$R = -\frac{V_{\Phi\Phi}\Phi}{V_\Phi} \tag{2.14}$$

ist ein Maß für intertemporale relative Risikoaversion.

[29] Intuitiv beschreibt dieses Optimum für den Investor also denjenigen Punkt, an dem eine extra Einheit Konsum mit den „Kosten" entgangener Vermögensakkumulation übereinstimmt. Vgl. Faria and McAdam (2013), S. 439

[30] Vgl. Huang and Litzenberger (1988), S. 196. Eine detaillierte Beschreibung des Envelope Theorems ist bei Dixit (1990), S. 57 zu finden.

[31] Im Folgenden werden diese Summanden auch als Hedge-Komponenten bezeichnet. Dabei werden sie fortlaufend mit h_1, h_2, h_3 und h_4 gekennzeichnet.

[32] Das Arrow-Pratt-Maß ist ein Maß für die Risikoaversion eines Entscheidungsträgers. Es wurde nach Kenneth Arrow und John W. Pratt benannt, wobei grundsätzlich zwischen der absoluten Risikoaversion und der relativen Risikoaversion unterschieden wird. Vgl. Pratt (1964) und Arrow (1965).

Aus der Optimalitätsbedingung (2.10) folgt $V_\Phi > 0$. Durch Anwendung der Kettenregel auf $V_\Phi = u'(C^*)$ ergibt sich $V_{\Phi\Phi} = u_{CC}C_\Phi < 0$, so dass für die intertemporale relative Risikoaversion $R > 0$ gilt. Aus der Beziehung $V_{\Phi\Phi} = u_{CC}C_\Phi$ ist auch unmittelbar der Zusammenhang zum Arrow-Pratt-Maß der relativen Risikoaversion ersichtlich. Die Krümmung der intertemporalen Wertfunktion V hängt zusätzlich von C_Φ ab, der Wirkung des Vermögens auf den optimalen Konsum. Der Koeffizient R lässt sich somit alternativ folgendermaßen darstellen:

$$R = -\frac{V_{\Phi\Phi}\Phi}{V_\Phi} = -\frac{u_{CC}C_\Phi\Phi}{u_C} = -\frac{u_{CC}C}{u_C}\frac{C_\Phi\Phi}{C}.$$

D. h. die intertemporale relative Risikoaversion entspricht dem Produkt aus dem Arrow-Pratt-Maß der relativen Risikoaversion und der Vermögenselastizität des Konsums.[33]

Die Terme $\frac{s}{\frac{V_{\Phi\Phi}\Phi}{V_{\Phi s}}}$ und $\frac{f}{\frac{V_{\Phi\Phi}\Phi}{V_{\Phi f}}}$ aus dem dritten bzw. vierten Summanden der rechten Seite von Gleichung (2.10) lassen sich durch die Erweiterung mit dem Grenznutzen des Vermögens V_Φ ebenfalls in Abhängigkeit von der intertemporalen relativen Risikoaversion R darstellen:

$$\frac{s}{\frac{V_{\Phi\Phi}\Phi}{V_{\Phi s}}} = \frac{\frac{V_{\Phi s}s}{V_\Phi}}{R} \qquad \text{bzw.} \qquad \frac{f}{\frac{V_{\Phi\Phi}\Phi}{V_{\Phi f}}} = \frac{\frac{V_{\Phi f}f}{V_\Phi}}{R}.$$

Entsprechend ihrer Struktur lassen sich die Quotienten $\frac{V_{\Phi s}s}{V_\Phi}$ und $\frac{V_{\Phi f}f}{V_\Phi}$ wie folgt determinieren:

[33] Bei der Elastizität handelt es sich um eine dimensionslose Größe, die die Sensitivität einer Variablen in Hinblick auf eine andere Variable misst. Als Kennzahl gibt sie dabei die relative Änderung an, die an einer (zu erklärenden) abhängigen Variable als Reaktion auf eine relative Änderung einer unabhängigen (erklärenden) Variablen eintritt. Ein bekanntes Beispiel aus der Volkswirtschaftslehre ist die Preiselastizität der Nachfrage, welche die relative Änderung der nachgefragten Menge eines Gutes infolge einer relativen Änderung des Preises des entsprechenden Gutes angibt. Eine quantitative Kategorisierung der Elastizität erfolgt anhand der Basis eins. Ist der Elastizitätskoeffizient größer (gleich, kleiner) als dieser Wert, wird der Terminus elastisch (einheitselastisch, unelastisch) verwendet. Vgl. Chiang (1984), S. 191, Pindyck and Rubinfeld (2011), S. 65 und Tomann (2005), S. 55.

Definition 2. *Die Terme*

$$R_s = -\frac{V_{\Phi s} s}{V_\Phi} \qquad bzw. \qquad R_f = -\frac{V_{\Phi f} f}{V_\Phi} \tag{2.15}$$

sind Preiselastizitäten des Grenznutzens des Vermögens. Sie geben an, mit welchen relativen Veränderungen der Grenznutzen des Vermögens V_Φ marginal auf relative marginale Veränderungen des Kassakurses s bzw. Futureskurses f reagieren.

Unter Berücksichtigung der Definitionen 1 und 2 kann die optimale Hedge-Rate (2.11) alternativ folgendermaßen dargestellt werden:

$$h^* = \frac{\sigma_s \sqrt{f}}{\sigma_f \sqrt{s}} \rho_{sf} - \frac{\kappa_f(\theta_f - f)}{\sigma_f^2 R} + \frac{R_f}{R} + \frac{\sigma_s \sqrt{f}}{\sigma_f \sqrt{s}} \rho_{sf} \frac{R_s}{R}. \tag{2.16}$$

Insgesamt besteht die Hedge-Rate aus einer präferenzfreien (h_1), einer spekulativen (h_2) und zwei preiselastischen (h_3 und h_4) Hedge-Komponenten.[34] Aus jeder Komponente der Hedge-Rate lassen sich essentielle Erkenntnisse gewinnen, die im Folgenden genauer beschrieben werden.

Der erste Term der optimalen Hedge-Rate h_1 wird auch als purer Hedge-Term bezeichnet und erfasst das reine Hedgemotiv des Investors. h_1 ist unabhängig von der intertemporalen relativen Risikoaversion R, sowie den intertemporalen Preiselastizitäten R_s bzw. R_f und somit die einzige präferenzfreie Komponente der optimalen Hedge-Rate. Durch die Abhängigkeit von dem Kassakurs s und dem Futureskurs f ist der pure Hedge-Term h_1 implizit zeitabhängig.[35] Das Vorzeichen von h_1 ist wegen $s, f, \sigma_s, \sigma_f > 0$ von dem korrelativen Zusammenhang ρ_{sf} abhängig. Ist der Kovarianzterm zwischen dem Wertpapier s und dem Futures f positiv (negativ), so ist der Hedge-Term h_1 ebenfalls positiv (negativ). Isoliert betrachtet nimmt der Investor

[34] Die Zerlegung der Hedge-Rate in eine Absicherungs- und eine Spekulationskomponente ist bereits bei Working (1953) und Johnson (1960) vorzufinden. Der Fachterminus der dritten bzw. vierten Hedge-Komponente wird durch die Definition 2 motiviert.

[35] Die durch $\sqrt{s}$ und $\sqrt{f}$ implizit gegebene Dynamik resultiert aus den unterstellten Wurzel-Diffusionsprozessen für den Kassakurs s bzw. dem Futureskurs f, insbesondere aus den Diffusionstermen $\sigma_s \sqrt{s}$ und $\sigma_f \sqrt{f}$. Bei Broll and Wahl (2012), S. 113 ff., folgen Kassa- und Futureskurs annahmegemäß einer geometrisch Brownschen Bewegung, wodurch sich eine rein statische Hedge-Komponente ergibt.

in diesem Fall eine kurze (lange) Position in Terminkontrakten ein.[36] Darüber hinaus entspricht der pure Hedge-Term dem sogenannten Minimum-Varianz-Hedge[37]. Ziel des varianzminimalen Hedges ist es, eine bestimmte Anzahl[38] von Futures zu verkaufen, die gewährleistet, dass die Vermögensentwicklung zum Zeitpunkt t ein möglichst geringes Schwankungsrisiko aufweist.[39] Mit der Vermögensentwicklung $d\Phi$ aus Gleichung (2.4) lautet das zu lösende Optimierungsproblem

$$\min \mathbb{V}ar[d\Phi]. \tag{2.17}$$

Die Lösung des Problems entspricht der Hedge-Komponente h_1. Dieses Resultat liefert der folgende Satz:

Satz 2.2.2. *Die unbedingte Minimierung der Varianz der Vermögensentwicklung* $d\Phi$ (2.4) *ergibt den puren Hedge-Term* h_1.

Beweis: Aus Gleichung (2.4) ergibt sich

$$\mathbb{V}ar[d\Phi] = \Phi^2\sigma_\Phi^2 dt. \tag{2.18}$$

Die Minimierung von Gleichung (2.18) hinsichtlich der Hedge-Rate h ergibt folgende notwendige Bedingung:

$$\begin{aligned} \frac{\partial \mathbb{V}ar[d\Phi]}{\partial h} &= 0 \\ \Leftrightarrow \qquad 0 &= \Phi^2\left(\frac{-2\sigma_s\sigma_f\rho_{sf}}{\sqrt{sf}} + 2h\frac{\sigma_f^2}{f}\right) \\ \Leftrightarrow \qquad h &= \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}. \end{aligned}$$

[36] Vgl. Broll et al. (2010), S. 23.
[37] Vgl. Briys and Solnik (1992).
[38] Alternativ kann auch der Anteil, also die Hedge-Ratio betrachtet werden.
[39] Vgl. Albrecht and Maurer (2008), S. 579.

Für die zweite Ableitung gilt

$$\frac{\partial^2 \mathbb{V}ar[d\Phi]}{\partial h^2} = 2\frac{\sigma_f^2}{f}\Phi^2 > 0,$$

so dass die hinreichende Bedingung für die Existenz eines Minimums erfüllt ist. □

Der Verkauf (Kauf) von Futures im Umfang von $H_1 = h_1\Phi$ bewirkt unter der Annahme $\rho_{sf} > 0$ ($\rho_{sf} < 0$) eine Minderung des Risikos der Vermögensentwicklung $d\Phi$ im Vergleich zur Nicht-Absicherung:

$$\mathbb{V}ar[d\Phi]|_{h=h_1} = \Phi^2\frac{\sigma_s^2}{s}(1-\rho_{sf}^2)dt < \Phi^2\frac{\sigma_s^2}{s}dt = \mathbb{V}ar[d\Phi]|_{h=0}\,. \tag{2.19}$$

Der Risikoabsicherungsgrad ist direkt von der Hedging-Effektivität[40] abhängig. Je stärker (schwächer) der korrelative Zusammenhang zwischen Kassa- und Futureskurs, desto kleiner (größer) wird der Term auf der linken Seite der Ungleichung.[41] Darüber hinaus belegt Satz 2.2.2 das allgemein bekannte Resultat, dass bei einem Cross-Hedge ein Full-Hedge nicht optimal ist. In diesem Modell ist ein Full-Hedge sogar nahezu ausgeschlossen.[42] Die Ursache liegt in der Unvollkommenheit der Risikomärkte, resultierend aus der imperfekten Korrelation der Wiener Prozesse von Kassa- und Terminkurs.

Wird der stochastische Zusammenhang zwischen Kassa- und Terminkurs durch die Kovarianz erfasst, lässt sich eine alternative Darstellung für den Minimum-Varianz-Hedge erzielen:

Korollar 2.2.1. *Eine äquivalente Darstellung der Hedge-Rate h_1 ist gegeben durch:*[43]

$$h_1 = \frac{\mathbb{C}ov[\frac{ds}{s}, \frac{df}{f}]}{\mathbb{V}ar[\frac{df}{f}]}. \tag{2.20}$$

[40] Die Hedging-Effektivität misst die durch das Hedging realisierbare Varianzverringerung bezogen auf die Varianz ohne Hedging. Vgl. Broll and Wahl (2012), S. 52.

[41] Vgl. Albrecht and Maurer (2008), S. 580.

[42] Aufgrund der Parameterkonstellation und der Zeitabhängigkeit kann ohne konkrete Spezifikation der Kursprozesse von s und f keine Aussage über den Umfang der Hedge-Komponente h_1 getroffen werden. Für die Spezifikation der MR-Kursprozesse ist insbesondere die Kenntnis der Parameter$\kappa_s, \theta_s, \sigma_s, \kappa_f, \theta_f, \sigma_f$ sowie des Korrelationskoeffizienten ρ_{sf} erforderlich.

[43] $\mathbb{C}ov\,[\cdot]$ kennzeichnet den Kovarianzoperator.

Beweis: Für die rechte Seite von (2.20) gilt

$$\frac{\mathbb{C}ov[\frac{ds}{s}, \frac{df}{f}]}{\mathbb{V}ar[\frac{df}{f}]} = \frac{\frac{\sigma_s \sigma_f}{\sqrt{sf}}\rho_{sf}}{\frac{\sigma_f^2}{f}} = \frac{\sigma_s \sqrt{f}}{\sigma_f \sqrt{s}}\rho_{sf},$$

wodurch unmittelbar die Behauptung folgt. □

Der Minimum-Varianz-Hedge ist demzufolge von der Kopplung zwischen den relativen Änderungen von Futures- und Kassakurs sowie den relativen Änderungen des Futureskurses abhängig. Der Umfang des puren Hedge-Terms ist dabei umso kleiner (größer), je größer (kleiner) die Varianz der Futuresrendite und je kleiner (größer) die Kovarianz zwischen Kassa- und Futuresrendite ist.

Der zweite Term der optimalen Hedge-Ratio h_2 erfasst das spekulative Motiv des Investors. Das Vorzeichen von h_2 wird eindeutig von dem Vorzeichen des Zählers bestimmt, weil der Nenner aufgrund der positiven intertemporalen relativen Risikoaversion R und dem positiven Diffusionsparameter σ_f^2 immer positiv ist. Das Vorzeichen des Zählers hängt von der Situation auf dem Terminmarkt ab. Diesbezüglich ist zwischen drei möglichen Konstellationen zu unterscheiden:[44]

Definition 3. *Der Terminmarkt wird als MR-unverzerrt bezeichnet, wenn der Terminpreis dem gegebenen Gleichgewichtsniveau θ_f entspricht, also $f = \theta_f$ gilt. Der Terminkurs enthält keine Risikoprämie.*

Liegt der Terminpreis unterhalb des Gleichgewichtsniveaus, $f < \theta_f$, so ist der Terminmarkt durch MR-Backwardation gekennzeichnet. Es existiert eine positive Risikoprämie.

Liegt der Terminpreis oberhalb des Gleichgewichtsniveaus, $f > \theta_f$, so ist der Terminmarkt durch MR-Contango gekennzeichnet. Es existiert eine negative Risikoprämie.

[44] Den originären Begriff Backwardation verwendete Keynes (1930) um Situationen zu beschreiben, in denen der aktuelle Terminkurs unter dem aktuellen Kassakurs liegt. Als Normal Backwardation wurden Situationen bezeichnet, in denen der aktuelle Terminkurs geringer ist als der erwartete Terminkurs bei Lieferung. Seit den 60er Jahren wird hierfür abkürzend der Begriff Backwardation verwendet. Vgl. Duffie (1989), S. 101 f.

Der Driftterm $\kappa_f(\theta_f - f)$ gibt die erwartete absolute Risikoprämie des Futures pro infinitesimaler Zeitänderung (dt) an.[45] Sie ist positiv (negativ), wenn sich der Terminmarkt in einer MR-Backwardation (MR-Contango) Situation befindet, d. h. der Kurs des Futures f liegt unterhalb (oberhalb) des Gleichgewichtsniveaus θ_f. Dadurch weist h_2 insgesamt ein negatives (positives) Vorzeichen auf. Ein risikoaverser Investor berücksichtigt die Situation auf dem Terminmarkt und weicht von dem Minimum-Varianz-Hedge h_1 ab. Bei positiver (negativer) Risikoprämie wählt der Investor, ohne Berücksichtigung der Hedge-Terme h_3 und h_4, eine Hedge-Rate unterhalb (oberhalb) des puren Hedge-Terms. Die Fachtermini der daraus resultierenden Hedging-Strategien werden folgendermaßen spezifiziert:

Definition 4. *Sichert ein Investor die Entwicklung seines Vermögens risikominimierend ab, $h = h_1$, wird dies als reines Hedging bezeichnet. Wenn der Anteil verkaufter Futures am Vermögen geringer ist als die reine Hedge-Rate ($h < h_1$), entscheidet sich der Investor für Underhedging. Overhedging liegt vor, wenn der Investor eine höhere Hedge-Rate wählt als beim reinen Hedging ($h > h_1$).*

Die Begriffe Underhedging und Overhedging werden immer hinsichtlich einer Benchmark, in diesem Fall das reine Hedging, determiniert. In der Literatur wird häufig das Full-Hedging[46] als Benchmark gewählt.[47] Underhedging (Overhedging) bezeichnet dann die Absicherung des Basisobjekts mit einem geringeren (höheren) Betrag als eins. In dem hier betrachteten Hedging-Modell wird aufgrund der imperfekten Korrelation die varianzminimale Absicherung durch das reine Hedging erzielt, so dass die Hedge-Komponente h_1 als Basis dient.

Der konkrete Umfang der spekulativen Hedge-Rate h_2 ist proportional zu dem Verhältnis von dem Driftterm $\kappa_f(\theta_f - f)$ zu dem Diffusionsparameter σ_f^2. Diese Relation ist von fundamentaler Bedeutung und resultiert aus dem unterstellten Wurzeldiffusionsprozess (2.2). Bei der Charakterisierung hilft der nachfolgende Begriff:

[45] Es gilt $\mathbb{E}[df] = \kappa_f(\theta_f - f)\,dt$.
[46] Eine alternative Bezeichnung lautet Vollabsicherung bzw. perfekter Hedge.
[47] Vgl. Duffie (1989), S. 226.

Definition 5. *Bezeichnet x_t, $t \geq 0$, einen Wurzel-Diffusionsprozess mit den Parametern $\kappa_x \geq 0, \theta_x, \sigma_x > 0$ und einem Wiener Prozess z_x, der die stochastische Differentialgleichung*

$$dx = \kappa_x(\theta_x - x)dt + \sigma_x \sqrt{x} dz_x \tag{2.21}$$

löst, dann gibt der Quotient

$$MRSR_x = \frac{\mathbb{E}[\frac{dx}{x}]/dt}{\mathbb{V}ar[\frac{dx}{x}]/dt} = \frac{\kappa_x(\theta_x - x)}{\sigma_x^2} \tag{2.22}$$

das durch den stochastischen Prozess x induzierte Verhältnis zwischen dem erwarteten relativen Zuwachs und der Varianz des Zuwachses (jeweils pro marginaler Zeitänderung dt) an und wird als Mean-Reversion-Sharpe-Ratio (kurz: MR-Sharpe-Ratio) des Kursprozesses x bezeichnet.

Der Begriff Mean-Reversion-Sharpe-Ratio verdeutlicht, dass durch die Gleichung (2.22) implizit eine Kennzahl gegeben ist, welche das Verhältnis des erwarteten relativen Zuwachses pro Risikoeinheit angibt.[48] Als Maß für das Risiko wird in diesem Fall die Varianz der relativen Zuwächse gewählt. Erst die Einbeziehung der Varianz ermöglicht in diesem Kontext eine sinnvolle Gegenüberstellung von Ertrag und Risiko eines Wertpapiers, welches dem Prozess (2.21) folgt, weil in diesem Fall beide Größen in Relation zum Zeitinkrement dt gegeben sind. Ausschlaggebend hierfür ist der Wiener-Prozess z_x. Das Inkrement dz_x über ein Zeitintervall dt ist definiert als $dz_x = \varepsilon_t \sqrt{dt}$, wobei ε_t eine normalverteilte Zufallsvariable mit einem Erwartungswert von null und einer Standardabweichung von eins ist. Demzufolge ist die Varianz proportional zum Zeitinkrement dt.[49] Diese Zeitproportionalität überträgt sich auf den Wurzel-Diffusionsprozess (2.21), so dass die MR-Sharpe-Ratio (2.22) tatsächlich eine geeignete Kennzahl darstellt, die einen zeitkongruenten Vergleich zwischen er-

[48] Die nach William F. Sharpe benannte Sharpe-Ratio ist ein risikoadjustiertes Performance-Maß, welches die Überrendite eines Wertpapiers je Einheit Risiko misst. Aus diesem Grund wurde die Kennzahl von Sharpe (1966) ursprünglich als Reward-to-Variability-Ratio bezeichnet. Vgl. auch Sharpe (1975) und Sharpe (1994).

[49] Vgl. Dixit and Pindyck (1994), S. 67.

wartetem Ertrag und Risiko zulässt. Darüber hinaus ist zu beachten, dass es sich aufgrund der Dynamik von x um eine zeitabhängige Kennzahl handelt.

Die MR-Sharpe-Ratio des Futures $MRSR_f$ ist Bestandteil der zweiten Hedge-Komponente und verdeutlicht das Spekulationsmotiv des Investors: Das Risiko wirkt negativ, die Risikoprämie positiv auf den Umfang dieser Hedge-Ratio.[50] Der konkrete Umfang von h_2 ist proportional zu der MR-Sharpe-Ratio $MRSR_f$. Der Proportionalitätsfaktor ist die intertemporale relative Risikotoleranz des Investors, d. h. der Kehrwert der intertemporalen relativen Risikoaversion R.

Ein ähnliches Resultat ist bereits aus der einfachen Portefeuille-Theorie bekannt. Besitzt ein Anleger die Möglichkeit sein Vermögen in ein risikoloses und ein risikobehaftetes Wertpapier aufzuteilen, so investiert er nur dann in das riskante Wertpapier, sofern dieses eine positive erwartete Risikoprämie aufweist.[51] Die Risikoprämie ist in diesem Kontext definiert als Differenz zwischen der erwarteten Rendite des riskanten Wertpapiers und dem risikolosen Zinssatz.[52] Die Nachfrage des Investors nach dem risikobehafteten Wertpapier ist unter der Annahme normalverteilter Renditen proportional zum Verhältnis von erwarteter Risikoprämie zum Risiko, gemessen anhand der Varianz der Wertpapierrendite. Der Proportionalitätsfaktor ist in diesem Fall die globale relative Risikotoleranz des Investors.[53,54] Beträgt diese null, d. h. weist der In-

[50] Die MR-Sharpe-Ratio des Futures ist in diesem Modell auch für erwartete negative relative Zuwächse von f aussagekräftig. In dieser Situation befindet sich der Terminmarkt in einer MR-Contango Situation, d. h. der Kurs des Futures f liegt oberhalb des Gleichgewichtsniveaus θ_f. Isoliert betrachtet nimmt der Investor durch die spekulative Hedge-Rate h_2 eine kurze Position in Terminkontrakten ein. Insofern stellt eine negative MR-Sharpe-Ratio in Verbindung mit einer Short-Position ein Komplement zu einer positiven MR-Sharpe-Ratio in Verbindung mit einer Long-Position dar. Die klassische „Sharpe-Ratio“ ist dagegen im negativen Bereich nicht aussagekräftig, da dann ein höheres Risiko zu einer besseren (weniger negativen) Sharpe-Ratio führt.

[51] Vgl. Gollier (2001), S. 54.

[52] Vgl. Arrow (1971), S. 99 ff. oder Eeckhoudt et al. (2005), S. 66 f.

[53] Huang and Litzenberger (1988), S. 100 f.

[54] Bezeichnet r_f den Zinssatz für die risikolose Anlage und $\tilde{r}_1$ die stochastische Rendite des risikobehafteten Wertpapiers und W_0 das Anfangvermögen des Investors. Der Anleger investiert einen wertmäßigen Anteil von x_1 in das risikobehaftete Wertpapier und $1 - x_1$ in das risikolose Wertpapier, so dass das stochastische Endvermögen durch $\tilde{W} = W_0(1 + r_f + x_1(\tilde{r}_1 - r_f))$ gegeben ist. Für das Optimierungsproblem $\max_{x_1} \mathbb{E}[u(\tilde{W})]$ ist die Bedingung $\mathbb{E}[u'(\tilde{W})(\tilde{r}_1 - r_f)] = 0$ aufgrund der Konkavität von u sowohl notwendig als auch hinreichend. Mit Hilfe des Kovarianzoperators kann dieser Ausdruck alternativ geschrieben werden als $\mathbb{E}[u'(\tilde{W})]\mathbb{E}[\tilde{r}_1 - r_f] = -\mathbb{C}ov[u'(\tilde{W}), \tilde{r}_1]$. $\tilde{W}$ und $\tilde{r}_1$ sind bivariat normalverteilt, so dass mit dem Lemma von Stein $\mathbb{E}[u'(\tilde{W})]\mathbb{E}[\tilde{r}_1 - r_f] = -\mathbb{E}[u''(\tilde{W})]\mathbb{C}ov[\tilde{W}, \tilde{r}_1]$ folgt. Durch Einsetzen von $\tilde{W}$ in den Kovarianzterm und Umformen nach dem wertmäßigen Anteil x_1 ergibt

vestor eine unendlich hohe Risikoaversion auf, legt der Investor das gesamte Vermögen risikolos an.

Im Rahmen einer Portefeuille-Planung mit zwei riskanten Titeln untersucht Ross (1981) die Aufteilung des Vermögens eines Investors in Abhängigkeit seiner Risikoaversion. Er kommt zu dem Ergebnis, dass strenger risikoaverse Investoren Portefeuilles mit geringerem Risiko präferieren.[55] Für den Grenzfall der unendlich hohen Risikoaversion bedeutet dies, dass das gesamte Vermögen in das Minimum-Varianz-Portfeuille investiert wird.

Eine weitere Verwandtschaft des hier vorgestellten Hedging-Modells zur Portefeuille-Theorie ergibt sich hinsichtlich der Markowitz-Effizienz. Ein bekanntes Resultat aus der Portefeuille-Theorie ist, dass Investoren nur risiko-effiziente Portefeuilles halten. Ein Portefeuille heißt risiko-effizient, wenn es kein anderes Portefeuille gibt, das bei mindestens gleichem Ertrag ein geringeres Risiko oder höchstens gleichem Risiko einen höheren Ertrag aufweist.[56] Dieser Terminus motiviert eine dynamisch effiziente Hedge-Rate wie folgt zu definieren:

Definition 6. *Eine Hedge-Rate heißt momentan-effizient, wenn es keine andere Hedge-Rate gibt, die bei mindestens gleicher erwarteter Vermögensentwicklung ein geringeres Risiko oder höchstens gleichem Risiko der Vermögensentwicklung einen höheren Ertrag aufweist.*

Die Bezeichnung momentan akzentuiert in diesem Zusammenhang die Dynamik der effizienten Hedge-Rate. Diese Dynamik wird durch die Kursprozesse s (2.1) und f (2.2) induziert und überträgt sich auf die erwartete Vermögensentwicklung (2.5) bzw. Varianz der Vermögensentwicklung (2.6). Daraus resultiert die Zeitabhängigkeit der effizienten Hedge-Raten.

Um die konkrete Spezifikation der momentan-effizienten Hedge-Rate für das vorliegende Modell zu bestimmen, wird diejenige Hedge-Rate gesucht, die die Varianz der

sich $x_1 = \frac{\mathbb{E}[\tilde{r}_1 - r_f]}{R_g \mathbb{V}ar[\tilde{r}_1]}$, wobei $R_g = -\frac{\mathbb{E}[u''(\tilde{W})]}{\mathbb{E}[u'(\tilde{W})]} W_0$ die globale relative Risikoaversion des Investors bezeichnet.

[55] Vgl. Ross (1981), S. 631 ff.

[56] Vgl. Markowitz (1991), S. 22 und Ingersoll (1987), S. 87.

Vermögenszuwächse für ein gegebenes Niveau erwarteter Vermögenszuwächse minimiert. Formal lässt sich das Entscheidungsproblem durch das Optimierungsproblem

$$\min \mathbb{V}ar[d\Phi] \; u.d.N. \; \mathbb{E}[d\Phi] = \overline{EW}dt \tag{2.23}$$

darstellen, wobei $d\Phi$ die Vermögensentwicklung aus Gleichung (2.4) bezeichnet. Zur Lösung dieses Optimierungsproblems bietet sich das Lagrangeverfahren an.[57]

Satz 2.2.3. *In dem hier vorgestellten Hedging-Modell ist eine momentan-effiziente Hedge-Rate gegeben durch*

$$h_{eff} = \frac{\sigma_s \sqrt{f}}{\sigma_f \sqrt{s}} \rho_{sf} - \frac{\lambda}{2} \frac{\kappa_f(\theta_f - f)}{\sigma_f^2}, \tag{2.24}$$

d. h. h_{eff} *löst das Optimierungsproblem* (2.23). *Der Lagrange-Multiplikator* λ *misst dabei die Sensitivität des Varianzminimums der Vermögensentwicklung hinsichtlich Änderungen im Niveau der Restriktion.*

Beweis: Es wird diejenige Hedge-Rate gesucht, die die Varianz der Vermögenszuwächse für ein gegebenes Niveau erwarteter Vermögenszuwächse minimiert. Mit dem Lagrange-Multiplikator λ und der Vermögensentwicklung $d\Phi$ aus Gleichung (2.4) lautet die über h und λ zu minimierende Lagrange-Funktion:

$$L(h, \lambda) = \mathbb{V}ar[d\Phi] - \lambda\Big(\mathbb{E}[d\Phi] - \overline{EW}dt\Big). \tag{2.25}$$

Durch Einsetzen von $\mathbb{V}ar[d\Phi] = \Phi^2 \sigma_\Phi^2 dt$ und $\mathbb{E}[d\Phi] = -\kappa_\Phi(\theta_\Phi - \Phi)dt$ in die Lagrange-Funktion (2.25) ergibt sich für die Bedingung erster Ordnung bzgl. h:

$$\begin{aligned} \frac{\partial L}{\partial h} &= 0 \\ \Leftrightarrow \quad 0 &= -2\frac{\sigma_s \sigma_f}{\sqrt{sf}} \rho_{sf} + 2h\frac{\sigma_f}{f} - \lambda\Big(-\frac{\kappa_f(\theta_f - f)}{f}\Big). \end{aligned}$$

[57] Das Verfahren eignet sich für Optimierungsprobleme, bei denen Nebenbedingungen in Form von Gleichungen einzuhalten sind. Vgl. Chiang (1984), S. 369 ff.

Auflösen nach der Hedge-Rate h liefert die Darstellung:

$$h = \frac{\sigma_s \sqrt{f}}{\sigma_f \sqrt{s}} \rho_{sf} - \frac{\lambda}{2} \frac{\kappa_f(\theta_f - f)}{\sigma_f^2}.$$

Die hinreichende Bedingung für die Existenz eines Extrempunktes an der Stelle h wird mit Hilfe des Vorzeichenwechselkriteriums nachgewiesen. Sei $\varepsilon > 0$, dann gilt für den Funktionswert von $\frac{\partial L}{\partial h}$ an der Stelle $h - \varepsilon$:

$$\begin{aligned} \left.\frac{\partial L}{\partial h}\right|_{h-\varepsilon} &= -2\frac{\sigma_s \sigma_f}{\sqrt{sf}} \rho_{sf} + 2h\frac{\sigma_f}{f} - 2\varepsilon\frac{\sigma_f}{f} - \lambda\left(-\frac{\kappa_f(\theta_f - f)}{f}\right) \\ \Leftrightarrow \quad \left.\frac{\partial L}{\partial h}\right|_{h-\varepsilon} &= -2\varepsilon\frac{\sigma_f^2}{f} < 0. \end{aligned}$$

Analog ergibt sich für den Funktionswert von $\frac{\partial L}{\partial h}$ an der Stelle $h + \varepsilon$:

$$\left.\frac{\partial L}{\partial h}\right|_{h+\varepsilon} = 2\varepsilon\frac{\sigma_f}{f} > 0.$$

Der Nachweis des Vorzeichenwechsels für die Funktion $\frac{\partial L}{\partial h}$ an der Stelle h von minus nach plus liefert die hinreichende Bedingung und kennzeichnet den Extrempunkt als Minimum.

Die Interpretation des Langrange-Multiplikators ergibt sich über folgende allgemeine Darstellung der partiellen Ableitung der Langrange-Funktion nach der Hedge-Rate h:

$$\begin{aligned} \frac{\partial L(h,\lambda)}{\partial h} &= \frac{\partial \mathbb{V}ar[d\Phi]}{\partial h} - \lambda\frac{\partial \mathbb{E}[d\Phi]}{\partial h} = 0 \\ \Leftrightarrow \qquad \lambda &= \frac{\frac{\partial \mathbb{V}ar[d\Phi]}{\partial h}}{\frac{\partial \mathbb{E}[d\Phi]}{\partial h}}. \end{aligned}$$

D. h. λ gibt die Sensitivität der Varianz der Vermögensentwicklung bzgl. Änderungen im Niveau der Restriktion an. □

Satz 2.2.3 zeigt, dass eine effiziente Hedge-Rate h_{eff} aus der Summe des varianzminimalen Hedge-Terms h_1 und einer mit der intertemporalen relativen Risikotoleranz gewichteten MR-Sharpe-Ratio des Futures besteht.

Die effiziente Hedge-Rate ist demnach implizit in der optimalen Hedge-Rate (2.11) enthalten. Diese Aussage gilt unabhängig von den individuellen Präferenzen des Investors. Lediglich der Proportionalitätsfaktor in der spekulativen Hedge-Rate h_2 ist von der individuellen intertemporalen relativen Risikoaversion R abhängig. Im Gegensatz zur klassischen Portefeuille-Theorie, wo ein optimales Portefeuille auch immer ein effizientes Portefeuille darstellt,[58] ist in dem hier betrachteten Hedging-Modell die optimale Hedge-Rate nicht zwingend eine effiziente Hedge-Rate. Formal ist dies leicht durch die Existenz der beiden Hedge-Terme h_3 und h_4 in der optimalen Hedge-Rate (2.11) zu begründen. Die Funktion dieser beiden Terme wird im Folgenden näher analysiert.

Ein risikoaverser Investor berücksichtigt, dass sowohl der Kassakurs s als auch der Futureskurs f unsicher sind. Diese Unsicherheit beeinflusst explizit das Vermögen Φ, welches die Basis für den zukünftigen Konsum bildet.[59] Für den Investor nachteilige Entwicklungen bzgl. des Wertpapierkurses s bzw. Futureskurses f verringern das Vermögen und folgerichtig den zukünftigen Konsum. Um eine Glättung des Konsumprozesses über den Planungshorizont $[0, T]$ zu erreichen, weicht der Investor von der effizienten Hedge-Rate ab. Der folgende Satz zeigt, dass die preiselastischen Hedge-Terme $h_3 = \frac{f}{\frac{V_{\Phi\Phi\Phi}}{V_{\Phi f}}} = \frac{R_f}{R}$ und $h_4 = \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{s}{\frac{V_{\Phi\Phi\Phi}}{V_{\Phi s}}} = \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{R_s}{R}$ essentiell sind, um das Konsumrisiko hinsichtlich der intertemporalen Konsumstrategie zu minimieren.

[58] Vgl. Ingersoll (1987), S. 87.
[59] Vgl. Optimalitätsbedingung (2.10).

Satz 2.2.4. *Die optimale Hegde-Rate h^* eines risikoaversen Investors ist o. B. d. A. nicht momentan-effizient. Der Grad der Abweichung von der momentan-effizienten Hedge-Rate h_{eff} entspricht der Summe der preiselastischen Hedge-Terme h_3 und h_4, die implizit eine Minimierung des Risikos der Konsumentwicklung für ein gegebenes Niveau erwarteter Vermögenszuwächse bewirken.*

Beweis: Für den Nachweis wird diejenige Hedge-Rate gesucht, die das Konsumrisiko, gemessen anhand der Varianz der Zuwächse des optimalen Konsumpfades, für ein gegebenes Niveau erwarteter Vermögenszuwächse minimiert. Mit dem Lagrange-Multiplikator λ und der Vermögensentwicklung $d\Phi$ aus Gleichung (2.4) lautet die über h und λ zu maximierende Lagrange-Funktion:

$$L(h, \lambda) = \frac{1}{2}\mathbb{V}ar\Big[dC^*\Big] - \lambda\Big(\mathbb{E}[d\Phi] - (-\kappa_\Phi(\Theta_\Phi - \Phi)dt)\Big). \tag{2.26}$$

Für die optimale Konsumentwicklung $dC^*(\Phi, s, f, t)$ ergibt sich durch die Anwendung des Lemmas von Itô:[60]

$$\begin{aligned} dC^* &= C_t dt + C_\Phi d\Phi + C_s ds + C_f df \\ &+ \frac{1}{2}C_{\Phi\Phi}(d\Phi)^2 + \frac{1}{2}C_{ss}(ds)^2 + \frac{1}{2}C_{ff}(df)^2 \\ &+ C_{\Phi s}\Phi s\, d\Phi ds + C_{\Phi f}\Phi f\, d\Phi df + C_{sf} sf\, dsdf. \end{aligned} \tag{2.27}$$

[60] Die folgenden partiellen Ableitungen des Konsums C auf der rechten Seite von (2.27) beziehen sich auf den optimalen Konsumprozess. Aus Gründen der Übersichtlichkeit wird jedoch auf eine gesonderte *-Kennzeichnung verzichtet.

Für die Varianz der optimalen Konsumentwicklung gilt demnach:[61]

$$\begin{aligned}
\mathbb{V}ar[dC^*] &= \mathbb{V}ar[C_\Phi\, d\Phi + C_s\, ds + C_f\, df] \\
&= C_\Phi^2\, \mathbb{V}ar[d\Phi] + C_s^2\, \mathbb{V}ar[ds] + C_f^2\, \mathbb{V}ar[df] \\
&+ 2\Big(C_\Phi\, C_s\, \mathbb{C}ov[d\Phi, ds] + C_\Phi\, C_f\, \mathbb{C}ov[d\Phi, df] + C_s\, C_f\, \mathbb{C}ov[ds, df]\Big) \\
&= \Big(C_\Phi^2 \Phi^2 \sigma_\Phi^2 + C_s^2 \sigma_s^2 s + C_f^2 \sigma_f^2 f \\
&\quad + 2\Big(C_\Phi C_s \Phi\Big(\sigma_s^2 - h\sigma_s\sigma_f \frac{\sqrt{s}}{\sqrt{f}}\rho_{sf}\Big) \\
&\quad + C_\Phi C_f \Phi\Big(\sigma_s\sigma_f \frac{\sqrt{f}}{\sqrt{s}}\rho_{sf} - h\sigma_f^2\Big) \\
&\quad + C_s C_f \sigma_s \sigma_f \sqrt{sf}\rho_{sf}\Big)\Big)\, dt.
\end{aligned}$$

Nach Einsetzen von $\mathbb{V}ar[dC^*]$ und $\mathbb{E}[d\Phi]$ in die Lagrange-Funktion (2.26) ergibt sich für die Bedingung erster Ordnung bzgl. h:

$$\begin{aligned}
\frac{\partial L}{\partial h} &= 0 \\
\Leftrightarrow \quad 0 &= 2C_\Phi^2\, \Phi^2 \Big(\frac{hs\sigma_f^2 - \sqrt{sf}\sigma_s\sigma_f\rho_{sf}}{sf}\Big) \\
&\quad - 2\Big(C_\Phi C_s \Phi \sigma_s \sigma_f \tfrac{\sqrt{s}}{\sqrt{f}}\rho_{sf} - C_\Phi C_f \Phi \sigma_f^2\Big) - \lambda\Phi\tfrac{\kappa_f(\theta_f - f)}{f}.
\end{aligned}$$

Auflösen nach h liefert die folgende Darstellung der Hedge-Rate:

$$h = \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf} + \frac{\lambda}{2C_\Phi^2\Phi}\frac{\kappa_f(\theta_f - f)}{\sigma_f^2} + \frac{f}{\frac{C_\Phi\Phi}{C_f}} + \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{s}{\frac{C_\Phi\Phi}{C_s}}. \tag{2.28}$$

[61] Da bis auf $C_\Phi\, d\Phi, C_s\, ds$ und $C_f\, df$ alle übrigen Terme in Gleichung (2.27) Terme der Ordnung $O(dt)$ sind, werden diese bei der Varianz- bzw. Kovarianzbildung elimiert, d. h. die Varianz der Entwicklung des optimalen Konsumprozesses ist unmittelbar gegeben durch $\mathbb{V}ar\Big[dC^*\Big] = \mathbb{V}ar\Big[C_\Phi\, d\Phi + C_s\, ds + C_f\, df\Big]$.

Analog zum Beweis von Satz 2.2.3 wird anhand des Vorzeichenwechselkriteriums überprüft, ob es sich bei der Extremstelle um ein Minimum oder Maximum handelt. Mit $\varepsilon > 0$ ergibt sich als Funktionswert von $\dfrac{\partial L}{\partial h}$ an der Stelle $h - \varepsilon$:

$$\begin{aligned}\left.\frac{\partial L}{\partial h}\right|_{h-\varepsilon} &= 2C_\Phi^2\,\Phi^2\left(\frac{hs\sigma_f^2 - \sqrt{sf}\sigma_s\sigma_f\rho_{sf}}{sf}\right) - 2\varepsilon C_\Phi^2\,\Phi^2\,s\sigma_f^2 \\ &- 2\Big(C_\Phi C_s\Phi\sigma_s\sigma_f\frac{\sqrt{s}}{\sqrt{f}}\rho_{sf} - C_\Phi C_f\Phi\sigma_f^2\Big) - \lambda\Phi\frac{\kappa_f(\theta_f - f)}{f} \\ &= -2\varepsilon C_\Phi^2\,\Phi^2\,s\sigma_f^2 < 0.\end{aligned}$$

Für den Funktionswert von $\dfrac{\partial L}{\partial h}$ an der Stelle $h + \varepsilon$ ergibt sich analog:

$$\left.\frac{\partial L}{\partial h}\right|_{h+\varepsilon} = 2\varepsilon C_\Phi^2\,\Phi^2\,s\sigma_f^2 > 0.$$

Für die Funktion $\frac{\partial L}{\partial h}$ liegt an der Stelle h ein Vorzeichenwechsel von minus nach plus vor, d. h. die hinreichende Bedingung für die Existenz eines Minimums ist erfüllt.

Im zweiten Teil des Beweises ist zunächst die Identität der dritten und vierten Hedge-Komponente aus den Gleichungen (2.11) und (2.28) nachzuweisen. Aus der Optimalitätsbedingung (2.10) ergeben sich durch Anwendung der Kettenregel die partiellen Ableitungen $V_{\Phi\Phi} = u_{CC}C_\Phi, V_{\Phi s} = u_{CC}C_s$ und $V_{\Phi f} = u_{CC}C_f$. Mit den daraus resultierenden Beziehungen

$$\frac{V_{\Phi\Phi}}{V_{\Phi s}} = \frac{C_\Phi}{C_s} \tag{2.29}$$

sowie

$$\frac{V_{\Phi\Phi}}{V_{\Phi f}} = \frac{C_\Phi}{C_f} \tag{2.30}$$

ergibt sich unmittelbar die Identität der dritten und vierten Hedge-Komponenten der Gleichungen (2.11) und (2.28).

Aus Satz 2.2.3 ist bereits bekannt, dass die Summe der Hedge-Komponenten h_1 und h_2 von Gleichung (2.28), d. h. die Summe aus dem varianzminimalen Hedge-Term und einer mit der intertemporalen relativen Risikotoleranz gewichteten MR-Sharpe-Ratio

des Futures, einer momentan-effizienten Hedge-Rate entspricht. Während diese beiden Hedge-Komponenten also die Varianz des Vermögens für ein gegebenes Niveau erwarteter Vermögensrendite minimieren, wird durch die Hinzunahme der Hedge-Komponenten h_3 und h_4 implizit die Minimierung der Varianz des Konsumprozesses erreicht. □

Das Ziel des Investors ist eine Verstetigung seines Konsumprofils. Dies erreicht der Investor, indem er sich gegen Risiken absichert, die durch die Zustandsvariablen s und f induziert werden. Satz 2.2.4 verdeutlicht die Funktionsweise der zustandsabhängigen Hedge-Terme hinsichtlich der Minimierung des Konsumrisikos. Aus Gleichung (2.11) wird deutlich, dass der Umfang der Hedge-Komponenten h_3 und h_4 insbesondere von der intertemporalen relativen Risikoaversion R sowie den intertemporalen Preiselastizitäten R_s und R_f abhängt. Während R immer größer null ist, sind die Vorzeichen von R_s und R_f zustandsabhängig, so dass universal keine Aussage getroffen werden kann, ob die Terme h_3 bzw. h_4 die Hedge-Rate ceteris paribus erhöhen oder verringern.

Der Zähler der dritten Hegde-Komponente $R_f = -\frac{V_{\Phi f} f}{V_\Phi}$ gibt die Preiselastizität des Futures f an. Die partielle Ableitung $V_{\Phi f} = \partial V_\Phi / \partial f$ beschreibt in diesem Kontext den Effekt einer Änderung des Futures-Kurses f auf den Grenznutzen des Vermögens. Für $V_{\Phi f} > 0$ (< 0) gilt, dass ein Anstieg (Rückgang) des Futures-Kurses als eine Verminderung (Zunahme) des Vermögens Φ aufgefasst werden kann, welches ebenfalls zu einem höheren (geringeren) Grenznutzen führt. Welches Vorzeichen $V_{\Phi f}$ annimmt, hängt von der Situation auf dem Terminmarkt ab.

Satz 2.2.5. *Ist der Terminmarkt durch MR-Backwardation (MR-Contango) gekennzeichnet, ergibt sich bei gegebener intertemporaler relativer Risikoaversion $R > 0$ ein positiver (negativer) Hedge-Term h_3.*

Beweis: Durch Einsetzen der partiellen Ableitungen $V_{\Phi\Phi} = u_{CC}C_\Phi, V_{\Phi s} = u_{CC}C_s$ und $V_{\Phi f} = u_{CC}C_f$ in die Gleichung (2.11) ergibt sich folgende Darstellung der optimalen Hedge-Rate:

$$h^* = \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf} + \frac{\kappa_f(\theta_f - f)}{\sigma_f^2\frac{u_{CC}}{u_C}C_\Phi\Phi} + \frac{C_f f}{C_\Phi\Phi} + \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{C_s s}{C_\Phi\Phi}. \tag{2.31}$$

Wird mit $\eta_C = -\frac{u_{CC}}{u_C}C$ die Elastizität des Grenznutzens definiert, so kann nach einigen Äquivalenzumformungen folgende Darstellung für die erwartete relative Risikoprämie des Futures erzielt werden:

$$\frac{\kappa_f(\theta_f - f)}{f} = \frac{\sigma_f^2}{f}\left[\left(\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf} - h^*\right)\frac{C_\Phi\Phi}{C} - \frac{C_f f}{C} - \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{C_s s}{C}\right]\eta_C. \tag{2.32}$$

Durch die Anwendung von Lemma A.1.1 aus Anhang A.1 ergibt sich folgender Zusammenhang:

$$\frac{\kappa_f(\theta_f - f)}{f} = \mathbb{C}ov\left(\frac{dC}{C}, \frac{df}{f}\right)\eta_C. \tag{2.33}$$

Für eine von Neumann-Morgenstern Nutzenfunktion u gilt $u_C > 0$ und $u_{CC} < 0$ und damit für die Elastizität des Grenznutzens $\eta_C > 0$. Demnach weisen der Kovarianzterm auf der rechten Seite der Gleichung und die relative Risikoprämie des Terminmarktes das gleiche Vorzeichen auf. Befindet sich der Terminmarkt in MR-Backwardation (MR-Contango) gilt $f < (>)\ \theta_f$ und damit

$$sign\left(\frac{\kappa_f(\theta_f - f)}{f}\right) = sign\left(\mathbb{C}ov\left(\frac{dC}{C}, \frac{df}{f}\right)\right) > (<)\ 0.$$

Diese Gleichung impliziert für eine MR-Backwardation (MR-Contango) Situation $C_f > (<)\ 0$. Dadurch wird der Ausdruck $V_{\Phi f} = U_{CC}C_f$ negativ (positiv). Die intertemporale Preiselastizität $R_f = -\frac{V_{\Phi f}f}{V_\Phi}$ besitzt wegen $f, V_\Phi > 0$ folglich ein entgegengesetztes Vorzeichen zu $V_{\Phi f}$. Insgesamt gilt demnach, dass die Hedge-Komponente h_3

positive (negative) Werte annimmt, wenn der Terminmarkt durch MR-Backwardation (MR-Contango) gekennzeichnet ist. □

Eine weitere Implikation von Satz 2.2.5 ist, dass die beiden Hedge-Komponenten h_2 und h_3 für dieselbe Situation auf dem Terminmarkt entgegengesetzte Vorzeichen aufweisen. In einer MR-Backwardation (MR-Contango) Situation weist die spekulative Komponente h_2 ein negatives (positives) Vorzeichen auf, h_3 nimmt dagegen positive (negative) Werte an. Während sich beide Hedge-Terme bzgl. des Vorzeichens diametral entgegenstehen, stellen sie jedoch gleichzeitig funktionale Komplemente dar. Liegt der Futureskurs unterhalb (oberhalb) des Gleichgewichtsniveaus θ_f, erwartet der Investor aufgrund des unterliegenden MR-Prozesses eine Rückkehr zum Gleichgewichtsniveau θ_f und damit einen steigenden (fallenden) Terminkurs. D. h. bei Existenz einer positiven (negativen) Risikoprämie auf dem Terminmarkt, verringert (erhöht) ein risikoaverser Investor durch die spekulative Komponente h_2 ceteris paribus seine Hedge-Rate, um von dieser Situation zu profitieren. Tatsächlich gilt die Entwicklung des Futureskurses in Richtung des Gleichgewichtsniveaus jedoch nur im Erwartungswert. Demnach berücksichtigt der Investor das durch die Zustandsvariable f induzierte Risiko, dass sich der Futureskurs zukünftig für ihn potentiell nachteilig entwickelt. In Analogie zu Merton (1973), wird eine für den Investor nachteilige bzw. ungünstige Entwicklung einer Zustandsvariable durch die nachfolgende Definition charakterisiert.

Definition 7. *Die Änderung der Zustandsvariablen $j \in \{s, f\}$ wird für den Investor als nachteilig oder ungünstig bezeichnet, wenn der Grenznutzen des Vermögens ansteigt.*

Motiviert wird diese Definition durch den fundamentalen Zusammenhang zwischen dem Vermögen Φ und dem Grenznutzen des Vermögens V_Φ. Führt die Veränderung einer Zustandsvariablen[62] zu einem höheren Grenznutzen des Vermögens, kann dies in gewisser Hinsicht als eine Verminderung des Vermögens Φ aufgefasst werden, welches ebenfalls zu einem höheren Grenznutzen führt. Für den Futureskurs f kann die Art der

[62] Prinzipiell kann damit sowohl ein Anstieg als auch ein Rückgang vom Kassakurs s oder Futureskurs f gemeint sein.

Veränderung in Abhängigkeit von der Situation auf dem Terminmarkt präzise spezifiziert werden.

Korollar 2.2.2. *Ist der Terminmarkt durch MR-Backwardation (MR-Contango) gekennzeichnet, so ist ein Rückgang (Anstieg) des Futureskurses f eine für den Investor nachteilige Entwicklung.*

Beweis: In dem Beweis von Satz 2.2.5 wurde gezeigt, dass in einer MR-Backwardation (MR-Contango) Situation $C_f > 0$ ($C_f < 0$) gilt. Die partielle Ableitung des Grenznutzens bzgl. des Futures-Kurses $V_{\Phi f} = U_{CC}C_f$ besitzt demnach ein negatives (positives) Vorzeichen. Ein Rückgang (Anstieg) des Futureskurses bewirkt in einer MR-Backwardation (MR-Contango) Situation folglich einen Anstieg des Grenznutzens des Vermögens V_Φ. □

In einer MR-Backwardation (MR-Contango) Situation führt ein Rückgang (Anstieg) des Futureskurses wegen $C_f > 0$ ($C_f < 0$) zu einer Verringerung des Konsums. Gegen dieses, durch den Zustand von f induzierte Risiko, sichert sich der Investor durch die zustandsabhängige Hedge-Komponente h_3 ab. Diese erhöht (verringert) in einer MR-Backwardation (MR-Contango) Situation ceteris paribus den Umfang der optimalen Hedge-Rate h^*. Isoliert betrachtet nimmt der Investor durch die Hedge-Komponente h_3 eine kurze (lange) Position in Terminkontrakten ein. Der Investor sichert sich in Abhängigkeit von der Situation auf dem Terminmarkt gegen fallende (steigende) Futureskurse ab und erreicht dadurch eine Stabilisierung seines Konsumprofils.[63]

Nach Definition 7 sind nachteilige Entwicklungen prinzipiell für beide Zustandsvariablen möglich, d. h. nicht nur für den Futureskurs f, sondern auch für den Kassakurs s. Trotz des unterliegenden MR-Prozesses lässt sich für die Zustandsvariable s jedoch keine Aussage über die konkrete Ausprägung der Entwicklung machen. D. h. es kann nicht in Abhängigkeit des Zustands[64] konstatiert werden, dass ein Anstieg bzw. Rückgang des Kassakurses s ceteris paribus einen Anstieg des Grenznutzens

[63] Vgl. Satz 2.2.4.

[64] In Analogie zu Definition 3 ist damit gemeint, dass sich für den Zustand des Kassakurses s drei qualitative Ausprägungen ergeben: Er kann größer, kleiner oder gleich dem langfristigen Gleichgewichtsniveau θ_s sein.

des Vermögens V_Φ bewirkt. Infolgedessen lässt sich auch das Vorzeichen der Hedge-Komponente h_4 nicht in Abhängigkeit des Zustands spezifizieren. Die Ursache hierfür liegt in der Beschaffenheit des Hedging-Modells, in dem der Investor keine Investitionsentscheidung hinsichtlich seines Vermögens trifft.[65] Der Investor legt das jeweils zu Beginn einer Periode zur Verfügung stehende Kapital vollständig in ein Wertpapier an, welches dem Prozess (2.1) folgt.

Neben der Analyse der Vorzeichen der Hedge-Terme h_3 und h_4 ist bereits bekannt, dass die Terme essentiell sind, um eine Verstetigung des Konsumprofils zu erreichen. Im Folgenden liegt der Fokus auf den Ausprägungen dieser Hedge-Komponenten. Sie hängen maßgeblich von den intertemporalen Preiselastizitäten R_f bzw. R_s ab, sowie von der intertemporalen relativen Risikoaversion R. Eine Interpretation hinsichtlich der Funktionalität der zustandsabhängigen Hedge-Terme lautet folgendermaßen:[66]

Satz 2.2.6. *Für die Zustandsvariable $j \in \{s, f\}$ gibt der Term*

$$-\frac{V_{\Phi j}}{V_{\Phi\Phi}} = -\frac{R_j}{R}\frac{\Phi}{j} \tag{2.34}$$

die erforderliche Vermögenskompensation zur Aufrechterhaltung des gegenwärtigen Grenznutzens des Konsums an.

Beweis: Durch Anwendung des Impliziten Funktionentheorems[67] auf den Grenznutzen des Vermögens $V_\Phi(\Phi, s, f, t)$, hinsichtlich der Zustandsvariablen $j \in \{s, f\}$, ergibt sich der Ausdruck

$$\left.\frac{\partial\Phi}{\partial j}\right|_{V_\Phi} = -\frac{V_{\Phi j}}{V_{\Phi\Phi}} = -\frac{R_j}{R}\frac{\Phi}{j}.$$

[65] Damit ist die Aufteilung des Vermögens auf zwei oder mehr Wertpapiere gemeint. Diese Fragestellung wird im Rahmen dieses Hedging-Modells nicht betrachtet. Aus diesem Grund existiert auch keine Entscheidungsvariable, die ein ähnliches Vorgehen wie in Satz 2.2.5 ermöglichen würde.

[66] Im Rahmen eines intertemporalen Optimierungsmodells zeigt Merton (1973), dass Investoren Markowitz-effiziente Portefeuilles mit weiteren Portfeuilles kombinieren, welche höchstmöglich mit den Zustandsvariablen korreliert sind. Investoren sichern sich dadurch gegen nachteilige Entwicklungen hinsichtlich der Investitionsmöglichkeiten ab. Aus diesem Grund wurden sie von Merton als Hedge-Portefeuilles bezeichnet. Erzielt wurde dieses Resultat unter der Prämisse der Aufrechterhaltung des gegenwärtigen Grenznutzens des Konsums. Dies stellt gleichzeitig den Ausgangspunkt des nachfolgenden Satzes dar.

[67] Vgl. Chiang (1984), S. 210 ff.

Unter Berücksichtigung der Optimalitätsbedingung (2.10) ergibt sich unmittelbar die Behauptung. □

Der Ausdruck (2.34) gibt die erforderliche Vermögenskompensation hinsichtlich der Zustandsvariablen $j \in \{s, f\}$ an, um den gegenwärtigen Grenznutzen des Vermögens aufrechtzuerhalten.[68] In der Volkswirtschaftslehre ist dieser Sachverhalt als Grenzrate der Substitution (GRS) bekannt.[69] Die Grenzrate der Substitution entspricht in diesem Fall dem Verhältnis der partiellen Ableitungen des Grenznutzens des Vermögens V_Φ nach der Zustandsvariablen j und dem Vermögen Φ. Als substanzieller Bestandteil der Hedge-Komponenten h_3 und h_4 bewirkt sie für den Investor eine Absicherung gegenüber nachteiligen Entwicklungen der Zustandsvariablen. Dieser Absicherungseffekt wird im Folgenden für den Futureskurs f hinsichtlich der Vermögensentwicklung (2.3) präzise analysiert.

In einer MR-Backwardation (MR-Contango) Situation gilt $V_{\Phi f} < (>)\ 0$, d. h. ein Rückgang (Anstieg) des Futureskurses führt zu einem höheren Grenznutzen des Vermögens.[70] Wegen $V_{\Phi f} < (>)\ 0$ und $V_{\Phi\Phi} < 0$ ergibt sich in diesem Fall ein negativer (positiver) Wert für die erforderliche Vermögenskompensation $\Phi_f = -\frac{V_{\Phi f}}{V_{\Phi\Phi}}$. Unter Berücksichtigung eines positiven Futureskurses stimmt das Vorzeichen von Φ_f somit mit dem Vorzeichen des Hedge-Volumens $-H_3 = \Phi_f f$ überein. Um die Auswirkungen auf die Vermögensentwicklung zu analysieren, ist zusätzlich der Einfluss die Entwicklung des Futureskurses in Gleichung (2.3) zu beachten. Korrespondierend mit einem Rückgang (Anstieg) des Futureskurses ergibt sich für die relative Änderung df/f ein negativer (positiver) Wert. Dadurch realisiert der Investor einen Vermögenszuwachs. Er ist gegenüber nachteiligen Entwicklungen des Futureskurses abgesichert.

Dieses Ergebnis steht im Einklang mit der Folgerung aus Satz 2.2.5, dass die zweite und dritte Hedge-Komponente für dieselbe Terminmarkt-Situation entgegengesetzte Vorzeichen aufweisen. Während der Investor in einer MR-Backwardation (MR-

[68] Wegen der Optimalitätsbedingung (2.10) gibt Gleichung (2.34) gleichzeitig auch die erforderliche Vermögenskompensation an, um den Grenznutzen des Konsums aufrechtzuerhalten.
[69] Vgl. z. B. Varian (2010), S. 48 ff.
[70] Nach Definition 7 wird hierdurch eine nachteilige Entwicklung des Futureskurses charakterisiert.

Contango) Situation durch die Spekulationsposition h_2 isoliert betrachtet eine lange (kurze) Position in Terminkontrakten einnimmt, ergibt sich durch die Hedge-Rate h_3 bei separater Betrachtung ein Terminverkauf (Terminkauf). Demnach erzielt der Investor bei einem Anstieg (Rückgang) des Terminkurses durch die Spekulationsposition einerseits einen positiven (negativen) Vermögenszuwachs, wohingegen durch die Komponente h_3 eine Minderung (Erhöhung) des Vermögens realisiert wird.

In den vorherigen Ausführungen ist bereits implizit der Zusammenhang zwischen der erforderlichen Vermögenskompensation Φ_j ($j \in \{s, f\}$) und der korrespondierenden Hedge-Komponente bzw. dem korrespondierenden Hedge-Volumen diskutiert worden. Der optimale Hedge-Term h_3, der das Zustandsrisiko bzgl. f minimiert, entspricht dabei betragsmäßig dem Produkt aus der erforderlichen Vermögenskompensation Φ_f und dem Skalierungsfaktor f/Φ. Letzterer erfasst die Relation zwischen den Niveaus des Futureskurses f und des Vermögens Φ. Die Argumentation für die Hedge-Komponente h_4 verläuft zunächst analog. Der Term besteht ebenfalls aus der für die Zustandsvariable s erforderlichen Vermögenskompensation Φ_s multipliziert mit dem Skalierungsfaktor s/Φ. Von entscheidender Bedeutung ist in diesem Zusammenhang jedoch die Multiplikation mit dem Faktor $\frac{\sigma_s \sqrt{f}}{\sigma_f \sqrt{s}} \rho_{sf}$, welcher der puren Hedge-Ratio h_1 entspricht. Der Grund hierfür liegt im indirekten Hedging; die Risikosteuerung der Zustandsvariablen s erfolgt über das Futures-Hedging. Das in diesem Fall entstehende Risiko infolge der imperfekten Korrelation zwischen Kassakurs s und Futureskurs f ist ein typisches Beispiel für ein endogenes, nicht absicherbares Risiko.[71] Insofern wird die Hedge-Komponente h_4 zusätzlich durch den Minimum-Varianz-Hedge determiniert. Dabei lässt sich das Risiko der Zustandsvariablen s umso effektiver steuern, je weniger imperfekt Kassakurs und Futureskurs korreliert sind. Dagegen erfolgt die Absicherung des Zustandsrisikos f durch die Hedge-Rate h_3 direkt durch das Futures-Hedging. Dies impliziert eine perfekte Korrelation. Der daraus resultierende Multiplikator eins wird nicht explizit dargestellt.

[71] Das Basisrisiko entsteht, weil das abzusichernde Kursrisiko nicht dem zu Grunde liegenden Basisinstrument (Underlying) des Futures entspricht. Formal wird das Basisrisiko durch die imperfekte Korrelation zwischen Kassa- und Futureskurs erfasst. Vgl. Saunders (2002), S. 562.

Alternativ kann für die erforderliche Vermögenskompensation Φ_j eine konsumbasierte Interpretation aufgezeigt werden. Werden die partiellen Ableitungen $V_{\Phi j}$ und $V_{\Phi\Phi}$ durch $u_{CC}C_j$ bzw. $u_{CC}C_\Phi$ substituiert, wird der intertemporale Substitutionseffekt durch den Term C_j/C_Φ erfasst. Demnach sind die zustandsabhängigen Hedge-Komponenten h_3 und h_4 derart disponiert, dass der Investor den gegenwärtigen Grenznutzen des Konsums aufrechterhalten kann. Diese Aussage wird im Folgenden anhand des Futures analysiert. Eine aus Sicht des Investors nachteilige Entwicklung des Futureskurses[72] im Umfang von df erhöht das Vermögen des Investors durch die korrespondierende Hedge-Komponente h_3 ceteris paribus um $-H_3 d_f/f = -\frac{C_f}{C_\Phi}df = d\Phi_3$.[73] Diese Vermögenserhöhung führt zu einem Konsumanstieg im Umfang von $C_\Phi d\Phi_3 = -C_f df$. Demzufolge wird die nachteilige Entwicklung des Futureskurses vollständig kompensiert. Als Konsequenz hieraus ergibt sich, dass der Investor trotz einer nachteiligen Entwicklung des Futureskurses den gegenwärtigen Grenznutzen des Konsums ceteris paribus aufrechterhalten kann.[74] Die Argumentation für die Hedge-Rate h_4 erfolgt analog.[75]

Eine weitere Möglichkeit mit dem die Funktionsweise der zustandsabhängigen Hedge-Terme h_3 und h_4 aufgezeigt werden kann ist der Ansatz der Elastizität. Werden Zähler und Nenner der Terme $V_{\Phi f}f/V_{\Phi\Phi}\Phi$ und $V_{\Phi s}s/V_{\Phi\Phi}\Phi$ jeweils mit V_Φ erweitert, ergeben sich für diese Hedge-Komponenten die bereits aus Gleichung (2.16) bekannten Darstellungen $\frac{R_f}{R}$ und $\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{R_s}{R}$. Die Terme R, R_s und R_f sind allesamt Elastizitäten. Die intertemporale relative Risikoaversion R entspricht der Elastizität des Grenznutzens des momentanen Vermögens. Sie beschreibt die Wirkung einer marginalen Erhöhung des Vermögens auf den marginalen Grenznutzens des

[72] Im Fall einer MR-Backwardation (MR-Contango) Situation ist damit ein Rückgang (Anstieg) des Futureskurses gemeint. Vgl. Definition 7.

[73] $d\Phi_3$ kennzeichnet die Vermögensänderung, die aus der Änderung des Futureskurses df unter Berücksichtigung des Hedge-Volumens H_3 resultiert.

[74] Die bei einer Änderung des Futureskurses auftretenden Interdependenzen hinsichtlich der anderen Hedge-Komponenten werden an dieser Stelle nicht berücksichtigt.

[75] Zusätzlich ist zu beachten, dass die Risikosteuerung der Zustandsvariablen s über das Futures-Hedging erfolgt. Insofern ist der Minimum-Varianz-Hedge h_1 elementarer Bestandteil der Hedge-Komponente h_4. Die Absicherung gegenüber dem Zustandsrisiko s ist umso effektiver, je höher der Grad der Korrelation zwischen Kassakurs und Futureskurs ist.

Vermögens. Dagegen sind die Preiselastizitäten R_s und R_f ein Maß für die Sensitivität der relativen Veränderung des Grenznutzens des Vermögens aufgrund relativer Änderungen im Kassakurs s bzw. Futureskurs f. Insofern haben die Terme R_s/R und R_f/R in den Hedge-Komponenten h_3 und h_4 die Funktion, unterschiedliche Elastizitäten in einem zweckentsprechenden Verhältnis auszugleichen. Ist die Preiselastizität R_j, $j \in \{s, f\}$, größer als (gleich, kleiner als) die intertemporale relative Risikoaversion R, so bewirkt eine relative Änderung der Zustandsvariablen j eine betragsmäßig größere (gleiche, geringere) relative Änderung des Grenznutzens des Vermögens V_Φ als eine relative Änderung von R, die sich aufgrund einer relativen Änderung des Vermögens Φ ergibt. Der Term R_j/R stellt somit einen dynamischen Proportionalitätsfaktor hinsichtlich der Sensitivität des Grenznutzens des Vermögens dar, der die Sensitivität der Preiselastizität der Zustandsvariablen j in Relation zur intertemporalen relativen Risikoaversion R misst.

Durch Substitution der partiellen Ableitungen $V_{\Phi j}$ und $V_{\Phi\Phi}$ durch $u_{CC}C_j$ bzw. $u_{CC}C_\Phi$, lässt sich auch für den Ansatz der Elastizitäten eine konsumbasierte Interpretation anführen. Für die Quotienten R_s/R und R_f/R ergeben sich in diesem Fall die Darstellungen $\frac{C_s s}{C_\Phi \Phi}$ und $\frac{C_f f}{C_\Phi \Phi}$. Werden Zähler und Nenner dieser Terme jeweils mit C erweitert, erhält man unmittelbar $\frac{\partial C}{C}\frac{s}{\partial s}/\frac{\partial C}{C}\frac{\Phi}{\partial \Phi}$ bzw. $\frac{\partial C}{C}\frac{f}{\partial f}/\frac{\partial C}{C}\frac{\Phi}{\partial \Phi}$, d. h. eine Relation von Elastizitäten. Konkret ergibt sich demnach ein Verhätnis der Preiselastizität[76] des Konsums zu der Vermögenselastizität des Konsums, wodurch implizit eine Normierung der Preiselastizitäten erfolgt.

Der Ausgangspunkt der vorangegangenen Untersuchungen hinsichtlich der Hedge-Komponenten h_3 und h_4 war deren charakteristische Funktionsweise den gegenwärtigen Grenznutzen des Vermögens aufrechtzuerhalten. In einem ähnlichen Modellrahmen bewertet Breeden (1984) diesen Ansatz als myopische Sichtweise auf das Hedging, weil lediglich der Grenznutzen des gegenwärtigen Konsums in Betracht gezogen wird. Ziel des Investors ist es jedoch, seinen Konsum und die Hedge-Rate derart zu konstituieren, so dass der erwartete Nutzen des lebenslangen Konsumstroms maximiert

[76] Hinsichtlich des Kassakurses s bzw. des Futureskurses f.

wird. Insofern wählt der Investor sein intertemporales Absicherungsverhalten mit dem Ziel, den Erwartungsnutzen des lebenslangen Konsumstroms $V(\Phi, s, f, t)$ vollständig gegenüber stochastischen Variationen der Zustandsvariablen zu immunisieren.[77] Demnach sind in der optimalen Hedge-Rate des Investors Terme enthalten, die Vermögenskompensationen bzgl. Veränderungen der Zustandsvariablen darstellen, um den erwarteten lebenslangen Nutzen aufrechtzuerhalten. Die Terme werden demzufolge die Form $-V_s/V_\Phi$ und $-V_f/V_\Phi$ haben. In diesem Kontext wird in Analogie zu Breeden (1984) die Annahme getroffen, dass die prozentualen Vermögenskompensationen hinsichtlich einer Änderung der Zustandsvariablen $j \in \{s, f\}$ unabhängig vom Vermögensniveau sind, d. h. $\frac{\partial}{\partial \Phi}\left[\frac{V_j}{V_\Phi \Phi}\right] = 0$ gilt. Unter dieser Voraussetzung lässt sich folgendes Resultat zeigen:

Satz 2.2.7. *Die erforderliche Vermögenskompensation zur Aufrechterhaltung des Erwartungsnutzens des lebenslangen Konsumstroms bzgl. der Zustandsvariablen $j \in \{s, f\}$ lautet:*

$$\Pi_j = -\frac{V_j}{V_\Phi \Phi}$$

und es gilt

$$-\frac{V_{\Phi j}}{V_{\Phi\Phi}} = \Phi\left(\frac{R-1}{R}\right)\Pi_j. \tag{2.35}$$

Beweis: Bzgl. der getroffenen Annahme $\frac{\partial}{\partial \Phi}\left[\frac{V_j}{V_\Phi \Phi}\right] = 0$ gilt unter Anwendung der Quotientenregel

$$0 = \frac{V_{j\Phi} V_\Phi \Phi - V_j(V_{\Phi\Phi}\Phi + V_\Phi)}{(V_\Phi \Phi)^2},$$

woraus unmittelbar die Bedingung

$$V_{j\Phi} V_\Phi \Phi = V_j V_{\Phi\Phi} \Phi + V_j V_\Phi$$

folgt. Dividiert man beide Seiten durch $V_\Phi \Phi V_{\Phi\Phi}$ ergibt sich

$$\frac{V_{j\Phi}}{V_{\Phi\Phi}} = \frac{V_j}{V_\Phi} + \frac{V_j}{V_{\Phi\Phi}\Phi}$$

[77] Breeden definiert dies als perfekten Hedge. Vgl. Breeden (1984), S. 288.

resp.

$$-\frac{V_{j\Phi}}{V_{\Phi\Phi}} = -\Phi\left(1 + \frac{1}{\frac{V_{\Phi\Phi\Phi}}{V_\Phi}}\right)\frac{V_j}{V_\Phi\Phi}.$$

Durch Einsetzen der intertemporalen relativen Risikoaversion $R = \frac{V_{\Phi\Phi}\Phi}{V_\Phi}$ und Anwendung des Satzes von Schwarz[78,79] folgt die Behauptung. □

Die erforderliche Vermögenskompensation Π_j wird insbesondere von den partiellen Ableitungen der Wertfunktion V hinsichtlich der Zustandsvariablen j und Φ beeinflusst. Diese geben an, wie eine zusätzliche Einheit der Zustandsvariablen bewertet wird, d. h. die Auswirkung einer marginalen Änderung der Zustandsvariablen auf den Erwartungsnutzen des lebenslangen Konsumstroms. Aus diesem Grund werden die partiellen Ableitungen V_s, V_f und V_Φ auch als Schattenpreise[80] bezeichnet. Aus der Optimalitätsbedingung (2.10) ist einerseits bekannt, dass der Schattenpreis des Vermögens V_Φ positiv ist.[81] Andererseits entspricht der Schattenpreis dem Grenznutzen des Konsums. Unter der Prämisse einer konkaven Nutzenfunktion korrespondiert somit ein hoher Schattenpreis des Vermögens mit einem geringen Konsumniveau. Über die Vorzeichen der Schattenpreise V_s und V_f kann ohne weitere Annahmen keine Aussage getroffen werden. Diesbezüglich wird im Folgenden der Einfluss der intertemproalen relativen Risikoaversion R untersucht.

Der Term $R - 1/R$ auf der rechten Seite von Gleichung (2.35) kann als Gewichtungsfaktor der erforderlichen Vermögenskompensation Π_j aufgefasst werden. Sowohl der

[78] Der Satz von Schwarz besagt, dass unter gewissen Voraussetzungen die Reihenfolge bei einer Differentiation höherer Ordnung keine Rolle spielt. Die Reihenfolge der gemischten partiellen Ableitung k-ter Ordnung darf vertauscht werden, sofern die partiellen Ableitungen k-ter Ordnung wiederum stetige Funktionen sind. In der Literatur wird dieser Sachverhalt auch häufig als Young-Theorem bezeichnet, vgl. Chiang (1984), S. 313.

[79] Die Voraussetzung für die Anwendbarkeit des Satzes von Schwarz auf die Wertfunktion V ist erfüllt, d. h. es gilt $V_{\Phi j} = V_{j\Phi}$ für $j \in \{s, f\}$.

[80] Vgl. Dixit (1990), S. 162. Im Kontext der Dynamischen Optimierung wird der Begriff Schattenpreis häufig in Verbindung mit einer sogenannten „Kozustandsvariablen“ verwendet. Die Kozustandsvariable wird in das Problem der optimalen Steuerung im Rahmen der Hamilton-Funktion (kurz: Hamiltonian) eingeführt. Diese stellt den Ausgangspunkt des Maximum-Prinzips von Pontryagin dar, der zentralen Methode zur Lösung von Problemstellungen in der Optimalen Kontrolltheorie. Dieses Prinzip ist eine dynamische Verallgemeinerung der statischen Lagrange-Methode. Der Multiplikator der korrespondierenden Hamilton-Funktion wird in diesem Zusammenhang als Kozustandsvariable bezeichnet, vgl. Chiang (1992), S. 167 ff. Für eine genaue Beschreibung des Maximum-Prinzips von Pontryagin vgl. insbesondere Pontryagin et al. (1962).

[81] D. h. bei einer marginalen Zunahme des Vermögens erhöht sich der Erwartungsnutzen des lebenslangen Konsumstroms um V_Φ Nutzeneinheiten.

Umfang als auch das Vorzeichen des Gewichtungsfaktors hängen von dem Grad der intertemporalen relativen Risikoaversion R ab. Als Referenzwert für R bietet sich der Wert eins an, d. h. es wird ein Investor mit logarithmischer Nutzenfunktion betrachtet.[82] Für beide Zustandsvariablen s und f ergibt sich in diesem Fall jeweils eine erforderliche Vermögenskompensation von null. Damit einhergehend besteht die optimale Hedge-Rate (2.11) nur aus den Komponenten h_1 und h_2. Demnach sichert sich ein Investor mit logarithmischer Nutzenfunktion nicht gegen das durch die Zustandsvariablen s und f induzierte Risiko ab. Er wählt eine momentan-effiziente Risikopolitik, die als Konsequenz einen stärker schwankenden Konsumprozess hat.[83] Satz 2.2.7 belegt somit im Rahmen dieses Hedging-Modells die Gültigkeit für das bereits aus der Portefeuille-Optimierung bekannte Resultat, dass ein Investor mit logarithmischer Nutzenfunktion myopisch agiert.[84] Je größer die Abweichung der intertemporalen relativen Risikoaversion R von dem Referenzwert ist, desto stärker wird die Vermögenskompensation Π_j gewichtet. Für den Extremfall einer unendlich hohen intertemporalen relativen Risikoaversion nimmt der Gewichtungsfaktor $R-1/R$ den Wert eins an.[85] Dagegen gehen geringere Grade der intertemporalen relativen Risikoaversion[86] mit größeren negativen Werten für den Gewichtungsfaktor $R - 1/R$ einher.[87] Darüber hinaus verdeutlicht Gleichung (2.35) den Zusammenhang der erforderlichen Vermögenskompensationen aus unterschiedlichen Zeitperspektiven: Der Aufrechter-

[82] Vgl. z. B. Ingersoll (1987), S. 315.

[83] Dies ist eine unmittelbare Implikation von Satz 2.2.4, in dem gezeigt wurde, dass die zustandsabhängigen Hedge-Terme h_3 und h_4 implizit eine Minimierung des Risikos der intertemporalen Konsumstrategie bewirken.

[84] In der Portefeuille-Theorie wird der Terminus „Myopie“ als Synonym für ein kurzfristig orientiertes Verhalten verwendet. Liegt eine mehrperiodige Portefeuille-Planung vor, so verhält sich ein Investor myopisch, wenn die von ihm getroffenen Entscheidungen unabhängig vom Planungshorizont sind, d. h. wenn er jede Periode als seine letzte Planungsperiode ansieht. Demzufolge werden Informationen über zukünftige Perioden nicht berücksichtigt. In dem Spezialfall einer logarithmischen Nutzenfunktion handelt ein Investor stets myopisch, vgl. Ingersoll (1987), S. 177 ff. In der zeit-diskreten Modellierung hat Samuelson (1969) ein myopisches Verhalten zudem für Nutzenfunktionen aus der CRRA-Klasse nachgewiesen, sofern die Renditen der Wertpapiere unabhängig und identisch verteilt sind. Zu dem gleichen Ergebnis kommt Merton (1969) für den Fall der zeit-stetigen Portefeuille-Optimierung.

[85] Es gilt $\lim_{R\to\infty} 1 - \frac{1}{R} = 1$.

[86] Unter der Voraussetzung, dass $R < 1$ gilt.

[87] Es gilt $\lim_{R\to 0} 1 - \frac{1}{R} = -\infty$.

haltung des Erwartungsnutzens vom gegenwärtigen bzw. lebenslangen Konsumstrom. Insofern weist der Gewichtungsfaktor $R - 1/R$ eine Art Transformationsfunktion auf.

Wird von der speziellen Annahme einer logarithmischen Nutzenfunktion abgesehen, nimmt die intertemporale relative Risikoaversion R einen Wert ungleich eins an. Nach Gleichung (2.35) ergeben sich demzufolge von null verschiedene erforderliche Vermögenskompensationen hinsichtlich der Zustandsvariablen s und f. Ob diese Vermögenskompensationen positive oder negative Werte annehmen, wird durch die Schattenpreise V_s bzw. V_f bestimmt. Obwohl im Allgemeinen ohne konkrete Spezifikation der Wertfunktion V die partiellen Ableitungen V_s und V_f nicht explizit dargestellt werden können, lässt sich mittels der Gleichung (2.35) zumindest für den Schattenpreis des Futures in Abhängigkeit von der Situation auf dem Terminmarkt und der intertemporalen relativen Risikoaversion eine qualitative Aussage über das Vorzeichen treffen. Aus dem Beweis von Satz 2.2.5 ist bereits bekannt, dass der Term $V_{\Phi f}$ in einer MR-Backwardation (MR-Contango) Situation kleiner (größer) als null ist. Der konkave Verlauf der Wertfunktion garantiert $V_{\Phi\Phi} < 0$, wodurch die linke Seite von (2.35) kleiner (größer) null ist. Unter Berücksichtigung des positiven (erwarteten) Grenznutzens des Vermögens V_Φ nimmt die erforderliche Vermögenskompensation Π_f für $R > 1$ ($R < 1$) das gleiche (umgekehrte) Vorzeichen an. Damit ist der Schattenpreis V_f bei MR-Backwardation (MR-Contango) positiv (negativ).[88]

Dieses Resultat ist ökonomisch von hoher Relevanz. Für dieselbe Terminmarktsituation ergeben sich in Abhängigkeit der intertemporalen relativen Risikoaversion R unterschiedliche Vorzeichen für den Schattenpreis des Futures. Für Investoren, deren Grad der intertemporalen relativen Risikoaversion größer (kleiner) als eins ist, ergibt sich in MR-Backwardation ein positiver (negativer) Wert für V_f. Dieser Wert gibt an, wie eine zusätzliche Einheit der Zustandsvariable bewertet wird: Ein marginal höherer Futureskurs führt zu einer Erhöhung (Verringerung) des Erwartungsnutzens des optimalen intertemporalen Konsumstroms. Alternativ lässt sich V_f als derjenige Preis

[88] Analog zu der Interpretation der Hedge-Komponente h_4 auf Seite 43 kann aufgrund der Charakteristika des Modells ohne konkrete Spezifikation der Wertfunktion V keine Aussage über das Vorzeichen des Schattenpreises V_s getroffen werden.

auffassen, den ein Investor für eine marginale Erhöhung des Futureskurses bereit ist zu zahlen.[89] Seitens des Investors erfolgt damit implizit eine simultane Bewertung hinsichtlich der Vorteilhaftigkeit diverser Effekte, die mit einer Zustandsänderung des Futures einhergehen. Dies tangiert einerseits den Konsum, andererseits die Ertrags-Risikokonstellation des Futures. Investoren, deren Grad der intertemporalen relativen Risikoaversion größer als eins ist, bewerten einen Anstieg des Futures in einer MR-Backwardation Situation hinsichtlich des Erwartungsnutzens des lebenslangen Konsumstroms ceteris paribus als vorteilhaft. Genau die entgegengesetzte Bewertung ist bei Investoren wiederzufinden, die eine intertemporale relative Risikoaversion kleiner als eins aufweisen. In diesem Fall ist der Schattenpreis V_f negativ; ein Anstieg des Futures wird hinsichtlich des Erwartungsnutzens des lebenslangen Konsumstroms ceteris paribus als nachteilig empfunden. Für eine MR-Contango Situation erfolgt die Argumentation analog: Für Investoren, deren Grad der intertemporalen relativen Risikoaversion größer (kleiner) als eins ist, ergibt sich ein negativer (positiver) Wert Schattenpreis V_f. In diesem Fall führt ein marginal niedrigerer Futureskurs[90] zu einer Erhöhung (Verringerung) des Erwartungsnutzens des optimalen intertemporalen Konsumstroms.

Aus den vorangegangenen Untersuchungen ergibt sich unmittelbar die Fragestellung, ob die von dem Investor geforderte Vermögenskompensation hinsichtlich der Zustandsvariablen f größer sein kann als die auf dem Terminmarkt eingegangene spekulative Position. Formal bedeutet dies, dass der Umfang der Hedge-Komponente h_3 betragsmäßig größer ist, als der Umfang des spekulativen Hedge-Terms h_2. Dies ist genau dann der Fall, wenn der Investor das durch den Futures f induzierte Konsumrisiko höher einschätzt als das Gewinnpotential in Hinblick auf die aktuelle Risikoprämie des Terminmarkts. Sofern diese Eigenschaft auf die optimale Hedging-Politik eines Investors zutrifft, liegt es nahe, folgende Definition anzuführen:

[89] Unter der Prämisse, dass sich der Investor für den Rest des Planungshorizonts optimal verhält.

[90] D. h. in Analogie zu der MR-Backwardation Situation ein Wert, der näher an dem Gleichgewichtsniveau θ_f liegt.

Definition 8. *Herrscht im Terminmarkt MR-Backwardation (MR-Contango) vor, so heißt ein Investor in einem Zustand (universal) spekulationsavers, wenn die von ihm mit dem Faktor $R-1$ gewichtete erforderliche Vermögenskompensation bzgl. dem (jedem) Zustand f größer (kleiner) ist als die MR-Sharpe-Ratio des Futures $MRSR_f$, d. h. es gilt:*

$$(R-1)\frac{V_f}{V_\Phi \Phi} > (<) \frac{\kappa_f(\theta_f - f)}{\sigma_f^2 f}. \tag{2.36}$$

In Abhängigkeit von der Situation auf dem Terminmarkt ergibt sich aus der Ungleichung (2.36) für MR-Backwardation (MR-Contango) die Bedingung $h_2 + h_3 > (<) 0$. Gilt diese Ungleichung auch unter Einbeziehung der Hedge-Komponente h_4, entscheidet sich der Investor für Overhedging (Underhedging).[91] Dieser Sachverhalt wird zusammenfassend wie folgt formuliert:

Satz 2.2.8. *In dem hier vorgestellten Hedging-Modell existieren Situationen, in denen sich ein spekulationsaverser Investor in MR-Backwardation (MR-Contango) für eine Absicherungspolitik entscheidet, bei der der Anteil verkaufter Futures am Vermögen höher ist als beim reinen Hedging.*

Beweis: Angenommen der Kassakurs s und der Terminkurs f sind unkorreliert, d. h. es gilt $\rho_{sf} = 0$. Die optimale Hedge-Rate lautet in diesem Fall $h^* = h_2 + h_3 = -\frac{\kappa_f(\theta_f - f)}{\sigma_f^2 R} + \frac{R_f}{R}$. Da es sich annahmegemäß um einen spekulationsaversen Investor handelt, ist in einer MR-Backwardation (MR-Contango) Situation folgende Ungleichung erfüllt:

$$(R-1)\frac{V_f}{V_\Phi \Phi} > (<) \frac{\kappa_f(\theta_f - f)}{\sigma_f^2 f}. \tag{2.37}$$

Werden beide Seiten der Ungleichung mit dem positiven Futureskurs f sowie der positiven intertemporalen relativen Risikotoleranz $1/R$ multipliziert, ergibt sich als äquivalente Ungleichung

$$\frac{R-1}{R}\frac{V_f f}{V_\Phi \Phi} > (<) \frac{\kappa_f(\theta_f - f)}{\sigma_f^2 R} \tag{2.38}$$

[91] D. h. in MR-Backwardation (MR-Contango) gilt $h^* = h_2 + h_3 + h_4 > (<) 0$.

bzw. nach Satz 2.2.7

$$\frac{V_{\Phi f} f}{V_{\Phi\Phi}\Phi} > (<) \frac{\kappa_f(\theta_f - f)}{\sigma_f^2 R}. \tag{2.39}$$

Aus der Ungleichung (2.39) und unter Berücksichtigung der Unkorreliertheit zwischen Kassa- und Terminkurs resultiert unmittelbar $h^* = h_2 + h_3 > (<) 0$. Demnach entscheidet sich der Investor in einer MR-Backwardation (MR-Contango) Situation für Overhedging (Underhedging). □

Als Implikation von Satz 2.2.5 ist bereits bekannt, dass der spekulative Hedge-Term h_2 und die preiselastische Hedge-Komponente h_3 für dieselbe Terminmarkt-Situation entgegengesetzte Vorzeichen aufweisen. Satz 2.2.8 beschreibt in diesem Zusammenhang Konstellationen, in denen h_3 betragsmäßig größer ist als h_2. Trotz positiver (negativer) Risikoprämie des Terminmarkts ist Overhedging (Underhedging) optimal. Fundamental kann dieses Verhalten damit begründet werden, dass ein Investor das durch den Zustand von f induzierte Konsumrisiko höher einschätzt als das Gewinnpotential in Hinblick auf die aktuelle Risikoprämie des Terminmarkts.

Die vorherigen Ausführungen haben gezeigt, dass unter gewissen Voraussetzungen das aus der Literatur des Futures-Hedgings allgemein bekannte Resultat, indem es in einer MR-Backwardation (MR-Contango) Situation stets zu einem Underhedge (Overhedge) kommt, seine Gültigkeit verliert.[92] Im Rahmen des vorliegenden Hedging-Modells ist es z. B. ausreichend von unkorrelierten Kassa- und Terminkursen auszugehen und durch die Spekulationsaversion eine strengere Annahme an die Präferenzen eines Investors zu stellen. Um die Auswirkungen auf die optimale Absicherungspolitik konkret analysieren und das Ausmaß des Overhedgings (Underhedgings) spezifizieren zu könnnen, sind zusätzliche Annahmen hinsichtlich der Wertfunktion eines Investors unabdingbar. Dies ist allerdings nicht Gegenstand der aktuellen Untersuchung.

[92] Vgl. z. B. Briys et al. (1990), Briys et al. (1993), Broll and Wong (2002) und Wahl and Broll (2003).

2.3 Intertemporales Konsumprofil

Der Einfluss der Absicherungsmöglichkeit des Kursrisikos mittels Futures auf das optimale intertemporale Konsumprofil des Investors ist Kernpunkt der nachfolgenden Untersuchungen. In der neoklassischen Wachstumstheorie wird das optimale intertemporale Konsumprofil durch die Keynes-Ramsey-Regel beschrieben, die die Bereitschaft des Anlegers ausdrückt Konsum in der Gegenwart gegen Konsum in der Zukunft zu tauschen.[93] Unter den hier vorliegenden stochastischen Rahmenbedingungen wird eine modifizierte Keynes-Ramsey-Regel hergeleitet und anschließend analysiert. Insbesondere interessiert die Fragestellung unter welchen Umständen Futures-Hedging die Bereitschaft der intertemporalen Konsumsubstitution beeinflusst und wie sensibel der Anleger auf Veränderungen einzelner Parameter reagiert.

Die klassische Keynes-Ramsey-Regel beschreibt die optimale Wachstumsrate des Konsums als Differenz des Kapitalmarktzinses und der Zeitpräferenzrate.[94] Eine positive Zeitpräferenzrate des Anlegers bedeutet, dass zukünftiger Konsum weniger hoch bewertet wird als gegenwärtiger Konsum.[95] Demnach lässt sie sich als Mindestprämie interpretieren, die der Anleger fordert, damit er hinsichtlich der Reduktion des Gegenwartkonsums um eine Einheit zugunsten einer entsprechenden Erhöhung des Zukunftskonsums gerade indifferent ist.

Dagegen lässt sich der Zinssatz als Prämie auffassen, die der Anleger erhält, wenn er seinen Gegenwartskonsum um eine Einheit einschränkt, um den Zukunftskonsum zu erhöhen. Die Wachstumsrate des Konsums ist positiv, wenn der Sicherheitszins größer als die Zeitpräferenzrate ist. In diesem Fall wird Konsum intertemporal substituiert,

[93] Vgl. Frenkel and Hemmer (1999) S. 81.

[94] Der Name geht zurück auf den englischen Mathematiker Frank P. Ramsey, der die Regel in dem Artikel Ramsey (1928) hergeleitet hat. Dieser Artikel enthält gleichzeitig eine intuitive und ökonomische Erklärung der Regel von John M. Keynes, so dass sich die Bezeichnung Keynes-Ramsey-Regel etablierte.

[95] Vgl. Christiaans (2004), S. 16.

d. h. es wird heutiger Konsum in die Zukunft verlagert.[96] Der erwartete Konsum bleibt unverändert, wenn die Zeitpräferenzrate und der risikofreie Zins übereinstimmen.[97]

Das vorliegende Modell grenzt sich in mehrfacher Hinsicht von den Modellen der neoklassischen Wachstumstheorie ab. Zum einen ermöglicht Futures-Hedging dem risikoaversen Investor Risiken seines Wertpapierbestands abzuwälzen. Zum anderen wird die Unvollkommenheit der Risikomärkte für Derivate berücksichtigt. Diese resultiert aus der imperfekten Korrelation zwischen Kassa- und Terminkurs. Des Weiteren ist das hier betrachtete Modell dadurch gekennzeichnet, dass neben der Hedge-Entscheidung keine zusätzliche Investitionsentscheidung gegeben ist; insbesondere existiert keine risikolose Anlagemöglichkeit. Aufgrund dieser fundamentalen Unterschiede wird die Beschreibung der Konsumwachstumsrate in dem hier vorliegenden Modellrahmen eine wesentlich komplexere Form haben. Der nächste Satz gibt die modifizierte Keynes-Ramsey-Regel an:

Satz 2.3.1. *Sichert der Investor das Kursrisiko mittels Futures ab, die einem CIR-Prozess folgen, so ergibt sich unter Einbeziehung der optimalen Hedge-Rate nach Gleichung* (2.11) *folgende modifizierte Keynes-Ramsey-Regel für imperfekt korrelierte Risiken:*

$$\frac{\mathbb{E}[-dU'(C^*)/dt]}{U'(C^*)} = \left(\frac{\kappa_s(\theta_s - s)}{s} - \frac{\kappa_f(\theta_f - f)}{f}\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\right) - \tau - (R + R_s)\frac{\sigma_s^2(1-\rho_{sf}^2)}{s}. \tag{2.40}$$

[96] Und umgekehrt.
[97] Vgl. Broll and Wahl (2012), S. 128.

Beweis: Die optimale Konsumumschichtung hat auf Grundlage der Bellman-Gleichung zu erfolgen. Die Differentiation von Gleichung (2.9) nach Φ ergibt unter Berücksichtigung der optimalen Konsumentscheidung (2.10):

$$\begin{aligned} 0 &= V_\Phi\left(\frac{\kappa_s(\theta_s - s)}{s} - \tau\right) \\ &+ V_{\Phi t} + V_{\Phi\Phi}\Phi\left(\frac{\kappa_s(\theta_s - s) + \sigma_s^2}{s} - \frac{C}{\Phi} - h\left(\frac{\kappa_f(\theta_f - f)}{f} + \frac{\sigma_s\sigma_f}{\sqrt{sf}}\rho_{sf}\right)\right) \\ &+ V_{\Phi s}\left(\kappa_s(\theta_s - s) + \sigma_s^2\right) + V_{\Phi f}\left(\kappa_f(\theta_f - f) + \sigma_s\sigma_f\frac{\sqrt{f}}{\sqrt{s}}\rho_{sf}\right) \\ &+ \frac{1}{2}V_{\Phi\Phi\Phi}\Phi^2\left(\frac{\sigma_s^2}{s} + h^2\frac{\sigma_f^2}{f} - 2h\frac{\sigma_s\sigma_f}{\sqrt{sf}}\rho_{sf}\right) + \frac{1}{2}V_{\Phi ss}\sigma_s^2 s + \frac{1}{2}V_{\Phi ff}\sigma_f^2 f \\ &+ V_{\Phi\Phi s}\Phi\left(\sigma_s^2 - h\sigma_s\sigma_f\frac{\sqrt{s}}{\sqrt{f}}\rho_{sf}\right) + V_{\Phi\Phi f}\Phi\left(\sigma_s\sigma_f\frac{\sqrt{f}}{\sqrt{s}}\rho_{sf} - h\sigma_f^2\right) + V_{\Phi sf}\sigma_s\sigma_f\sqrt{sf}\rho_{sf}. \end{aligned} \tag{2.41}$$

Durch Anwendung der Itô Regeln für die partielle Ableitung V_Φ der Wertfunktion wird eine vereinfachte Darstellung der Gleichung (2.41) erzielt. Für das totale Differential von V_Φ gilt:

$$\begin{aligned} dV_\Phi &= V_{\Phi t}dt + V_{\Phi\Phi}d\Phi + V_{\Phi s}ds + V_{\Phi f}df \\ &+ \frac{1}{2}V_{\Phi\Phi\Phi}(d\Phi)^2 + \frac{1}{2}V_{\Phi ss}(ds)^2 + \frac{1}{2}V_{\Phi ff}(df)^2 \\ &+ V_{\Phi\Phi s}d\Phi ds + V_{\Phi\Phi f}d\Phi df + V_{\Phi sf}dsdf. \end{aligned} \tag{2.42}$$

Einsetzen der Kursentwicklungen (2.1) von s bzw. (2.2) von f sowie der Vermögensentwicklung $d\Phi$ (2.4) in die Darstellung (2.42) liefert:

$$\begin{aligned} dV_\Phi &= \Big[V_{\Phi t} + V_{\Phi\Phi}\Phi\Big(\frac{\kappa_s(\theta_s - s)}{s} - \frac{C}{\Phi} - h\frac{\kappa_f(\theta_f - f)}{f}\Big) \\ &\quad + V_{\Phi s}\kappa_s(\theta_s - s) + V_{\Phi f}\kappa_f(\theta_f - f) \\ &\quad + \frac{1}{2}V_{\Phi\Phi\Phi}\Phi^2\Big(\frac{\sigma_s^2}{s} + h^2\frac{\sigma_f^2}{f} - 2h\frac{\sigma_s\sigma_f}{\sqrt{sf}}\rho_{sf}\Big) + \frac{1}{2}V_{\Phi ss}\sigma_s^2 s + \frac{1}{2}V_{\Phi ff}\sigma_f^2 f \\ &\quad + V_{\Phi\Phi s}\Phi\Big(\sigma_s^2 - h\sigma_s\sigma_f\frac{\sqrt{s}}{\sqrt{f}}\rho_{sf}\Big) + V_{\Phi\Phi f}\Phi\Big(\sigma_s\sigma_f\frac{\sqrt{f}}{\sqrt{s}}\rho_{sf} - h\sigma_f^2\Big) \\ &\quad + V_{\Phi sf}\sigma_s\sigma_f\sqrt{sf}\rho_{sf}\Big]dt \\ &+ \quad V_{\Phi\Phi}\Phi\Big(\frac{\sigma_s}{\sqrt{s}}dz_s - h\frac{\sigma_f}{\sqrt{f}}dz_f\Big) + V_{\Phi s}\sigma_s\sqrt{s}dz_s + V_{\Phi f}\sigma_f\sqrt{f}dz_f. \end{aligned} \tag{2.43}$$

Wird Gleichung (2.41) mit dt multipliziert und in das Ergebnis die Darstellung (2.43) eingesetzt, ergibt sich eine stochastische Differentialgleichung des Vermögens:

$$\begin{aligned} -dV_\Phi &= \Big[V_\Phi\Big(\frac{\kappa_s(\theta_s - s)}{s} - \tau\Big) + V_{\Phi\Phi}\Phi\Big(\frac{\sigma_s^2}{s} - h\frac{\sigma_s\sigma_f}{\sqrt{sf}}\rho_{sf}\Big) \\ &\quad + V_{\Phi s}\sigma_s^2 + V_{\Phi f}\sigma_s\sigma_f\frac{\sqrt{f}}{\sqrt{s}}\rho_{sf}\Big]dt \\ &- \quad V_{\Phi\Phi}\Phi\Big(\frac{\sigma_s}{\sqrt{s}}dz_s - h\frac{\sigma_f}{\sqrt{f}}dz_f\Big) - V_{\Phi s}\sigma_s\sqrt{s}dz_s - V_{\Phi f}\sigma_f\sqrt{f}dz_f. \end{aligned}$$

Unter Verwendung der Bedingung $dU'(C^*) = dV_\Phi$, die sich unmittelbar aus der Optimalitätsbedingung (2.10) ableiten lässt, und der optimalen Hedge-Rate h^* nach (2.11) folgt für das intertemporale Konsumprofil:

$$\begin{aligned} -dU'(C^*) &= \Big[U'(C^*)\Big(\frac{\kappa_s(\theta_s - s)}{s} - \frac{\kappa_f(\theta_f - f)}{\sqrt{sf}}\frac{\sigma_s}{\sigma_f}\rho_{sf} - \tau\Big) \\ &\quad + \Big(V_{\Phi\Phi}\frac{\Phi}{s} + V_{\Phi s}\Big)\sigma_s^2(1 - \rho_{sf}^2)\Big]dt \\ &- \quad \Big(\Phi V_{\Phi\Phi}\frac{\sigma_s}{\sqrt{s}} + V_{\Phi s}\sigma_s\sqrt{s}\Big)dz_s + \Big(HV_{\Phi\Phi}\frac{\sigma_f}{\sqrt{f}} + V_{\Phi f}\sigma_f\sqrt{f}\Big)dz_f. \end{aligned} \tag{2.44}$$

Nach Division durch den Faktor dt und den Grenznutzen $U'(C)$ sowie Anwendung des Erwartungswertoperators ergibt sich schließlich folgende Keynes-Ramsey-Regel und damit die Behauptung:

$$\begin{aligned}
\frac{\mathbb{E}[-dU'(C^*)/dt]}{U'(C^*)} &= \left(\frac{\kappa_s(\theta_s - s)}{s} - \frac{\kappa_f(\theta_f - f)}{f}\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\right) - \tau \\
&+ \left(\frac{V_{\Phi\Phi}}{V_\Phi}\frac{\Phi}{s} + \frac{V_{\Phi s}}{V_\Phi}\right)\sigma_s^2(1-\rho_{sf}^2) \\
&= \left(\frac{\kappa_s(\theta_s - s)}{s} - \frac{\kappa_f(\theta_f - f)}{f}\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\right) - \tau \\
&- \left(R + R_s\right)\frac{\sigma_s^2(1-\rho_{sf}^2)}{s}.
\end{aligned}$$

□

Gleichung (2.40) gibt die optimale Konsumstrategie des Investors an, der das Kursrisiko mittels Futures absichert. Diese wird deutlich durch die unterliegende Dynamik der Kursprozesse sowie der imperfekten Korrelation zwischen Kassa- und Terminkurs beeinflusst und weist somit eine wesentlich komplexere Form als die klassische Keynes-Ramsey-Regel auf. Die Bedeutung der einzelnen Terme der Gleichung wird im Folgenden analysiert.

Der erste Summand $d_s = \frac{\kappa_s(\theta_s - s)}{s}$ in den runden Klammern auf der rechten Seite von (2.40) gibt die erwartete relative Kursänderung des Wertpapiers s an. Diese ist positiv (negativ), wenn sich der Kassakurs oberhalb (unterhalb) des langfristigen Gleichgewichtsniveaus θ_s befindet. Der zweite Summand $d_f = \frac{\kappa_f(\theta_f - f)}{f}\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}$ gibt die mit dem reinen Hedge-Term gewichtete erwartete relative Risikoprämie des Futures an. Befindet sich der Terminmarkt in einer MR-Backwardation (MR-Contango) Situation, resultiert eine positive (negative) Risikoprämie. Einschließlich des Minuszeichens wird dieser Ausdruck in MR-Backwardation (MR-Contango) negativ (positiv).[98] Zusammengenommen folgt für den in runden Klammern stehenden Ausdruck $d_{sf} = d_s - d_f$, dass er kein eindeutiges Vorzeichen aufweist, weil er maßgeblich von

[98] Unter der Prämisse positiver Korreliertheit zwischen Kassa- und Futureskurs.

den Zuständen des Kassa- und Terminkurses determiniert wird. Von diesem Ausdruck wird die (positive) Zeitpräferenzrate subtrahiert. Die klassische Keynes-Ramsey-Regel beschreibt die optimale Konsumstrategie als Differenz zwischen risikolosem Zins und Zeitpräferenzrate. Demnach scheint der Ausdruck in den runden Klammern hinsichtlich des intertemporalen Konsumprofils eine Art Kompensation des risikolosen Zinses darzustellen, mit deren Hilfe ein Abgleich mit der Zeitpräferenzrate stattfindet. Der letzte Term auf der rechten Seite von (2.40) stellt ein Produkt aus zwei Faktoren dar. Der erste Faktor ist die Summe aus der intertemporalen relativen Risikoaversion und der Preiselastizität des Wertpapiers s. Während die Risikoaversion per definitionem größer null ist, wurde bereits in Abschnitt 2.2.2 diskutiert, dass sich das Vorzeichen der Preiselastizität im Rahmen dieses Modells nicht in Abhängigkeit des Zustands spezifizieren lässt. Der zweite Faktor erfasst das Risiko des Wertpapiers und die Hedging-Effektivität. Je weniger imperfekt das Futures-Hedging ist, desto geringer ist der Einfluss des letzten Terms auf die optimale Konsumstrategie.

Aus den vorherigen Ausführungen wird deutlich, dass ohne zusätzliche Annahmen die Keynes-Ramsey-Regel (2.40) schwierig zu interpretieren ist. In den nachfolgenden Untersuchungen werden somit spezielle Szenarien bzw. konkrete Parameterkonstellationen unterstellt, um Aussagen über den Verlauf der optimalen Konsumstrategie des Investors treffen zu können. Zu berücksichtigen ist, dass die Aussagen nur im Erwartungswert gelten, also den erwarteten zukünftigen Konsum betreffen.

Korollar 2.3.1. *In Situationen, in denen die Differenz aus der erwarteten relativen Kassakursänderung und der mit dem reinen Hedge-Term gewichteten relativen Änderung des Futureskurses kleiner als die Zeitpräferenzrate ist und der Grad der intertemporalen relativen Risikoaversion den Betrag der Preiselastizität R_s übersteigt, nimmt der erwartete Konsum eindeutig ab.*

Beweis: Die Behauptung folgt unmittelbar aus der verallgemeinerten Keynes-Ramsey-Regel (2.40). Die Realisationen des Kassa- bzw. Terminkurses sind nach Voraussetzung derart gegeben, dass der Term in den runden Klammern auf der rechten Seite

von (2.40) einen geringeren Wert als die Zeitpräferenzrate aufweist, wodurch der Ausdruck $d_{sf} - \tau$ negativ wird. Der letzte Term auf der rechten Seite von (2.40) wird zunächst ohne Einbeziehung des negativen Vorzeichens betrachtet. Laut Annahme übersteigt der Grad der intertemporalen relativen Risikoaversion den Betrag der Preiselastizität R_s, so dass der Faktor $(R + R_s)$ auf jeden Fall positiv ist. Der nachfolgende Bruch ist wegen $s > 0, \sigma_s \geq 0$ und $\rho_{sf}^2 \leq 1$ größer gleich null. Insgesamt gilt somit $\frac{\mathbb{E}[-dU'(C^*)/dt]}{U'(C^*)} < 0$, der erwartete Konsum nimmt eindeutig ab. □

Die erste Voraussetzung des Korrolars hängt von den Zuständen des Kassa- und Terminkurses sowie von sämtlichen exogenen Parametern ab. Demnach lässt sich erst nach einer Realisation der stochastischen Prozesse (2.1) und (2.2) der Wert der Differenz der erwarteten relativen Kassakursänderung und der mit dem reinen Hedge-Term gewichteten relativen Änderung des Futureskurses bestimmen und mit der Zeitpräferenzrate vergleichen. Unabhängig davon lassen sich Konstellationen identifizieren, die diese Voraussetzung erfüllen. Liegen der Kassakurs oberhalb und der Terminkurs unterhalb ihres korrespondierenden langfristigen Gleichgewichtsniveaus, gilt $d_{sf} < 0$ und somit $d_{sf} < \tau$. In diesem Fall erwartet der Investor einen sinkenden Kassakurs sowie einen steigenden Futureskurs, wodurch der Gegenwartskonsum zulasten des Zukunftskonsums begünstigt wird. Befindet sich der Kassakurs unterhalb seines Gleichgewichtsniveaus und ist die erwartete relative Kursänderung kleiner als die Summe aus der Zeitpräferenzrate und der mit dem reinen Hedge-Term gewichteten relativen Änderung des Futureskurses, gilt ebenfalls $d_{sf} < \tau$, d. h. das erwartete zukünftige Konsumniveau nimmt ab. Es existieren weitere Situationen, die sich negativ auf das Sparverhalten des Investors auswirken.

Die zweite Voraussetzung betrifft das Ausmaß der Risikoaversion und die Preiselastizität R_s. Das Vorzeichen der Summe dieser beiden Größen bestimmt gleichzeitig das Vorzeichen des Produkts, weil der zweite Faktor für den Fall imperfekt korrelierter Risiken immer größer null ist.[99] Ist der Grad der intertemporalen relativen

[99] Der Grad der intertemporalen relativen Risikoaversion R und die Preiselastizität R_s werden maßgeblich durch die Wertfunktion determiniert. Insofern lässt sich ohne eine konkrete Spezifikation der Wertfunktion keine Aussage darüber treffen, welcher der beiden Terme größer ist.

Risikoaversion größer als der Betrag der Preiselastizität R_s, wirkt sich dies negativ auf das Sparverhalten des Investors aus. Mit abnehmender (zunehmender) Hedging-Effektivität (Volatilität des Kassakurses) verstärkt sich ceteris paribus diese Tendenz.

Die Aussagen des letzten Absatzes lassen sich ohne weiteres auf den Spezialfall übertagen, in dem die Differenz aus der erwarteten relativen Kassakursänderung und einer gewichteten relativen Änderung des Futureskurses mit der Zeitpräferenzrate übereinstimmt.

Korollar 2.3.2. *In Situationen, in denen die Differenz aus der erwarteten relativen Kassakursänderung und der mit dem reinen Hedge-Term gewichteten relativen Änderung des Futureskurses der Zeitpräferenzrate entspricht sowie der Grad der intertemporalen relativen Risikoaversion den Betrag der Preiselastizität R_s übersteigt, nimmt der erwartete Konsum infolge imperfekt korrelierter Kursrisiken eindeutig ab.*

Beweis: Die Behauptung folgt unmittelbar aus der verallgemeinerten Keynes-Ramsey-Regel (2.40). Die Realisationen des Kassa- bzw. Terminkurses sind nach Voraussetzung derart gegeben, dass der Term in den runden Klammern auf der rechten Seite von (2.40) gleich der Zeitpräferenzrate ist, wodurch der Ausdruck $d_{sf} - \tau$ null wird. In dieser Situation wird die optimale Konsumstrategie allein durch den letzten Term auf der rechten Seite von (2.40) bestimmt. Laut Annahme übersteigt der Grad der intertemporalen relativen Risikoaversion den Betrag der Preiselastizität R_s, so dass der Faktor $(R + R_s)$ auf jeden Fall positiv ist. Der nachfolgende Bruch ist wegen $s > 0, \sigma_s \geq 0$ und der imperfekten Korrelation $\rho_{sf}^2 < 1$ größer null. Unter Einbeziehung des negativen Vorzeichens gilt somit $\frac{\mathbb{E}[-dU'(C^*)/dt]}{U'(C^*)} < 0$. Der erwartete Konsum nimmt eindeutig ab. Diese Tendenz verstärkt sich mit abnehmender (zunehmenden) Hedging-Effektivität (Schwankungen des Kassakurses). □

In den bisherigen Untersuchungen wurden stets imperfekt korrelierte Risiken betrachtet und der Einfluss der Hedging-Effektivität auf das intertemporale Konsumprofil analysiert. Die Auswirkung des perfekten Hedgings auf die optimale Konsumstrategie beschreibt das nachfolgende Korollar:

Korollar 2.3.3. *Sind Kassa- und Terminkurs perfekt positiv korreliert ergibt sich die optimale Konsumstrategie aus der Differenz der erwarteten relativen Kassakursänderung und der mit dem reinen Hedge-Term gewichteten relativen Änderung des Futureskurses, abzüglich der Zeitpräferenzrate. Insbesondere ist die Wachstumsrate unabhängig von der Preiselastizität R_s und wird durch folgende vereinfachte Keynes-Ramsey-Regel beschrieben:*

$$\frac{\mathbb{E}[-dU'(C^*)/dt]}{U'(C^*)} = \left(\frac{\kappa_s(\theta_s - s)}{s} - \frac{\kappa_f(\theta_f - f)}{f}\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\right) - \tau. \tag{2.45}$$

Beweis: Die Behauptung folgt unmittelbar durch Einsetzen der Bedingung $\rho_{sf}^2 = 1$ in die verallgemeinerte Keynes-Ramsey-Regel (2.40). □

Das Vorzeichen des Ausdrucks d_{sf} auf der rechten Seite von Gleichung (2.45) hängt von den Ausprägungen der Summanden d_s und d_f ab. Der Term d_s ist positiv (negativ), wenn sich der Kassakurs s unterhalb (oberhalb) des langfristigen Gleichgewichtsniveaus θ_s befindet. In dieser Situation ist die erwartete relative Änderung des Wertpapiers s positiv (negativ). Je größer (kleiner) ceteris paribus die erwartete Wertänderung ist, desto größer (kleiner) ist die erwartete Konsumwachstumsrate. Die Argumentation für den Summanden d_f bzw. die Situation auf dem Terminmarkt erfolgt analog. In MR-Backwardation (MR-Contango) ist die erwartete relative Änderung des Futureskurses f positiv (negativ). Unter der Prämisse perfekt positiv korrelierter Kursrisiken gilt $d_f > 0$. Durch das negative Vorzeichen von d_f ergibt sich ein genau diametraler Einfluss für die optimale Konsumstrategie des Investors. D. h., je größer (kleiner) ceteris paribus die erwartete relative Wertänderung des Futureskurses f ist, desto kleiner (größer) ist die erwartete Konsumwachstumsrate. Die Aussage des Korollars lässt sich ohne weiteres auf perfekt negativ korrelierte Kursrisiken übertragen.[100]

[100] Die Gültigkeit dieser Aussage wird dadurch fundiert, dass der Investor für den Fall negativ korrelierter Kursrisiken eine lange Position in Terminkontrakten einnimmt. Der Term d_{sf} ergibt sich in diesem Fall aus der Summe der erwarteten relativen Kassakursänderung und der mit dem reinen Hedge-Term gewichteten erwarteten relativen Änderung des Futureskurses, d. h. es gilt $d_{sf} = d_s + d_f$. Insofern behält das Korollar auch bei perfekt negativer Korrelation seine Gültigkeit.

Die Differenz $d_{sf} - \tau$ entspricht der rechten Seite von Gleichung (2.45) und gibt die Sparquote des Investors an. Sofern der Term d_{sf} größere Werte annimmt als die Zeitpräferenzrate, resultieren positive Werte. Der Investor verlagert heutigen Konsum in die Zukunft. Dieser Effekt verstärkt sich, je größer (kleiner) ceteris paribus der (die) Term d_{sf} (Zeitpräferenzrate τ) ist.[101] Negative Grenzerträge des Sparens korrespondieren mit Werten von d_{sf}, die kleiner als die Zeitpräferenzrate sind. Der Investor präferiert heutigen Konsum, der erwartete Konsum nimmt ab. Für $d_{sf} = \tau$ ist die erwartete Konsumwachstumsrate konstant. In dieser Situation ist der Investor indifferent bezüglich der Reduktion des Gegenwartskonsums um eine Einheit zugunsten einer Steigerung des Zukunftskonsums. Aufgrund der Vielzahl von unterliegenden Parametern und der Zustandsabhängigkeit bzgl. s und f ist eine derartige Konstellation jedoch nahezu ausgeschlossen.[102] Die obigen Ausführungen lassen sich komprimiert wie folgt darstellen:

$$\frac{\mathbb{E}[-dU'(C^*)/dt]}{U'(C^*)} \begin{cases} < 0, & \text{falls } d_{sf} < \tau, \\ = 0, & \text{falls } d_{sf} = \tau, \\ > 0, & \text{falls } d_{sf} > \tau. \end{cases}$$

Die letzte Analyse des Konsumprofils untersucht den Spezialfall einer deterministischen Kassakursentwicklung. Das Wertpapier s ist in diesem Szenario risikolos;[103] eine Absicherung gegen die Kursentwicklung ist daher nicht erforderlich. Die Auswirkungen auf das intertemporale Konsumprofil werden wie folgt beschrieben:

Korollar 2.3.4. *Ist die Kassakursentwicklung risikolos, d. h. gilt $\sigma_s = 0$, entspricht die optimale Wachstumsrate des Konsums der erwarteten relativen Kassakursänderung abzüglich der Zeitpräferenzrate. Insbesondere ist die optimale Wachstumsrate un-*

[101] Gilt ceteris paribus für größere (kleinere) Werte von d_s (d_f).

[102] Bei den getroffenen Aussagen ist stets zu beachten, dass sie nur im Erwartungswert gelten, d. h. für den erwarteten zukünftigen Konsum zutreffen.

[103] In diesem Fall besitzt die Differentialgleichung (2.1) keinen Diffusionsterm.

abhängig von der Preiselastizität R_s und wird durch folgende vereinfachte Keynes-Ramsey-Regel beschrieben:

$$\frac{\mathbb{E}[-dU'(C^*)/dt]}{U'(C^*)} = \frac{\kappa_s(\theta_s - s)}{s} - \tau. \tag{2.46}$$

Beweis: Die Behauptung folgt unmittelbar durch Einsetzen der Bedingung $\sigma_s = 0$ in die verallgemeinerte Keynes-Ramsey-Regel (2.40). □

Die optimale Konsumstrategie (2.46) ergibt sich aus der Differenz der relativen Änderung des Wertpapiers s und der Zeitpräferenzrate. Ist die Differenz positiv, wird heutiger Konsum in spätere Perioden verlagert. Dieser Effekt verstärkt sich umso mehr, je größer (kleiner) ceteris paribus die relative Wertänderung des Wertpapiers (Zeitpräferenzrate) ist. Zu berücksichtigen ist, dass die Konsumwachstumsrate aufgrund der deterministischen Kursentwicklung nach einer bestimmten Zeitperiode immer negativ wird. Befindet sich der Kassakurs unterhalb des langfristigen Gleichgewichtsniveaus ist die relative Änderung des Wertpapiers positiv, ist jedoch eine in s fallende Funktion. Ist also zu einem bestimmten Zeitpunkt die relative Änderung kleiner als die Zeitpräferenzrate, so gilt dies auch automatisch für alle nachfolgenden Zeitpunkte. Die Aussage gilt sogar immer, wenn sich der Kassakurs oberhalb des langfristigen Gleichgewichtsniveaus befindet.[104] Festzuhalten ist außerdem, dass bei einer deterministischen Kassakursentwicklung die optimale Konsumstrategie weder durch die Situation auf dem Terminmarkt noch durch die Hedging-Effektivität beeinflusst wird.

2.4 Fazit und Ausblick

Die Ausführungen in den vorangegangenen Abschnitten haben elemantare Erkenntnisse im Rahmen des Futures-Hedgings geliefert. Zunächst wurde im Rahmen des Modells die Gültigkeit des Separationstheorems hinsichtlich der optimalen Konsum-

[104] Für diese Zustände ist die relative Änderung des Wertpapiers negativ, so dass die rechte Seite von (2.46) negative Werte aufweist.

und Hedgeentscheidung nachgewiesen. Ein Investor trifft seine dynamische Konsumentscheidung zu jedem Zeitpunkt auf Basis einer modifizierten Grenznutzenregel, der sogenannten Envelope-Bedingung. Bei optimaler Konsumstrategie stimmt der Grenznutzen des Konsums mit dem erwarteten Grenznutzen des Vermögens in jedem Zeitpunkt und Zustand überein.[105] Das Vermögen stellt dabei die Kapitalakkumulation für den optimalen zukünftigen Konsum dar. Seine optimale Absicherungspolitik trifft ein Investor unter Berücksichtigung verschiedener Faktoren: Dem Grad der Korrelation zwischen Kassa- und Futureskurs, der Situation auf dem Terminmarkt sowie den relevanten Zustandsvariablen. Formal lässt sich die optimale Hedge-Rate in eine präferenzfreie, eine spekulative und zwei preiselastische Komponenten zerlegen. Während die präferenzfreie Komponente das reine Absicherungsmotiv des Investors erfasst, wird die spekulative Komponente durch die MR-Sharpe-Ratio des Futures in Abhängigkeit von der intertemporalen relativen Risikoaversion determiniert. Die beiden preiselastischen Hedge-Terme sind essentiell, um das Konsumrisiko hinsichtlich der intertemporalen Konsumstrategie zu minimieren. Hierdurch sichert sich ein Investor gegenüber Risiken ab, die durch die beiden Zustandsvariablen induziert werden.

Weitere Analysen ergaben, dass der spekulative Hedge-Term und die futurespreiselastische Hedge-Komponente für dieselbe Terminmarktsituation entgegengesetzte Vorzeichen aufweisen und somit funktionale Komplemente in der Absicherungspolitik des Investors darstellen. Dagegen ist aufgrund des hohen Komplexitätsgrades des Modells ohne konkrete Spezifikation der Wertfunktion leider keine Vorzeichenanalyse hinsichtlich der (kassa-)preiselastischen Hedge-Komponente möglich. In Analogie zu Merton (1973) bzw. Breeden (1984) wurden die preiselastischen Hedge-Komponenten als erforderliche Vermögenskompensationen zur Aufrechterhaltung des gegenwärtigen Grenznutzens des Konsums bzw. des Erwartungsnutzens des lebenslangen Konsumstroms aufgefasst und sowohl formal als auch interpretatorisch voneinander abgegrenzt. Von erheblicher Relevanz war in diesem Zusammenhang der Terminus Schattenpreis, der die Auswirkung

[105] Vgl. Wahl and Broll (2003), S. 173.

einer marginalen Änderung einer Zustandsvariablen auf den Erwartungsnutzen des lebenslangen Konsumstroms angibt.

Der enge Zusammenhang zwischen dem spekulativen Hedge-Term und der futurespreiselastischen Hedge-Komponente motivierte die Fragestellung, ob die von dem Investor geforderte Vermögenskompensation hinsichtlich der Zustandsvariablen f größer sein kann als die auf dem Terminmarkt eingegangene spekulative Position. In diesem Fall ist das Ausmaß der Absicherung gegen das durch den Futures induzierte Konsumrisiko größer als das Ausmaß der Spekulationstätigkeit auf dem Terminmarkt. Investoren, deren Absicherungspolitik durch diese Eigenschaft gekennzeichnet ist, werden als spekulationsavers charakterisiert. Im Rahmen dieses Hedging-Modells wurde für spekulationsaverse Investoren das aus der Literatur des Futures-Hedgings allgemein gültige Resultat, indem es in einer MR-Backwardation (MR-Contango) Situation stets zu einem Underhedge (Overhedge) kommt, widerlegt. Diesbezüglich war es ausreichend, eine spezielle Korrelationsstruktur zwischen Kassa- und Futureskurs zu unterstellen. Um eindeutige Aussagen über das Ausmaß des Overhedgings (Underhedgings) abzuleiten, sind weitere Einschränkungen der Präferenzen notwendig. Dies stellt einen interessanten Ansatzpunkt für weitere Untersuchungen dar.

Des Weiteren wurde der Einfluss der Absicherungsmöglichkeit des Kassakursrisikos mittels Futures auf das optimale intertemporale Konsumprofil des Investors untersucht. Die klassische Keynes-Ramsey-Regel verliert ihre Gültigkeit; als Resultat ergibt sich eine stochastische Version der Keynes-Ramsey-Regel für die optimale Kapitalakkumulation. Diese berücksichtigt das Niveau des Kassa- und Terminkurses, das Ausmaß der Risikoaversion bzw. der (Kassa-)Preiselastizität sowie die Hedging-Effektivität. Ohne zusätzliche Annahmen ist die Interpretation der stochastischen Keynes-Ramsey-Regel (2.40) sehr diffizil. Insofern wurden für die Untersuchungen spezielle Szenarien bzw. konkrete Parameterkonstellationen unterstellt, um qualitative Aussagen hinsichtlich der optimalen Konsumstrategie treffen zu können. Für die Realisation des Kassa- und Futureskurses war insbesondere relevant, ob sie sich ober- oder unterhalb des langfristigen Gleichgewichtsniveaus befinden. In Abhängigkeit davon lässt sich

eine Aussage darüber treffen, ob durch die beiden ersten Terme auf der rechten Seite von Gleichung (2.40) die Sparquote des Investors ceteris paribus erhöht oder verringert wird. Die Zeitpräferenzrate geht negativ in die optimale Konsumwachstumsrate ein. Sie stellt einen Indikator für die Ungeduld hinsichtlich des Nutzens dar. Je größer (kleiner) die Zeitpräferenzrate, desto geringer ist die Bereitschaft des Investors Konsum in der Gegenwart gegen Konsum in der Zukunft zu tauschen. Der letzte Term der stochastischen Keynes-Ramsey-Regel berücksichtigt einerseits das Ausmaß der Risikoaversion sowie die (Kassa-)Preiselastizität, andererseits die Hedging-Effektivität und das Kassakursrisiko. Der Einfluss dieses Terms auf das optimale Konsumprofil ist umso geringer, je weniger imperfekt das Hedging bzw. je kleiner die Volatilität des Kassakurses ist. Der Grad der intertemporalen relativen Risikoaversion und (Kassa-) Preiselastizität werden maßgeblich durch die Wertfunktion determiniert. Diese wird im Rahmen des hier vorliegenden Modells nicht konkret spezifiziert, so dass eine eindeutige Aussage über den Einfluss der korrespondierenden Summe in der Gleichung (2.40) nicht möglich ist.

Chapter 3

Forecasting Multivariate Portfolio-Value-at-Risk using Smooth Nonparametric Bernstein Vine Copulas

3.1 Introduction

Following the growing criticism of elliptical models, copulas have emerged both in insurance and risk management as a powerful alternative for modeling the complete dependence structure of a multivariate distribution. In a nutshell, a copula is a multivariate cumulative distribution function (cdf) with uniform marginals that combines univariate marginal cdfs to a joint multivariate cdf. As the marginal cdfs do not necessarily have to be in the same distributional family, copulas allow for an extremely flexible modeling of multivariate distributions. Mainly due to this flexibility, the literature on copulas and their use in risk management applications has grown exponentially since the introduction of copulas to the field of finance by (among others) Li (2000) and Embrechts et al. (2002) with several studies concentrating on statistical inference and model selection for copulas (see, e. g., Kim et al., 2007, Genest et al., 2009b) as

well as applications (see, e. g., Chan and Kroese, 2010, Grundke and Polle, 2012, Ye et al., 2012).[106]

In practice, however, risk managers are often faced with the problem of modeling the joint loss distribution of large portfolios. As evidenced by the financial crisis of 2007-2008, misspecified dependence structures between financial time series can lead to disastrous consequences. Poon et al. (2004) show that the correct identification of the dependence structure is crucial in financial applications and use a nonparametric model for identifying and modeling the joint-tail distribution of stock returns. Copula models constitute an alternative approach to adequately capture the extreme tail dependence structures of multivariate financial data. But as simple parametric copulas are often not flexible enough to model the complex dependence structures, recent works by Joe (1997), Bedford and Cooke (2002) and Whelan (2004) have proposed copula models which are highly flexible but at the same time still tractable even in higher dimensions. Most notably, vine copulas (also called pair-copula constructions, PCC in short) have emerged as the most promising tool for modeling dependence structures in high dimensions.[107] Vine copulas consist of a cascade of conditional bivariate copulas (so called pair-copulas) that are combined hierarchically to yield a copula (and thus a cdf) that can be used for building a joint distributional model for a data set. In contrast to simple copulas like, e. g., the Gaussian copula inherent in a multivariate normal distribution, however, a vine copula consists of a set of pair-copulas which can each be chosen from a different parametric copula family. As a result, vine copulas are extremely flexible yet still tractable even in high dimensions as all computations necessary in statistical inference are performed on bivariate data sets (see Aas et al., 2009, for a first discussion of vine copulas in an applied setting).

[106] A literature review with a special emphasis on finance-related papers using copulas is given by Genest et al. (2009a). An overview of the different branches of the copula literature is given by Embrechts (2009).

[107] Competing modeling concepts like nested and hierarchical Archimedean copulas are analyzed by Aas and Berg (2009) as well as Fischer et al. (2009). They conjecture that vine copulas should be preferred over nested or hierarchical Archimedean copulas.

Similar to the bivariate case,[108] the correct selection of the parametric constituents of the vine, i. e., the pair-copulas, is crucial for the correct specification of a vine copula model. In case of the popular C- and D-vine specifications, the calibration and estimation of a d-dimensional vine requires to solve the problem of selecting and estimating $d(d-1)/2$ different pair-copulas from the set of candidate bivariate parametric copula families. Thus, a vine model's increased flexibility only comes at the expense of an increased model risk.

As a remedy, recent studies have suggested to select the parametric pair-copulas based on graphical data inspection and goodness-of-fit tests (Aas et al., 2009) and to employ sequential heuristics based on Akaike's Information Criterion (AIC) (see, e. g., Brechmann et al., 2012, Dissmann et al., 2013). Kurowicka (2011) and Brechmann et al. (2012) propose strategies for simplifying vines by replacing certain pair-copulas by the independence copula (yielding a *truncated* vine copula) or the Gaussian copula (yielding a *simplified vine*). Finally, Hobæk-Haff and Segers (2012) propose the use of empirical pair-copulas in vine models to circumvent the problem of selecting parametric pair-copulas.

In this paper, we propose a simple alternative to the selection of the pair-copulas of a vine copula model from a set of parametric copula families. We use the recently proposed nonparametric Bernstein copulas (Sancetta and Satchell, 2004, Pfeifer et al., 2009, Diers et al., 2012) as pair-copulas (thus circumventing the need to choose parametric ones) yielding smooth nonparametric vine copula models that do not require the specification of parametric families. Thus, we extend the ideas laid out by Hobæk-Haff and Segers (2012) by using an approximation to the empirical pair-copulas. In contrast to their work, however, we approximate the pair-copulas not only non-parametrically but also by the use of continuous functions.[109] In addition, especially the Bernstein

[108] See Genest et al. (2009b) and Weiß (2013) for discussions of the problem of selecting the best fitting parametric copula.

[109] Using smooth functions to approximate the true underlying dependence structure is in line with our intuition. The superiority of smooth nonparametric approximations of the copula over simple empirical copulas, however, is also found by Shen et al. (2008). They argue that the improved approximation by the linear B-spline copulas is due to their Lipschitz continuity.

copula has recently attracted attention in insurance modeling (Diers et al., 2012) and has already proven its merits in an applied setting. Therefore, the contributions of the proposed smooth nonparametric vine copulas are twofold: First, the use of Bernstein copulas completely obviates the need for the error-prone selection of pair-copulas from pre-specified sets of parametric copulas. The resulting smooth and nonparametric vine copulas do not only constitute extremely flexible tools for modeling high-dimensional dependence structures, they are also characterized by a smaller model risk than their parametric counterparts. Second, Bernstein copulas have been shown to improve on the estimation of the underlying dependence structure by competing nonparametric empirical copulas.[110] The modeling of a vine model's pair-copulas by the use of smooth approximating functions is thus a natural extension of recently proposed (highly discontinuous) empirical pair-copulas.

By means of a simulation study, we illustrate that the proposed Bernstein vine model is equally well suited for the task of approximating the true dependence structure of a data set than a benchmark heuristic in low dimensions and outperforms the benchmark in higher dimensions. In our empirical study, we show that our proposed model accurately forecasts the Value-at-Risk (VaR) for multivariate portfolios consisting of commodities, stocks, and bonds. Thus, this paper contributes significantly to the current state of the art by using copula models to measure dependence structures between international equity markets (see, e. g., Rodriguez, 2007) and the cross-market linkages between equity and commodity markets (see, e. g., Delatte and Lopez, 2013).[111] In contrast to the heuristic benchmark, our Bernstein vine model works reliably without running into the numerical problems of the heuristic benchmark (i. e., non-convergence of the maximization of the log-likelihood function).

[110] For example, Bernstein copulas provide a higher rate of consistency than other common nonparametric estimators and do not suffer from boundary bias (Kulpa, 1999, Sancetta and Satchell, 2004, Diers et al., 2012). Similarly, other approximations as, e. g., linear B-spline copulas have also been shown to yield lower average squared approximation errors than competing discrete approximations (Shen et al., 2008).

[111] In a multitude of studies, copula models have also been used to examine the co-movements between the equity market and the foreign exchange market (see, e. g., Ning, 2010).

The results presented in this study show that our proposed vine copula model with smooth nonparametric Bernstein pair-copulas outperforms the benchmark model with parametric pair-copulas in higher dimensions with respect to the accuracy and numerical stability of the approximation to the true underlying dependence structure. While our nonparametric vine copula model yields slightly worse average squared errors than a benchmark vine copula calibrated by selecting parametric pair-copulas based on AIC values in lower dimensions (e. g., $d = 3, 5, 7$) in our simulations, this result is reversed in higher dimensions. For random vectors of dimension $d = 11$ and higher, the parametric benchmark did not produce any results in more than 50% of the simulations as the algorithms used for optimizing the log-likelihood functions did not converge. In higher dimensions (i. e., the main field of application of vine copulas), our nonparametric modeling approach is thus clearly superior to a parametric vine copula model. Our risk management application, however, shows that even in lower dimensions ($d = 5$) our nonparametric model yields VaR-forecasts that are not rejected by a range of formal statistical backtests. Consequently, the slightly worse approximation errors of our nonparametric model in lower dimensions do not seem to affect the modeling of a given dependence structure too severely thus underlining the usefulness of our proposed model.

The remainder of this article is structured as follows. Section 3.2 introduces vine copulas and Bernstein copulas. Section 3.3 presents the results of a simulation study on the approximation errors of both our nonparametric Bernstein vine copula model as well as a heuristically calibrated parametric benchmark model. In Section 3.4, we conduct an empirical analysis for a five-dimensional financial portfolio. Section 3.5 concludes.

3.2 Problem formulation

3.2.1 Portfolio Value-at-Risk

Consider an investor holding a portfolio of d financial assets such as stocks, bonds and commodities. It is a well-known stylized fact that both the univariate and joint distributions of the returns on the financial assets will usually not be normal. However, risk managers and investors are highly interested in modeling the multivariate joint distribution of the portfolio returns to evaluate portfolio risk and construct trading and hedging strategies based on the portfolio's risk exposure.

Therefore, let $\mathbf{X} = (X_1, \ldots, X_d)$ be a vector of d univariate continuous random variables X_i ($i = 1, \ldots, d$). In a risk management setting, the variables X_i are commonly interpreted as the return (or loss) of the ith financial asset in a portfolio. Then, $F_i(x_i)$ is the cumulative distribution function of the random variable X_i and $f(x_i)$ denotes the corresponding (marginal) probability density function (pdf) of X_i. The joint distribution function F is given by

$$F(\mathbf{x}) = P(X_1 \leq x_1, \ldots, X_d \leq x_d), \qquad \forall\, \mathbf{x} = (x_1, \ldots, x_d), \tag{3.1}$$

with the corresponding density being

$$f(\mathbf{x}) = \frac{\partial^d F(\mathbf{x})}{\partial x_1 \ldots \partial x_d}. \tag{3.2}$$

The ultimate objective of our paper is to accurately forecast the Value-at-Risk of a portfolio $Y = \sum_{i=1}^{d} \omega_i X_i$ composed of d assets for a given significance level $\alpha \in (0, 1)$ and given portfolio weights ω_i with $\omega_i \geq 0$ and $\sum_{i=1}^{d} \omega_i = 1$. The Value-at-Risk of the portfolio is defined by

$$VaR_\alpha(Y) = Q_\alpha = \inf\{y \in \mathbb{R} : F(y) \geq \alpha\}, \tag{3.3}$$

where Q_α denotes the α-quantile of Y. In order to estimate this quantile, we need to specify the joint distribution of the portfolio constituents $\mathbf{X} = (X_1, \ldots, X_d)$.[112] Next, we shortly introduce the fundamentals of copulas and describe our proposed model for the joint distribution of $\mathbf{X}$.

3.2.2 Copulas and vine copulas

Copulas can be used to separate the stochastic modeling of a multivariate distribution into the tasks of fitting the univariate marginals and finding a suitable copula to represent the dependence structure. The fundamental result in copula theory is given by Sklar's Theorem (1959).[113] The key idea is to split a d-dimensional distribution function F in two parts, the marginal distribution functions F_i and a copula C which is a d-variate cdf on $[0;1]^d$ with uniform marginals and which fully describes the dependence structure of the distribution, i. e., Sklar's theorem states that

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{3.4}$$

Similarly, the joint density can be represented by

$$f(\mathbf{x}) = c(F_1(x_1), \ldots, F_d(x_d)) \prod_{i=1}^{d} f_i(x_i), \tag{3.5}$$

where $c(u_1, \ldots, u_d)$ is the d-variate copula density given by $\frac{\partial C(u_1,\ldots,u_d)}{\partial u_1 \ldots \partial u_d}$, $u_1, \ldots, u_d \in [0;1]$, and f_i $(i = 1, \ldots, d)$ are the marginal densities.

We now describe the special case of vine copulas which constitute one possibility for decomposing the multivariate density in (3.5) into a product of conditional bivari-

[112] One could think about using a coherent risk measure like, e. g., the Expected Shortfall or measures like the Omega score (see, e. g., Brown et al., 2009) in our study instead of the portfolio's Value-at-Risk. However, Value-at-Risk continues to be a mandatory risk measure for banks in the discussions on the Third Basel Accord ("Basel III") even though regulators now require banks to calculate a so-called "stressed VaR" to account for an inadequate modeling of risks using VaR during times of market stress. We therefore restrict our analysis to the forecasting of the VaR of the portfolio.

[113] See Rüschendorf (2013) for a complete formulation of Sklar's theorem.

ate copulas[114] and marginal densities. Starting point is the observation that a joint probability density function of dimension d can be expressed as follows

$$f(\mathbf{x}) = f(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_1, x_2) \cdot \ldots \cdot f(x_d|x_1, \ldots, x_{d-1}). \tag{3.6}$$

Given the commonly used assumption that conditional copulas do not depend on the values of variables that are conditioned on,[115] each factor in this product can then be decomposed further using a conditional copula, i. e.,

$$f(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \tag{3.7}$$

$$f(x_3|x_1, x_2) = c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_3(x_3) \tag{3.8}$$

and more generally

$$f(x|\mathbf{v}) = c_{xv_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})) \cdot f(x|\mathbf{v}_{-j}) \tag{3.9}$$

with $c_{12}(\cdot)$ being the (in this case unconditional) copula density of (x_1, x_2), $c_{23|1}(\cdot)$ being the conditional copula density of (x_2, x_3) given x_1, and $\mathbf{v}$ being a d-dimensional vector, where v_j is an arbitrarily chosen component of $\mathbf{v}$ and $\mathbf{v}_{-j}$ denotes the v-vector, excluding this component.

For dimension $d = 3$, substituting the elements of the initial decomposition in (3.6) with the conditional copulas yields the representation

$$\begin{aligned} f(x_1, x_2, x_3) = {} & c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \\ & \cdot \; c_{12}(F_1(x_1), F_2(x_2)) \\ & \cdot \; c_{13}(F_1(x_1), F_3(x_3)) \\ & \cdot \; f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \end{aligned} \tag{3.10}$$

[114] The conditional copula of $(X_1, X_2)|Y = y$, where $X_1|Y = y \sim F_{1|Y}(\cdot|y)$ and $X_2|Y = y \sim F_{2|Y}(\cdot|y)$, is the joint distribution function of $U_1 = F_{1|Y}(X_1|y)$ and $U_2 = F_{2|Y}(X_2|y)$ given $Y = y$.

[115] This simplifying assumption is made to keep PCCs tractable for inference and was previously studied by Hobæk-Haff et al. (2010).

with c_{12}, c_{13} and $c_{23|1}$ as *pair-copulas*. Note that as there are several possible decompositions of the conditional distributions, there exist an exponential number of different vine models with which the joint density of $\mathbf{X}$ can be represented depending on the variables one chooses to condition on.

Two popular classes of vines are given by the so-called C- and D-vines,[116] which will be used later on in our application to financial market data. We concentrate on these vine types as they are the most frequently used vine models in applications (see, e. g., Chollete et al., 2009, Heinen and Valdesogo, 2009, Weiß and Supper, 2013, for different applications of vines in asset pricing and risk management). The basic idea to use nonparametric Bernstein copulas as pair-copulas, however, can also be applied to any vine type. In our study, we employ both C- and D-vines to analyze the potential effect of the vine type on our main results.

For a Canonical or C-vine, the joint density is represented by

$$f(\mathbf{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,\ldots,j-1}(F(x_j|x_1,\ldots,x_{j-1}), F(x_{j+i}|x_1,\ldots,x_{j-1})), \quad (3.11)$$

while the corresponding representation with a D-vine is given by

$$f(\mathbf{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}(F(x_i|x_{i+1},\ldots,x_{i+j-1}), F(x_{i+j}|x_{i+1},\ldots,x_{i+j-1})). \quad (3.12)$$

Note that the sets $\{i,\ldots,j\}$ and $\left\{x_i,\ldots,x_j\right\}$ with $i > j$ are set to empty, i. e., the corresponding copula densities and cdfs are unconditional.

Examples of possible decompositions of a five-dimensional random vector via a C- and D-vine copula are shown in Figure 3.1 and 3.2, respectively.[117]

[116] The classes of C- and D-vines are subsets of the so-called regular vines (or R-vines in short, see Brechmann et al., 2012). We do not consider other types of R-vines in this paper but note that our proposed use of smooth Bernstein and B-spline copulas as pair-copulas can also be extended to other subsets of R-vines.

[117] Both figures illustrate the fact that the possible decompositions of a d-dimensional joint density via a vine model can be represented as a nested set of trees (see Bedford and Cooke, 2002).

Figure 3.1: **Five-dimensional C-vine copula.**

The figure shows an example of a five-dimensional C-vine copula with five random variables, four trees and ten edges. The nodes in the first tree correspond to the five random variables that are being modeled and each edge corresponds to a bivariate conditional or unconditional pair-copula.

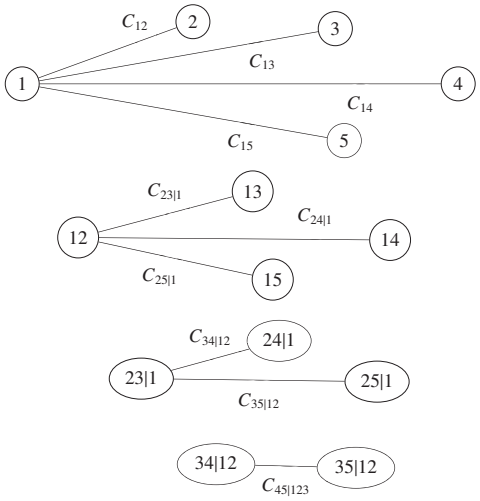
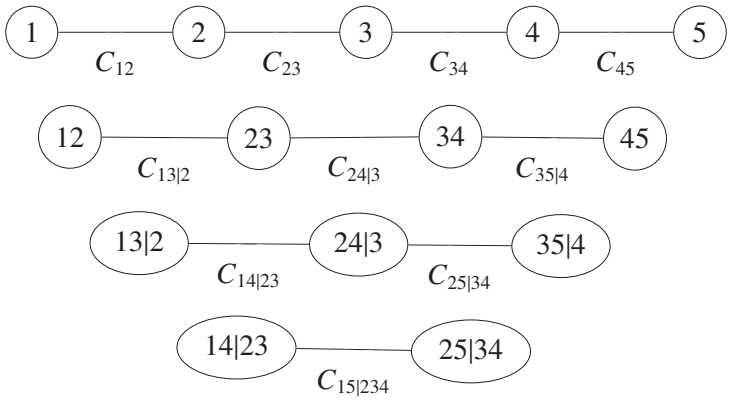


Figure 3.2: **Five-dimensional D-vine copula.**

The figure shows an example of a five-dimensional D-vine copula with five random variables, four trees and ten edges. The nodes in the first tree correspond to the five random variables that are being modeled and each edge corresponds to a bivariate conditional or unconditional pair-copula.



As is clear from (3.11) and (3.12) and from Figures 3.1 and 3.2, respectively, the dependence structure in a data sample is captured in the bivariate pair-copulas in a vine model. As no assumptions are placed on the parametric forms of the pair-copulas, vine copulas allow for the modeling of highly complex dependence structures in high dimensions.[118]

[118] See also the excellent studies by Joe (1997) and Bedford and Cooke (2002) for further properties of vine copulas.

We now demonstrate how to fit a vine copula model to a given dataset in three separate steps. First, one needs to select the tree structure of the vine model. Although C- and D-vines possess a fixed tree structure, a different permutation of the variable indices in Figures 3.1 and 3.2 will yield different copulas. Consequently, the selection of the tree structure for a C- or D-vine amounts to the selection of a permutation of the indices $1, \ldots, d$ of the random variables. As such, for a d-dimensional random vector their exist $d!/2$ different C- and D-vines, respectively (see Aas et al., 2009).[119] Once a permutation has been chosen, the structure of the vine is fully specified. Second, the statistician has to select $d(d-1)/2$ bivariate pair-copulas from candidate copula families. In the last step, the parameters of the pair-copulas have to be estimated. To select the optimal tree structure, Dissmann et al. (2013) propose a heuristic procedure in which the tree structure is chosen via a maximum spanning tree algorithm that maximizes the sum of the absolute empirical Kendall's τ of all possible variable pairs on a given level of the tree. For the selection of the parametric pair-copulas, Brechmann and Czado (2013) and Dissmann et al. (2013) propose a sequential heuristic which selects the best fitting parametric copula family for each pair-copula based on the candidate copulas' AIC values. Although this sequential selection of parametric pair-copulas using AIC does not necessarily yield a globally optimal AIC value for the vine, Brechmann and Czado (2013) show that this heuristic yields considerably better results than a selection algorithm based on copula goodness-of-fit tests.[120]

In the following, we concentrate on the problem of selecting the bivariate pair-copulas. It is obvious that the accuracy of the vine model critically depends on the correct specification of the pair-copulas. Our key idea is to substitute the (parametric) bivariate pair-copulas with smooth nonparametric estimates of the underlying (pairwise) dependence structures. As a benchmark to our proposed nonparametric method, we employ the sequential heuristic by Brechmann and Czado (2013) and Dissmann

[119] Choosing the best fitting tree structure manually thus quickly becomes unfeasible in higher dimension (see Dissmann et al., 2013).

[120] Results by Grundke and Polle (2012) underline this finding as their empirical results cast additional doubt on the ability of copula goodness-of-fit tests to identify stressed risk dependencies.

et al. (2013), which first fits a fixed number of previously chosen candidate parametric copula families to the data. Then, for each pair-copula, the candidate parametric copula family that yields the lowest AIC value is chosen. In contrast, in our proposed approach, no parametric copula family needs to be chosen for a given pair-copula as all pair-copulas are modeled nonparametrically. Thus, we expect our nonparametric approach to improve on the heuristic benchmark for several reasons. First, the nonparametric modeling of the pair-copulas completely eliminates the model risk of choosing an incorrect parametric family for a given pair-copula. Second, Grønneberg and Hjort (2014) prove that the use of AIC as a model selection criterion is not correct in case rank-transformed pseudo-observations are used (as is common in almost all applications of copulas in finance).[121] Third, as already hinted at by Dissmann et al. (2013), the incorrect specification of the parametric pair-copulas in the upper levels of a vine's tree structure can lead to a propagation and amplification of rounding errors causing the heuristic to yield extremely inaccurate results.

In the next subsection, we define and discuss Bernstein copulas which we use as smooth nonparametric estimates of the pair-copulas in a vine model.

3.2.3 Bernstein copulas

As a nonparametric candidate for the pair-copulas in (3.11) and (3.12), we consider the recently proposed Bernstein copulas. In the following, we briefly state some basic mathematical facts on Bernstein polynomials and Bernstein copulas, respectively. The Bernstein polynomials of degree m are defined as

$$B_{m,k}(z) = \binom{m}{k} z^k (1-z)^{m-k}, \tag{3.13}$$

where $k = 0, \ldots, m \in \mathbb{N}$ and $0 \leq z \leq 1$.

[121] Given a data sample $\left\{x_{ij}\right\}_{i=1,\ldots,d;\, j=1,\ldots,n}$ of a d-dimensional random vector and sample size n, a pseudo-sample of observations of the underlying copula is given by $\left\{u_{ij}\right\}_{i=1,\ldots,d;\, j=1,\ldots,n}$ with $u_{ij} = R_{ij}/(n+1)$ where R_{ij} is the rank of x_{ij} among $x_{i1}, \ldots, x_{in}$ (see Genest et al., 2009b).

As our focus lies on the nonparametric modeling of pair-copulas in vines, we restrict our analysis in the following to bivariate Bernstein copulas. Let $\mathbf{U} = (U_1, U_2)$ denote a discrete bivariate random vector with uniform margins over $T_i := \{0, 1, \dots m_i - 1\}$ with grid size $m_i \in \mathbb{N}$ and $i = 1, 2$. In our analysis, we later choose $m_1 := m_2 := m = const$. We then define

$$p(k_1, k_2) := P\left(\bigcap_{i=1}^{2}\{U_i = k_i\}\right), \quad (k_1, k_2) \in \bigtimes_{i=1}^{2} T_i. \tag{3.14}$$

Then

$$c(u_1, u_2) := \sum_{k_1=0}^{m_1-1}\sum_{k_2=0}^{m_2-1} p(k_1, k_2)\prod_{i=1}^{2} m_i B_{m_i-1,k_i}(u_i), \quad (u_1, u_2) \in [0, 1]^2 \tag{3.15}$$

defines the density of a two-dimensional copula which we refer to as a Bernstein copula. More precisely, we refer to $c(u_1, u_2)$ as the Bernstein copula density induced by $\mathbf{U}$. Pfeifer et al. (2009) show by integrating expression (3.15) that the Bernstein copula induced by $\mathbf{U}$ is itself given by

$$\begin{aligned} C(v_1, v_2) &:= \int_0^{v_2}\int_0^{v_1} c(u_1, u_2)du_1du_2 \\ &= \sum_{k_1=0}^{m_1}\sum_{k_2=0}^{m_2} P\left(\bigcap_{i=1}^{2}\{U_i < k_i\}\right)\prod_{i=1}^{2} B_{m_i,k_i}(v_i) \end{aligned} \tag{3.16}$$

for $(v_1, v_2) \in [0, 1]^2$.[122]

The general definition of a Bernstein copula given in equation (3.16) assumes that the copula is induced by an arbitrary vector $\mathbf{U}$ on a grid $\bigtimes_{i=1}^{2} T_i$ and the Bernstein polynomials are used to approximate a set of observations of $\mathbf{U}$ on the grid points (k_1, k_2). To approximate the true copula underlying a given data sample, a natural choice for the sample of $\mathbf{U}$ is a pseudo-sample of the (discrete) empirical copula process as defined, e. g., by Deheuvels (1979).

[122] Note that in order to smoothly approximate the distribution or density of a copula in (3.15) and (3.16), very high degrees for the Bernstein polynomials have to be chosen.

Let $\left\{u_{ij}\right\}_{i=1,2;j=1,\ldots,n}$ be a bivariate pseudo-sample of the copula inherent in a given data set as defined above. The empirical copula process of Deheuvels (1979, 1981) is then defined as

$$C_n(u_1, u_2) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\left(u_{1j} \leq u_1; u_{2j} \leq u_2\right) \tag{3.17}$$

for $(u_1, u_2) \in [0; 1]^2$. The corresponding empirical copula frequency (density) is then given by

$$c_n\left(\frac{i}{n}, \frac{j}{n}\right) = \begin{cases} 1/n, & \left(\frac{i}{n}, \frac{j}{n}\right) \in \left\{u_{ij}\right\}, \\ 0, & \text{otherwise}, \end{cases} \tag{3.18}$$

for $i, j = 1, \ldots, n$. Similar to the empirical cdf, the empirical copula process given in (3.17) is a consistent (discrete) estimator of the true underlying copula in a data sample (Deheuvels, 1979, 1981). Consequently, a continuous approximation of the true underlying copula of a sample is given by the Bernstein copula induced by a pseudo-sample of the empirical copula process. Note that, in contrast to popular copulas like the Gaussian or Student's t copula, the Bernstein copula based on the empirical copula process is completely nonparametric as one only requires the rank statistics of the original sample and both the Bernstein copula density and the Bernstein copula themselves do not depend on any parameters that need to be estimated.

To approximate the dependence structure in a given data sample of size n, we first compute the pseudo-sample $\left\{u_{ij}\right\}_{i=1,2;j=1,\ldots,n}$ and the values $c_n\left(\frac{i}{n}, \frac{j}{n}\right)$ of the empirical copula frequency. Next, the unity square is segmented into a grid of $m \times m$ cells and the values $c_n\left(\frac{i}{n}, \frac{j}{n}\right)$ are inserted into the respective cells according to the arguments of c_n. The frequencies in each cell are then summed up to yield the entries $[a_{kl}]_{k,l=1,\ldots,m}$ of a $m \times m$ matrix which will be referred to as a contingency table. As C_n equals the true copula C of the data only asymptotically, the discrete approximation via the contingency table $[a_{kl}]$ will not necessarily have uniform marginals.[123] To circumvent this problem, Pfeifer et al. (2009) propose to transform the contingency table $[a_{kl}]$

[123] Of course, this problem is more pronounced the smaller the sample size n is.

into a (possibly suboptimal) new contingency table $[x_{kl}]$ with uniform marginals via a Lagrange optimization approach yielding

$$x_{kl} = a_{kl} - \frac{a_{\cdot l}}{m} - \frac{a_{k\cdot}}{m} + \frac{2}{m^2} \qquad \text{for } k, l = 1, \dots, m, \tag{3.19}$$

where the index $\cdot$ denotes summation. Note that the quality of the Lagrange solution is reduced by an increasing number of the sample size n.[124] We therefore chose to employ a different optimization strategy to correct for the non-uniform distribution of the marginals.

Consequently, we calculate the approximation $[x_{kl}]$ to the contingency table $[a_{kl}]$ by solving the following optimization problem:

$$\sum_{k=1}^{m} \sum_{l=1}^{m} (x_{kl} - a_{kl})^2 \longrightarrow \min \tag{3.20}$$

subject to

$$\sum_{k=1}^{m} x_{kj} = \sum_{l=1}^{m} x_{il} = \frac{1}{m} \qquad \text{and} \qquad x_{ij} \geq 0 \quad \text{for } i, j = 1, \dots, m. \tag{3.21}$$

To solve for the $[x_{kl}]$, we make use of the quadratic optimization algorithm of Goldfarb and Idnani (1982). In preliminary tests, the found solutions to this optimization problem yielded significantly lower quadratic errors than the procedure initially proposed by Pfeifer et al. (2009) thus confirming the need for a more refined optimization strategy. Finally, the optimized contingency table $[x_{kl}]$ is used to define the joint distribution of the discrete random vector $\mathbf{U}$ and to induce the Bernstein copula in equation (3.16).

To use Bernstein copulas as pair-copulas both in our simulation study and the empirical application, we require efficient algorithms for simulating and evaluating the density and distribution of a given vine copula. To this end, we adapt the algorithms initially proposed by Aas et al. (2009) by substituting the parametric h-hunctions (i. e.,

[124] In unreported results, the optimization strategy of Pfeifer et al. (2009) proved to yield only suboptimal results.

the partial derivatives of the copula densities) in these algorithms by the partial derivatives of the fitted bivariate Bernstein copulas.

3.2.4 Fitting and simulating from a vine copula

For the purpose of simulating observations from a vine copula, we require the so-called h-function of two uniform variables x and v. The h-function $h(x, v, \boldsymbol{\theta})$ represents the conditional distribution function of a bivariate copula and can be written as

$$h(x, v, \boldsymbol{\theta}) = F(x|v) = \frac{\partial C_{xv}(x, v)}{\partial v}, \tag{3.22}$$

with $\boldsymbol{\theta}$ being the set of parameters for the copula of the joint distribution function of x and v. The second parameter of $h(\cdot)$ always corresponds to the conditioning variable. This representation can be generalized further to the case of a vector of conditioning variables yielding

$$F(x|\mathbf{v}) = \frac{\partial C_{xv_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})}, \tag{3.23}$$

where $C_{ij|k}$ is a bivariate distribution function.

In the following, we first focus on the simulation algorithm initially proposed by Aas et al. (2009), with the objective of sampling d observations $u_1, \ldots, u_d$ from a given vine copula.[125] Note that simulating from a d-dimensional vine model requires a fully specified PCC, i. e., a given tree structure and a set of pair-copulas. In other words, in addition to a permutation of the d random variables that determines the tree structure of the vine, $d(d-1)/2$ bivariate copulas must be specified parametrically or nonparametrically. In a parametric vine model, the pair-copulas are selected from a predefined set of candidate copula families and a related vector of parameters $\boldsymbol{\theta}$. In the nonparametric case, the pair-copulas are modeled using the Bernstein copula with corresponding contingency tables.

[125] See Bedford and Cooke (2002), and Kurowicka (2011) for a review of the fundamentals of vine simulations.

We start by describing the main characteristics of the generic algorithm that is applicable to all vine types (Algorithm 1). The respective implementations for sampling from a canonical vine and for a D-Vine are given in Algorithm 3 (Appendix B.1) and in Algorithm 4 (Appendix B.2), respectively. In particular, the crucial evaluation of the conditional distribution functions $F(x|v)$ by using the h-function is shown. For a pair-copula in tree $j+1$ ($j \in \{1, \dots, d-1\}$), one has to evaluate $F(x|v)$ for a j-dimensional vector v. This is achieved by repeatedly applying $h(x|v, \boldsymbol{\theta}) = F(x|v)$ (see (3.22)). Note that in (3.23), v_j is an arbitrary component of v and v_{-j} is the $(j-1)$-dimensional vector v excluding v_j. Also note that the required copulas $C_{xv_j|v_{-j}}$ in (3.23) for tree $j+1$ are bivariate copulas with parameters/contingency tables determined in tree j.

Next, we demonstrate how to fit a vine model to a given dataset. In the following, we assume the vine type and the tree structure (i. e., the permutation of the data variables) to be known and concentrate on the selection of the pair-copulas. We first focus on a parametric model fitting, i. e., we try to find the best $d(d-1)/2$ suitable bivariate parametric copulas for describing the observed dependence structures. For each pair-copula, we need to select a parametric copula family (e. g., the Gaussian or Student's t-copula) and then estimate its parameters in a second step. Both steps are implemented simultaneously by the sequential heuristic proposed by Brechmann and Czado (2013) and Dissmann et al. (2013). In the following, we refer to this procedure as $biparfit(\cdot)$ and it can be summarized as follows: For each edge i ($i = 1, \dots, d-j$) (bivariate copula) of tree T_j ($j = 1, \dots, d-1$) of the vine, the parameters of each copula in a prespecified set of candidate parametric copula families are estimated using the original data $\mathbf{x}$ (tree T_1) or data that have been transformed via the use of the h-function in the previous tree T_{j-1} via maximum likelihood estimation. Then, for each of the $d(d-1)/2$ edges, the best fitting parametric copula (the copula family as well as its parameter(s)) is selected based on Akaike's Information Criterion (AIC). Brechmann and Czado (2013) and Dissmann et al. (2013) show that the resulting vine copula PCC_{AIC} performs exceptionally well in simulations.

Algorithm 1: Simulating d observations from a C- or D-vine model

Input: d, fam, $\boldsymbol{\theta}$

Let d be the dimension of the vine. If the underlying vine is parametrically specified, fam denotes the set of $d(d-1)/2$ bivariate parametric copula families defining the PCC structure and $\boldsymbol{\theta}$ the set of the related parameters. In case of a Bernstein PCC, fam denotes $d(d-1)/2$ nonparametric Bernstein copulas and $\boldsymbol{\theta}$ contains the corresponding contingency tables.

Output: **u**

d dependent uniform $[0;1]$ variables $(u_1, \ldots, u_d) =:$ **u**

begin

Draw a sample $w_1, \ldots, w_d$ of size $d \in \mathbb{N}$, independent uniform on [0;1].

Set

$$\begin{aligned} u_1 &= w_1, \\ u_2 &= F^{-1}(w_2|u_1), \\ u_3 &= F^{-1}(w_3|u_1, u_2), \\ &\vdots \\ u_d &= F^{-1}(w_d|u_1, \ldots, u_{d-1}). \end{aligned} \tag{3.24}$$

Calculate the quantile function F^{-1} of the cdf F in the following way.

for $i = 2$ *to* d **do**

Use the pre-specified pair-copulas to determine $F(u_i|u_1, \ldots, u_{i-1})$ and the corresponding inverse function $F^{-1}(u_i|u_1, \ldots, u_{i-1})$, by applying the definition of the h-function in (3.22) and the relation (3.23) recursively for both vine structures.

The choice of the v_i variable in (3.23) depends on the vine type:

if *type* ="C" **then**

$$\text{Set } F(u_i|u_1, \ldots, u_{i-1}) = \frac{\partial C_{i,i-1|1,\ldots,i-2}(F(u_i|u_1, \ldots, u_{i-2}), F(u_{i-1}|x_1, \ldots, u_{i-2}))}{\partial F(u_{i-1}|u_1, \ldots, u_{i-2})}.$$

else *type* = "D"

$$\text{Set } F(u_i|u_1, \ldots, u_{i-1}) = \frac{\partial C_{i,1|2,\ldots,i-1}(F(u_i|u_2, \ldots, u_{i-1}), F(u_1|u_2, \ldots, u_{i-1}))}{\partial F(u_1|u_2, \ldots, u_{i-1})}.$$

end

end

In contrast, in our nonparametric approach, the bivariate copula densities in the decomposed joint distribution density in (3.11) or (3.12) are replaced by nonparametric Bernstein copulas. Note that our nonparametric approach does neither require the selection of the copula families nor the estimation of parameters for the pair-copulas. These key differences of our nonparametric approach are taken into account by the procedure $binonparfit(\cdot)$ which models each pair-copula nonparametrically. For each edge i ($i = 1, \ldots, d - j$) (bivariate copula) of tree T_j ($j = 1, \ldots, d - 1$) of the vine, the Bernstein copula density (3.15) is fitted to the original data $\mathbf{x}$ (tree T_1) or data that have been transformed via the use of the h-function in the previous tree T_{j-1} via the quadratic optimization algorithm in (3.20). This results in the nonprametric vine model $PCC_{Bernstein}$ that consists of $d(d-1)/2$ bivariate Bernstein copulas with corresponding contingency tables.

Again, we only present the main characteristics of the algorithm in the following and present the implementations for both C- and D-vines in algorithms 3 and 4 in the Appendix. Note that similar to Algorithm 1, evaluations of the conditional distribution functions $F(x|v)$ are required. Again, this is done by sequentially applying relation (3.23).

Algorithm 2: (Fitting a C- or D-vine copula)

Input: $class$, $type$, $data$, $bifit(\cdot)$

Let $class$ be a binary coded variable that indicates whether the model should be fitted parametrically (= 1) or nonparametrically (= 0). $type \in$ {"C", "D"} denotes the vine type (C-Vine or D-Vine), and $data = (\mathbf{x}_1, \ldots, \mathbf{x}_d)$ is a $n \times d$ sample being the basis for performing the model fit. For a parametric modeling of the vine ($class = 1$), the procedure $bifit(\cdot)$ is set to $biparfit(\cdot)$. In case the vine is modeled nonparametrically ($class = 0$), the procedure $binonparfit(\cdot)$ is used in place of $bifit(\cdot)$.

Output: PCC

PCC is an array of $d(d-1)/2$ bivariate fitted (parametric or nonparametric) copulas.

begin
- $k = 1$;
- Start in the first tree T_1;
- **for** *edge* $i = 1$ *to* $d - 1$ **do**
  - Determine the original data pairs subject to the vine type for edge i;
  - **if** $type =$"C" **then** Set $\mathbf{y} = (\mathbf{x}_1, \mathbf{x}_{i+1})$;
  - **else** Set $\mathbf{y} = (\mathbf{x}_i, \mathbf{x}_{i+1})$;
  - Put $\mathbf{y}$ into $bifit(\cdot)$ to get the best fitting bivariate copula PCC_k for edge i with copula parameter(s)/contingency table $\theta_{1,i}$;
  - $k = k + 1$;
  - Set $v_{1,i} = h(\mathbf{y}, \theta_{1,i})$ to transform the original data pair of edge i to calculate the observations $v_{1,i}$ required for tree T_2;
- **end**
- Run through the remaining trees in the following manner;
- **for** *trees* $j = 2$ *to* $d - 1$ **do**
  - **for** *edge* $i = 1$ *to* $d - j$ **do**
    - Estimate the best fitting copula PCC_k with copula parameter(s)/contingency table $\theta_{j,i}$ by applying $bifit(\cdot)$ to the transformed data pair $\mathbf{y} = (v_{j-1,i}, v_{j-1,i+1})$ from the previous tree T_{j-1};
    - k=k+1;
  - **end**
  - **if** $j < d$ **then**
    - Set $v_{j,i} = h(\mathbf{y}, \theta_{j,i})$ to calculate the data needed for tree T_{j+1};
  - **end**
- **end**
- **return** PCC

end

3.3 Simulations

In this section, we illustrate the superiority of the smooth nonparametric vine model over the sequential heuristic of Brechmann et al. (2012) and Dissmann et al. (2013) for selecting the pair-copulas in a vine parametrically. We first demonstrate the setup of the simulation study which follows a similar procedure laid out in Shen et al. (2008). The results of the simulations are given subsequently.

3.3.1 Design of the simulation study

We first describe the data generating process (DGP) for our simulation study. In our simulations, we repeat our analysis for both C-vines and D-vines to investigate the question whether our results are significantly affected by the type of vine. For each of the two vine types (and consequently for both sets of simulations), in each simulation a random sample of size n is drawn from a randomly calibrated PCC. That is, both the parametric form and the corresponding parameters of the vine are chosen randomly in each simulation run. As candidate parametric copula families from which the pair-copulas of the true vine models are chosen, we use the Gaussian, Student's t, Clayton, Gumbel, Survival Clayton, Survival Gumbel, the rotated Clayton copula (90 degrees) and the rotated Gumbel copula (90 degrees). The parameters of the pair-copulas are then chosen randomly within the domain of the respective copula's parameters. The randomly calibrated vine used as the DGP in each simulation run $l = 1, \dots, 1,000$ is referred to as $PCC^{(l)}_{DGP}$.

The simulated sample from $PCC^{(l)}_{DGP}$ is then used to fit a vine model parametrically ($\widehat{PCC}^{(l)}_{AIC}$) and nonparametrcially ($\widehat{PCC}^{(l)}_{Bernstein}$) using Algorithm 2. Based on the (known) DGP, the fit of both approximations is then evaluated based on $PCC^{(l)}_{DGP}$.

We consider two different sample sizes $n = 200$ and $n = 500$ to assess the decreasing effect of the sample size on the approximation error. Furthermore, we analyze the effect of the vine type (C- or D-vine) as well as the dimensionality d of the vine

model on the approximation errors. To be precise, we simulate random samples from vines of dimension $d = 3, 5, 7, 11, 13, 15$. As the dimension increases, so does the number of variables one has to condition on in the pair-copulas of the vine's lower trees. The pair-copulas in the lower trees of the vine, however, are generally more computationally difficult to estimate due to the increasing number of variables one has to condition on in the corresponding h-functions so that the accurate approximation of the pair-copulas on all levels of the vine constitutes a considerable challenge to our nonparametric approximation.[126] At the same time, the curse of dimensionality could additionally complicate the approximation of the pair-copulas thus making the comparison of our approximation for different dimensions a sensible exercise. Finally, we expect the propagation and amplification of rounding errors to increase in higher dimensions possibly leading to large approximation errors of the parametric heuristic or even the non-convergence of some of the involved log-likelihood optimizations.

For each sample size n, dimension d and vine type, we simulate $1,000$ random samples $PCC_{DGP}^{(l)}$ ($l = 1, \ldots, 1,000$) and approximate each sample with a vine copula using Bernstein copulas as pair-copulas.[127] As we are only interested in approximating the pair-copulas and not in the correct calibration of the tree structure, we calibrate our nonparametric vine $\widehat{PCC}_{Bernstein}^{(l)}$ employing the correct tree structure and using the procedure $binonparfit(\cdot)$ in algorithm 2. As a benchmark, we calibrate a second vine copula $\widehat{PCC}_{AIC}^{(l)}$ by using the sequential heuristic $biparfit(\cdot)$ in algorithm 2. Furthermore, we also compute the fraction of simulations in which the sequential procedure did not yield any result due to either the non-convergence of the maximization of the log-likelihood function and consequently the parameter estimation or due to the average squared error (ASE) of the approximation tending to infinity as a result of incorrectly chosen parametric pair-copulas.

[126] This is one reason why Aas et al. (2009), Brechmann et al. (2012) and Dissmann et al. (2013) propose to capture as much dependence of the joint distribution that is to be modeled in the first trees of a vine model. If these pair-copulas are modeled accurately, the remaining pair-copulas in the lower trees can then be truncated or simplified. Furthermore, the truncation and simplification of a vine on the lower levels of the vine's tree limits the potential propagation of rounding errors.

[127] Unreported tests showed that 1,000 simulations are sufficient for the results of the Monte Carlo study to converge.

As a measure for the approximation error in the lth simulation, we compare the data generating vine model with the parametric and nonparametric approximations and use the average squared error $ASE := ASE(PCC_{DGP}, \widehat{PCC}_{approx}; d, g_1, g_2)$ of the cdfs of all bivariate pair-copulas each taken at $g_1 \times g_2$ uniform grid points in $I^2 := [0,1]^2$, i. e.,

$$ASE^{(l)} := \frac{2}{d(d-1)} \frac{1}{g_1 \cdot g_2} \sum_{i=1}^{d(d-1)/2} \sum_{j=1}^{g_1} \sum_{k=1}^{g_2} \left(\hat{C}_i \left(\frac{j}{g_1+1}, \frac{k}{g_2+1} \right) - C_i \left(\frac{j}{g_1+1}, \frac{k}{g_2+1} \right) \right)^2, \tag{3.25}$$

where C_i is the cdf of the ith pair-copula of $PCC^{(l)}_{DGP}$ and $\hat{C}_i$ the cdf of the ith pair-copula of either $\widehat{PCC}^{(l)}_{Bernstein}$ or $\widehat{PCC}^{(l)}_{AIC}$. To compute the approximation error, we set $g_1 = g_2 = 100$ and compute the ASE based on 10,000 grid points. Finally, the ASE of all 1,000 simulations are averaged to compare the overall performance of the different models.

3.3.2 Results of simulations

Results from our simulation study are presented in Table 3.1.

The results shown in Table 3.1 present several interesting insights into the finite sample properties of both the heuristically calibrated parametric and our proposed nonparametric vine copula models. First, we can see from Table 3.1 that for lower dimensions (e. g., $d = 3$ and $d = 5$) the ASE of our nonparametric approach is considerably larger than for the parametric model calibrated by sequentially selecting the pair-copulas based on AIC values. With increasing dimension of the random vector, however, we can observe that the approximation error of the parametric model increases disproportionately compared to our proposed nonparametric model. Furthermore, the nonparametric model appears to be able to match the approximation error of the parametric approach for dimensions $d = 13$ and higher. Most importantly, the parametric modeling approach becomes highly numerically unstable in higher dimensions as a result of the non-convergence of several maximizations of the log-likelihood function. At the same time, our proposed nonparametric vine with Bernstein pair-copulas is extremely reliable yielding acceptable approximations to the true underlying depen-

Table 3.1: **Comparison of the Average Squared Error.**

The table presents a comparison of the Average Squared Errors (ASE) of the parametric and nonparametric approximation to randomly specified vine copulas as well as the fraction of times (in %) the sequential heuristic based on AIC did not produce any results due to the non-convergence of the maximization of the log-likelihood functions (in these cases, the heuristic did not produced a fitted vine) or the approximation error of the heuristic tending to infinity. The ASE is given in multiples of 10^{-3}. All results are given in averages of $1,000$ simulations.

	Sequential AIC		**Bernstein Pair-Copulas**	
	ASE	**No result (in %)**	**ASE**	**No result (in %)**
Dimension $d = 3$				
C-Vine ($n = 200$)	0.040336	6.90	2.741199	0.00
C-Vine ($n = 500$)	0.030566	7.70	2.595884	0.00
D-Vine ($n = 200$)	0.044312	7.00	2.772305	0.00
D-Vine ($n = 500$)	0.023562	7.40	2.531201	0.00
Dimension $d = 5$				
C-Vine ($n = 200$)	2.226120	17.30	5.063710	0.00
C-Vine ($n = 500$)	2.101471	17.80	4.864722	0.00
D-Vine ($n = 200$)	2.271577	18.30	5.103421	0.00
D-Vine ($n = 500$)	2.107821	18.90	4.912165	0.00
Dimension $d = 7$				
C-Vine ($n = 200$)	4.128765	30.20	6.176894	0.00
C-Vine ($n = 500$)	3.843362	26.90	6.070333	0.00
D-Vine ($n = 200$)	4.021383	27.00	6.108655	0.00
D-Vine ($n = 500$)	3.880709	28.70	6.083041	0.00
Dimension $d = 9$				
C-Vine ($n = 200$)	5.394219	37.40	6.816040	0.00
C-Vine ($n = 500$)	5.219997	33.00	6.782382	0.00
D-Vine ($n = 200$)	5.021444	36.40	6.741291	0.00
D-Vine ($n = 500$)	4.956846	37.10	6.736978	0.00
Dimension $d = 11$				
C-Vine ($n = 200$)	6.184304	45.00	7.308551	0.00
C-Vine ($n = 500$)	5.972068	39.60	7.187704	0.00
D-Vine ($n = 200$)	5.843754	46.50	7.164167	0.00
D-Vine ($n = 500$)	5.688924	42.70	7.124501	0.00
Dimension $d = 13$				
C-Vine ($n = 200$)	6.758818	50.00	7.544622	0.00
C-Vine ($n = 500$)	6.577154	46.60	7.533562	0.00
D-Vine ($n = 200$)	6.265431	54.50	7.421761	0.00
D-Vine ($n = 500$)	6.232211	50.60	7.404687	0.00
Dimension $d = 15$				
C-Vine ($n = 200$)	7.159409	51.90	7.749707	0.00
C-Vine ($n = 500$)	7.112975	54.00	7.826334	0.00
D-Vine ($n = 200$)	6.680398	56.80	7.630245	0.00
D-Vine ($n = 500$)	6.598233	55.00	7.584712	0.00

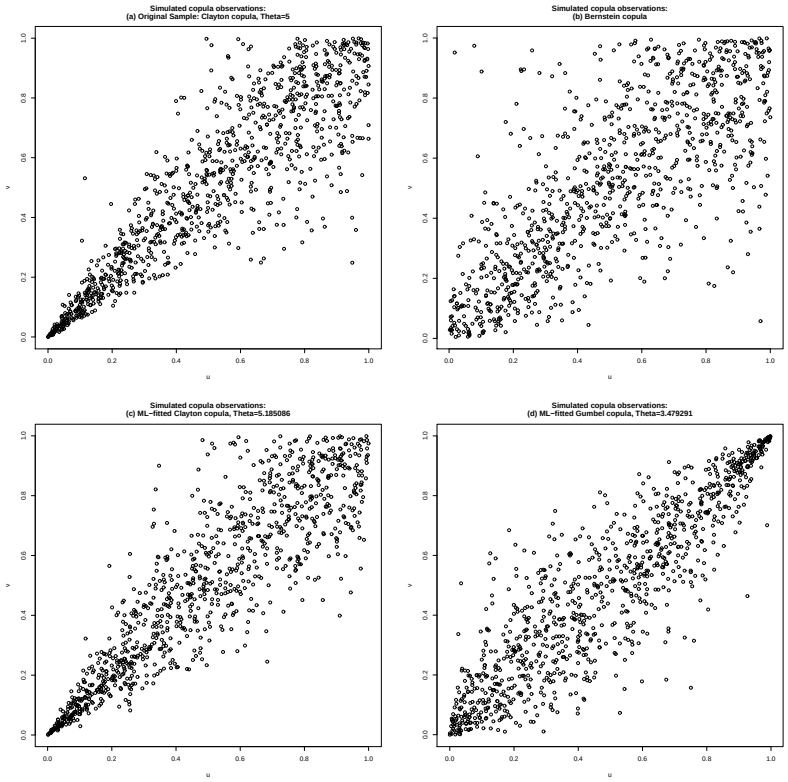
dence structure even for high-dimensional random vectors. The parametric approach, on the other hand, yielded a fitted vine in less than 50% of all simulations for dimension $d = 13$ and higher. In many of these cases, the infinite approximation error or the non-convergence of the algorithms used for maximizing the log-likelihood functions were caused by the wrong selection of several parametric families for the pair-copulas in the vine model. Concerning the type of the vine copula model, we find no significant differences between the average approximation errors of the C- or D-vines. As expected, we also find the average approximation error of both the parametric and nonparametric model to be decreasing in the sample size used for estimating both models.

To further illustrate the finding that the nonparametric model improves on the accuracy of a parametric vine especially in higher dimensions, we plot simulated samples from several parametrically and non-parametrically fitted copulas in Figure 3.3 where we assume that the true underlying dependence structure is given by a Clayton copula with parameter $\theta = 5$. From this copula, we simulate a random sample of size $n = 1,000$ and fit both a nonparametric Bernstein copula as well as a parametric Clayton and Gumbel copula via Maximum-Likelihood to the data. From all three fitted copulas, we again simulate a random sample and compare the plots of the simulated observations with the original sample.

The plots in Panels (a) and (b) in Figure 3.3 show how the Bernstein copula approximately captures the lower tail dependence of the original sample shown in Panel (a). The plot of the Bernstein copula in Panel (b) also shows, however, that the nonparametric approximation of the data sample coincides with a loss in information on the tail behaviour of the true underlying dependence structure. At the same time, Panel (c) underlines the notion that the nonparametric model is not superior to a parametric model in which the parametric copula family has been chosen correctly. If, however, the parametric copula family is chosen incorrectly like it is shown in Panel (d), the wrong selection of the parametric copula family can cause considerable approximation errors and a severely inaccurate modeling of the underlying tail dependence.

Figure 3.3: **Simulated observations from parametrically and nonparametrically fitted copulas.**

The figure shows plots of $n = 1,000$ observations simulated from different parametric and nonparametric copulas. Panel (a) shows the plot of the original observations simulated from a Clayton copula with parameter $\theta = 5$ which are used to calibrate a nonparametric Bernstein copula and two parametric Clayton and Gumbel copulas. Panel (b) shows a sample of simulated observations from a Bernstein copula which was calibrated based on the sample in Panel (a). Panels (c) and (d) show similar plots of simulated samples from a parametric Clayton and Gumbel copula which were fitted via Maximum-Likelihood using the original sample shown in Panel (a).



As the number of pair-copulas and trees in a vine model increases with the dimension of the data sample, the parametric benchmark model suffers due to the potential misspecification of the pair-copulas and the propagation of errors. Our nonparametric approach completely obviates the need for the error-prone selection of pair-copulas from pre-specified sets of parametric copulas. For this reason, given no prior information on the $d(d-1)/2$ parametric copula families, the nonparametric Bernstein copula vine model clearly improves on the fit of an inaccurately fitted parametric model. The nonparametric modeling of the pair-copulas thus seems to be a sensible approach especially when the number of pair-copulas that need to be selected from candidate parametric copula families increases (i. e., with increasing dimension).

3.4 Empirical study

3.4.1 Methodology

The purpose of our empirical study is to investigate the superiority of the proposed smooth nonparametric vine copula models over the competing calibration strategy based on a sequential selection of parametric pair-copulas via AIC with regard to the accurate forecasting of a portfolio's VaR. The simulation study presented in the previous section has highlighted the finding that our vine model with smooth nonparametric pair-copulas is especially well-suited for dependence modeling in higher dimensions as the selection of parametric pair-copulas becomes inaccurate due to error propagation and amplification. For low-dimensional problems, the heuristic selection of parametric copulas, however, seems to outperform our nonparametric approach with respect to the ASE of the vine copula's approximation. To show that our nonparametric model matches the results of the parametric heuristic even for low-dimensional problems, we concentrate in our empirical analysis on the VaR-forecasts of a five-dimensional portfolio.

Financial data are usually characterized by the presence of both conditional heteroscedasticity and asymmetric dependence in the log returns on financial assets. Therefore, we follow the vast majority of studies on copula models for VaR-estimation (Jondeau and Rockinger, 2006, Ausín and Lopes, 2010, Hafner and Reznikova, 2010) and employ standard GARCH(1,1)-models with Student's t-distributed innovations to model the marginal behaviour of our data. Although different specifications of the GARCH model are also possible, results found by Hansen and Lunde (2005) suggest that the choice of the order of a GARCH model is only of little importance for the model's forecasting accuracy.

Throughout the empirical study, we consider continuous log returns on financial assets with prices P_t ($t = 0, 1, \ldots, T$). The assets' log returns R_t are defined by $R_t :=$

$\log(P_t/P_{t-1})$ for $t \geq 1$. Our focus lies on modeling the joint distribution of the d assets, i. e., the joint distribution of the returns $R_{t1}, \ldots, R_{td}$.

The marginal behaviour of the assets is modeled by the use of GARCH(1,1)-models with t-distributed innovations. GARCH-models are without doubt the most popular class of models for describing the time series dynamics of financial return data, especially when one expects the data to exhibit time-varying volatility clustering. The marginal model is then given by

$$R_{tj} = \mu_j + \sigma_{tj} Z_{tj} \tag{3.26}$$

and

$$\sigma_{tj}^2 = \alpha_{0j} + \alpha_{1j} R_{t-1,j}^2 + \beta_j \sigma_{t-1,j}^2, \; j = 1, \ldots, d; \; t = 1, \ldots, T. \tag{3.27}$$

Equation (3.26) describes the evolution of the returns R_{tj} with μ_j being the expected return, σ_{tj} being the conditional return volatility and Z_{tj} being independent and identically t-distributed innovations. For the conditional volatility of the process, equation (3.27) assumes an autoregressive moving average model (ARMA-model) for the evolution of the volatilities σ_{tj}^2 dependent on previous values of the (squared) returns and variances with the parameters $\alpha_{0j}, \alpha_{1j}, \beta_j > 0$ governing the behaviour of the ARMA-model of volatilities.

The dependence structure between the d assets is introduced into the model by assuming the vector $\mathbf{Z}_t = (Z_{t1}, \ldots, Z_{td})$ $(t = 1, \ldots, T)$ of the innovations to be jointly distributed under a d-dimensional copula C with

$$F_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_d, \boldsymbol{\omega} | \mathcal{F}_{t-1}) = C\left[F_1(z_1; \boldsymbol{\nu}_1 | \mathcal{F}_{t-1}), \ldots, F_d(z_d; \boldsymbol{\nu}_d | \mathcal{F}_{t-1}); \boldsymbol{\omega}\right], \tag{3.28}$$

where $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_d$ are the parameter vectors of the innovations, C is a copula, $\mathcal{F}_{t-1}$ is the information set available to the investor at time $t-1$, and $\boldsymbol{\omega}$ is a vector of copula parameters (in case of the parametric model, otherwise $\boldsymbol{\omega}$ is simply empty).

The parameters of the univariate GARCH-models are estimated via Quasi-Maximum Likelihood Estimation. For the estimation of both the nonparametric vine model as well as the parametric model calibrated by using the pair-copulas' AIC values, we make use of rank-transformed pseudo-observations rather than the original sample as input data.[128] As the main results for copulas only hold for i.i.d. samples, we use the parameter estimates for the univariate GARCH models and transform the original observations into standardized residuals to yield (approximately) i.i.d. observations before computing the pseudo-observations (Dias and Embrechts, 2009).

In our empirical application, we consider an equally-weighted five-dimensional portfolio with returns $R_{p,t} = d^{-1} \sum_{j=1}^{d} R_{tj}$. The results from our simulation study underline the finding that our proposed vine copula model with Bernstein pair-copulas, on average, yields a better approximation to the empirical copula than the heuristically calibrated parametric model especially in higher dimensions. However, the parametric copula vine model could still outperform our proposed model with respect to the forecasting accuracy in low dimensions. In our empirical application, we therefore restrict our analysis to a portfolio consisting of five assets to additionally illustrate the nonparametric Bernstein vine copula model's superiority for low-dimensional problems.

To forecast the portfolio returns, we employ the algorithm presented in the study by Nikoloulopoulos et al. (2012) initially proposed for in-sample forecasting which was extended to out-of-sample forecasting by Weiß (2013).

The aim of the algorithm is the computation of a one-day-ahead forecast for the portfolio return $R_{p,t}$ via Monte Carlo simulation. In a first step, $K = 10,000$ observations $u_{T+1,1}^{(k)}, \ldots, u_{T+1,d}^{(k)}$ $(k = 1, \ldots, K)$ from the fitted (parametric or nonparametric) vine copula are simulated. Using the quantile function of the fitted marginal Student's t distributions, the simulated vine copula observations are then transformed into observations $z_{T+1,j}^{(k)}$ from the joint distribution of the innovations. In the next step, the simulated innovations are transformed into simulated returns $R_{T+1,j}^{(k)} = \hat{\mu}_j + \hat{\sigma}_{T+1,j} z_{T+1,j}^{(k)}$

[128] For a comparative study on the finite sample properties of different ML-based estimators for copulas, see Kim et al. (2007). The authors show that absent any information on the true distribution of the marginals, statistical inferences should be based on rank-transformed pseudo-observations.

where $\hat{\sigma}_{T+1,j}$ and $\hat{\mu}_j$ are the forecasted conditional volatility and mean values from the previously fitted marginal GARCH models. The MC-simulated forecasts of the portfolio return is then simply given by $R^{(k)}_{T+1,p} = d^{-1} \sum_{j=1}^{d} R^{(k)}_{T+1,j}$. Sorting the simulated portfolio returns for a given day in the forecasting period and taking the empirical one-day α percentile then yields the forecasted α%-VaR.

To backtest the results of our forecasting, we employ the test of conditional coverage proposed by Christoffersen (1998) as well as two duration-based tests discussed in Christoffersen and Pelletier (2004).[129]

All three backtests are based on the hit sequence of VaR-exceedances which is defined by

$$h_{t,\alpha} := \begin{cases} 1, & \text{if } R_{p,t} < \text{VaR}_\alpha(R_{p,t}) | \mathcal{F}_{t-1}, \\ 0, & \text{otherwise,} \end{cases}$$

with t being the time subscript and $\mathcal{F}_{t-1}$ being the set of available information. The test of conditional coverage by Christoffersen (1998) and Christoffersen and Pelletier (2004) jointly tests for the correct number of VaR-exceedances (unconditional coverage) and the serial independence of the violations over the complete out-of-sample (independence).[130] Under the null hypothesis of a correct number of VaR-exceedances that are independent over time, the hit sequence is simply distributed as (Christoffersen and Pelletier, 2004)

$$h_{t,\alpha} \sim i.i.d.\ Bernoulli(\alpha).$$

Then, let P be the length of the out-of-sample, P_1 be the number of VaR-exceedances and P_0 be the number of days on which the daily VaR-forecast was not exceeded,

[129] See Berkowitz et al. (2011) for an excellent review of different methods for backtesting Value-at-Risk forecasts. A comparison of different backtests can be found in the recent study by Escanciano and Pei (2012).

[130] The test of unconditional coverage has been implicitly incorporated in the Basel Accord for determining capital requirements for market risks, see Basel Committee on Banking Supervision (1996). Consequently, it has since become an industry standard in market risk management, see, e. g., Escanciano and Pei (2012).

respectively (and consequently $P = P_1 + P_0$). The likelihood function for the i.i.d. *Bernoulli* hit sequence with unknown parameter π_1 is

$$L(h_{t,\alpha}, \pi_1) = \pi_1^{P_1}(1-\pi_1)^{P-P_1} \tag{3.29}$$

and the Maximum-Likelihood estimate of π_1 is simply given by $\hat{\pi}_1 = P_1/P$. The test of unconditional coverage is then given by a likelihood ratio test based on

$$LR_{UC} = -2\left(\ln L(h_{t,\alpha}, \hat{\pi}_1) - \ln L(h_{t,\alpha}, \alpha)\right). \tag{3.30}$$

To test the hypothesis of independently distributed hits, the hit sequence is assumed to follow a first order Markov sequence with switching probability matrix

$$\Pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix} \tag{3.31}$$

with π_{ij} being the probability of an i on day $t-1$ being followed by a j on the next day t and $i, j \in \{1; 0\}$. Using the likelihood function

$$L(h_{t,\alpha}, \pi_{01}, \pi_{11}) = (1-\pi_{01})^{P_0-P_{01}}\pi_{01}^{P_{01}}(1-\pi_{11})^{P_1-P_{11}}\pi_{11}^{P_{11}},$$

where P_{ij} is the number of observations in $h_{t,\alpha}$ where a j follows an i and $i, j \in \{1; 0\}$ and the ML-estimates $\hat{\pi}_{01} = P_{01}/P_0$ and $\hat{\pi}_{11} = P_{11}/P_1$, the likelihood ratio test of the independence of hits is given by

$$LR_{ind} = 2\left(\ln L(h_{t,\alpha}, \hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(h_{t,\alpha}, \hat{\pi}_1)\right). \tag{3.32}$$

Both tests are then combined via $LR_{CC} = LR_{UC} + LR_{ind}$ to yield the test of conditional coverage. We note here that we do not rely on the asymptotic Chi-squared distribution of the test statistic. Although easy to implement, p-values derived under the assumption of the test statistic following a Chi-squared distribution are usually

incorrect due to the generally low sample sizes when using hit sequences. Instead, we follow Christoffersen and Pelletier (2004) in generating approximate p-values via Monte Carlo-simulation.

As an alternative to the test of conditional coverage, Christoffersen and Pelletier (2004) propose backtests based on the durations between VaR-exceedances. Then, let

$$D_i = t_i - t_{i-1} \tag{3.33}$$

be the duration of time (in trading days) between two subsequent VaR-exceedances where t_i is the time of the ith VaR-exceedance. Under the null hypothesis of a correctly specified VaR model, we would expect the process of no-hit durations to have no memory and mean $1/\alpha$. Consequently, the process D of durations should follow an exponential distribution with $f_{exp}(D;\alpha) = \alpha \exp(-\alpha D)$.[131] As an alternative hypothesis, Christoffersen and Pelletier (2004) propose to use the Weibull distribution for the process D with $f_W(D;a,b) = a^b b D^{b-1} \exp\left(-(aD)^b\right)$ which nests the exponential distribution from the null hypothesis for $b = 1$.

Although this test potentially captures higher order dependence in the hit sequence $h_{t,\alpha}$, the information from the temporal ordering of the no-hit durations is not exploited in the backtest. As a remedy, Christoffersen and Pelletier (2004) propose a conditional duration-based test based on the Exponential Autoregressive Conditional Duration (EACD) model of Engle and Russell (1998). In the standard EACD (1,0) model, the conditional expected duration $\mathbb{E}_{i-1}(D_i)$ is assumed to follow the process

$$\mathbb{E}_{i-1}(D_i) \equiv \psi_i = \omega + \beta D_{i-1}. \tag{3.34}$$

[131] See Christoffersen and Pelletier (2004) for details of the backtest and the motivation for using a continuous distribution for the discrete process D.

Again assuming an underlying exponential density with mean equal to one in the null hypothesis, the conditional distribution of the duration is is given by

$$f_{EACD}\left(D_i|\psi_i\right) = \frac{1}{\psi_i}\exp\left(-\frac{D_i}{\psi_i}\right). \tag{3.35}$$

The null hypothesis of independent no-hit durations is then given by $H_0 : \beta = 0$.

3.4.2 Data

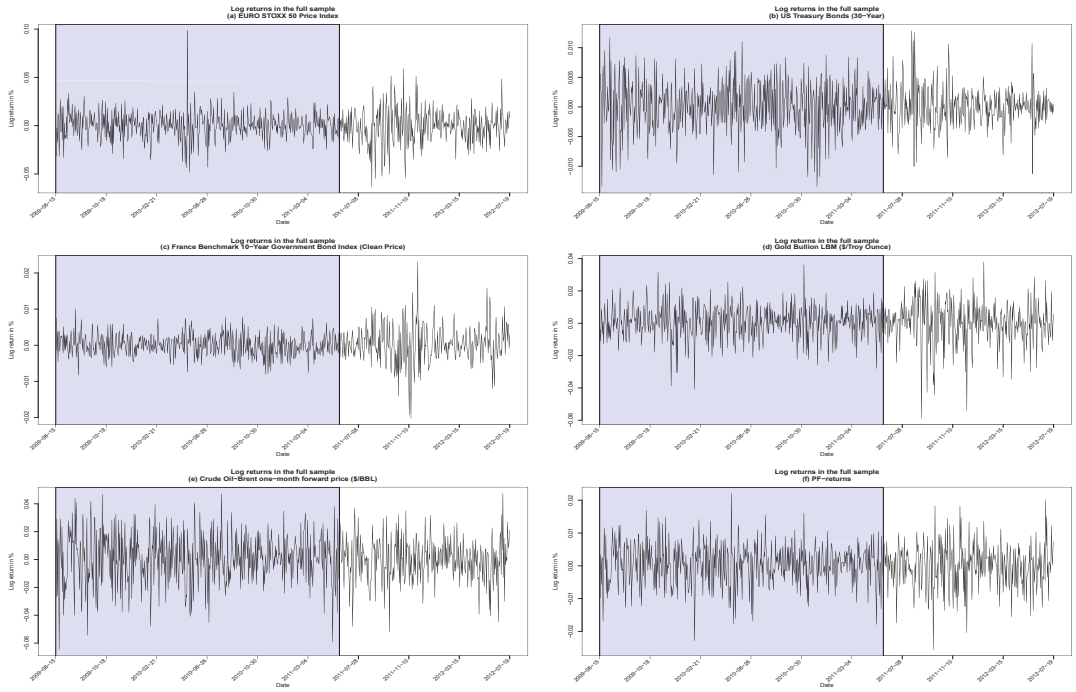
We consider a five-dimensional equal-weighted portfolio consisting of the returns on the EURO STOXX 50 Price Index, 30-year US Treasury Bonds, France Benchmark 10-year Government Bond Index, Gold Bullion LBM and one-month forward Crude Oil Brent. We obtain the data from the *Thomson Reuters Datastream* database. We follow the screening procedure proposed by Ince and Porter (2006) to control for known sources of data errors in *Datastream*. To be precise, we check whether our data include prices below $ 1 (which could lead to erroneous log returns due to *Datastream*'s practice of rounding prices) as well as log returns above 300% that are reversed within one month. Our five univariate time series do not suffer from any of these data errors.

Our sample covers a period of 800 trading days ranging from June 6, 2009 to July 19, 2012 and thus includes the aftermath of the default of Lehman Bros. as well as the onset of the Sovereign Debt Crisis. We use rolling windows with a length of 500 trading days for forecasting the one-day-ahead VaR on the following trading day. Our full out-of-sample spans a period of 300 trading days. Time series plots of the five portfolio constituents as well as the returns on the equal-weighted portfolio are shown in Figure 3.4. Panels (a) through (e) show the time series plots of the univariate returns, while Panel (f) shows the time series plot of the portfolio. The initial in-sample is shaded in grey to highlight the out-of-sample consisting of 300 trading days.

The plots in Figure 3.4 show several distinct features that can complicate VaR-forecasting. First, all plots exhibit the common stylized fact of volatility clusters, e. g., in Panels (a) and (c). Second, overall volatility of the univariate returns differs sig-

Figure 3.4: **Time series plots of log returns in the full sample.**

The figure shows plots of the log returns on the EURO STOXX 50 Price Index in Panel (a), US Treasury Bonds (30-year) in Panel (b), France Benchmark 10-Year Government Bond Index (Clean Price) in Panel (c), Gold Bullion LBM ($/Troy Ounce) in Panel (d), Crude Oil-Brent one-month forward ($/BBL) in Panel (e) and the returns on an equal-weighted portfolio consisting of the five individual assets in Panel (f). The sample covers the period from June 15, 2009 to July 19, 2012 (800 trading days). The plots show the log returns during our complete sample and are divided into the initial in-sample of 500 trading days (shaded in grey) and the out-of-sample of 300 trading days.



nificantly across our five portfolio constituents. For example, while the returns on the 30-year US treasury bonds are quite volatile in the in-sample and seem to calm in the out-sample, the opposite is true for the France Benchmark 10-year Government Bond Index which exhibits low volatility in the in-sample and a pronounced cluster of high volatility and extreme spikes around November 2011. This last result is, however, not surprising considering the fact that the Sovereign Debt Crisis experienced a climax at that time with the resignation of the Greek and Italian Prime Ministers, early elections in Spain and the expansion of the European Financial Stabilisation Mechanism (EFSM). Similarly, the price of Gold bullion became more volatile in the out-of-sample as well. The plot in Panel (f) shows that the combination of the five individual assets produces a portfolio which exhibits several phases of both high and low volatility as well as sudden extreme spikes in the portfolio's log returns.

3.4.3 Results

Following the methodology presented in Section 3.4.1, we compute the one-step-ahead forecasts of the portfolio-VaR for each day in the out-of-sample using rolling windows of 500 trading days. To analyze the differential effect of different confidence levels for the VaR on our models' forecasting accuracy, we forecast the 2%-, 5%- and 10%-VaR for a long and the 97.5%-VaR for a short position in the portfolio. Thus, we would expect 6, 15, 30 and 8 exceedances below the forecasted VaRs, respectively.[132] The VaR-forecasts as well as the realized portfolio returns for all four confidence levels are shown in Figures 3.5 and 3.6. In both figures, Panels (a) and (b) show the realized portfolio returns and the VaR-forecasts computed by the use of the nonparametric vine copula and the parametric vine copula model calibrated via the heuristic based on Akaike's Information Criterion, respectively.

The plots in Figures 3.5 and 3.6 show that both the nonparametric and the parametric model yield rather accurate forecasts of the portfolio's risk. While all VaR-forecasts are sufficiently close to the realized portfolio returns, exceedances of the VaR-forecasts occur only in case of large losses on the portfolio investment. Also, we can see that both the parametric and nonparametric model specifications yield quite similar VaR-forecasts. We would expect 6, 15, and 30 VaR-exceedances for the three significance levels of a long position in the portfolio and 8 VaR-exceedances for the significance level of the short portfolio, respectively. While the benchmark-PCC yields 6, 16, 34 and 3 violations, our nonparametric model forecasts portfolio losses in a very similar way with 5, 15, 37, and 3 violations. This means that the finding of comparable VaR-forecasts of both the nonparametric and parametric model remains valid for all four VaR confidence levels we consider. Figures 3.5 and 3.6 also underline our first impression from the simulation study that the benchmark model forecasts the VaR of the portfolio quite accurately. Otherwise, it highlights that our proposed nonparametric vine copula model seems to perform well even for low-dimensional portfolios. In

[132] For the short position, VaR-exceedances are defined as returns above the daily VaR-forecast.

Figure 3.5: **Comparison of 10%- and 97.5%-VaR forecasts and realized portfolio returns.**

The figure shows plots of the log returns on the five-dimensional portfolio we consider in our empirical study and the VaR-forecasts. Panel (a) presents the realized portfolio returns and the VaR-forecasts computed by the use of the nonparametric vine copula as the dependence model. Panel (b) shows a corresponding comparison of the portfolio returns and the VaR-forecasts estimated via a vine copula with the parametric pair-copulas chosen according to the sequential heuristic taken from the R-package *CDVine* based on Akaike's Information Criterion. For both models, the $(1-\alpha)$-VaR is computed for confidence levels $\alpha \in \{10\%; 97.5\%\}$. Both plots show results for the out-of-sample of 300 trading days. The portfolio consists of the returns on the EURO STOXX 50 Price Index, US Treasury Bonds (30-year), France Benchmark 10-Year Government Bond Index, Gold Bullion LBM and Crude Oil-Brent one-month forward.

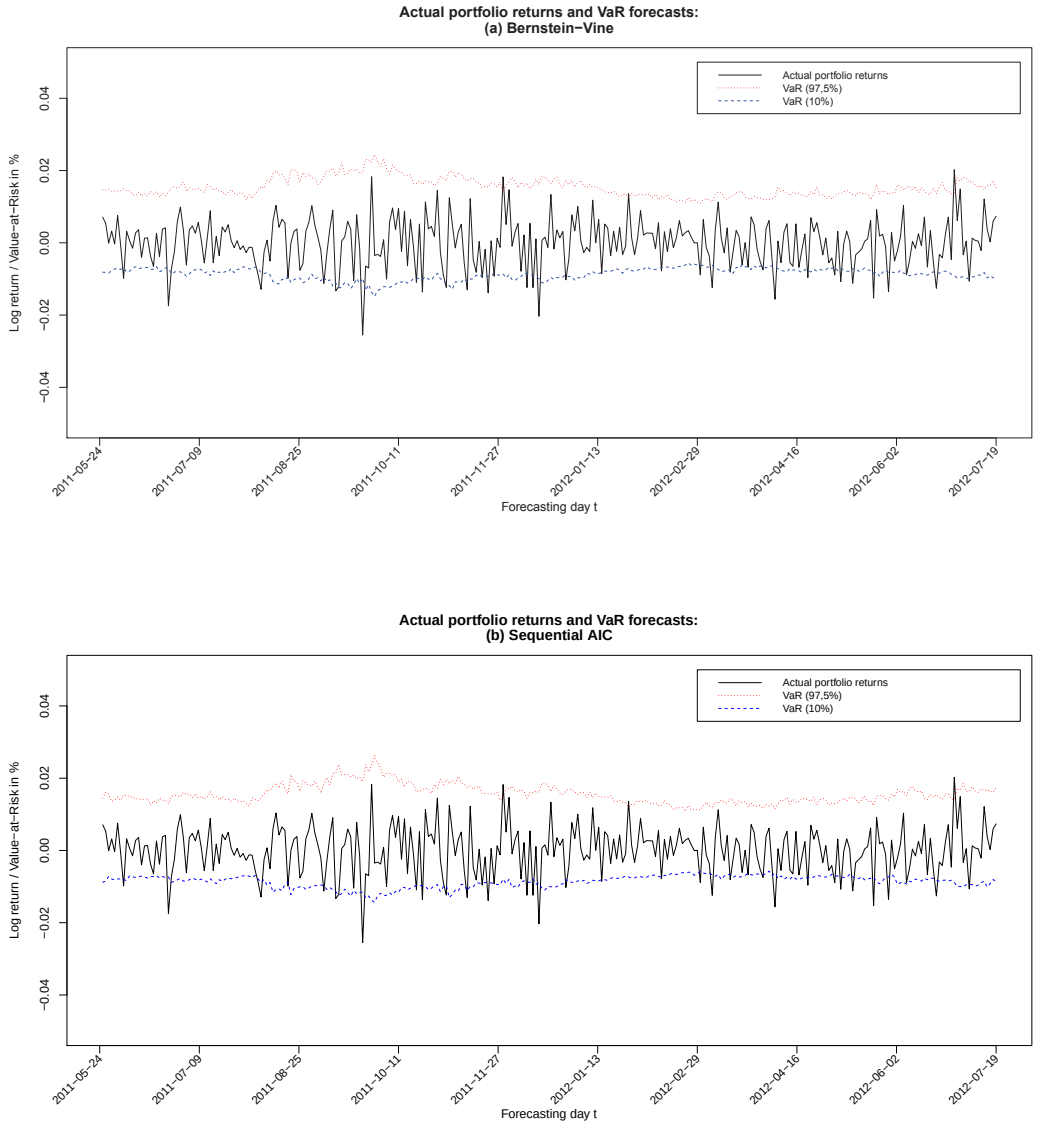
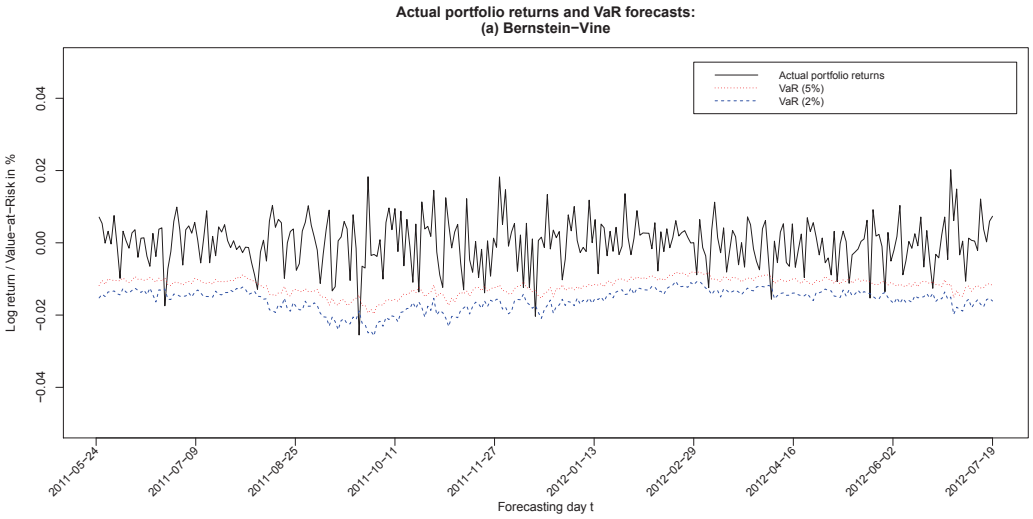
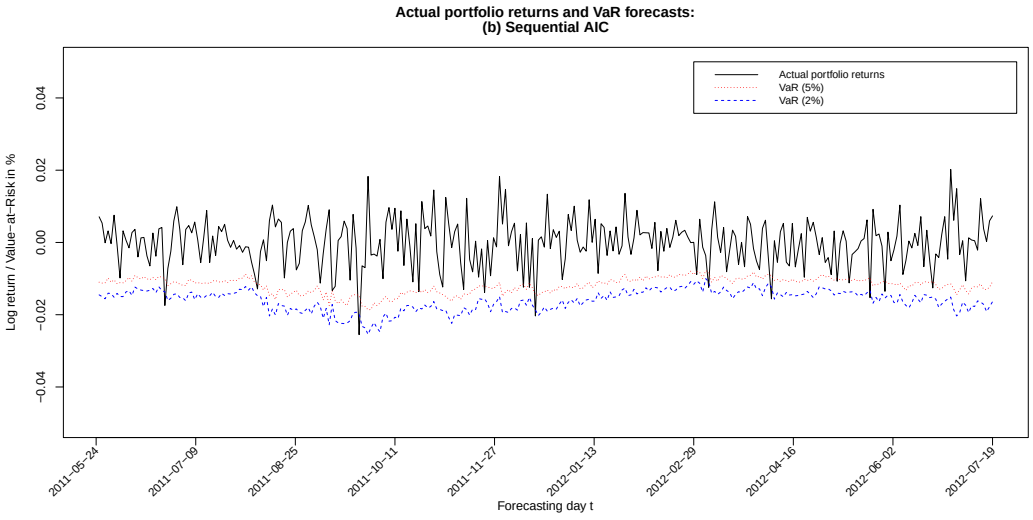


Figure 3.6: **Comparison of 2%- and 5%-VaR forecasts and realized portfolio returns.**

The figure shows plots of the log returns on the five-dimensional portfolio we consider in our empirical study and the VaR-forecasts. Panel (a) presents the realized portfolio returns and the VaR-forecasts computed by the use of the nonparametric vine copula as the dependence model. Panel (b) shows a corresponding comparison of the portfolio returns and the VaR-forecasts estimated via a vine copula with the parametric pair-copulas chosen according to the sequential heuristic taken from the R-package *CDVine* based on Akaike's Information Criterion. For both models, the $(1-\alpha)$-VaR is computed for confidence levels $\alpha \in \{2\%; 5\%\}$. Both plots show results for the out-of-sample of 300 trading days. The portfolio consists of the returns on the EURO STOXX 50 Price Index, US Treasury Bonds (30-year), France Benchmark 10-Year Government Bond Index, Gold Bullion LBM and Crude Oil-Brent one-month forward.

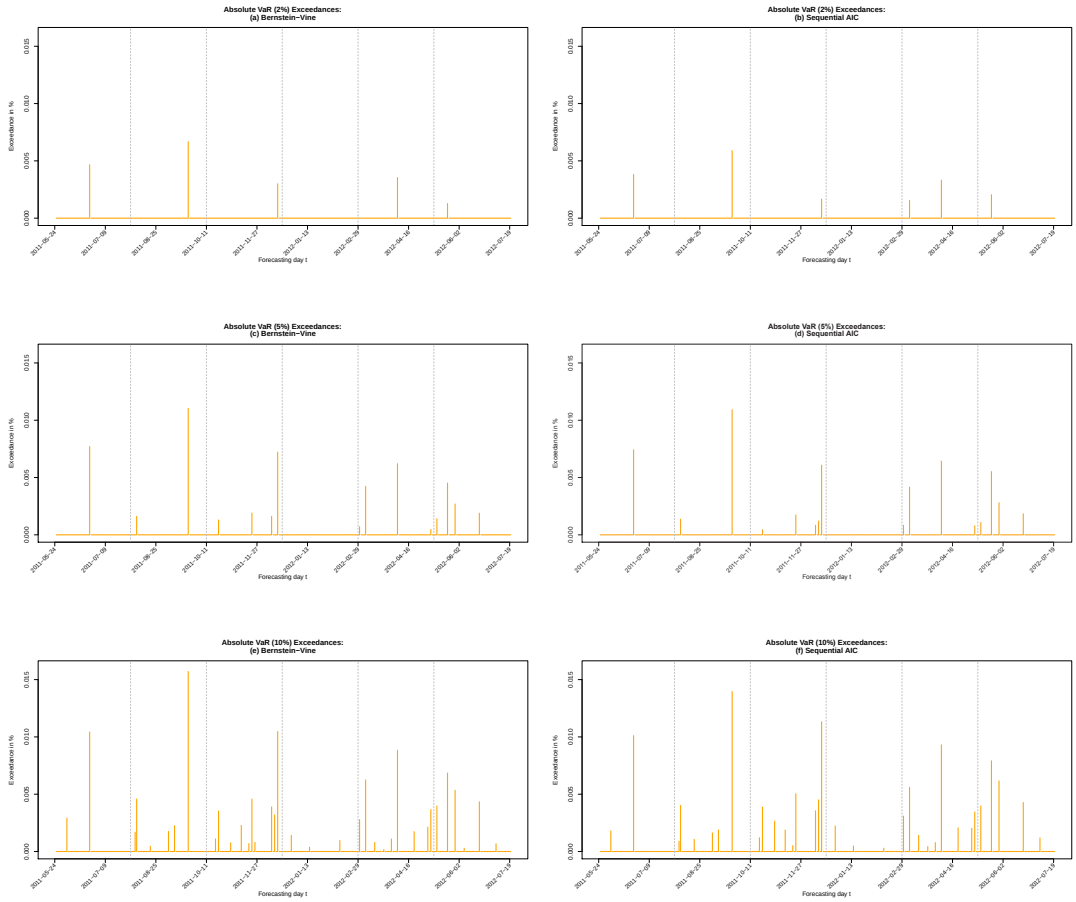




addition to this, the results presented in Figures 3.5 and 3.6 also show that both models adequately adapt the VaR-forecasts to changes in the portfolio returns' volatility.[133] To further assist in the interpretation of the results, Figure 3.7 highlights the VaR-exceedances for all models and the three confidence levels for a long position in the portfolio.

Figure 3.7: **Negative VaR-exceedances for the Bernstein Vine and the parametric benchmark.**

The figure shows plots of the negative VaR-exceedances (i. e., losses below the daily VaR-forecasts) computed from the nonparametric Bernstein vine copula model (Panels (a), (c) and (e)) and the parametric benchmark vine model with the parametric pair-copulas chosen according to the sequential heuristic taken from the R-package *CDVine* based on Akaike's Information Criterion (Panels (b), (d) and (f)). For both models, the $(1-\alpha)$-VaR is computed for confidence levels $\alpha \in \{2\%; 5\%; 10\%\}$. Both plots show results for the out-of-sample of 300 trading days. The portfolio consists of the returns on the EURO STOXX 50 Price Index, US Treasury Bonds (30-year), France Benchmark 10-Year Government Bond Index, Gold Bullion LBM and Crude Oil-Brent one-month forward. For ease of presentation the losses exceeding the VaR-forecasts are shown as positive real numbers.



[133] The flexible adjustment of both models to changes in return volatility also underlines the fact that a static dependence model in conjunction with dynamic marginal models suffices to model and forecast the dynamics of a multivariate return distribution.

The plots in Figure 3.7 underline our first impression from Figures 3.5 and 3.6 that both models forecast the VaR of the portfolio quite accurately. We can see from Figure 3.7 that not only do both models yield (approximately) correct numbers of VaR-exceedances for all three confidence levels, the exceedances also seem to occur randomly in time. Most importantly, however, our proposed nonparametric vine copula model with GARCH margins easily matches the heuristically calibrated parametric vine w.r.t. the accuracy of VaR-forecasting even for a relatively low-dimensional portfolio. To further substantiate this finding, we perform three formal backtests on the results of both the parametric and nonparametric vine models. The results of the three backtests are presented in Table 3.2.

Table 3.2: **Backtesting results.**

The table presents the results of three different backtests performed on the out-of-sample forecasts for the portfolio VaR estimated from the vine copula models calibrated parametrically via the sequential selection of pair-copulas via AIC and the nonparametric modeling of the pair-copulas using Bernstein copulas, respectively. The three backtests are the test of conditional coverage, the unconditional and the conditional duration-based tests proposed in Christoffersen and Pelletier (2004). The table reports the expected and the realized number of VaR-exceedances as well as the p-values for the three backtests. For both models, the backtesting results are reported for the $(1-\alpha)$-VaR for confidence levels $\alpha \in \{2\%; 5\%; 10\%; 97.5\%\}$. For the 97.5%-VaR, exceedances are given under the assumption of a short position in the portfolio.

		Sequential AIC		**Bernstein Pair-Copulas**	
	Exceedances (expected)	**P-value**	**Exceedances (realized)**	**P-value**	**Exceedances (realized)**
VaR $\alpha = 2\%$					
Conditional Coverage	6	0.994	6	0.850	5
Unconditional Duration	6	0.017	6	0.011	5
Conditional Duration	6	0.999	6	0.999	5
VaR $\alpha = 5\%$					
Conditional Coverage	15	0.437	16	0.573	15
Unconditional Duration	15	0.424	16	0.186	15
Conditional Duration	15	0.166	16	0.132	15
VaR $\alpha = 10\%$					
Conditional Coverage	30	0.152	34	0.042	37
Unconditional Duration	30	0.028	34	0.039	37
Conditional Duration	30	0.991	34	0.990	37
VaR $\alpha = 97.5\%$					
Conditional Coverage	8	0.116	3	0.117	3
Unconditional Duration	8	0.053	3	0.052	3
Conditional Duration	8	1.000	3	1.000	3

The backtesting results stress our finding that both models yield comparable results. For example, only one VaR-model is rejected at the 99% confidence level based on the test of conditional coverage. Although the results of the unconditional duration-based backtest imply a significantly worse forecasting accuracy of both models, the p-values for both the nonparametric and the parametric model are comparable for different confidence levels of the VaR. This indicates that neither model outperforms the other one based on our second backtest. If we use the conditional duration-based test of Christoffersen and Pelletier (2004) instead, none of the VaR-models is rejected. Turning to the number of VaR-exceedances, the results of our nonparametric vine copula model are slightly better for the 5%-VaR than those of the parametric benchmark while the opposite is true for the (for most practical uses too optimistic and thus unsuitable) 10%-VaR.

Our backtesting results indicate that both models yield acceptable VaR-forecasts for a relatively low-dimensional portfolio. One could conclude from this finding that in general using our nonparametric vine copula model does not yield significantly better VaR-forecasts. However, our empirical analysis was deliberately aimed at testing the hypothesis that the nonparametric model yields accurate VaR-forecasts in lower dimensions. In unreported tests of high-dimensional portfolios, the parametric benchmark suffered from the same drawbacks that were also observed in our simulation study. At the same time, our nonparametric model produced accurate VaR-forecasts even for high-dimensional portfolios.

3.5 Summary

In this paper, we propose to model the pair-copulas in a vine copula model nonparametrically by the use of Bernstein copulas. Our proposed model has the advantage of a significantly reduced model risk as it avoids the error-prone selection of pair-copulas from candidate parametric copula families. We test the approximation error of the smooth nonparametric Bernstein vine copula model against a parametric benchmark calibrated by the use of a sequential heuristic based on AIC. The superiority of our proposed model is exemplified in an empirical risk management application.

The results we find in our simulation study show that for low-dimensional problems, the parametric modeling approach outperforms our proposed nonparametric approach only marginally. However, the differences in the approximation error quickly vanish for higher dimensions with both models yielding comparable average approximations errors for dimensions $d = 13$ and higher. At the same time, our proposed nonparametric vine copula model does not suffer from numerical instability and error propagation which plagues the parametric benchmark due to an increasing number of wrongly selected parametric pair-copulas.

In the empirical risk management application, we test whether the differences in the average approximation error of the parametric and nonparametric vine copula models cause significant differences in both models' accuracy of forecasting the VaR of a low-dimensional asset portfolio. The results of our analysis show that even in lower dimensions ($d = 5$), our nonparametric vine copula model yields VaR-forecasts that cannot be rejected by several different formal backtests. The proposed nonparametric vine copula model thus seems to match the (good) results of a parametric vine copula model in lower dimensions and significantly outperforms this benchmark in higher dimensions.[134]

[134] In the applications we consider (i. e., financial risk management), the dimension d of the model is fixed and predetermined by the number of financial assets included in a portfolio. Consequently, we do no attempt to identify an optimal dimension d in our paper. However, we believe that the question for which dimension of the problem both approaches yield qualitatively the same results should be addressed in future research.

Chapter 4

Mixture Pair-Copula-Constructions

4.1 Introduction

It has become a stylized fact in both the finance as well as the risk management literature that elliptical models such as the multivariate Gaussian distribution cannot fully capture the dependence structures often found in financial asset returns. Starting with the work by Embrechts et al. (2002), several studies have criticized the inadequacy of correlation-based models for modeling the non-linear dependence in financial returns advocating the use of copulas instead. At the same time, elliptical copula models, which have become an industry standard in credit risk modeling following the influential study by Li (2000), have been found to perform just as poorly as their correlation-based counterparts due to their symmetric tail independence (see, e. g., Cherubini et al., 2004, Fischer et al., 2009). Especially in times of financial market turmoil, neglecting the tail dependence between financial time series can have disastrous effects on both banks and insurers as evidenced during the recent financial crisis.

While Archimedean copulas have been found to adequately model the lower tail dependence in bivariate financial portfolios (see, e. g., Weiß, 2011), simple parametric copula models are often not flexible enough to model the complex dependence structures of multivariate data. Consequently, studies by Joe (1996, 1997), Bedford and Cooke (2001, 2002) and Whelan (2004) have tried to construct high-dimensional cop-

ula models which are flexible enough to model complex multivariate data sets and yet at the same time tractable. The most prominent example of these high-dimensional copula models are the so-called vine copulas (also called pair-copula constructions, PCC).[135] Vine copulas are hierarchical in nature and only require the specification of bivariate copulas conditional on certain sets of variables (so called pair-copulas). The first application of vines in a risk management setting is due to Aas et al. (2009) and since then, vine copulas have emerged as the method of choice for modeling high-dimensional dependence structures due to their enormous flexibility.[136]

Although vine copulas are extremely flexible tools for modeling multivariate data, the increase in flexibility comes at the price of a greatly increased model risk. To fully specify a d-dimensional vine copula, one needs to choose $d(d-1)/2$ different pair-copulas from a set of candidate bivariate parametric copula families. Similar to the case of non-nested copulas, the question of how to select the optimal parametric copula family for the pair-copulas remains unanswered.[137] Methods for selecting the pair-copulas in vine copulas include the selection based on graphical data inspection and goodness-of-fit tests (Aas et al., 2009), sequential heuristics based on Akaike's Information Criterion (Brechmann et al., 2012) as well as tests based on empirical pair-copulas (Hobæk-Haff and Segers, 2012). A different approach is taken by Weiß and Scheffer (2012), who substitute the parametric pair-copulas in vines by smooth nonparametric Bernstein copulas to circumvent the otherwise necessary selection of parametric copula families for the pair-copulas. Finally, Kurowicka (2011) and Brechmann et al. (2012) propose strategies for simplifying vines by replacing certain pair-copulas by the independence copula (yielding a *truncated* vine copula) or the Gaussian copula (yielding a *simplified vine*).

[135] Overviews of different nested and hierarchical copula models in high dimensions are given in Aas and Berg (2009) and Fischer et al. (2009). A recent application of hierarchical Marshall-Olkin copula models to the estimation of the systemic risk of different countries is due to Baglioni and Cherubini (2013).

[136] For different applications of vines in asset pricing and risk management see Chollete et al. (2009), Heinen and Valdesogo (2009) and Weiß and Supper (2013).

[137] This well known problem in copula modeling has been addressed, e. g., by Breymann et al. (2003), Kole et al. (2007), Fischer et al. (2009) and Weiß (2011, 2013).

In this article, we propose a new approach to solve the problem of selecting the pair-copulas in a vine model. As pair-copulas, we employ mixture copulas (i. e., convex combinations of parametric copulas) yielding so-called *mixture pair-copula-constructions* (*Mixture-PCCs* in short). Our modeling strategy is related to the work of Kim et al. (2013) and their idea of using mixtures of D-vines but tackles the problem of constructing a vine model from a different perspective. While their study is concerned with the use of mixtures of D-vines (yielding multivariate models of even higher flexibility), we construct C- and D-vines with mixture pair-copulas to minimize the possibility of misspecifying a vine model. The use of bivariate mixture copulas as pair-copulas is beneficial for two reasons: First, the use of mixture copulas instead of simple bivariate parametric copulas increases the flexibility of a vine model even further. As all mixture copulas are bivariate and thus tractable in estimation, and as pruning strategies like simplifying or truncating the vine can also be applied to Mixture-PCCs, the increase in flexibility is not canceled out by an increase in computational complexity. Second, mixture-pair-copula constructions completely obviate the need for the error-prone selection of pair-copulas from pre-specified sets of parametric copulas. Thus, this paper contributes significantly to the current state of the art by proposing an extremely flexible yet still tractable model for high-dimensional dependence structures that lacks the model risk of current vine model specifications.

We illustrate the superiority of our proposed model by performing both a simulation and an empirical study on the in-sample and out-of-sample Value-at-Risk forecasting accuracy of Mixture-PCCs and heuristically calibrated vine copulas. The results from our simulation study show that our proposed Mixture-PCCs produce Value-at-Risk estimates that possess a comparable and especially in higher dimensions a better in-sample fit than the heuristic benchmark from the recent literature. In the empirical study of a four-dimensional financial portfolio, our proposed Mixture-PCC is characterized by a significantly better out-of-sample fit than the benchmark which overestimates portfolio risk. Our model thus helps risk managers to reduce regulatory risk capital. At the same time, portfolio losses are satisfactorily forecasted.

The remainder of this article is structured as follows. Section 4.2 introduces the basic properties of vine and mixture copulas and outlines the idea to combine both of them into Mixture-PCCs. Section 4.3 presents the results of our simulation study on the in-sample fit of Mixture-PCCs for Value-at-Risk forecasting. In Section 4.4, we discuss the results of our empirical study in which we compare the out-of-sample fit of a Mixture-PCC model with that of a heuristic benchmark from the literature based on the Value-at-Risk forecasts for a four-dimensional financial portfolio. Section 4.5 concludes.

4.2 Combining mixture and vine copulas

The purpose of this section is to shortly review the basic properties of vine copulas, mixture copulas, and the combination of the two, respectively.

4.2.1 Pair-copula constructions

In essence, copulas can be used to separate a multivariate distribution into its marginals and the dependence structure which is fully captured by the copula. This idea is formalized in Sklar's theorem (1959) which states that a d-dimensional cumulative distribution function (cdf) F can be split in two parts, the marginal distribution functions F_i and a copula C which is a d-variate cdf on $[0;1]^d$ with uniform marginals and which fully describes the dependence structure inherent in F:[138]

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)), \tag{4.1}$$

with $\mathbf{x} = (x_1, x_2, \ldots, x_d)$. Similarly, the joint multivariate density f can be represented by

$$f(\mathbf{x}) = c(F_1(x_1), \ldots, F_d(x_d)) \prod_{i=1}^{d} f_i(x_i), \tag{4.2}$$

[138] See Rüschendorf (2013) for a complete formulation of Sklar's theorem.

where $c(u_1, \ldots, u_d)$ is the d-variate copula density given by $\frac{\partial C(u_1,\ldots,u_d)}{\partial u_1 \ldots \partial u_d}$ and f_i $(i = 1, \ldots, d)$ are the marginal densities.

Building on this basic result, pair-copula constructions present an extremely flexible way of decomposing the multivariate density f into a cascade of bivariate copula densities.[139] We start our discussion of PCCs by observing that a joint probability density function of dimension d can be decomposed into its unconditional and conditional marginal densities via

$$f(x_1, \ldots, x_d) = f(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_1, x_2) \cdot \ldots \cdot f(x_d|x_1, \ldots, x_{d-1}). \tag{4.3}$$

Each conditional marginal density in this product can then be decomposed further using copula densities, e. g., via the relation

$$f(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2), \tag{4.4}$$

with $F_i(\cdot)$ being the cumulative distribution function (cdf) of x_i $(i = 1, \ldots, d)$ and $c_{12}(\cdot)$ being the unconditional copula density of (x_1, x_2). Repeating this representation for $f(x_3|x_1, x_2)$, we get the decomposition

$$f(x_3|x_1, x_2) = c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_3(x_3), \tag{4.5}$$

with $c_{23|1}(\cdot)$ being the conditional copula of (x_2, x_3) given x_1. The factorization of conditional densities can be repeated in an iterative manner yielding a decomposition

[139] For a rigorous examination of the statistical properties of vine copulas, see Joe (1996, 1997) and Bedford and Cooke (2001, 2002).

of the d-dimensional unconditional density in equation (4.3). For dimension $d = 3$, this decomposition is explicitly given by

$$\begin{aligned} f(x_1, x_2, x_3) &= c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \\ &\cdot\ c_{12}(F_1(x_1), F_2(x_2)) \\ &\cdot\ c_{13}(F_1(x_1), F_3(x_3)) \\ &\cdot\ f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3), \end{aligned} \tag{4.6}$$

with c_{12}, c_{13} and $c_{23|1}$ as *pair-copulas*. It is important to stress that this representation of the density is only one of several possible ways of decomposing $f(x_1, x_2, x_3)$ using marginal densities and pair-copulas depending on the variables one chooses to condition on.

The different possible decompositions of a d-dimensional joint density can be represented as nested sets of trees where two nodes are joined by an edge in tree $j + 1, j = 1, \ldots, d - 1$, only if the corresponding edges in tree j share a common node (see Bedford and Cooke, 2001, 2002). Consequently, there are $d - 1$ trees, where tree j has $d + 1 - j$ nodes and $d - j$ edges with each edge corresponding to a pair-copula density, i. e., a density of a conditional bivariate parametric copula. Two well-known classes of such representations are called C- and D-vines with the number of distinct C- and D-vines being $d!/2$. In a Canonical or C-vine, each tree has a unique node (without loss of generality this is node 1) that is connected to all other nodes yielding the representation

$$f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}(F(x_i|x_{i+1}, \ldots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1})), \tag{4.7}$$

where the subscript j identifies the tree, while i runs over all edges in each tree. In contrast to a C-vine, no node in any tree T_j is connected to more than two edges in a D-vine yielding the decomposition

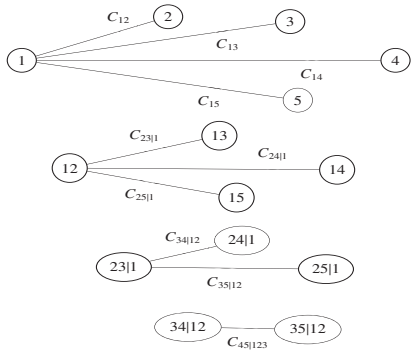
$$f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,\ldots,j-1}(F(x_j|x_1, \ldots, x_{j-1}), F(x_{j+i}|x_1, \ldots, x_{j-1})). \tag{4.8}$$

In Figure 4.1, we illustrate the hierarchical nature of vine copulas by plotting the tree structure of both a C- and a D-vine model.

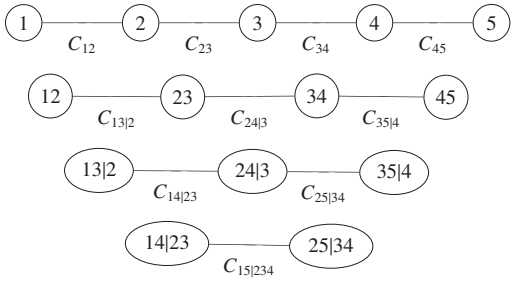
Figure 4.1: **Five-dimensional C- and D-vine copulas.**

The figure shows examples of a five-dimensional (a) C-vine and (b) D-vine copula with five random variables, four trees and ten edges. The nodes in the first tree correspond to the five random variables that are being modeled and each edge corresponds to a bivariate conditional or unconditional pair-copula. In both plots, C_{ij} is an unconditional bivariate copula while $C_{ij|x,y,\ldots}$ is the conditional bivariate copula of variables i and j given variables $x, y, \ldots$.

(a) Five-dimensional C-vine copula



(b) Five-dimensional D-vine copula



For the purpose of fitting a vine copula model in section 4.2.2, note that the marginal conditional distributions included in the decompositions (4.7) and (4.8) can be represented via

$$F(x|\mathbf{v}) = \frac{\partial C_{xv_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})}, \tag{4.9}$$

where $\mathbf{v}$ denotes a d-dimensional vector and $C_{ij|\mathbf{k}}$ is a bivariate copula distribution function. In this context, v_j is an arbitrarily chosen component of the vector v and $\mathbf{v}_{-j}$ describes the vector that excludes this component. If the vector v is univariate, equation (4.9) simplifies to

$$F(x|v) = \frac{\partial C_{xv}(F_x(x), F_v(v))}{\partial F_v(v)}. \tag{4.10}$$

Later, the estimation of a vine copula model will require the use of so-called h-function which is defined for two uniform variables x and v (i. e., $F_x(x) = x$ and $F_v(v) = v$) as the conditional distribution function of x given v, i. e.,

$$h(x, v, \boldsymbol{\theta}) = F(x|v) = \frac{\partial C_{xv}(x, v)}{\partial v}, \tag{4.11}$$

with $\boldsymbol{\theta}$ being the set of parameters for the copula of the joint distribution function of x and v. The second parameter of $h(\cdot)$ always corresponds to the conditioning variable.

Pair-copula constructions, and C- and D-vines in particular, have become widely used in empirical applications due to the enormous flexibility of these models. This flexibility is due to the fact that the $d(d-1)/2$ different pair-copulas in equations (4.7) and (4.8) can be selected from different parametric copula families. Furthermore, the sheer number of $d!/2$ possible C- and D-vine decompositions enable the statistician to choose from a wide range of flexible models for describing the dependence structure in high-dimensional distributions. Yet, the increased modeling flexibility of pair-copula constructions only comes at the cost of a greatly increased model risk as both the selection of pair-copulas and the choice of a vine decomposition are known to be error-prone. One possibility to alleviate the problem of selecting the best-fitting pair-copulas which we consider in this paper is the use of mixtures of parametric copulas as pair-copulas. In addition to C- and D-vines, recent studies on PCCs have proposed the use of so-called regular vines (R-vines in short) which put less restrictions on the vine trees and which include C- and D-vines as subsets (see Dissmann et al., 2013). Although R-vines are even more flexible than C- and D-vines and are thus becoming increasingly popular, their use is hindered by the same problem of selecting the right parametric pair-copulas one faces when using a C- or D-vine. In the following, we concentrate in our analysis on C- and D-vines to limit the computational complexity of our simulations but note that our results are readily applicable to R-vines as well.

4.2.2 Mixture copulas

We now turn our attention briefly to the basic definition of a mixture copula, focusing on bivariate copula densities c_i ($i = 1, \ldots, g$) with g being a fixed number of bivariate mixture copula components. Note that, in many applications of mixture models, not only the parametric forms of the components c_i but also the number g of components itself is unknown and has to be estimated from the available data of the underlying problem.

Let X_r and X_s be two univariate random variables (rvs) (we assume $r, s = 1, \ldots, d, r \neq s$) with continuous cdfs F_r and F_s, respectively. The probability integral transforms of X_r and X_s are given by $U_r = F_r(X_r)$ and $U_s = F_s(X_s)$, respectively. We intend to model the (unique) bivariate copula $C(u_r, u_s)$ of the joint distribution (X_r, X_s) by using a mixture (i. e., a convex combination) of g parametric copulas as components. With $c_1(u_r, u_s; \boldsymbol{\theta}_1), \ldots, c_g(u_r, u_s; \boldsymbol{\theta}_g)$ being the densities of the g parametric copulas, the density of the corresponding *mixture copula* is given by

$$c(u_r, u_s; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i c_i(u_r, u_s; \boldsymbol{\theta_i}), \tag{4.12}$$

where $\boldsymbol{\theta}_i$ is the vector of unknown parameters for the ith copula component of the mixture and $\boldsymbol{\Psi}$ denotes the vector including all unknown parameters, i. e., $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_g, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g)$.[140] The mixing proportions (or weights) π_i are nonnegative quantities that sum to one:

$$0 \leq \pi_i \leq 1 \qquad (i = 1, \ldots, g) \tag{4.13}$$

and

$$\sum_{i=1}^{g} \pi_i = 1. \tag{4.14}$$

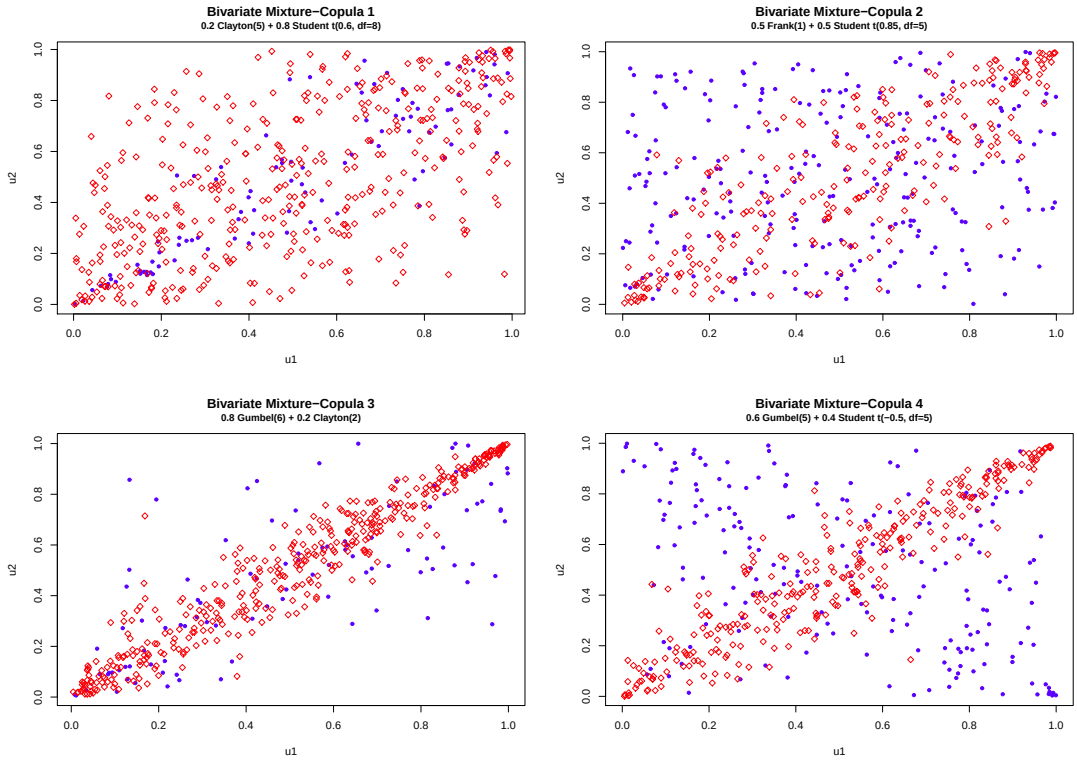
[140] Note that the mixture copula itself is given by $C(u_r, u_s; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i C_i(u_r, u_s; \boldsymbol{\theta_i})$.

As the components $c_1(u_r, u_s; \boldsymbol{\theta_1}), \ldots, c_g(u_r, u_s; \boldsymbol{\theta_g})$ are copula densities, it easily follows that (4.12) again defines a bivariate copula density (see Nelsen, 2006). We will refer to c_i as the ith *component copula density* of the mixture.

To illustrate the effect of mixing a second parametric copula to another one (with possibly different tail dependence), we plot simulated samples of size $T = 500$ from several different bivariate mixture copulas in Figure 4.2.

Figure 4.2: **Scatter plots of different mixture copulas.**

The figure shows plots of $T = 500$ observations simulated from four different mixture copulas. Panel (1) shows the plot of simulated observations from a mixture of a Clayton copula ($\theta = 5; \pi = 0.2$) in blue dots and a Student's t copula ($\rho = 0.6; df = 8; 1 - \pi = 0.8$) in red diamonds. Panel (2) shows the plot of simulated observations from a mixture of a Student's t copula ($\rho = 0.85; df = 5; 1 - \pi = 0.5$) in red diamonds and a Frank copula ($\theta = 1; \pi = 0.5$) in blue dots. Panel (3) shows the plot of simulated observations from a mixture of a Gumbel copula ($\theta = 6; \pi = 0.8$) in red diamonds and a Clayton copula ($\theta = 2; 1 - \pi = 0.2$) in blue dots. Panel (4) shows the plot of simulated observations from a mixture of a Gumbel copula ($\theta = 5; \pi = 0.6$) in red diamonds and a Student's t copula ($\rho = -0.5; df = 5; 1 - \pi = 0.4$) in blue dots.



The upper left plot in Figure 4.2 shows a convex combination of a Clayton and a Student's t copula. Due to the large impact of the Student's t copula on the data, the simulated sample exhibits a significantly symmetric tail dependence structure although

the Clayton copula with a weight of 0.2 introduces some lower tail dependence into the data. The upper right plot shows a sample of an equally weighted combination of a Frank copula and a Student's t copula. Here, the plot clearly reveals the symmetric tail dependence structure in the data caused by both mixture constituents. In the lower left plot, the upper tail dependence stemming from the dominant Gumbel copula in the mixture is quite evident while the lower tail dependence in the data caused by the introduction of the Clayton copula into the mixture is far less obvious. Finally, the lower right plot in Figure 4.2 presents a mixture of a Student's t copula with a negative correlation parameter and an upper tail dependent Gumbel copula.[141]

The simulations from different bivariate mixture copulas given in Figure 4.2 underline the importance to account for diverse tail dependence patterns in applied copula modeling. Conversely, the plots of all four mixture copulas highlight the flexibility with which these models can be used to describe the dependence structure in a given data set. Also, even in the bivariate case, simple parametric copula families could prove inadequate to model dependence structures in real life applications while mixture copulas produce a significantly better model fit. Modeling the pair-copulas in a PCC via convex combinations thus seems to be a sensible approach especially if one keeps in mind that the fit of the vine model critically depends on the fit of all $d(d-1)/2$ pair-copulas that need to be specified in a d-dimensional vine model.

4.2.3 Incomplete data problem and EM algorithm

When using finite mixture models, the parameter vector $\boldsymbol{\Psi}$ cannot be estimated via classical maximum likelihood estimation (MLE) as the likelihood function is unbounded due to the incomplete structure of the data (see, e. g., Dempster et al., 1977, McLachlan and Peel, 2000). Although some studies in the finance literature like, e. g., Rodriguez (2007) and Ruenzi and Weigert (2013) employ MLE, the identification problem of the mixture model should be taken into account by using the Expectation-Maximization

[141] Note that, strictly speaking, the parameter ρ of the Student's t copula is not exactly equal to the linear correlation of the data. However, we still refer to it as the "correlation parameter" of the t copula.

(EM) algorithm for estimating the mixture parameters to avoid otherwise biased parameter estimates. In the following, we shortly restate some well-known facts on the estimation of mixture-copulas using the EM algorithm (for details, see, e. g., Hu, 2006, Li et al., 2011).

Assume that we are given a bivariate data sample $\mathbf{x} = ((x_r^1, x_s^1), \ldots, (x_r^T, x_s^T))$ of size T. A pseudo-sample of the underlying copula is given by

$$\mathbf{u} = (\mathbf{u_r}, \mathbf{u_s}) = ((u_r^1, u_s^1), \ldots, (u_r^T, u_s^T)), \tag{4.15}$$

where $u_r^t = \hat{F}_r(x_r^t)$, $u_s^t = \hat{F}_s(x_s^t)$ $(t = 1, \ldots, T)$, and $\hat{F}_r, \hat{F}_s$ are the empirical cdfs of the marginal distributions.[142] The observed data vector $\mathbf{u}$ is treated as being incomplete and the tth observation (u_r^t, u_s^t) is assumed to be drawn from one of the g mixture elements.

Next, we define T component random variables Z_t of dimension g which specify for each observation in the sample the component-copula in (4.12) from which the observed tuple (u_r^t, u_s^t) $(r \neq s)$ are assumed to have been drawn. The ith element of Z_t, $(Z_t)_i := Z_{it}$ $(i = 1, \ldots, g)$ is an indicator variable that takes on the value 1 whenever the tth observation is drawn from the ith component copula, and 0 otherwise. For a random pseudo-sample $\mathbf{u}$ of a copula, the random vectors $Z_1, \ldots, Z_T$ are assumed to be unconditionally multinomially distributed with probabilities $\pi_1, \ldots, \pi_g$, i. e.,

$$Z_1, \ldots, Z_T \overset{i.i.d.}{\sim} \text{Multinomial}_g(1; \pi_1, \ldots, \pi_g), \tag{4.16}$$

with $z_1, \ldots, z_T$ denoting the associated vector realizations of $Z_1, \ldots, Z_T$. Consequently, the unobservable vector of component-indicator variables

$$\mathbf{z} = (z_1, \ldots, z_T) \tag{4.17}$$

[142] The idea to transform the marginal series non-parametrically using the empirical cdfs goes back to Genest et al. (1995) and aims at solving the problem of misspecifying the marginal models. The effects of misspecified marginals on VaR estimates is illustrated by Fantazzini (2009) and Kim et al. (2013) show that estimating copula parameters using pseudo-observations is superior to the use of the Full-Maximum-Likelihood and Inference-for-Margins methods in finite samples.

complements the observed data vector $\mathbf{u}$ to yield the complete data vector

$$\mathbf{u_c} = (\mathbf{u}, \mathbf{z}). \tag{4.18}$$

Hence, the distribution of the incomplete-data vector $\mathbf{u}$ is included in the distribution of the complete-data vector $\mathbf{u_c}$. Under these conditions, the complete-data log-likelihood for $\boldsymbol{\Psi}$, $\log L_c(\boldsymbol{\Psi})$, can be factorised into the product of the marginal densities of Z_t and the conditional copula densities of (u_r^t, u_s^t) given the z_t, where the data vectors $(u_r^1, u_s^1), \ldots, (u_r^T, u_s^T)$ are assumed to be conditionally independent given $z_1, \ldots, z_T$:

$$\log L_c(\boldsymbol{\Psi}) := \sum_{i=1}^{g} \sum_{t=1}^{T} z_{it}(\log \pi_i + \log c_i(u_r^t, u_s^t; \boldsymbol{\theta_i})). \tag{4.19}$$

The EM algorithm is a generic method for computing the parameters in an incomplete-data problem by treating the z_{it} as missing data and has become the most commonly used method for fitting mixture distributions (Dempster et al., 1977). We shortly describe the two steps of each iteration of the EM algorithm, the expectation step (E-step) and the maximization step (M-step), which are successively repeated until convergence.

The E-step handles the addition of the unobservable data, i. e., computing the conditional expectation of the complete-data log-likelihood $L_c(\boldsymbol{\Psi})$, given the observed pair $(\mathbf{u_r}, \mathbf{u_s})$ and using the current estimate for $\boldsymbol{\Psi}$. Starting with an initial parameter vector $\boldsymbol{\Psi}^{(0)}$, the calculation of the conditional expectation of $L_c(\boldsymbol{\Psi})$ can be written as follows:

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(0)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(0)}}\{\log L_c(\boldsymbol{\Psi})|(\mathbf{u_r}, \mathbf{u_s})\}. \tag{4.20}$$

Consequently, on the $(m + 1)$th iteration, the E-step requires the computation of $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(m)})$, where $\boldsymbol{\Psi}^{(m)}$ denotes the estimate of $\boldsymbol{\Psi}$ after the mth EM-iteration. Due to the fact that $\log L_c(\boldsymbol{\Psi})$ is linear in z_{it}, the computation of the conditional expectation

of (4.19) given $\mathbf{u}$ on the $(m+1)$th iteration leads to the calculation of the conditional expectation of Z_{it}:

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\Psi}^{(m)}}[Z_{it}|(\mathbf{u_r}, \mathbf{u_s})] &= P_{\boldsymbol{\Psi}^{(m)}}[Z_{it} = 1|(\mathbf{u_r}, \mathbf{u_s})] && (4.21)\\
&= \frac{\pi_i^{(m)} c_i(u_r^t, u_s^t; \boldsymbol{\theta_i}^{(m)})}{\sum_{h=1}^{g} \pi_h^{(m)} c_h(u_r^t, u_s^t; \boldsymbol{\theta_h}^{(m)})} && (4.22)\\
&=: \tau_i((u_r^t, u_s^t); \boldsymbol{\Psi}^{(m)}). && (4.23)
\end{aligned}$$

While the mixing proportion π_i can be regarded as the prior probability that the tuple (u_r^t, u_s^t) corresponds to the ith mixture weight, the quantity $\tau_i((u_r^t, u_s^t)$ can be seen as the posterior probability that the observed data (u_r^t, u_s^t) belongs to the ith component of the mixture copula. Thus, using equation (4.22), the computation of the conditional expectation (4.19) is then equivalent to

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(m)}) = \sum_{i=1}^{g} \sum_{t=1}^{T} \tau_i((u_r^t, u_s^t); \boldsymbol{\Psi}^{(m)})(\log \pi_i + \log c_i(u_r^t, u_s^t; \boldsymbol{\theta}_i)). \tag{4.24}$$

The M-step on the $(m+1)$th iteration calculates the updated estimate $\boldsymbol{\Psi}^{(m+1)}$. This is achieved by maximizing $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(m)})$ with respect to $\boldsymbol{\Psi}$. Note that the computation of $\boldsymbol{\Psi}^{(m+1)}$ is decomposed into two parts, because the updated mixing proportions $\pi_i^{(m+1)}$ are independent of the updated estimates $\boldsymbol{\theta}^{m+1} = (\boldsymbol{\theta_1}^{m+1}, \ldots, \boldsymbol{\theta_g}^{m+1})$. Using the current conditional expectation $\tau_i((u_r^t, u_s^t); \boldsymbol{\Psi}^{(m)})$, the updated estimate of π_i is given by

$$\pi_i^{(m+1)} = \sum_{t=1}^{T} \tau_i(u_r^t, u_s^t; \boldsymbol{\Psi}^{(m)})/T \qquad (i = 1, \ldots, g). \tag{4.25}$$

Observe that each observation u_r^t, u_s^t affects the ith mixture component $\pi_i^{(m+1)}$ via the current estimate of its posterior probability. The update of $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_g})$ of the $(m+1)$th iteration is directly given by (4.24), so that $\boldsymbol{\theta}^{(m+1)}$ must satisfy the following equation:

$$\sum_{i=1}^{g} \sum_{t=1}^{T} \tau_i(u_r^t, u_s^t; \boldsymbol{\Psi}^{(m)}) \frac{\partial \log c_i(u_r^t, u_s^t; \boldsymbol{\theta_i})}{\partial \boldsymbol{\theta}} = 0. \tag{4.26}$$

Dempster et al. (1977) show that the incomplete-data likelihood function $L(\boldsymbol{\Psi})$ does not decrease after an EM iteration, i. e.,

$$L(\boldsymbol{\Psi}^{(m+1)}) \geq L(\boldsymbol{\Psi}^{(m)}). \tag{4.27}$$

In addition to $L(\boldsymbol{\Psi})$ being non-decreasing in successive EM iterations, an upper bound for the likelihood assures convergence of the EM algorithm.

4.2.4 Mixture-pair-copula constructions

In this part of our analysis, we combine both concepts to yield so-called mixture-pair-copula constructions. The basic idea behind Mixture-PCCs is to use mixtures of parametric copulas as pair-copulas in a PCC. Obviously, the g components of the mixture-pair-copulas do not necessarily have to be identical throughout the vine, nor does the number of components g itself. In case both the number and the parametric form of the component copulas are fixed, we will refer to this as a *simple Mixture-PCC*. If both the number and the parametric forms are not prespecified but rather estimated prior to the EM algorithm, we will call such a model an *extended Mixture-PCC*. It is clear that an extended Mixture-PCC offers even more flexibility for modeling dependence structures than a simple Mixture-PCC as it imposes less restrictions on the parametric form of the mixture-pair-copulas. However, its specification requires the additional selection of both g and the component copulas separately for all $d(d-1)/2$ pair-copulas of the Mixture-PCC. Model selection for mixture models can be done by using a criterion for the fit of the model like, e. g., Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC) or the Consistent AIC (CAIC) (see Kim et al., 2013, McLachlan and Peel, 2000).

Due to the additional computational complexity of estimating an extended Mixture-PCC, we concentrate in the rest of our study on the analysis of simple Mixture-PCCs and their performance in Monte Carlo simulations and an empirical application. Never-

theless, we note that the results for extended Mixture-PCCs should be identical or even better than the results presented in the following based on simple Mixture-PCCs.[143]

Next, we demonstrate how to fit a Mixture-PCC model to a given multivariate data set as well as a known vine type with a given tree structure (i. e., the permutation of the data variables). Note that usually not only the bivariate copulas and the corresponding parameters but also the tree structure need to be estimated.[144] In our paper, we focus on techniques for fitting the building blocks (bivariate copulas) in the vine model. Thus, the vine type and the tree structure are assumed to be known.

When modeling the dependence structure of a multivariate data set by the use of a Mixture-PCC, two major problems have to be addressed. First, the true parametric families of the bivariate mixture copula densities describing the dependence structure are unknown. Hence, for a d-dimensional vine copula, $d(d-1)/2$ suitable bivariate mixture copulas must be selected. Second, we require a tool for estimating the weights as well as the parameters of the selected copula families for each mixture pair-copula. Since we are focusing on simple Mixture-PCCs, the number g and the parametric form of each bivariate component copula is prespecified, i. e., the first problem is not taken into further consideration in our study. The second task is handled by using the EM algorithm sequentially. This method fits each mixture pair-copula (density) separately by estimating $\mathbf{\Psi}$, the vector including the mixing weights $\Pi_1, \ldots, \Pi_g$ as well as the vectors of the unknown copula parameters $\theta_1, \ldots, \theta_g$, in equation (4.12). More precisely, we propose the following modified version of the sequential estimation procedure outlined in Brechmann and Czado (2013) and Dissmann et al. (2013) to fully specify a d-dimensional Mixture-PCC:

[143] Selecting an extended Mixture-PCC will usually require the estimation of a model selection criterion like, e. g., Akaike's Information Criterion for several Mixture-PCC specifications. As such, the simple Mixture-PCC could be used as a benchmark which could be improved upon by estimating further extended Mixture-PCCs.

[144] See Brechmann and Schepsmeier (2013) for a discussion of the problem of selecting the optimal tree structure of a vine copula model.

Algorithm 1 *(Fitting a d-dimensional Mixture-PCC sequentially)*
For tree $j = 1, \ldots, d-1$ *in equations* (4.7) *(C-vine) or* (4.8) *(D-vine), perform the following steps:*

1. *For each edge* i ($i = 1, \ldots, d-j$) *(bivariate mixture copula) of tree* T_j *of the vine, the parameter vector* $\boldsymbol{\Psi}$ *of each mixture pair-copula is estimated using the original data (tree* T_1*) corresponding to the variables* i *and* $i+j$ *for a C-vine or* j *and* $j+i$ *for a D-vine, or the transformed data (estimated in Step 2) from the previous tree* T_{j-1} *via the EM algorithm.*

2. *Transform the observations used in this iteration by using the estimated mixture copula from tree* T_j *and the h-function defined in equation* (4.11) *to compute the observations (i. e., conditional distribution functions) of* $(F(x_i|x_{i+1}, \ldots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1}))$ *(C-vine) or* $(F(x_j|x_1, \ldots, x_{j-1}), F(x_{j+i}|x_1, \ldots, x_{j-1}))$ *(D-vine) for the next tree. If* $j = 1$*, use the original data to calculate the unconditional distribution functions* $(F(x_i), F(x_{i+j}))$ *(C-vine) or* $(F(x_j), F(x_{j+i}))$ *(D-vine), respectively.*

The algorithm highlights the sequential manner of the procedure by exploiting the tree-by-tree structure of vine copulas. In the first tree, the parameter vector $\boldsymbol{\Psi}$ of each bivariate mixture copula is estimated using the original data via the EM algorithm. The transformed variables for the second tree are computed subsequently using the mixture's h-function (4.11) as shown in Section 4.2.1.[145] With these transformed data, each bivariate mixture copula of the second tree is estimated. This process is repeated until the last tree is reached and the final mixture pair-copula is estimated.

Due to the sequential estimation procedure, our mixture approach does not guarantee to find a global optimum with respect to the accuracy of the weights as well as the parameters of each mixture pair-copula. Performing a full estimation over all parameters simultaneously, might provide a better model fit. Nevertheless, our sequential

[145] Note that the h-function of the mixture copula (distribution function) is given by the convex combination of the h-functions of the mixture components (i. e., the copula cdfs). The numerical computation of the mixture's h-function can then be achieved quite easily by (explicitly or numerically) solving the h-functions of the mixture components.

mixture procedure is nevertheless a beneficial approach as it considerably limits the computational burden of the estimations.

Admittedly, modeling the dependence structure by using Mixture-PCCs goes hand in hand with additional cost of computation. In a simple Mixture-PCC with one-parameter copula components, $g - 1$ mixture weights $\pi_1, \ldots, \pi_{g-1}$ and g copula parameters $\theta_1, \ldots, \theta_g$ have to be estimated for each building block in the vine model.[146] Thus, for a fully specified vine model, we have $d(d-1)/2 \cdot (2g-1)$ parameters in total that need to be estimated. However, all estimations are performed sequentially on bivariate data. Consequently, our model is still tractable even despite the increased number of parameters that need to be estimated and the gain in flexibility prevails the additional cost of computation.

4.3 Simulation study

In this section, we compare the performance of our proposed Mixture-PCC with that of the sequential heuristic of Brechmann et al. (2012) and Dissmann et al. (2013) for selecting the pair-copulas in a vine from prespecified parametric copula families. In our simulations, we mimic the frequent problem in financial applications of fitting a multivariate distribution to a time series of asset returns and subsequently estimating the Value-at-Risk of a portfolio consisting of these assets. More precisely, we are interested in assessing the quality of both model approaches by backtesting the vine copula models' VaR-estimates in the in-sample of our simulated data.

4.3.1 Simulation design

In the following, we shortly describe the data generating process (DGP) used in our simulation study. To simulate d asset price trajectories over a prespecified number of observations T, we employ standard GARCH(1,1) models to mimic frequently ob-

[146] Since the mixing weights sum to unity, one of them is redundant. Here, we arbitrarily omitted the gth mixing weight π_g.

served stylized facts in asset returns like volatility clustering, heavy tails, and serial correlation. Let z_{tj} ($t = 1, \ldots, T$; $j = 1, \ldots, d$) denote an independent, identically distributed sample of a multivariate distribution with properly scaled Student's t marginals and a random vine copula describing the dependence structure.[147] Then, the log-return time series is modeled by $\epsilon_{tj} = \sigma_{tj} z_{tj}$. The dynamic variance is characterized by the equation $\sigma_{tj}^2 = \omega_j + \beta_j \epsilon_{t-1,j}^2 + \gamma_j \sigma_{t-1,j}^2$, where all GARCH parameters are also chosen randomly within their respective domain of definition. In this context, β_j denotes the first lag ARCH parameter (state memory factor) and γ_j the first GARCH parameter (variance memory factor) for the log-return time series of asset j, respectively. Thus, $\omega_j/(1 - \alpha_j - \beta_j)$ is the unconditional variance for time-series j.

The randomly calibrated vine used as the DGP in each simulation run $l = 1, \ldots, 100$ is referred to as $PCC_{DGP}^{(l)}$. The simulated sample from $PCC_{DGP}^{(l)}$ is then used to fit a benchmark vine model ($\widehat{PCC}_{AIC}^{(l)}$) via the sequential heuristic proposed by Brechmann and Czado (2013) and Dissmann et al. (2013) and a Mixture-PCC ($\widehat{PCC}_{Mix}^{(l)}$) using Algorithm 4.2.4. Based on the (known) DGP, the fit of both approximations is then evaluated by in-sample forecasting.

4.3.2 Estimation of PCCs

Next, we discuss the multivariate model that is fitted to the simulated data. As financial return data is frequently characterized by conditional heteroskedasticity and asymmetric dependence, we resort to GARCH-type models both as data generating processes as well as models that are fitted to the data. Therefore, we employ GARCH(1,1) models with Student's t-distributed innovations (see Jondeau and Rockinger, 2006, Fantazzini,

[147] In this context, a random vine is given by a vine in which the vine type, the ordering of the variables, the parametric copula families for the pair-copulas, and the parameters are all chosen randomly in each simulation run. As candidate parametric copula families from which the pair-copulas of the true vine models are chosen, we use the Gaussian, Student's t, Clayton, Gumbel, Survival Clayton, Survival Gumbel, the rotated Clayton copula (90 degrees) and the rotated Gumbel copula (90 degrees). The parameters of the pair-copulas are then chosen randomly within the domain of the respective copula's parameters.

2009, Liu and Luger, 2009, Aas and Berg, 2009, Ausín and Lopes, 2010, Hafner and Reznikova, 2010) to describe the time series' marginal behavior.[148]

Then, let P_t $(t = 0, 1, \ldots, T)$ denote the price of a financial asset at time t. The asset's log return R_t is defined by $R_t := log(P_t/P_{t-1})$ $(t \geq 1)$. In our simulation study, R_t is simulated and results from our data generating process whereas in the empirical study, real prices of financial assets are used. The return time series are modeled using

$$R_{t,j} = \mu_j + \sigma_{t,j} Z_{t,j}, \tag{4.28}$$

$$\sigma^2_{t,j} = \alpha_{0,j} + \alpha_{1,j}\left(R_{t-1,j} - \mu_j\right)^2 + \beta_j \sigma^2_{t-1,j}, \quad j = 1, \ldots, d; t = 1, \ldots, T, \tag{4.29}$$

where Z_{tj} are independent and identically t-distributed innovations $Z_{t,j}$.

The vector $Z_t = (Z_{t,1}, \ldots, Z_{t,d})$ $(t = 1, \ldots, T)$ of the GARCH innovations captures the dependence structure between the d financial assets and is assumed to be jointly distributed with distribution

$$F_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_d, \boldsymbol{\Psi} | \mathcal{F}_{t-1}) = C^{Mix}\left[F_1(z_1; \boldsymbol{\nu}_1 | \mathcal{F}_{t-1}), \ldots, F_d(z_d; \boldsymbol{\nu}_d | \mathcal{F}_{t-1}); \boldsymbol{\Psi}\right], \tag{4.30}$$

where $\boldsymbol{\Psi}$ denotes the parameter vector of a d-dimensional copula C and $\nu_1 \ldots, \nu_d$ are the parameter vectors of the innovations. The parameters of the univariate GARCH-models are estimated via quasi-maximum likelihood. Using these parameter estimates, we convert the original observations into standardized residuals which in turn are transformed into a pseudo-sample using equation (4.15). The pseudo-observations are then used to fit the Mixture-PCC and the heuristic benchmark-PCC. As a benchmark model, we calibrate a PCC using the sequential heuristic proposed by Brechmann and Czado (2013) and Dissmann et al. (2013). Here, for each of the $d(d-1)/2$ pair copulas, all candidate parametric copulas are fitted to the data using maximum likelihood estimation

[148] Hansen and Lunde (2005) compare different versions of the original GARCH model and show that the majority of GARCH model variants are outperformed by the simple GARCH(1,1) specification. Consequently, the GARCH(1,1) time series filter with Student's t-distributed innovations is the model of choice for the marginal behavior of our data.

sequentially and the candidate that yields the optimal value of Akaike's Information Criterion (AIC) is chosen.

After both PCCs have been fitted, for each day in the in-sample, $K = 10,000$ observations $u^{(k)}_{T+1,1}, \ldots, u^{(k)}_{T+1,d}$ $(k = 1, \ldots, K)$ from the fitted parametric vine copula are simulated. This sample is then converted into observations $z^{(k)}_{T+1,1}$ from the distributions of the innovations via the inverse of the symmetric Student t-distributions. Then $\hat{\mu}_j + \hat{\sigma}_{T+1,j} z^{(k)}_{T+1,j}$ $(j = 1, \ldots, d)$ transforms the simulated observations into simulated returns, where $\hat{\sigma}_{T+1,j}$ and $\hat{\mu}_j$ are the conditional volatility forecasts and mean values stemming from the fitted marginal GARCH(1,1) models. The portfolio return is then easily calculated via $R^{(k)}_{T+1,pf} = d^{-1} \sum_{j=1}^{d} R^{(k)}_{T+1,j}$ (assuming an equal-weighted portfolio) and an estimate for the portfolio's α%-VaR can be calculated by ranking the K simulated returns. Finally, VaR-violations can be identified by comparing the α%-VaR with the observed return of $R_{T+1,pf}$.

In the simulation study, we consider two different sample sizes $T = 300$, and $T = 500$ to assess the effect of different sample sizes on the fit of the pair-copula constructions. Furthermore, we repeat our simulations for two dimensions $d \in \{3; 5\}$. We expect the overall fit of the PCCs to suffer at least to a small degree with increasing dimensionality. As the number of pair-copulas and trees in a vine model increases with the dimension of the data sample, rounding and estimation errors in the upper trees of a vine could be propagated thereby decreasing the fit of pair-copulas in the lower trees. Consequently, Aas et al. (2009), Brechmann et al. (2012) and Dissmann et al. (2013) propose to capture as much dependence of the joint distribution in the upper trees of a vine model so that the remaining pair-copulas in the lower trees can be truncated or simplified. Finally, we repeat all simulations for two different vine types (C- and D-vine) to test the robustness of our results to a change in vine type. See section 4.3.1 for the candidate parametric copula families from which the pair-copulas of the true vine models are chosen.

For each sample size, dimension, and vine type, we simulate 100 random samples and fit to each sample a simple Mixture-PCC and a heuristically calibrated PCC fol-

lowing the procedure laid out in Brechmann et al. (2012). In the Mixture-PCCs, each pair-copula is modeled as a convex combination of a Clayton, a Frank, a Gumbel and a Student's t copula.[149] We chose these particular parametric copula families as constituents of the mixture pair-copulas because they account for both dependence and independence in the tails of the distribution. The Mixture-PCC is then fitted using the EM algorithm. For the benchmark approach, we use a widely extended set of candidate parametric copula families. On the one hand, the full range of parametric copula families used in the simulation design in section 4.3.1 is taken into account. On the other hand, the Frank, Joe, BB1, BB6, BB7, BB8 and again rotated versions of these are provided additionally. Each candidate parametric pair-copula is fitted to the data via maximum likelihood estimation. Then, the candidate that yields the optimal AIC value is chosen. Using the fitted Mixture- and benchmark PCCs, we then estimate the 0.1%, 2.5%, and 5% Value-at-Risk of the simulated data and assess both models' fit by performing the two-tailed backtest of Christoffersen (1998) and the duration-based Weibull test of independent VaR-exceedances proposed by Christoffersen and Pelletier (2004) on the resulting sequences of VaR-violations. The outline of our simulation study can be summarized as follows:

1. For each sample size ($T = 300$; $T = 500$), dimension ($d = 3$, $d = 5$), and vine type (C- or D-vine) generate $l = 1, \ldots, 100$ random samples of size T from a randomly calibrated vine $PCC^{(l)}_{DGP}$.

2. Fit a benchmark vine model ($\widehat{PCC}^{(l)}_{AIC}$) via the sequential heuristic and a Mixture-PCC ($\widehat{PCC}^{(l)}_{Mix}$) to the data sample assuming the the correct vine type to be known.

3. For each day in the in-sample of size T, simulate $10,000$ observations both from $\widehat{PCC}^{(l)}_{AIC}$ and $\widehat{PCC}^{(l)}_{Mix}$.

[149] To limit the computational cost of the parameter estimations, we restrict the mixture copulas in the Mixture-PCCs to four parametric copula constituents. Note that while the estimation of the parameters of the benchmark PCC required about ten minutes in total for all 100 simulations, the estimation of our Mixture-PCCs required about five times longer.

4. Compute VaR-estimates for both vine models using the $10,000$ simulated returns. Compare the estimated VaR of both models with the return in the orignal sample in Step 1.

The results of the simulations are presented in Table 4.1.

Table 4.1 presents the average number of expected and realized VaR-exceedances as well as the percentage of simulations in which the VaR-backtest was rejected at the 5% significance level for both our Mixture-PCC and the benchmark PCC.

Starting with the results for a data generating process of dimension $d = 3$ in Panel (a), several findings are noteworthy. First, both models yield similar results for the smallest VaR-level of 0.1% with almost no VaR-violations and a perfect 0% of simulation runs in which neither model is rejected by the backtest. This result holds for both sample sizes and both C- and D-vines. Second, results start to differ from model to model for VaR-levels of 2.5% and 5%. While both models appear to overestimate portfolio risk (as evidenced by the average number of violations which is always below the expected number of VaR-violations), our proposed Mixture-PCCs perform significantly better than the benchmark for a VaR-level of 5% as shown by the smaller percentages of simulations in which the backtest is rejected. For a VaR-level of 2.5%, the rejection percentages of the Mixture-PCCs are comparable or slightly worse than the corresponding averages of the benchmark. Third, the results for both models do not differ significantly across the vine types and sample sizes although the percentages of the backtest rejections increase slightly for the 2.5% VaR with an increase in sample size.

Panel (b) presents the corresponding results for the simulations of data samples of dimension $d = 5$. Regarding the 0.1%-VaR-level, both approaches yield acceptable model fits with both models having a rejection percentage of zero in backtesting. Again, we find both approaches to be conservative and to overestimate portfolio risk to a certain degree as shown by the average number of VaR-violations. In contrast to the simulations for dimension $d = 3$, however, the results of the simulations for $d = 5$ clearly show that the benchmark PCC is outperformed by our Mixture-PCC. For all

Table 4.1: **Results of the simulation study - long position.**

The table presents the results of the simulation study on the in-sample fit of our proposed Mixture-PCC and a heuristically calibrated PCC. The data generating process is given by a multivariate GARCH model in which the marginal distributions follow GARCH(1,1) processes with Student's t distributed innovations and where the joint distribution of the innovations is characterized by a randomly calibrated vine copula (see Nikoloulopoulos et al., 2012, for details). The in-sample fit of both competing models is assessed by performing the two-tailed backtest of conditional coverage proposed in Christoffersen (1998) (CC) as well as the duration-based Weibull test of independence by Christoffersen and Pelletier (2004) (WB) on the in-sample estimates of the portfolio VaR. The table reports the expected and the realized number of VaR-exceedances as well as the rejection rates for a significance level for the backtest of 5%. The VaR-level α is set to 0.1%, 2.5%, and 5%, respectively. Results are presented separately for sample sizes $T = 300$ and $T = 500$ and for dimension $d = 3$ and $d = 5$, respectively.

				Sequential AIC			**Mixture-PCC**		
			Exceedances	**P-value < 5% (no. of rejections)**		**∅-Exceedances**	**P-value < 5% (no. of rejections)**		**∅-Exceedances**
Pair-copula construction	**sample size**	**VaR-Level**	**(expected)**	**CC**	**WB**	**(realized)**	**CC**	**WB**	**(realized)**
Panel (a): Dimension d=3									
C-Vine	T=300	0.1%	0.3	0.00	0.00	0.00	0.00	0.00	0.00
		2.5%	7.5	0.36	0.18	3.23	0.39	0.18	2.97
		5%	15	0.31	0.07	9.19	0.09	0.12	11.58
	T=500	0.1%	0.5	0.00	0.00	0.02	0.00	0.00	0.00
		2.5%	12.5	0.47	0.09	5.96	0.59	0.07	5.41
		5%	25	0.44	0.06	16.01	0.06	0.09	20.36
D-Vine	T=300	0.1%	0.3	0.00	0.00	0.02	0.00	0.00	0.00
		2.5%	7.5	0.40	0.15	3.44	0.41	0.16	3.46
		5%	15	0.23	0.05	9.53	0.06	0.03	13.45
	T=500	0.1%	0.5	0.00	0.00	0.06	0.00	0.00	0.00
		2.5%	12.5	0.48	0.02	5.83	0.57	0.02	5.59
		5%	25	0.45	0.03	15.35	0.05	0.11	21.44
Panel (b): Dimension d=5									
C-Vine	T=300	0.1%	0.3	0.00	0.00	0.02	0.00	0.00	0.00
		2.5%	7.5	0.36	0.22	3.32	0.28	0.17	3.54
		5%	15	0.23	0.07	9.66	0.01	0.08	14.3
	T=500	0.1%	0.5	0.00	0.00	0.04	0.00	0.00	0.00
		2.5%	12.5	0.52	0.07	5.69	0.52	0.05	5.97
		5%	25	0.46	0.10	15.32	0.05	0.09	22.37
D-Vine	T=300	0.1%	0.3	0.00	0.00	0.03	0.00	0.00	0.00
		2.5%	7.5	0.26	0.15	3.58	0.17	0.11	4.17
		5%	15	0.35	0.06	9.13	0.04	0.09	14.61
	T=500	0.1%	0.5	0.00	0.00	0.05	0.00	0.00	0.00
		2.5%	12.5	0.51	0.06	5.65	0.31	0.04	6.59
		5%	25	0.44	0.09	15.36	0.03	010	24.45

three VaR-levels, the Mixture-PCC is rejected in approximately the same or a smaller number of simulation runs than the benchmark PCC. For example, the in-sample VaR-estimates of the Mixture-PCC are only rejected in 5% (D-vine) and 3% (C-vine) of all simulations for a sample size of $T = 500$. In contrast, the benchmark PCC is rejected in 46% and 44% of all simulation runs, respectively. Moreover, the average number of violations produced by the Mixture-PCC is significantly closer to the expected number of violations. Turning to the analysis of the duration-based Weibull backtest of independent VaR-exceedances, we find similar results as for the test of conditional coverage with the exception that all estimated models (benchmark and Mixture-PCC) are rejected in far less simulations by the Weibull test than by the CC test. Table 4.2 shows corresponding results for the VaR-levels of 95%, 97.5%, and 99.9% for a short position in the portfolio.

Again, based on the results of the average number of exceedances and the results of the CC test, we find our Mixture-PCC to outperform the benchmark not only for a long but also for a short position in the portfolio.[150]

The results emphasize that both models yield acceptable Value-at-Risk forecasts. However, performing the two-tailed backtest of Christoffersen (1998) indicates a more appropriate model fit of our mixture approach for both long and short positions in the portfolio. This result is remarkable for the following reason. Since no mixtures are used to simulate the data, the Benchmark-PCC should in principle be able to perfectly recover the complete tree structure. However, due to the finite sample size, the Benchmark-PCC selects several pair-copulas incorrectly on average in our simulations. In contrast, our Mixture-PCCs appear to be significantly more flexible and to provide a better model fit even in this setting. As a consequence, our simulations provide us with ample evidence that especially for higher dimensions, the Mixture-PCC model yields significantly better in-sample VaR-estimates than the benchmark.

[150] The results of the Weibull backtest for the short position in the portfolio are mixed and do not allow for a clear interpretation.

Table 4.2: **Results of the simulation study - short position.**

The table presents the results of the simulation study on the in-sample fit of our proposed Mixture-PCC and a heuristically calibrated PCC. The data generating process is given by a multivariate GARCH model in which the marginal distributions follow GARCH(1,1) processes with Student's t distributed innovations and where the joint distribution of the innovations is characterized by a randomly calibrated vine copula (see Nikoloulopoulos et al., 2012, for details). The in-sample fit of both competing models is assessed by performing the two-tailed backtest of conditional coverage proposed in Christoffersen (1998) (CC) as well as the duration-based Weibull test of independence by Christoffersen and Pelletier (2004) (WB) on the in-sample estimates of the portfolio VaR. The table reports the expected and the realized number of VaR-exceedances as well as the rejection rates for a significance level for the backtest of 5%. The VaR-level α is set to 95%, 97.5% and 99.9%, respectively. Exceedances are given under the assumption of a short position in the portfolio. Results are presented separately for sample sizes $T = 300$ and $T = 500$ and for dimension $d = 3$ and $d = 5$, respectively.

				Sequential AIC			Mixture-PCC		
			Exceedances	P-value < 5% (no. of rejections)		∅-Exceedances	P-value < 5% (no. of rejections)		∅-Exceedances
Pair-copula construction	**sample size**	**VaR-Level**	**(expected)**	**CC**	**WB**	**(realized)**	**CC**	**WB**	**(realized)**
Panel (a): Dimension d=3									
C-Vine	T=300	99.9%	0.3	0.00	0.00	0.02	0.00	0.00	0.01
		97.5%	7.5	0.42	0.16	3.18	0.32	0.15	3.63
		95%	15	0.44	0.07	8.29	0.11	0.06	11.27
	T=500	99.9%	0.5	0.00	0.01	0.09	0.00	0.00	0.02
		97.5%	12.5	0.40	0.15	6.04	0.33	0.11	6.67
		95%	25	0.52	0.08	14.80	0.13	0.11	19.55
D-Vine	T=300	99.9%	0.3	0.01	0.00	0.07	0.00	0.00	0.07
		97.5%	7.5	0.47	0.17	3.21	0.56	0.18	2.95
		95%	15	0.45	0.13	8.44	0.29	0.07	9.41
	T=500	99.9%	0.5	0.00	0.00	0.04	0.01	0.01	0.16
		97.5%	12.5	0.54	0.08	5.38	0.67	0.18	4.63
		95%	25	0.74	0.03	13.71	0.47	0.06	15.53
Panel (b): Dimension d=5									
C-Vine	T=300	99.9%	0.3	0.00	0.00	0.05	0.00	0.00	0.02
		97.5%	7.5	0.38	0.16	3.17	0.16	0.11	4.46
		95%	15	0.45	0.07	7.96	0.04	0.11	13.39
	T=500	99.9%	0.5	0.00	0.00	0.08	0.00	0.00	0.03
		97.5%	12.5	0.62	0.07	5.13	0.19	0.09	7.29
		95%	25	0.66	0.10	13.75	0.05	0.09	23.03
D-Vine	T=300	99.9%	0.3	0.00	0.00	0.04	0.01	0.00	0.21
		97.5%	7.5	0.39	0.11	3.00	0.15	0.09	4.45
		95%	15	0.41	0.09	8.11	0.07	0.06	13.10
	T=500	99.9%	0.5	0.00	0.00	0.07	0.01	0.02	0.24
		97.5%	12.5	0.61	0.07	5.18	0.27	0.08	7.28
		95%	25	0.69	0.10	13.78	0.10	0.06	21.71

4.4 Empirical study

The results of our simulation study show that our proposed Mixture-PCCs perform exceptionally well in-sample especially for higher dimensions. In this section, we support our main finding by performing an empirical study on the out-of-sample forecasting accuracy of our Mixture-PCC and the heuristic benchmark.

4.4.1 Methodology and data

In our empirical study, we assess the accuracy of both the Mixture-PCC and the heuristically calibrated benchmark-PCC in out-of-sample portfolio-VaR forecasting. To this end, we analyze a four-dimensional portfolio consisting of three selected stocks and the price of Gold bullions. To be precise, we study the time series of daily log-returns $R_{t,pf} = 4^{-1} \sum_{j=1}^{4} R_{t,j}$ of an equal-weighted four-dimensional portfolio and forecast the Value-at-Risk of this portfolio. The multivariate GARCH model that we fit to the data resembles the one proposed by Nikoloulopoulos et al. (2012) and which we also employ in the simulation study (see Section 4.3). In contrast to our simulation study, however, we are now interested in forecasting the portfolio's Value-at-Risk out-of-sample.[151]

Our example portfolio consists of the stocks of Citigroup, General Electric, and Deutsche Bank, as well as the price for gold bullion LBM. We selected these four investments for our example portfolio as to cover a diverse set of asset classes (gold and stocks), industries (banks and industrial conglomerate) as well as regions (Europe and United States). While we expect the stocks of the two banks in our portfolio to experience strong co-movements in stock returns, the addition of the stock of General Electric and gold bullion should allow for ample diversification effects. Moreover, we opted for a four-dimensional portfolio which suffices to exemplify the advantages of our proposed modeling approach while at the same time limiting the computational

[151] See Weiß (2013) for details of the extension of the model of Nikoloulopoulos et al. (2012) to out-of-sample forecasting.

workload. Time-series data for the four assets are retrieved from *Thomson Reuters Financial Datastream*. To account for known data errors in *Datastream*, we follow Ince and Porter (2006) and apply the following data filters to our sample. First, we control for extreme log returns above 300% that are reversed within one month and exclude these returns from our data. Second, we exclude returns of prices below \$1 to avoid incorrect log returns that could otherwise arise from *Datastream*'s practice of rounding prices. Neither of our four price time series exhibits any of these data errors.

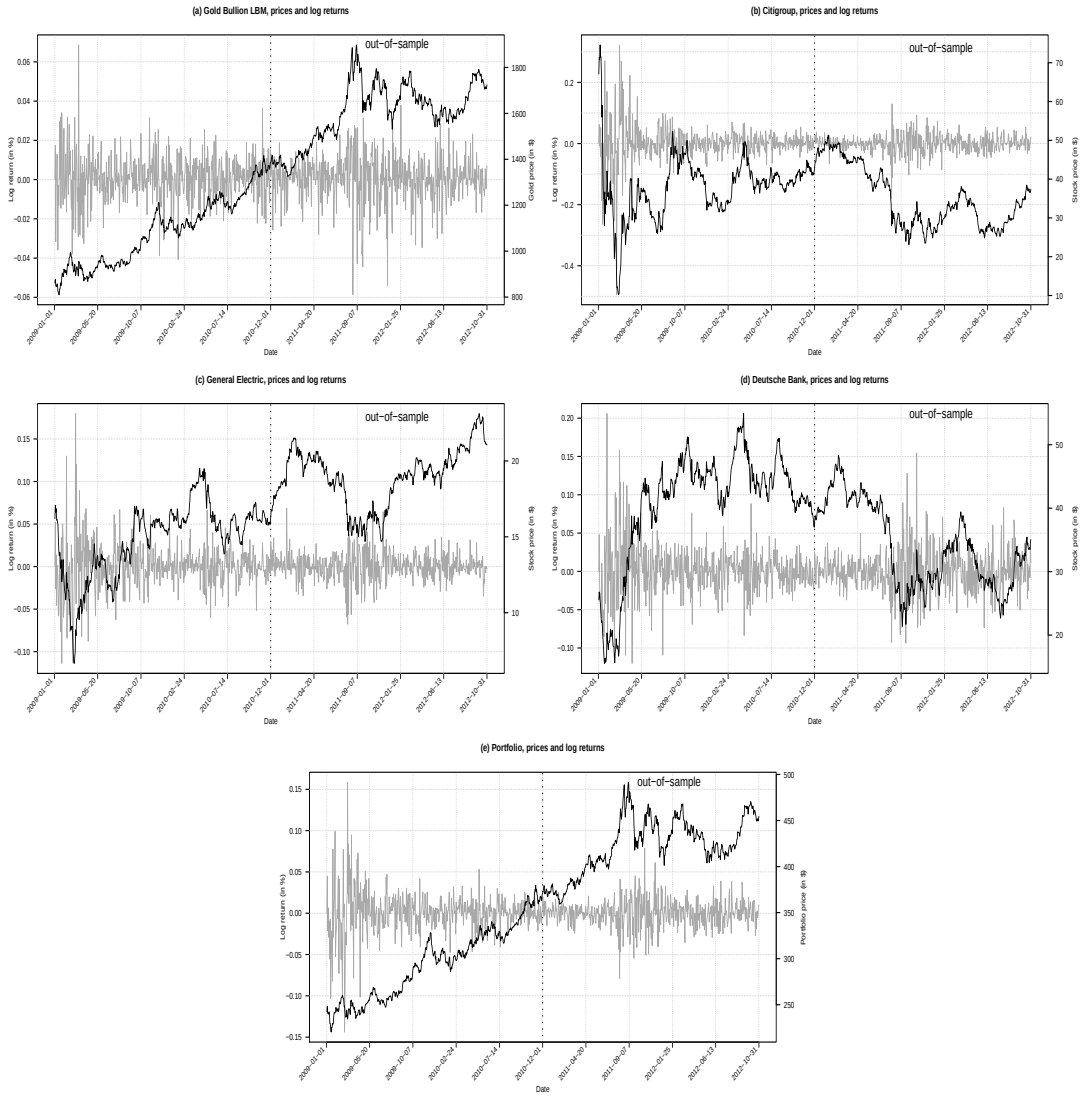
The dependence is strongest between gold bullion and the individual stocks. Thus, we believe it is appropriate to model the empirical data using a canonical vine model with gold bullion as the so-called *root node*. Consequently, the pair-copulas in the first tree of the vine model the bivariate distributions of gold bullion-Citigroup, gold bullion-Deutsche Bank, and gold bullion-General Electric. In the benchmark model, we use the same set of candidate parametric copula families as in our simulation study presented in Section 4.3.1. Again, in our mixture approach, each pair-copula is modeled as a convex combination of a Clayton, a Frank, a Gumbel, and a Student's t copula.[152] Thus, both models allow for a wide range of possible dependence structures.

The data for our empirical study covers the time period from January 1, 2009 to October 31, 2012. Our sample period thus starts in the post Lehman Bros. era and covers the climax of the European Sovereign Debt crisis. In total, our sample includes $1,000$ trading days. We forecast the portfolio's Value-at-Risk by employing rolling windows of 500 trading days. In turn, the out-of-sample period consists of 500 trading days. Figure 4.3 shows plots of the log returns and the quotes on the four financial assets (Panel (a) through (d)) as well as the portfolio (Panel (e)), respectively. Each plot includes a vertical line to separate the initial in-sample and the out-of-sample.

[152] Note that the pre-specified set of parametric copulas from which the mixture pair-copulas are built can easily be extended to more than four candidate parametric copulas. Preliminary tests in our study, however, showed that the four parametric copulas we use suffice to model the dependence structure in the data.

Figure 4.3: **Time series plots of quotes and log returns of individual assets and the portfolio used in the empirical study.**

The figure shows plots of the log returns (grey lines) and the quotes (black lines) on the financial assets considered in the empirical study. Panel (a) shows returns and quotes for Gold Bullion LBM ($/Troy Ounce), Panel (b) for the stock of Citigroup, Panel (c) for the stock of General Electric, Panel (d) for the stock of Deutsche Bank, and Panel (e) shows the quotes and log returns on an equal-weighted portfolio consisting of the four individual assets. The sample covers the period from January 1, 2009 to October 31, 2012 (1000 trading days). Each plot includes a vertical line to separate the initial in-sample and the out-of-sample. The data are taken from *Thomson Reuters Financial Datastream*.



The plots of the four portfolio constituents highlight several challenging features with respect to VaR-forecasting. First, all four assets possess different evolutions of their time series' volatility during the in- and the out-of-sample. While the returns of General Electric and Citigroup seem to calm down after a high volatility phase at the beginning of the in-sample (and thus, during the financial crisis), the time series of the

stock of Deutsche Bank exhibits high (and clustered) volatility in the out-sample. Most probably, the increase in stock return volatility for Deutsche Bank was caused by the renewal of the European Sovereign Debt Crisis in the summer of 2011. Throughout our sample, the gold price increases steadily with volatility spiking in the out-of-sample. Complementing the plots of the individual assets, Panel (e) in Figure 4.3 plots the time series prices and log returns of an equal-weighted portfolio of the four assets. While the time evolution of the price of the portfolio is almost identical to that of the gold bullion, the log returns on the portfolio exhibit both calm and extremely volatile phases. Moreover, the log returns also exhibit extreme positive and negative spikes at the start of the in-sample during the financial crisis.

To further illustrate the properties of our time series, we plot histograms of the log returns of the four individual assets and the portfolio in the full sample in Figure 4.4.

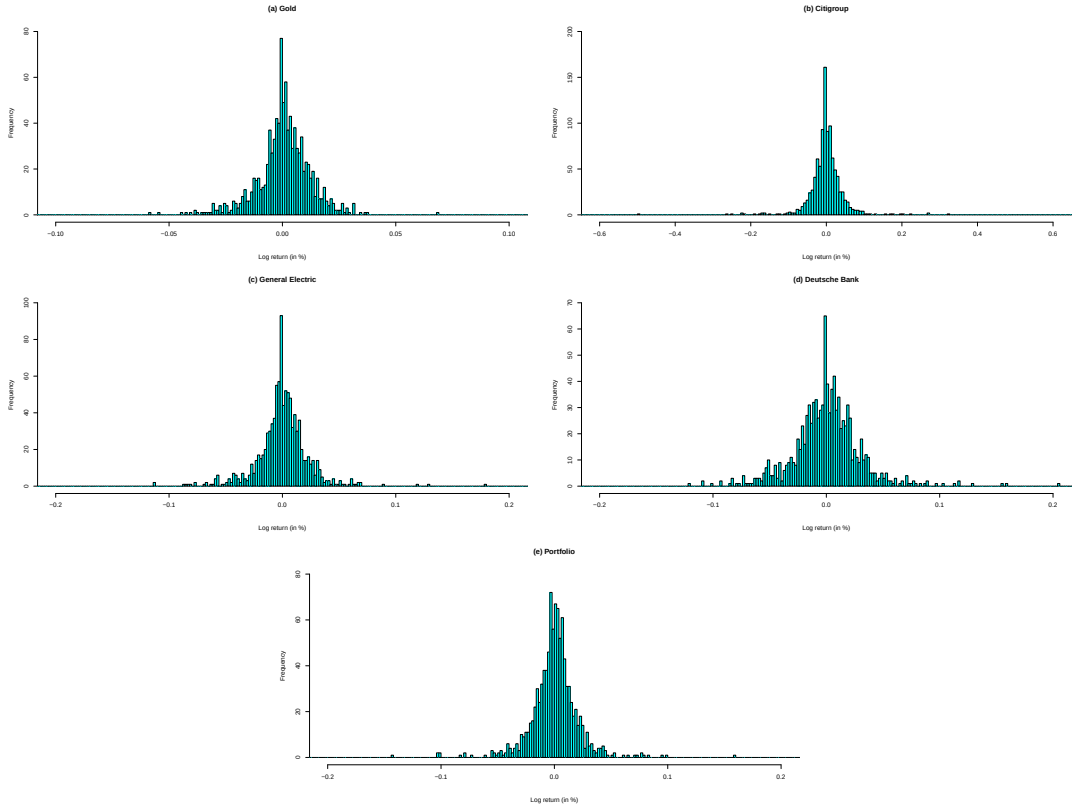
The plots given in Figure 4.4 underline the heavy left tail of the portfolio returns as well as the existence of several extreme returns in our data.

While this evidence prompts the use of copula models for modeling the joint distributions' behaviour in the tails, symmetric marginals might not suffice to model asymmetries in both the distribution of the residuals as well as in conditional volatilities. We therefore employ both the Jarque-Bera and the Kolmogorov-Smirnov test to check the hypothesis of normally distributed residuals. Both tests are rejected. Furthermore, performing the D'Agostino normality test on the residuals indicates strong deviations from normality due to kurtosis. Finally, we check the adequacy of our assumption of Student-t distributed residuals using Quantile-Quantile-plots. The results underline the adequacy of our marginal models and the fact that asymmetric marginal models are not required for our particular sample.

In order to check for asymmetries in conditional volatility, we jointly conduct the Sign Bias Test, the Negative Size Bias Test as well as the Positive Size Bias Test as proposed by Engle and Ng (1993). More precisely, we test the null that the squared residuals of the return series cannot be predicted by the sign and the magnitude of return shocks. The resulting p-values of the tests are quite large for all series, indicating

Figure 4.4: **Histograms of the log returns of individual assets and the portfolio used in the empirical study.**

The figure shows histograms of the log returns on the financial assets considered in the empirical study. Panel (a) shows returns for Gold Bullion LBM ($/Troy Ounce), Panel (b) for the stock of Citigroup, Panel (c) for the stock of General Electric, Panel (d) for the stock of Deutsche Bank, and Panel (e) shows the log returns on an equal-weighted portfolio consisting of the four individual assets. The sample covers the period from January 1, 2009 to October 31, 2012 (1000 trading days). The data are taken from *Thomson Reuters Financial Datastream*.



that there is no predictive power in the shocks with regard to the squared residuals. We thus find no evidence of asymmetric conditional volatility in any of the return series.

4.4.2 Results

We use the methodology presented in Section 4.4.1 to compute the one-step-ahead forecasts of the portfolio-VaR for each day in the out-of-sample using rolling windows of 500 trading days. To analyze the differential effect of different confidence levels for the VaR on our models' forecasting accuracy, we forecast the 0.1%-, 1%-, 2.5%- and 5%-VaR for a long position in the portfolio and the 95%-, 97.5%-, 99%-, and 99.9%-VaR for a short position, respectively. The selection of these VaR-levels is in

part motivated by the results of the simulation study presented in the previous section. On the one hand, it has been highlighted that both models, the heuristically calibrated parametric benchmark as well as our Mixture-PCC, seem to be well-suited for producing accurate 0.1%-VaR-forecasts. This result is quite remarkable given the fact that the 0.1%-VaR-level is of high importance for the calculation of regulatory capital requirements implemented in the Basel III framework. On the other hand, especially the heuristic benchmark was significantly outperformed in in-sample backtesting by the Mixture-PCC for higher significance levels.[153]

First, we present a comparison of the estimated pair-copula families and parameters for the benchmark model as well as the estimated weights and parameters in the mixture model for the dates for which the difference in the VaR-forecasts of both models is largest.[154] For the 0.1%-VaR-level, the largest difference between the forecasts of the two models is nearly 6% in portfolio-VaR (date: 2011-08-23). For the benchmark model, the Student's t (0.02; 13.11 degrees of freedom), the Survival Clayton (0.11) and the Rotated Joe-Clayton (1.06) copula are chosen in the first tree. In the second tree the Rotated Clayton-Gumbel copula is twice selected (1.71 and 1.53, respectively). The only copula in the third tree is the Gaussian (0.38).

The largest differences for the 1%- and 2.5%-VaR-levels are given on the same date (2011-08-11) and amount to 2.31% and 1.8% in portfolio-VaRs, respectively. The fitted benchmark model differs only marginally in the first two trees from the previous model. More precisely, the same copula families are chosen and the estimated parameters show only little variation. In the third tree, the Survival Clayton-Gumbel copula is chosen (2.19; 0.86). For the 5%-VaR-level the largest difference in forecasted VaRs is 1.34% (date: 2011-11-02). Here, in the first tree of the benchmark vine model, three Student's t pair-copulas ($-0.05, 0.01$ and -0.02; $6.59, 13.55$ and 6.70 degrees of freedom) are used. Again, in the second tree the Rotated Clayton-Gumbel copula is se-

[153] The estimation of the benchmark PCC required less than ten minutes, on average, while the simulation of the returns and the forecasting of the VaR required less than one hour, on average. We again find the estimation of our Mixture-PCC to take about three to five times longer than the heuristic benchmark.

[154] In the following, parameter estimates are presented in brackets.

lected twice (1.81 and 1.61 respectively). In the third tree the Survival Clayton-Gumbel copula is chosen (1.63; 0.96).

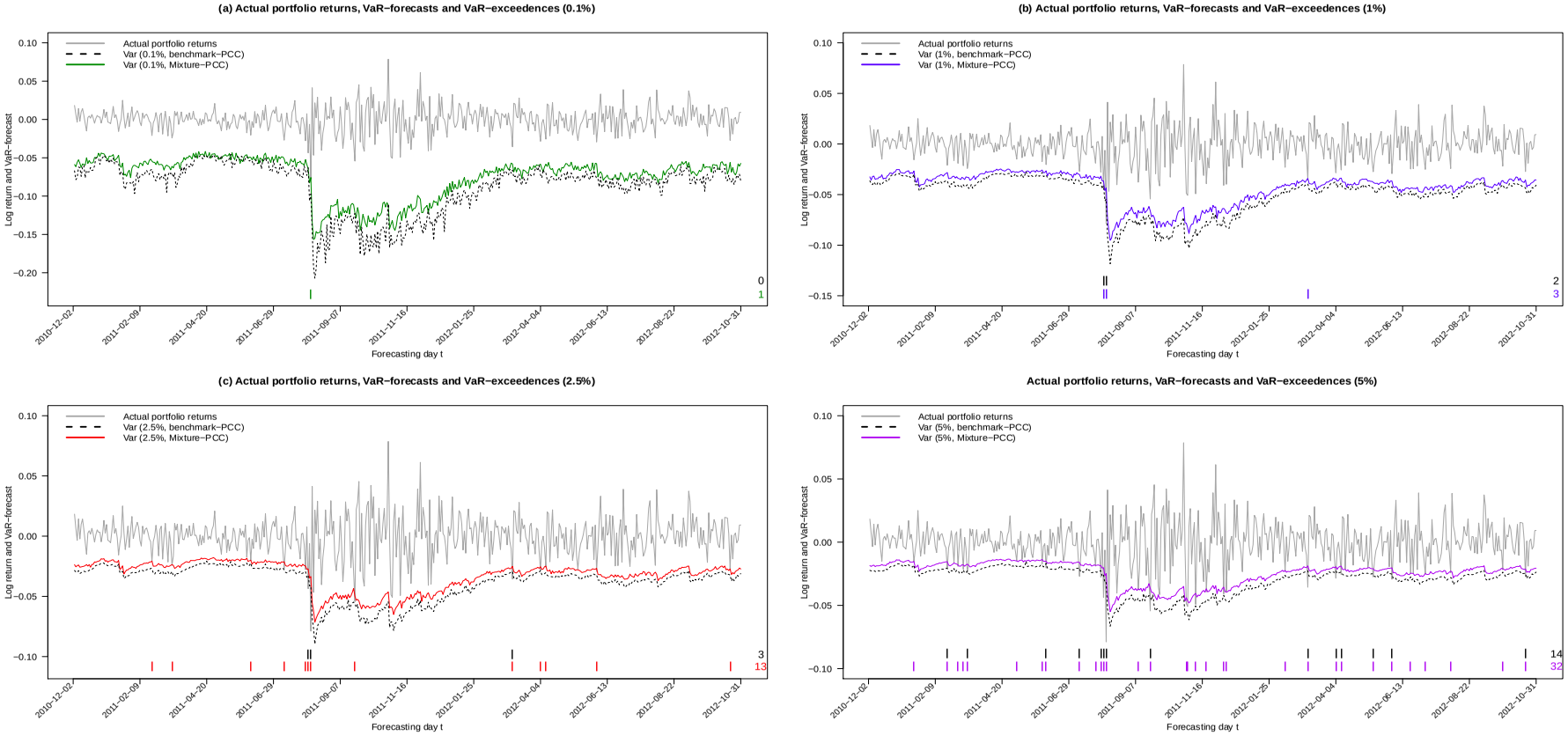
For all fitted Mixture-PCCs, the key component is the Student's t copula ($[-0.19, 0.25]$; degrees of freedom $\in [1.67, 14.37]$) with a percentage share of at least 31.36% and not more than 46.38% in all mixtures. The Frank copula ($[-89.04, 0.99]$) is the second main component and models the outcomes with strong negative dependence. However, the dependence in the tails of the Frank copula itself tends to be relatively weak. The weights are in the range of 23.27% to 29.50%. Furthermore, the Mixture-PCCs are significantly influenced by the Clayton copula $[1.33, 4.70]$ thus inducing significant lower tail dependence in the pair-copulas. The percentage share of this copula is in the range of 17.14% to 22.39%. Only a minor impact is given by the Gumbel Copula ($[2.33, 5.64]$), which captures the upper tail dependence in the model. Here, the weights are in the range between 10.62% and 21.64%.

Figure 4.5 shows the out-of-sample VaR-forecasts as well as the realized portfolio returns for all significance levels of a long position in the portfolio. Each of the Panels (a), (b), (c) and (d) presents the realized portfolio returns and the VaR-forecasts for the corresponding confidence levels highlighting the results of our proposed Mixture-PCC and the simple PCC calibrated via the heuristic based on Akaike's Information Criterion, respectively. Corresponding results for the short position in the portfolio are illustrated in Figure 4.6.

The plots in Figure 4.5 highlight our previous finding that both the Mixture-PCC and benchmark PCC adequately forecast portfolio losses. Both models seem to adapt well to the specific evolution of the realized portfolio returns for all VaR-levels. Additionally, note that exceedances of the VaR-forecasts occur only in periods of high volatility combined with large losses in the long position of the portfolio investment. We would expect 0.5, 5, 12.5 and 25 VaR-exceedances for the four significance levels. While the benchmark-PCC yields 0, 2, 3 and 14 violations, our Mixture-PCC forecasts portfolio losses more accurately with 1, 3, 13 and 32 violations.

Figure 4.5: **Comparison of 0.1%-, 1%-, 2.5% and 5%-VaR forecasts and realized portfolio returns.**

The figure shows plots of the log returns on the four-dimensional portfolio considered in the empirical study and the out-of-sample VaR forecasts. Panel (a) presents the realized portfolio returns (grey line) and the VaR forecasts at the 0.1% significance level computed by the use of the benchmark pair-copula construction calibrated following the procedure laid out in Brechmann et al. (2012) (black line) and our mixture pair-copula-construction (colored line). Panels (b), (c) and (d) show similar comparisons for the 1%-, 2.5% and 5%-VaR. VaR-exceedances are indicated by vertical strokes at the bottom of each plot with the total number of exceedances given in the lower right corner of each plot. The size of the out-of-samples is $T = 500$. The equal-weight portfolio consists of the returns on Gold Bullion LBM and the stocks of Citigroup, General Electric, and Deutsche Bank. The data are taken from *Thomson Reuters Financial Datastream*.



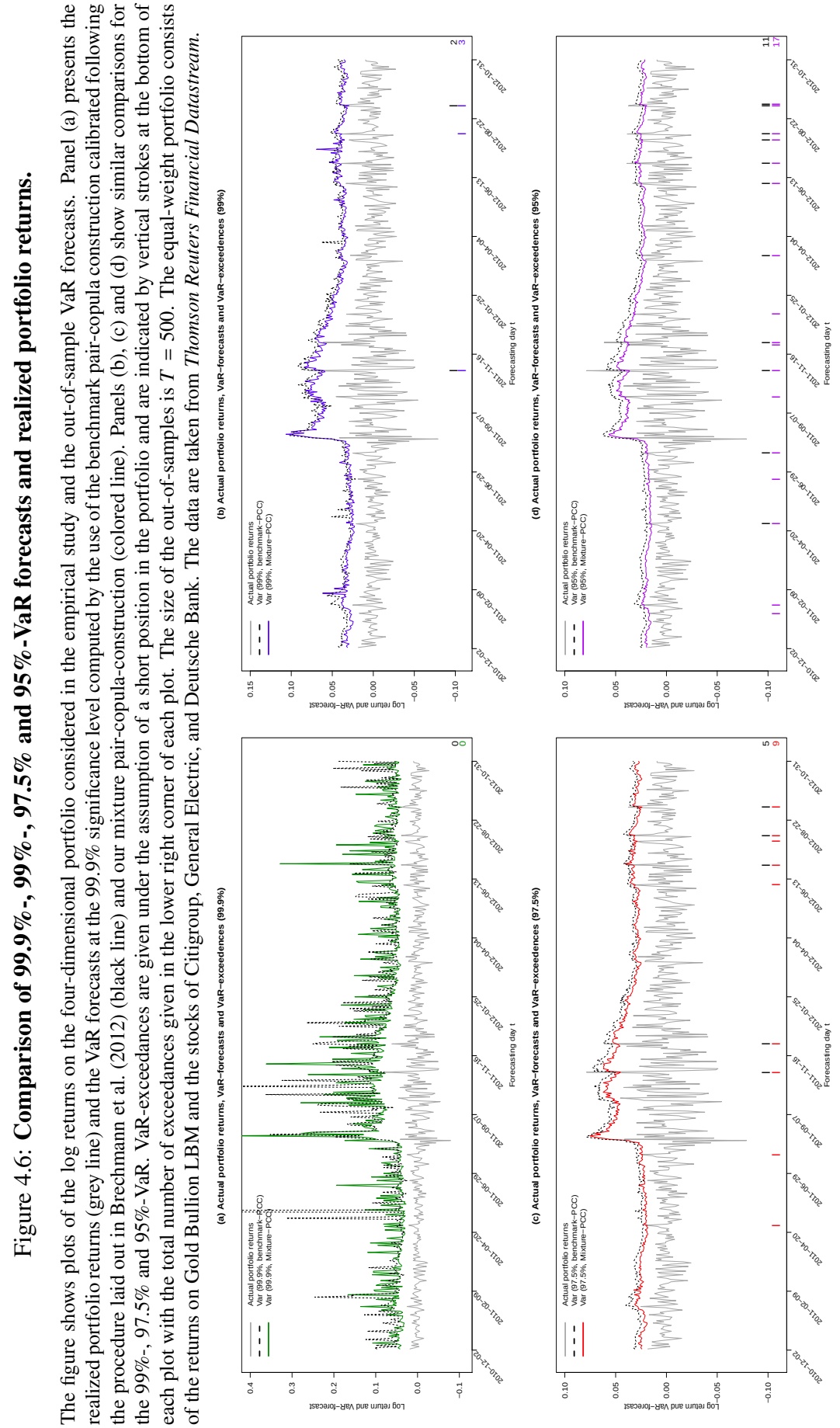


Figure 4.6: **Comparison of 99.9%-, 99%-, 97.5% and 95%-VaR forecasts and realized portfolio returns.**

The figure shows plots of the log returns on the four-dimensional portfolio considered in the empirical study and the out-of-sample VaR forecasts. Panel (a) presents the realized portfolio returns (grey line) and the VaR forecasts at the 99.9% significance level computed by the use of the benchmark pair-copula construction calibrated following the procedure laid out in Brechmann et al. (2012) (black line) and our mixture pair-copula-construction (colored line). Panels (b), (c) and (d) show similar comparisons for the 99%-, 97.5% and 95%-VaR. VaR-exceedances are given under the assumption of a short position in the portfolio and are indicated by vertical strokes at the bottom of each plot with the total number of exceedances given in the lower right corner of each plot. The size of the out-of-samples is $T = 500$. The equal-weight portfolio consists of the returns on Gold Bullion LBM and the stocks of Citigroup, General Electric, and Deutsche Bank. The data are taken from *Thomson Reuters Financial Datastream*.

Panel (a) in Figure 4.5 underlines our first impression from the simulation study that both models forecast the 0.1%-VaR of the portfolio quite accurately. In this case 0 and 1 realized exceedances for the benchmark model and our proposed Mixture-PCC, respectively, indicate an accurate forecast of the expected number of exceedances (0.5). Panel (b) shows similar VaR-forecasts for both models at the 1% significance level. Furthermore, both models again yield similar numbers of VaR-violations. Additionally, Panel (c) provides further interesting insights. On the one hand, the benchmark model again appears to forecast losses accurately. On the other hand, the mixture model yields almost the expected number of VaR-exceedances while the benchmark is significantly more conservative. Also, note that the plots of the VaR-forecasts of our Mixture-PCC are always above the plot of the forecasts of the benchmark. These findings are underlined by the plots shown in Figure 4.6 for the short position in the portfolio.

The plots in Figures 4.5 and 4.6 underline the impression that both models forecast the VaR of the portfolio quite accurately and yield (approximately) correct numbers of VaR-exceedances for the first three confidence levels. The key difference, however, is that our proposed Mixture-PCC model does not overestimate portfolio risk to such an extent as the benchmark model does. Our model can thus help risk managers to reduce regulatory risk capital.

To further substantiate these findings, we perform the two-tailed conditional coverage backtest proposed in Christoffersen (1998) on the out-of-sample VaR-forecasts of both models. The results of the formal backtest are presented in Table 4.3.[155]

[155] We also computed VaR-forecasts using historical simulations as a benchmark. In unreported results, this benchmark model produced six VaR-exceedances for the 0.1%, 1%, and 2.5% significance levels, and eleven exceedances for the 5% VaR-level (long position). Consequently, this naive benchmark produced VaR-forecasts that were regularly off target with the model significantly overestimating portfolio risk for most VaR-levels.

Table 4.3: **Backtesting results - Empirical study.**

The table presents the results of the two-tailed conditional coverage backtest proposed in Christoffersen (1998) performed on the out-of-sample forecasts for the portfolio-VaR estimated from the vine copula models calibrated by the sequential heuristic of Brechmann and Czado (2013) and Dissmann et al. (2013) and the Mixture-PCCs, respectively. The table reports the expected and the realized number of VaR-exceedances as well as the p-values of the two-tailed backtest of conditional coverage proposed in Christoffersen (1998) (CC) and the duration-based Weibull test of independence by Christoffersen and Pelletier (2004) (WB). For both models, the backtesting results are reported for the $(1-\alpha)$-VaR for significance level $\alpha \in \{0.1\%; 1\%; 2.5\%; 5\%\}$. For the 95%-, 97.5%-, 99%- and 99.9%-VaR, exceedances are given under the assumption of a short position in the portfolio. The out-of-sample consists of $T = 500$ trading days. The equal-weight portfolio consists of the returns on Gold Bullion LBM and the stocks of Citigroup, General Electric, and Deutsche Bank. The full sample covers the period from January 1, 2009 to October 31, 2012 (1000 trading days) with the data being taken from *Thomson Reuters Financial Datastream*.

		Sequential AIC			**Mixture-PCC**		
	Exceedances (expected)	**CC test (P-value)**	**WB test (P-value)**	**Exceedances (realized)**	**CC test (P-value)**	**WB test (P-value)**	**Exceedances (realized)**
VaR $\alpha = 0.1\%$ Cond. Cov.	0.5	0.2285	1.0000	0	1.0000	1.0000	1
VaR $\alpha = 1\%$ Cond. Cov.	5	0.2231	0.0402	2	0.4250	1.0000	3
VaR $\alpha = 2.5\%$ Cond. Cov.	12.5	0.0032	1.0000	3	0.8211	0.7887	13
VaR $\alpha = 5\%$ Cond. Cov.	25	0.0274	0.7695	14	0.3113	0.1764	32
VaR $\alpha = 95\%$ Cond. Cov.	25	0.0016	0.9389	11	0.2048	0.3362	17
VaR $\alpha = 97.5\%$ Cond. Cov.	12.5	0.0348	0.7567	5	0.4279	0.7065	9
VaR $\alpha = 99\%$ Cond. Cov.	5	0.2264	0.0385	2	0.4126	0.9116	3
VaR $\alpha = 99.9\%$ Cond. Cov.	0.5	0.2343	1.0000	1	0.2322	1.0000	0

The results of the backtest given in Table 4.3 again show that both models yield accurate VaR-forecasts for a significance level of 0.1%. The p-values differ considerably between both models, but both are not rejected at the 5% significance level. This is in line with our previous findings of the simulation study for the 0.1%-VaR-level. For the 1% significance level of the VaR, again, both models are not rejected by the formal backtest of conditional coverage. However, the formal backtest supports the results presented earlier for the $\alpha = 2.5\%$-Value-at-Risk forecasts. While the benchmark PCC is clearly rejected by the backtest due to its overestimation of portfolio risk, our Mixture-PCC is not rejected. This result also holds for the $\alpha = 5\%$-Value-at-Risk forecasts even though the number of the realized exceedances of the Mixture-PCC is higher than the expected number of exceedances. The VaR-levels of the short portfo-

lios (95%, 97.5%, 99%, and 99.9%, respectively) permit a similar interpretation. The benchmark PCC is clearly rejected whereas our Mixture-PCC is not rejected.[156] Consequently, the results of the formal backtest support our main finding that Mixture-PCCs forecast losses of financial portfolios significantly more accurately than the heuristic benchmark PCC.

A note of caution is, however, in order when interpreting our results. The use of mixture copulas as pair-copulas increases the (already large) number of parameters that need to be estimated to fully specify a pair-copula-constructions. As a result, the parameter uncertainty and the risk of overfitting the data (especially when the sample size is small) increase in parallel with the model's increased flexibility. As the Mixture-PCCs produce accurate and reliable results in both our simulations and our empirical study, we believe parameter uncertainty to be of lesser concern in the particular setting of our study. However, the potential bias due to a (too) small sample size and a (too) high number of mixture components could significantly impact results when using Mixture-PCCs in different settings.

4.5 Conclusion

In this study, we propose the use of mixture copulas in d-dimensional Pair-Copula-Constructions as a new strategy to circumvent the otherwise necessary (and error-prone) selection of parametric forms for the $d(d-1)/2$ bivariate pair-copulas. While previous studies in the literature have tried different approaches to select optimally fitting pair-copulas from parametric copula families characterized by different tail dependence (e. g., goodness-of-fit tests, graphical tools, etc.), we propose to use convex combinations of these bivariate copulas for each pair-copula in a vine model. Each mixture pair-copula is then estimated using the well-known EM-algorithm yielding a fully specified vine model in which no parametric copula needs to be selected as all parametric candidate copulas can be included in the mixture pair-copulas. After

[156] Complementing the test of conditional coverage, the results of the Weibull backtest show that with one exception no model is rejected based on the assessment of the independence of the exceedances.

outlining our proposed Mixture-PCC, we test the performance of our new model in comparison with a benchmark PCC in which each pair-copula is chosen by computing the Akaike's Information Criterion for each candidate parametric copula and selecting the copula with the optimal AIC value. We perform both a simulation study on the in-sample fit of both models as well as an empirical study in which we assess both models' out-of-sample forecasting accuracy.

Our main result can be summarized as follows: in our simulations and in the empirical study, both models yield acceptable Value-at-Risk forecasts. However, we show that our proposed Mixture-PCC yields better results in backtesting for at least the 2.5% significance level while the benchmark overestimates portfolio risk. Especially for higher dimensions, our Mixture-PCC model seems to approximate portfolio losses better than the benchmark which is far too conservative in many cases. Consequently, our model can help risk managers to save on regulatory risk capital while at the same time satisfactorily bounding possible portfolio losses.

For future research, one could think of an analysis of extended Mixture-PCCs in which the number of mixture pair-copulas is not fixed as it is done in this study. Furthermore, Mixture-PCCs could be combined with pruning strategies to truncate or simplify some of the mixture pair-copulas in lower trees to limit the computational cost. We expect all these extensions to lead to further improvements on the forecasting accuracy of our Mixture-PCC in comparison to models from the related literature and intend to address them in a future study.

Chapter 5

Extreme Dependence in Investor Attention to Bank stocks

5.1 Introduction

The effect of investor attention on financial markets has been of long-standing interest to economists. Traditionally, measures of attention were restricted to indirect proxies like trading volume, media news, and abnormal returns. Recent research in financial econometrics literature illustrates the usefulness of Google search data as a direct measure for retail investors' attention. Starting with the work by Da et al. (2011), most applications of Google search data in financial applications focus on asset pricing implications and predicting dynamics of stock market volatility (see, e. g., Da et al., 2011, Hamid and Heiden, 2014, Mondria and Wu, 2011, Vozlyublennaia, 2014). In the course of these studies, the co-movement between investor attention and stock prices (volatility) has been investigated in detail.

In this study, we provide a statistical modeling framework for specifying, estimating, and testing time series of investor attention measured by internet search queries. More precisely, our paper is the first to present both a univariate as well as a multivariate econometric model for Google search data. We find that the dependence structure of high-dimensional Google search data is significantly non-linear and asymmet-

ric. Furthermore, we document the existence of extreme dependence in Google search data pairs and, particularly noteworthy, between stock returns and the corresponding Google search data. Finally, our main contribution is to show a striking similarity in the joint distributions of a multivariate bank stock portfolio and the corresponding portfolio of Google search queries, respectively. Following our results, we hypothesize that investor attention as measured by internet search data and stock returns reflect almost the same information.

Starting point of our paper is the proper modeling of internet search queries. While modeling of time series such as stock returns, foreign exchange rates, and CDS spreads has become common practice in financial econometrics, the application of statistical techniques to investor attention measured by internet search data is widely unexplored. To this end, we provide a comprehensive time series analysis and extract meaningful statistics and other characteristics of Google search data. We find both autoregressive dynamics and conditional heteroskedasticity in the underlying data set. Additionally, there is statistical evidence for specific distributional characteristics like the presence of distinct levels of skewness and kurtosis within Google search data. Our analysis shows that the first- and second-moment dependence are well captured by asymmetric ARMA-CS-GARCH models. Finally, we also find that the skewed t as well as the skewed generalized error distribution (*sged*) provide good fits to the Google search data residuals.

In our multivariate econometric framework, we aim to model the joint distribution of high-dimensional search query data in a flexible way. In this regard, we propose the implementation of a vine copula approach, which is especially appropriate for two reasons: First, it allows to capture both linear dependence as well as potential non-linearities in the dependence structure. In fact, we document the existence of strong non-linear and asymmetric dependence in the Google search data. Second, due to their hierarchical construction, vine copulas allow to model different dependence structures between pairs of variables. As a result, our study provides the first empirical evidence of significant tail dependence in Google search data.

Beside their usefulness in capturing inherent dependency patterns of high dimensional data sets, vine copulas provide a powerful tool to detect similarities in the joint distribution of different data sets. To be precise, we find a striking similarity in the joint distributions of a multivariate bank stock portfolio and the corresponding portfolio of Google search queries, respectively. The remarkable similarities necessitates a detailed investigation of the co-movement between investor attention and stock returns. In our analysis, we provide first empirical evidence for the existence of tail dependence between stock returns and the respective search query pairs. Furthermore, we document that stock returns and Google search data evolve concurrently in real time. Our results suggest that investor attention measured by internet search data and stock returns reflect almost the same information.

Our paper makes several major contributions. Firstly, our findings push forward significantly the knowledge of search data and its characteristics. The study of Dimpfl and Jank (forthcoming) indicates that Google search data collected within one country are characterized by specific properties that are usually common to financial time series, like non-normality and autocorrelation. In contrast to their work, however, we take the discussion further and provide a comprehensive time series analysis of worldwide Google search data. Based on our results, our paper is the first to present an econometric model that is well-suited for capturing first- and second-moment dependencies in univariate search queries.

Furthermore, we propose to model the joint distribution of Google search data by using regular vine (R-vine) copulas. In this regard, our paper complements several previous studies in the field of using vine copulas in financial econometrics and quantitative risk management (see, e. g., Aas et al., 2009, Min and Czado, 2010, Dissmann et al., 2013, Christoffersen et al., 2012, Oh and Patton, 2013). But, to the best of our knowledge, this article provides the first application of the vine copula concept to internet search data.

In our empirical application, we provide first empirical evidence of significant tail dependence in Google search data. Referring to this, we contribute to the current

state of research in documenting extreme dependencies in time series. While non-linear dependence has been shown to exist in financial time series like stock returns (see, e. g., Poon et al., 2004, Bollerslev and Todorov, 2011) and credit risk (see, e. g., Christoffersen et al., 2013), this paper is the first to confirm that investor attention measured by Google search data is characterized by strong tail dependence as well.

Finally, our multivariate econometric framework is also beneficial in the context of modeling co-dependencies between investor attention and stock returns. The idea to examine possible causal relations between investor attention and stock returns is related to several studies in the literature (see, e. g., Da et al., 2011, Dimpfl and Jank, forthcoming, Hamid and Heiden, 2014, Mondria and Wu, 2011, Vozlyublennaia, 2014). However, in the course of these works the co-movement between attention and stock prices (volatility) is being investigated by the classical tools of Vector Autoregressive models and Granger Causality tests, despite their limitations to capture non-linear and asymmetric dependencies in time and between the data series. In contrast, our vine copula modeling approach is especially appropriate for non-linear and asymmetric modeling between time series dependencies. In fact, our study is the first to document significant tail dependence between investor attention and stock returns.

The remainder of the paper is structured as follows. In Section 5.2, we specify our data sample and present descriptive statistics of the data sets. The marginal and multivariate models we employ to the data are presented and discussed in Section 5.3. Section 5.4 contains the empirical application and a comprehensive discussion of the economic importance of the empirical findings. Section 5.5 concludes.

5.2 Data sample and descriptive statistics

We start our analysis by selecting a suitable data sample which provides reliable data for both stocks as well as internet search data. Due to the high liquidity of bank stocks and the high level of attention paid to banks in recent years, we focus on the biggest banks in the world measured by total assets and market capitalization. For these banks we collect daily data from equity markets and Google search queries as a measure for retail investors' attention. Starting with a brief specification of the sample construction we then describe both data sources and present descriptive statistics of the data subsequently.

5.2.1 Sample Construction

We restrict our analysis to the 50 biggest banks in the world, as measured by total assets and market capitalization, and collect data on a daily basis over a three-year period ranging from January 2011 to December 2013. We obtain equity prices for the banking firms from *Thomson Reuters Datastream*. Collecting daily data of internet search queries is a more challenging task. We retrieve search query data from the public social trends analytics service *Google Trends*.[157] The use of Google search data is beneficial, because Google is the largest global search engine and has a very large market share in the countries where the considered banks are domiciled.[158] In general, an ambitious task of identifying search frequencies is the wide range of terms which are linked to a (banking) firm (see, e. g., Da et al., 2011, Dimpfl and Jank, forthcoming). But since we are only interested in the active attention that retail investors pay to these banks, we only collect search volume indexes based on the banking firm names and do not employ ticker symbols.[159] Note that professional traders will not search firm name information in Google since all relevant news is provided by trading platforms. Un-

[157] Source: *http://www.google.com/trends*.

[158] In the reporting period *Google's* market share is around 93.21% in Europe, 92.55% in Asia, 84.50% in North America, 96.17% in South America and 92.92% in Oceania. Source: StatCounter.

[159] See Da et al. (2011) for a detailed discussion of the merits and flaws of using ticker symbols for companies.

fortunately the search for several bank names results in insufficient data series. More precisely, in 21 out of 50 bank names *Google Trends* reveal search volumes only for greater time intervals, e. g. weekly and quarterly, respectively. To this end, we have to remove the respective banks from our sample and proceed the analysis with 29 remaining banks (see Table A in Appendix C.1 for a detailed description of the sample constituents).

5.2.2 Stock Returns

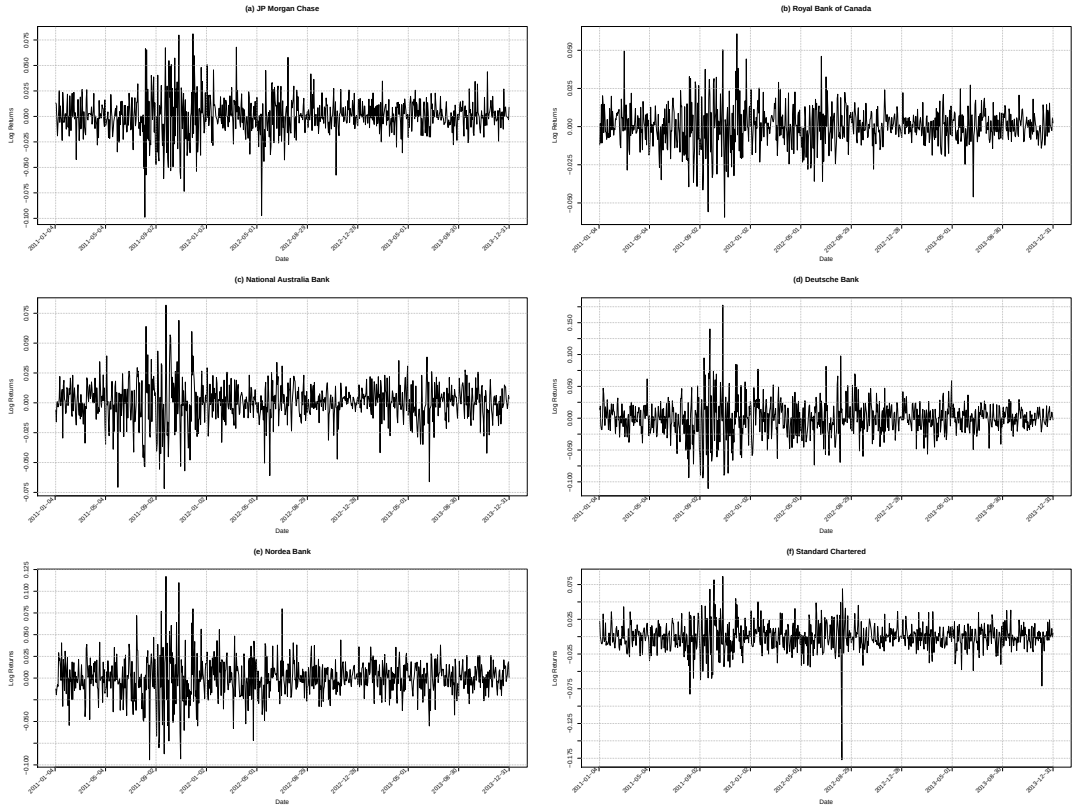
Equity prices for the remaining 29 banking firms are retrieved from *Datastream*. The relevant sample period covers 782 stock trading days and for each trading day and each of the banks, we calculate daily log-changes of stock prices. To control for known data errors in *Datastream*, we check for abnormal log returns above 300% that are reversed within one month. Note, that none of our 29 stock price time series exhibits this type of data error. Also note, during the course of 2011 - 12, four out of eighteen euro zone states[160] were in trouble to repay or refinance their government debt. Thus, we expect the stock returns of the banks to experience a strong co-movement due to their major investments in government bonds. This suspicion is confirmed by Figure 5.1 in which the time evolution of different log return series is illustrated. For increased readability, we restrict our graphical representation to 6 out of 29 time series.

The plots depict remarkable similarities in the time-evolution of the selected log return series. To put it more precisely, all plots exhibit the common stylized fact of volatility clusters, which becomes particularly clear for the periods of mid 2011 to early 2012 and the summer of 2012. As indicated above, the increased stock return volatility throughout our sample was caused by the renewal of the European Sovereign Debt Crisis. Note, after high volatility phases the log returns of all six banks seem to calm down simultaneously. Also note that the unreported plots of the other banking firms possess homogeneous evolutions of the respective time series' volatility.

[160] That is, Greece, Ireland, Portugal and Cyprus.

Figure 5.1: **Time evolution of stock returns.**

The figure shows plots of the log returns of six selected banks considered in the study. Panel (a) shows returns for JP Morgan, Panel (b) for the stock of Royal Bank of Canada, Panel (c) for the stock of National Australia Bank, Panel (d) for the stock of Deutsche Bank, Panel (e) for the stock of Nordea Bank, and Panel (f) shows the log returns of Standard Chartered. In each plot behind the bank name the number in brackets refers to the numbering in Table A. The sample covers the period from January 1, 2011 to December 31, 2013 (774 trading days). The stock return data are taken from *Thomson Reuters Financial Datastream*.



For further analysis, we now study more closely the cross-sectional variation in our log return data. Table 5.1 presents descriptive statistics on the cross-sectional distribution of daily log returns for our sample period spanning January 2011 to December 2013.

As becomes apparent from Table 5.1, the log returns are weakly negatively skewed and leptokurtic on average, with an average skewness and excess kurtosis of -0.1710 and 0.5490, respectively. Additionally we perform Ljung-Box tests for autocorrelation at up to the tenth lag and find significant autocorrelation (at the 0.05 level) in 16 out of 29 series of the log returns, and for all 29 series significant autocorrelation is found in the squared log returns. To put it briefly, we have to specify suitable econometric

Table 5.1: **Summary statistics for stock returns.**

The table reports descriptive statistics on the cross-sectional distribution of daily stock returns for the period from January 2011 to December 2013. The sample consists of the first 29 banks listed in Appendix C.1. We first calculate the time-series percentiles and moments for each bank in the sample, and then compute the cross-sectional percentiles and mean in a second step. That is, the columns present the percentiles and mean from the cross-sectional distribution of the measures listed in the rows.

		Percentiles						
	Min	5th	25th	Median	75th	95th	Max	Mean
				Stock returns				
Percentiles								
- Min	-0.2271	-0.1979	-0.1596	-0.1105	-0.0717	-0.0565	-0.0517	-0.1204
- 5th	-0.0633	-0.0589	-0.0447	-0.0337	-0.0262	-0.0209	-0.0171	-0.0368
- 25th	-0.0190	-0.0185	-0.0144	-0.0108	-0.0082	-0.0056	0.0000	-0.0114
- Median	-0.0008	-0.0000	0.0000	0.0000	0.0001	0.0008	0.0009	0.0001
- 75th	0.0000	0.0068	0.0094	0.0113	0.0142	0.0192	0.0206	0.0120
- 95th	0.0169	0.0193	0.0257	0.0342	0.0447	0.0553	0.0569	0.0356
- Max	0.0505	0.0562	0.0776	0.0928	0.1540	0.2048	0.2264	0.1113
Moments								
- Mean	-0.0014	-0.0006	-0.0003	0.0001	0.0001	0.0004	0.0007	-0.0000
- St. Dev.	0.0109	0.0122	0.0170	0.0207	0.0277	0.0362	0.0380	0.0229
- Skewness	-0.8107	-0.5190	-0.2812	-0.1781	-0.0369	0.1345	0.3051	-0.1710
- Exc. Kurt.	-2.1959	-1.5740	-0.7720	0.2370	1.4900	4.6810	6.7116	0.5490
							# of rejections	
LB test for returns							16	
LB test for squared returns							29	
LB test for absolute returns							29	

models for the conditional mean and variance to capture these features, as proposed in section 5.3.1 below.

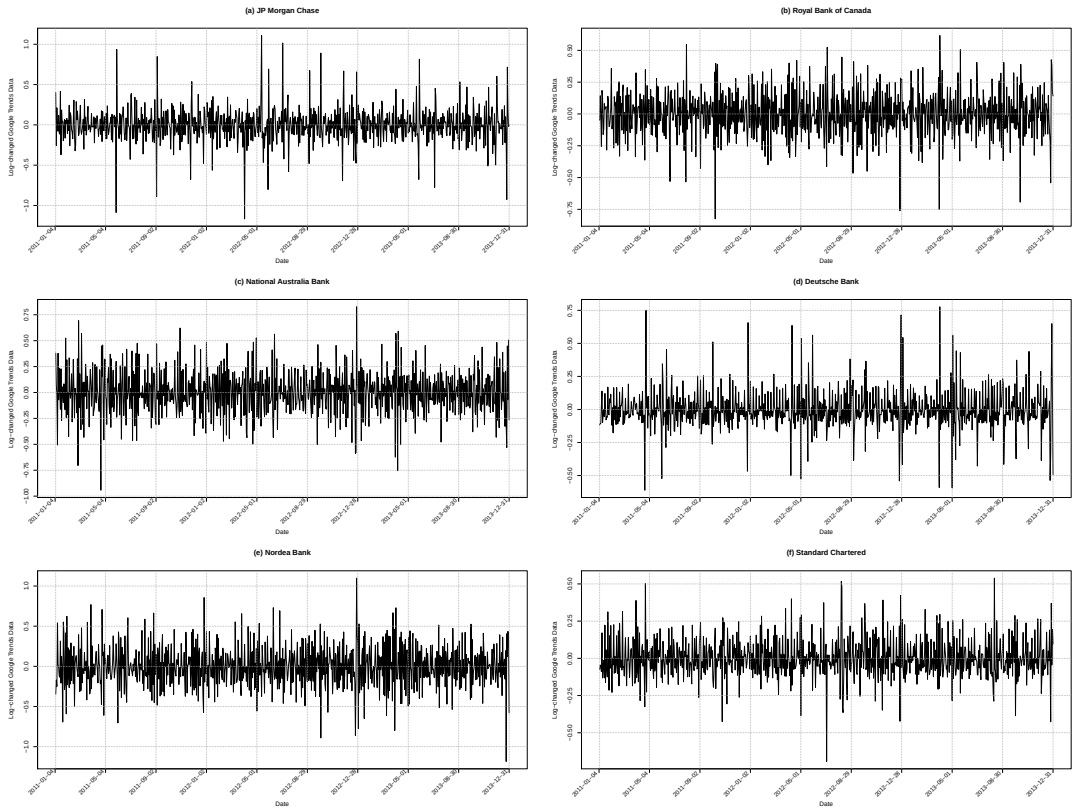
5.2.3 *Google Trends* data

We collect search query data for the names of the banks from January 2011 to December 2013 on a daily basis. *Google Trends* provides daily data for a time window up to a quarter. According to this, we download twelve quarters to cover the entire sample period. Note that the data in each quarter is scaled by the maximum of the search query data in the respective quarter. While this characteristic does not affect the calculation of the daily log-changes within a quarter, we are not able to compute the log-change of the search data from one quarter to the next seriously. Consequently,

we exclude the log-changes over consecutive quarters.[161] The log-changed time series (LC) for the other days are then calculated via $LC_t := \log(SQ_t) - \log(SQ_{t-1})$, where SQ_t ($t = 1, \ldots, T$) denotes the *Google Trends* search volume of a banking firm at time t. The time evolution of the log-changed search query for six selected banks are plotted in Figure 5.2.

Figure 5.2: **Time evolution of log-changed *Google Trends* data.**

The figure shows plots of the log-changed *Google Trends* data of six selected banks considered in the study. Panel (a) shows log-changed *Google Trends* search volume for JP Morgan, Panel (b) for the search volume of Royal Bank of Canada, Panel (c) for the search volume of National Australia Bank, Panel (d) for the search volume of Deutsche Bank, Panel (e) for the search volume of Nordea Bank, and Panel (f) shows the log-changed *Google Trends* search volume of Standard Chartered. In each plot behind the bank name the number in brackets refers to the numbering in Table A. The sample covers the period from January 1, 2011 to December 31, 2013 and is restricted to the respective 774 trading days of stock returns. The search volume data are taken from *Google Trends (http://www.google.com/trends/)*.



The plots highlight several challenging features with respect to modeling the marginal behavior of our *Google Trends* data sample. First, compared to the cor-

[161] Note, the time series of stock returns have been adjusted in the same way, i. e. we exclude stocks returns over two consecutive quarters to obtain homogeneous time series. Finally, both data series consist of 774 days.

responding stock return plots presented in Figure 5.1 large movements in investors attention resulting in greater value ranges of the log-changes become apparent. This finding holds throughout our whole sample. Moreover, we find extreme log-changes above 100% that are reversed within one week, as can be seen e. g. in Panels (a) and (e). To account for such extreme outliers in search queries, we follow Da et al. (2011) and winsorize each series at the 2.5% level in both tails. Second, the plots exhibit considerable differences in volatility, e. g., in Panels (a) and (c). Furthermore the time series seem to be characterized by seasonality, showing positive and negative jumps occurring successively.

To further analyze the properties of our time series, we present descriptive statistics on the cross-sectional distribution of the log-changed *Google Trends* data in Table 5.2.

Table 5.2: **Summary statistics for log-changed *Google Trends* data.**

The table reports descriptive statistics on the cross-sectional distribution of log-changed *Google Trends* data for the period from January 2011 to December 2013. The sample consists of the first 29 banks listed in Appendix C.1. We first calculate the time-series percentiles and moments for each bank in the sample, and then compute the cross-sectional percentiles and mean in a second step. That is, the columns present the percentiles and mean from the cross-sectional distribution of the measures listed in the rows.

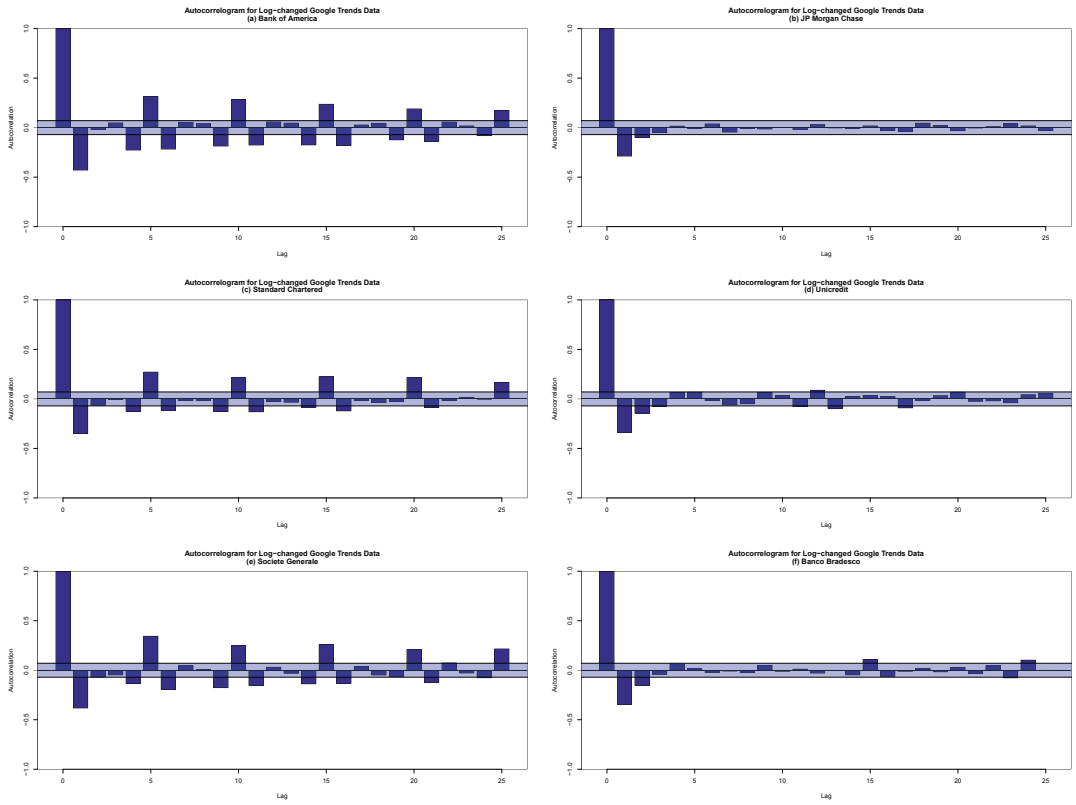
		Percentiles						
	Min	5th	25th	Median	75th	95th	Max	Mean
Log-changed *Google Trends* data								
Percentiles								
- Min	-1.5664	-1.2452	-1.1331	-0.8528	-0.6931	-0.5717	-0.4205	-0.8928
- 5th	-0.5808	-0.5612	-0.3525	-0.2624	-0.1879	-0.1185	-0.1100	-0.2962
- 25th	-0.2193	-0.2064	-0.1307	-0.0916	-0.0666	-0.0489	-0.0486	-0.1075
- Median	-0.0224	-0.0186	-0.0122	0.0000	0.0000	0.0071	0.0119	-0.0042
- 75th	0.0368	0.0429	0.0572	0.0870	0.1284	0.1938	0.2177	0.1041
- 95th	0.1138	0.1392	0.2201	0.2559	0.3758	0.5504	0.5616	0.3050
- Max	0.5390	0.6184	0.6931	0.8267	1.1087	1.4637	1.6422	0.9315
Moments								
- Mean	-0.0020	-0.0010	-0.0004	-0.0001	0.0002	0.0006	0.0026	-0.0001
- St. Dev.	0.0943	0.1164	0.1527	0.1911	0.2653	0.3147	0.3324	0.2057
- Skewness	-0.7554	-0.5572	-0.0762	0.0871	0.5952	1.1972	1.6642	0.2297
- Exc. Kurt.	-3.3793	-3.3198	-1.6075	2.1706	4.7905	9.8844	15.0345	2.4640
							# of rejections	
LB test for google							29	
LB test for squared google							27	
LB test for absolute google							28	

Our sample reveals positive skewness and is characterized by remarkable leptokurtosis. We therefore conclude that the underlying time series distributions are significantly non-normal.

To extend our investigation of the log-changed *Google Trends* data we now focus on the autocorrelation functions (ACFs) for the search queries and illustrate the corresponding ACF plots up to lag 25. For increased readability, we provide only six ACF plots in Figure 5.3, which are representative of all sample banks.

Figure 5.3: **Autocorrelation Functions (ACF) for log-changed *Google Trends* data.**

The figure shows auto-correlation plots of daily log-changed *Google Trends* data of six selected banks considered in the study. Panel (a) shows the ACF for the log-changed *Google Trends* search volume of Bank of America, Panel (b) for JP Morgan Chase, Panel (c) for Standard Chartered, Panel (d) for Unicredit, Panel (e) for Société Générale, and Panel (f) shows the ACF for the log-changed *Google Trends* search volume of Banco Bradesco. The shaded regions denote the 95% confidence interval around zero.



In each plot the shaded region denotes the 95% confidence interval. Thus, a bar extending beyond the shaded region indicates statistical significance at that lag. Consequently, the plots in Figure 5.3 underline our first impression from Figure 5.2 that the

log-changed search queries are characterized by significant autocorrelation. This result is also supported by Da et al. (2011), who find repeating hump-shaped patterns in log-changed *Google Trends* data. In the whole sample two different shapes of autocorrelation can be identified. On the one hand, Panels (a), (c) and (e) exhibit an alternating sequence of positive and negative spikes. In any of these cases the amplitude of these spikes slightly declines, but is not decaying to zero. On the other hand, Panels (b), (d) and (f) reveal significant negative autocorrelations up to lag 2. This finding is further supported by the Ljung-Box test, which is also based on the autocorrelation plot, but instead of testing randomness at each distinct lag, it checks the overall randomness based on a number of lags. For each log-changed search query we perform the Ljung-Box up to 10 lags and since the null hypothesis is rejected for all time series we conclude that the raw data are not random. Note, we are also able to reject the null of zero autocorrelation up to the tenth lag for the respective squared series. Motivated by the analysis so far we therefore need to specify econometric models that are available to cope with these issues.

5.3 Econometric Methodology

The purpose of this section is to present econometric models for the marginal distributions and the multivariate dependence structure. The modeling process is accomplished in two steps. First, we model the marginal densities of log-changed *Google Trends* data and stock returns, respectively. To capture the dependence between the marginals, we employ the concept of R-vine copulas in a second step.

5.3.1 Univariate Models for stock returns and log-changed *Google Trends* data

We first turn our attention to some fundamentals of time series analysis. To adequately model dependence characteristics of data, such as *Google Trends* data and stock returns, one important requirement of our univariate modeling approach is to generate

white-noise residuals. Thus, we are concerned with the statistical modeling of the dependence structure in the first and second moments of univariate time series. For this purpose, mean dynamics are modeled by using autoregressive moving average (ARMA) processes and based on the work by Bollerslev (1986) we then employ generalized autoregressive heteroskedastic (GARCH) processes to capture variance dynamics.

Stock returns

As financial data is frequently characterized by significant autocorrelation, we resort to AR processes for the mean dynamics of stock returns. Christoffersen et al. (2012) show that an AR process of relatively low order provides a reliable method to capture first-moment dependence. Furthermore, Oh and Patton (2013) use an AR model of order five and find the first three lags to be strongly significant. Following these findings we specify an AR(3) process to capture mean dynamics in univariate time series.[162]

With $R_{i,t}$ denoting the log return of bank i ($i = 1, \ldots, 29$) at time t ($t \geq 1$), the AR(3) process is formally estimated via

$$R_{i,t} = \mu_i + \Phi_{1,i} R_{i,t-1} + \Phi_{2,i} R_{i,t-2} + \Phi_{3,i} R_{i,t-3} + e_{i,t}.$$

The estimation method is based on minimizing the conditional least squares with the conditional mean function computed via

$$\mu_{i,t} = \mu_i + \Phi_{1,i} R_{i,t-1} + \Phi_{2,i} R_{i,t-2} + \Phi_{3,i} R_{i,t-3}.$$

In the next step, we apply GARCH-filtering techniques to the residuals $e_{i,t} = R_{i,t} - \mu_{i,t}$ to capture second-moment dependence in the time-series data. The vast majority of studies based on copula models (see, e. g., Jondeau and Rockinger, 2006, Fantazzini, 2009, Ausín and Lopes, 2010, Hafner and Reznikova, 2010) employ standard GARCH(1,1)-models to describe variance dynamics. In the financial econometrics literature, how-

[162] The notation AR(p) refers to an autoregressive model of order p.

ever, it has now become a stylized fact that volatility is asymmetric (see, e. g., Christie, 1982, Nelson, 1991). Asymmetry in volatility is commonly referred to as the leverage effect and stems from the fact that losses have a greater influence on future volatilities than do gains. Furthermore, there are several studies (see, e. g., Brailsford and Faff, 1996, Hansen and Lunde, 2005, Awartani and Corradi, 2010) showing that asymmetric volatility models outperform symmetric GARCH models regarding stock return volatility forecasting. To this end, we follow Oh and Patton (2013) and apply the GJR-GARCH model to the AR residuals.

Before proceeding, we focus on the adequate specification of the marginal distribution functions. Different empirical surveys show that the distribution of equity returns exhibits skewness and fat tails.[163] To handle these characteristics, we employ the skewed Student's t distribution of Fernandez and Steel (1998). In summary, we fit a GJR-GARCH(1,1) model to the AR residuals, $e_{i,t}$, so that

$$e_{i,t} = \sigma_{i,t}\varepsilon_{i,t}, \qquad \varepsilon_t|\mathcal{F}_{t-1} \sim iid\ skt(\nu_i, \gamma_i) \tag{5.1}$$

$$\sigma^2_{i,t} = \omega_i + \beta_i\sigma^2_{i,t-1} + \alpha e^2_{i,t-1} + \delta_i e^2_{i,t-1}\mathbb{1}_{(-\infty,0)}(e_{i,t-1}) \tag{5.2}$$

describes the evolution of the conditional volatility. In equation (5.2) $\mathbb{1}_{[\cdot,\cdot]}(\cdot)$ denotes the indicator function and the parameters α_i, β_i and δ_i must be positive. In equation (5.1) $\mathcal{F}_{t-1}$ represents the filtration containing all available information until time $t-1$ and $skt(\nu_i, \gamma_i)$ denominates the skewed t distribution with $\nu_i \in (2, \infty)$ degrees of freedom and skewness parameter $\gamma_i \in (0, \infty)$. The probability density function (pdf) f_{skt} of the skewed t distribution can be expressed by

$$f_{skt}(\varepsilon; \nu_i, \gamma_i) = \frac{2}{\gamma_i + \frac{1}{\gamma_i}}\left[f_t\left(\frac{\varepsilon}{\gamma_i}\right)\mathbb{1}_{[0,\infty)}(\varepsilon) + f_t(\gamma_i\varepsilon)\mathbb{1}_{(-\infty,0)}(\varepsilon)\right], \tag{5.3}$$

where f_t stands for the corresponding pdf of the univariate standard t distribution.

[163] See, e. g., Mandelbrot (1963) and Fama (1965) for early empirical evidence and Christoffersen et al. (2013) and Oh and Patton (2013) for more recent studies.

Google Trends data

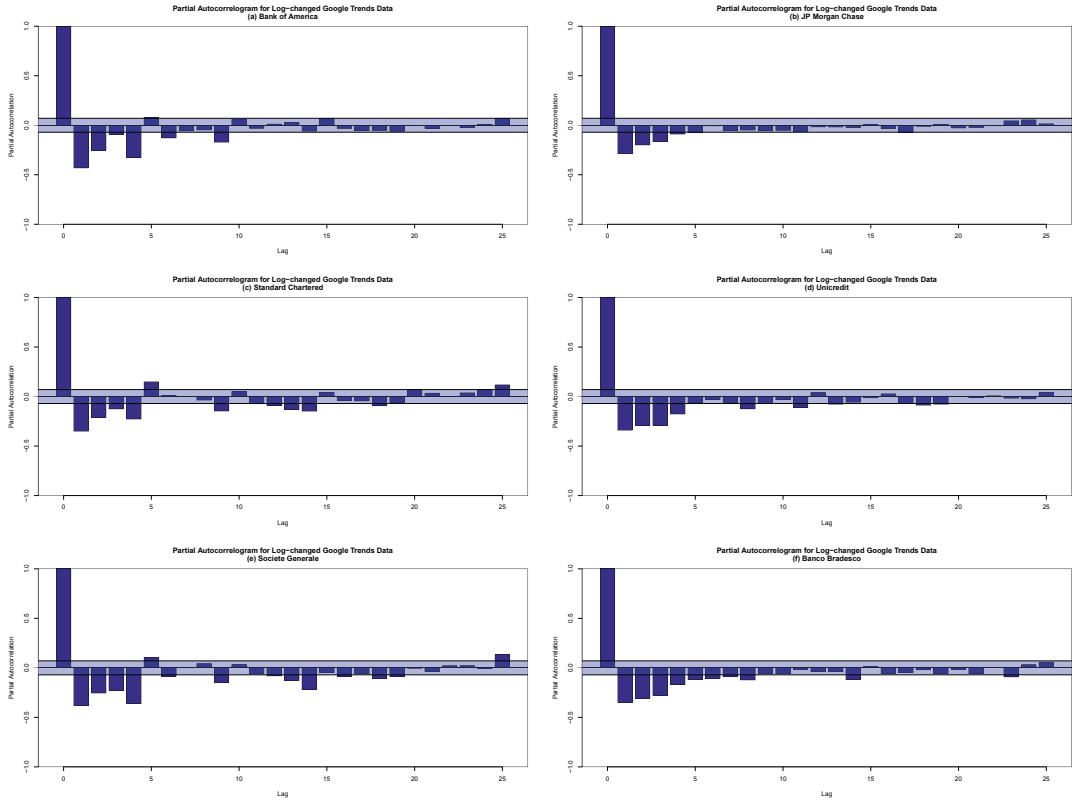
While there are plenty of studies on the characteristics of financial time-series data in the econometrics literature, modeling internet search queries has not yet been sufficiently investigated. At the model identification stage, we aim to detect seasonality, and consequently to adequately determine the order for the seasonal autoregressive and seasonal moving average terms. Therefore, we follow the classical method of model identification as described in Box and Jenkins (1970). In the analysis so far we used the ACF as a commonly-used tool for checking randomness in the observed data set. As pointed out in section 5.2.3 and supported by the Ljung-Box-Test, we find evidence of significant autocorrelation in *Google Trends* data. Remember, each correlogram shows remarkable autocorrelation between lags 1 to 5. Furthermore, Panels (a), (c) and (e) of Figure 5.3 reveal a seasonal pattern with spikes recurring always at the begin of a week due to the homogeneous occurrence of significant autocorrelation coefficients in the respective time periods, i. e. 1 to 5, 6 to 10, 11 to 15, etc. In contrast, Panels (b), (d) and (f) show significant negative autocorrelations only up to lag 2. Altogether, daily log-changes of *Google Trends* data reveal more autocorrelation than is commonly found for daily stock returns (e. g., the average first-order autocorrelation is -0.3668). Thus, the model for the conditional mean of our *Google Trends* data needs a more specific order structure than the commonly-used constant model for daily stock returns.

To assess the correct specification of our econometric models, we next analyze whether the higher-order lags are merely due to the propagation of the autocorrelation at lag 1. Therefore, we present partial autocorrelation functions (PACFs) of log-changed *Google Trends* search queries in Figure 5.4.[164] The partial autocorrelation coefficients are checked statistically at a significance level of 5%.

[164] Specifically, partial autocorrelations are useful in identifying the order of an autoregressive model.

Figure 5.4: **Partial Auto Correlation Functions (PACF) for log-changed *Google Trends* data.**

The figure shows partial auto-correlation plots of daily log-changed *Google Trends* data of the corresponding six banks used in Figure 5.3. Panel (a) shows the PACF for the log-changed *Google Trends* search volume of Bank of America, Panel (b) for JP Morgan Chase, Panel (c) for Standard Chartered, Panel (d) for Unicredit, Panel (e) for Société Générale, and Panel (f) shows the PACF for the log-changed *Google Trends* search volume of Banco Bradesco. The shaded regions denote the 95% confidence interval around zero.



The partial correlograms of Bank of America, Standard Chartered and Société Générale depicted on the left side of Figure 5.4 show the same pattern of statistical significance up to lag 5. The shaping of the following lags reveal only minor deviations. In Panel (a) the PACF displays only two further significant lags (6 and 9), but afterwards it remains well inside the 95% confidence interval. In contrast, Panels (c) and (e) show statistically significant coefficients even in all other 5 day time intervals. The significant coefficients between lags 6 and 25, however, are markedly less pronounced and thus do not seem to be particularly important. From the Panels on the right side we perceive similar PACFs. The PACF in the upper right Panel (b) only shows statistical significant spikes between the lags 1 to 5. Finally, in Panels (d) and

(f) partial autocorrelation is clearly present between lags 1 to 5, whereby at lag 5 in Panel (d) the coefficient of Unicredit corresponds exactly to the level of significance. Moreover, further individual significant coefficients reveal an only mild effect of the corresponding lags of the log-changed *Google Trends* data on itself. Altogether, it can be stated that in each PACF only a limited number of significant coefficients between lags 6 and 25 are given, which are considerably less pronounced and thus not being essential for our econometric framework. We thus conclude that an autoregressive model of order 5 (AR(5)) seems to capture exceptionally well the shape of the autocorrelation structure inherent in the log-changed *Google Trends* data. Conversely, the sample ACF is a useful tool for identifying the order of an moving average (MA) process. As discussed above, either the autocorrelation plots show a recurrent pattern and thus indicate a MA(5) model or only the first two lags reveal significant coefficients (and all others lie within the 95% confidence bounds) and thus suggests a MA model of order two for the data.

More formally, the two relevant ARMA specifications for log-changed *Google Trends* data $LC_i = \{LC_{i,t}\}_{t=1}^{T}$ $(i = 1, \ldots, 29)$ can be expressed via

$$LC_{i,t} = \mu_i + e_{i,t} + \sum_{j=1}^{5} \Phi_{j,i} LC_{i,t-j} + \sum_{k=1}^{s} \Psi_{k,i} e_{i,t-k}, \tag{5.4}$$

where $s \in \{2, 5\}$ denotes the number of moving-average terms. The estimation procedure is conducted via conditional least squares.

To pick up second-moment dependence inherent in the log-changed *Google Trends* time series we have to address the time-varying nature of volatility. Vozlyublennaia (2014) demonstrates that changes in returns significantly influence the level of investor attention and, more precisely, this impact is both short-term as well as long-term in nature. This finding is consistent with the theory of Barber et al. (2009), where investors create short-term price pressure, leading to higher volatility, whereas, from the long-term perspective a trend reversal is expected. Therefore, we employ the CS-GARCH model as proposed by Lee and Engle (1999) to decompose the conditional variance

into a transitory and permanent component regarding the short- and long-run movements of volatility. To additionally account for skewness and fat tails in the marginal distributions we use the method of Fernandez and Steel (1998) yielding skewed versions of the student t distribution and the generalized error distribution, respectively. According to this, we fit a CS-GARCH(1,1) model to the ARMA residuals, $e_{i,t}$. With $q_{i,t}$ denoting the permanent component of the conditional variance, the model can then be written as

$$e_{i,t} = \sigma_{i,t}\varepsilon_{i,t}, \qquad \varepsilon_{i,t}|\mathcal{F}_{i,t-1} \sim iid\ skt(\nu_i, \gamma_i) \vee sged(\kappa_i, \xi_i) \tag{5.5}$$

$$\sigma^2_{i,t} = q_{i,t} + \beta(\sigma^2_{i,t-1} - q_{i,t-1}) + \alpha(e^2_{i,t-1} - q_{i,t-1}) \tag{5.6}$$

$$q_{i,t} = \omega + \eta_1 q_{i,t-1} + \eta_2(e^2_{i,t-1} - \sigma^2_{i,t-1}), \tag{5.7}$$

where the difference between the conditional variance and its trend, $\sigma^2_{i,t} - q_{i,t}$ corresponds to the transitory component of the conditional variance. Analogously to the skewed t distribution we follow Fernandez and Steel (1998) to implement the skewed generalized error distribution $sged(\kappa, \xi)$ with shape parameter $\kappa \in (0, \infty)$ and skewness parameter $\xi \in (0, \infty)$. Letting f_{ged} represent the pdf of a univariate generalized error distribution, the pdf of $sged(\kappa, \xi)$, f_{sged}, is given by

$$f_{sged}(\varepsilon; \kappa_i, \xi_i) = \frac{2}{\xi_i + \frac{1}{\xi_i}} \left[f_{ged}\left(\frac{\varepsilon}{\xi_i}\right) \mathbb{1}_{[0,\infty)}(\varepsilon) + f_{ged}(\xi_i \varepsilon) \mathbb{1}_{(-\infty,0)}(\varepsilon) \right]. \tag{5.8}$$

Note that after applying ARMA-CS-GARCH filtering techniques we check the independence of the realized residuals, $e_{i,t}$, via the Ljung-Box-Test. The results are discussed subsequently in section 5.4.1. Also note that the distribution of the innovations crucially depends on the individual search queries. More precisely, for each series we test the model (5.6) with gaussian-, skewed-t- and skewed-ged innovations. However, the corresponding QQ-Plots provide clear evidence of fat tails indicating the inadequacy of the gaussian distribution. According to this, from both the skewed-t- and skewed-ged- distribution we select the distribution that fits each residual series best.

5.3.2 Dependence Modeling with R-vine Copulas

We now turn to the modeling of the joint distribution of stock returns and log-changed *Google Trends* data, respectively. As copulas allow splitting the analysis of the marginal distributions and the dependence structure, they are a reliable tool for capturing the dependence patterns inherent in our high-dimensional data sets. More precisely, we employ R-vine copula models to identify both potential asymmetric and non-linear dependence characteristics in an extremely flexible way. According to this, we shortly introduce the fundamentals of vine copulas and the specification of R-vines. The application of the model is presented subsequently.

Vine copulas

Mathematically speaking, a d-variate copula is a function on the unit cube $[0,1]^d$ which joins a multivariate distribution function to its one-dimensional marginals (see, e. g., Joe, 1997, Nelsen, 2006). According to the theorem of Sklar (1959), if $\mathbf{X} = (X_1, \ldots, X_d)$ denotes a vector of random variables with a joint density function $\mathbf{f} = (f_1, \ldots, f_d)$ it states that every multivariate distribution $\mathbf{F} = (F_1, \ldots, F_d)$ can be represented by

$$\mathbf{C}(u_1, \ldots, u_d) = \mathbf{F}(F_1^{-1}(x_1), \ldots, F_d^{-1}(x_d)) \tag{5.9}$$

for some appropriate d-dimensional copula $\mathbf{C}$. In this context, F_i^{-1} describes the generelized inverse of F_i and $u_i \in [0,1],\ i = 1, \ldots, d$. The corresponding joint multivariate density $\mathbf{f}$ can be expressed as follows

$$\mathbf{f}(x_1, \ldots, x_d) = \mathbf{c}(F_1(x_1), \ldots, F_d(x_d)) \prod_{i=1}^{d} f_i(x_i), \tag{5.10}$$

where the d-variate copula density $\mathbf{c}$ is calculated via $\frac{\partial C(u_1,\ldots,u_d)}{\partial u_1 \ldots \partial u_d}$. From equation (5.9) the key idea of copulas becomes apparent that the joint distribution includes implicitly a description of the marginals as well as their dependence structure.

Vine copulas offer an extreme flexible framework for high-dimensional dependence modeling. The fundamental idea is to decompose a multi-dimensional copula into a cascade of bivariate ones, so-called pair-copulas, as all computations necessary in statistical inference are more tractable on bivariate data sets.[165] As a result, by using these so called pair-copula constructions (PCC in short) we are able to factorize the multivariate density in (5.10) into a product of (unconditional and conditional) bivariate copulas and marginal densities as described below.

Starting point of our discussion on PCCs is the decomposition of a joint probability density function of dimension d into its unconditional and conditional marginal densities

$$f(\mathbf{x}) = f(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_1, x_2) \cdot \ldots \cdot f(x_d|x_1, \ldots, x_{d-1}). \tag{5.11}$$

As shown in Aas et al. (2009) each factor on the right side of (5.11) can then be expressed as a product of an appropriate pair-copula and a conditional marginal density via

$$f(x|\mathbf{v}) = c_{xv_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})) \cdot f(x|\mathbf{v}_{-j}), \tag{5.12}$$

where $\mathbf{v}$ denotes a d-dimensional vector and v_j is an arbitrarily chosen component of $\mathbf{v}$; thus $\mathbf{v}_{-j}$ characterizes the v-vector, excluding this component. Expressing all conditional densities in (5.11) by means of (5.12) we finally obtain a factorization of the d-dimensional copula density $\mathbf{f}$ that only consists pair-copulas and univariate marginals. Note, the way to factorize the density is very general and thus there exists a large number of possible pair-copula decompositions. We now illustrate the hierarchical construction of vine copulas for low dimensional cases. Regarding the base case of two dimensions the density function $\mathbf{f}(x_1, x_2)$ can easily expressed by

$$\mathbf{f}(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2). \tag{5.13}$$

[165] See the pioneering works of Bedford and Cooke (2001, 2002), also based on Joe (1996), for a probabilistic construction of multivariate distributions based on simple building blocks (pair-copulas).

From (5.13) we immediately obtain the conditional density of X_2 given X_1 by

$$f_{2|1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2). \tag{5.14}$$

For dimension $d = 3$, the decomposition of the joint probability density function (5.11) is explicitly given via

$$\mathbf{f}(x_1, x_2, x_3) = f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1,2}(x_3|x_1, x_2). \tag{5.15}$$

As can be seen from (5.14) the conditional density $f_{2|1}$ has already been expressed by terms of a pair-copula and a marginal distribution. On the other hand, choosing one of the conditioning variables for the third term on the right hand side of (5.15), e. g. x_1, we get via (5.12) the representation

$$f_{3|1,2}(x_3|x_1, x_2) = c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot f_{3|1}(x_3|x_1). \tag{5.16}$$

The decomposition consists of a pair-copula and, however, a conditional density. The latter one can be further factorized in an iterative manner yielding the full decomposition

$$\begin{aligned} \mathbf{f}(x_1, x_2, x_3) &= f_1(x_1) \\ &\cdot\ c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \\ &\cdot\ c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_3(x_3). \end{aligned} \tag{5.17}$$

As mentioned before, the representation (5.17) is particularly depending on the variable to condition on and thus is only one of several possible ways of decomposing $\mathbf{f}(x_1, x_2, x_3)$.

Based on the previous results it becomes apparent that pair-copula constructions are an extremely flexible but still tractable method of constructing multivariate distributions. After the seminal works of Joe (1996, 1997) and Bedford and Cooke (2001,

2002), the influential study of Aas et al. (2009) introduced pair-copula constructions in a risk management setting with applications to financial data. Since then, vine copulas have emerged as the most promising tool for modeling high-dimensional dependencies.[166]

The use of PCCs for our modeling strategy is beneficial for several reasons. On the one hand, one objective of our modeling framework is to capture the complex dependence structures of the considered banks network. Thus, to deal with these challenging structures and keeping the model still tractable, PCCs provide a proper handling by decomposing the high-dimensional distribution into bivariate building blocks. In the recent literature, a wide range of techniques for fitting these blocks have been provided and therefore, we are able to adequately cope the increased model risk of the vine model specifications.[167] Moreover, PCCs offer a clear graphical structure for describing multivariate distributions.

In line with the increasing popularity of vine copulas several types of PCCs have been emerged. In particular, canonical vines (C-vines), drawable vines (D-vines) and regular vines (R-vines), respectively, are commonly used vine specifications. In our modeling framework we focus on the more general class of regular vines, since it is less restrictive regarding the modeling of the dependence structure. We begin with the theoretical background of R-vines and discuss the fitting procedure subsequently.

R-vines

According to the pair-copula decomposition principle, there exist many different potential PCCs for a given multivariate distribution. In this context, the so-called (regular) vine methodology is one way to determine how the marginals of a particular PCC are

[166] For other important contributions related to different applications of vine copulas in asset pricing and risk management see Chollete et al. (2009), Heinen and Valdesogo (2009), Aas and Berg (2009), Min and Czado (2010, 2011), Weiß and Supper (2013) and Weiß and Scheffer (2015).

[167] While Aas et al. (2009) select the parametric pair-copulas based on graphical data inspection and goodness-of-fit tests, e. g. Brechmann et al. (2012) and Dissmann et al. (2013) employ sequential heuristics based on Akaike's Information Criterion (AIC). Another approach is given by Weiß and Scheffer (2015), who model the pair-copulas via convex combinations (so-called mixture pair-copulas) based on the EM algorithm. In contrast, Hobæk-Haff (2013) propose empirical pair-copulas as building blocks to circumvent the problem of selecting parametric pair-copulas.

coupled. This methodology dates back to Joe (1996) and was organized systematically by Bedford and Cooke (2001, 2002), who investigate a graphical representation that facilitates the description of the (un-)conditional specifications made for the joint distribution by a sequence of trees. To put it an a nutshell, a nested set of trees is called a regular vine and corresponds to a unique multivariate distribution, the so-called R-vine distribution.

We now proceed with a formal description of R-vines closely related to the work of Bedford and Cooke (2002). Let $T = \{N, E\}$ define a tree with nodes N and edges E, respectively, subject to the condition that E is a subset of unordered pairs of N without cycles. Further, for any $l, n \in N$ there has to exist a sequence $m_1, \ldots, m_k$ of elements of N such that $\{l, m_1\} \in E, \{m_1, m_2\} \in E, \ldots, \{m_k, n\}$. A set of linked trees $\mathcal{V} = \{T_1, \ldots, T_{d-1}\}$ is being equivalent to an R-vine on d elements if the following conditions are met:

(i) $T_1 = \{N_1, E_1\}$ is a tree with nodes $N_1 = \{1, \ldots, d\}$ and a set of edges E_1.

(ii) For $i = 2, \ldots, d$, $T_i = \{N_i, E_i\}$ is a connected tree with edge set E_i and node set $N_i = E_{i-1}$, with $\#N_i = d - (i - 1)$ and $\#N_i$ denoting the cardinality of N_i.

(iii) For $i = 2, \ldots, d - 1$ and two nodes $a = \{a_1, a_2\}, b = \{b_1, b_2\} \in N_i$ connected by an edge $e \in E_i$ must satisfy $\#a \cap b = 1$.

The proximity condition (property (iii)) implicitly ensures that two nodes are joined by an edge in tree T_{i+1} ($i = 1, \ldots, d-1$), only if the corresponding edges in tree T_i are adjacent, i. e. share a common node. Following Czado (2010), we denote the edges in tree T_i by $jk|D$ where $j < k$ and D identifies the conditioning set.[168] Then, two egdes with a common node are represented by $a = j(a), k(a)|D(a)$ and $b = j(b), k(b)|D(b)$ with

[168] Note that the order of the conditioned set $\{j, k\}$ is made to yield a uniquely determined order of the arguments of the bivariate copulas. Further note that $D = \varnothing$ is directly linked to the edge denoted by jk.

$U(a) := \{j(a), k(a), D(a)\}$ and $U(b) := \{j(b), k(b), D(b)\}$, respectively. The connection for these nodes a and b is given by edge $e = j(e), k(e)|D(e)$, where

$$\begin{aligned} j(e) &:= \min\{i : i \in (V(a) \cup V(b)) \setminus D(e)\}, \\ k(e) &:= \max\{i : i \in (V(a) \cup V(b)) \setminus D(e)\}, \\ D(e) &:= V(a) \cap V(b). \end{aligned}$$

To build up a statistical model on R-vine $\mathcal{V}$, each edge $e = j(e), k(e)|D(e) \in E_i$ in a vine $\mathcal{V}$ is associated with a bivariate copula density $c_{j(e),k(e)|D(e)}$. In this context, the nodes $j(e)$ and $k(e)$ are called the conditioned nodes, while $D(e)$ is the conditioning set. We then define an R-vine distribution as the distribution of the random vector $\mathbf{X}$ with marginal densities $f_1, \ldots, f_d$ and conditional densities $c_{j(e),k(e)|D(e)}$. As shown in Bedford and Cooke (2001), the joint density of X is uniquely determined and given via

$$\mathbf{f}(\mathbf{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e)|D(e)}(F(x_{j(e)}|\mathbf{x}_{D(e)}), (F(x_{k(e)}|\mathbf{x}_{D(e)})),$$

where $\mathbf{x}_{D(e)}$ denotes the subvector of $\mathbf{x}$ indicated by the indices contained in $D(e)$.

Estimation procedure for R-vine copulas

We now focus on fitting an R-vine copula model to a given dataset proceeding three separate tasks. First, one has to address the tree structure of the R-vine. Once a fixed structure is given, secondly, one needs to select $d(d-1)/2$ bivariate pair-copulas from candidate copula families. The third and final step is then to estimate the corresponding parameters of the chosen pair-copulas.

Given the sheer number of $d!/2 \cdot 2^{\binom{d-2}{2}}$ possible R-vines (see, e. g., Morales-Napoles et al., 2010) we need an estimation procedure that keeps the specification of the PCC tractable. To this end, we follow the idea of Dissmann et al. (2013), who propose a sequential model selection for vine copulas. More precisely, the stepwise method proceeds top-down and tree-by-tree in the vine. Initially, the structure of the

first tree $T_1 = (N_1, E_1)$ must be specified. Therefore, we identify the pairs of variables, $\{j, k\}$, $j, k = 1, \ldots, d$, which capture the strongest dependence based on pairwise Kendall's τ.[169] Then, we use the absolute values of Kendall's τ as edge weights to yield the tree, which maximizes the sum of edge weights among all possible trees. Recall that each edge of the resulting (first) tree corresponds to an unconditional pair-copula. In the following step, the selection of the appropriate parametric families and the estimation of the corresponding parameters are performed subsequently via Akaike Information Criterion (AIC).[170] By this means, for each of the $d - 1$ edges the best fitting candidate parametric copula is chosen.[171] Furthermore, we compute the transformed data (i. e. conditional distributions) for the second tree by using the fitted copulas in the first tree. In the sequel, the same procedure is applied to the transformed data and iterated until we yield a fully specified PCC.[172]

In order to obtain uniformly distributed data on which the R-vine copula modeling is based on, the GARCH processes and the corresponding distributions of innovations discussed above need to be specified correctly. Assuming this, the univariate marginal distributions are used to transform the GARCH residuals, $\varepsilon_{i,t}$, into pseudo-observations u_i by computing the ranks of the residuals $F_i(\varepsilon_i)$.

5.4 Empirical Application

In our empirical application, we implement our econometric modeling framework. Firstly, we apply the univariate models to both the stock returns as well as log-changed *Google Trends* data. Thereafter, the dependence structures of the filtered time series

[169] Note that Kendall's τ measures dependence independently of the assumed distribution and thus is a useful tool for combining different copula families. See, e. g., Mendes et al. (2010) and Czado et al. (2013) for alternative measures of dependence such as tail dependence and goodness of fit tests, respectively.

[170] See, e. g., Manner (2007) who investigates an AIC-based selection of bivariate copulas and finds that it is well-suited for identifying the correct copula family, while Brechmann and Czado (2013) and Dissmann et al. (2013) show in simulation studies that even PCCs fitted via AIC perform exceptionally well.

[171] That is, the copula family as well as its parameter(s) with the minimum AIC.

[172] Due to the sequential estimation procedure this approach does not guarantee to find a global optimum in terms of the model selection criterion, e. g., higher values of the likelihood function or lower values of the AIC/BIC.

are modeled by fitting R-vine copulas. We find striking evidence of commonality in the dependence structures of stock returns and the search queries for their names, respectively. Furthermore, we document the existence of significant tail dependence in search query pairs and between stock returns and the respective search queries.

5.4.1 Univariate Analysis

In this subsection, the results of the estimated marginal models are presented. Particularly, we study time series of daily log-return data for each of the 29 considered banks and the corresponding log-changed *Google Trends* data, respectively, for the period from January 2011 to December 2013.

Initially, we briefly summarize the main points given in section 5.3.1. The most prevalent stylized facts of stock returns are that they are characterized by significant autocorrelation and heavy tails, and hence being non-normally distributed. While Christoffersen et al. (2012) propose an AR model of order two to pick return dependence, we also find the third lag to be moderate significantly for most stock returns and thus include three lags in our AR specification. To account for asymmetry in volatility and as well as skewness and fat tails we employ the GJR-GARCH(1,1) filter as introduced by Glosten et al. (1993) and assume the resulting iid residuals to be skewed Student-t distributed following Fernandez and Steel (1998). Table 5.3 reports the cross-sectional distribution of the univariate AR(3)-GJR-GARCH(1,1)-Skew $t(\nu;\psi)$ models based on daily stock returns.

The estimated parameters shown in Table 5.3 indicate no distinctive abnormalities. All three AR lags are statistically significant, justifying the modeling approach to capture first-moment dependence. Moreover, further checks do not suggest the need for a higher order in the AR specification. The estimates of the conditional variance models highlight some important features of the underlying processes. First, the average value of parameter α is around 0.014, indicating that the lagged return shocks only have a limited impact on the current volatility. Of particular note is the cross-sectional

Table 5.3: **Cross-sectional distribution of parameter estimates.**

The table presents summary statistics of the estimated AR(3)-GJR-GARCH(1,1)-Skew $t(\nu;\psi)$ models estimated on log-difference of daily returns. For each of the 29 banks in the sample (see Appendix C.1) the models are estimated for the period from January 2011 to December 2013. The columns present descriptive statistics from the cross-sectional distribution of the parameters listed in the rows. The bottom panel shows the number of rejections (at the 0.05 level) across 29 banks from Ljung-Box tests for serial correlation up to 10 lags. The first row is for standardized residuals of log-difference of daily returns and the second row for squared standardized residuals. The bottom panel shows the number of rejections across 29 banks from the Kolmogorov-Smirnov test of the Skew $t(\nu;\psi)$ distribution used for the standardized residuals.

	Cross-sectional distribution										
		Percentiles						Moments			
	Min	5%	25%	Median	75%	95%	Max	Mean	St. Dev.	Skewness	Kurtosis
μ	-0.0007	-0.0006	-0.0002	0.0002	0.0004	0.0007	0.0009	0.0002	0.0004	-0.3549	-0.8333
Φ_1	-0.0762	-0.0661	-0.0286	0.0107	0.0371	0.0846	0.0921	0.0055	0.0482	0.0401	-1.0775
Φ_2	-0.0824	-0.0648	-0.0303	-0.0153	0.0032	0.0344	0.0444	-0.0139	0.0309	-0.1048	-0.4433
Φ_3	-0.1337	-0.1049	-0.0636	-0.0392	-0.0099	0.0153	0.0213	-0.0412	0.0393	-0.4346	-0.5805
ω	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	3.0838	10.0419
α	0.0000	0.0000	0.0000	0.0036	0.0176	0.0643	0.0838	0.0136	0.0218	2.0701	3.6026
β	0.0498	0.8860	0.9415	0.9488	0.9563	0.9665	0.9705	0.9143	0.1676	-4.7241	21.4322
γ	-0.0029	0.0139	0.0465	0.0615	0.0772	0.0914	0.1054	0.0609	0.0244	-0.7556	0.5625
ν	0.8528	0.8728	0.9249	0.9698	1.0006	1.0669	1.0752	0.9657	0.0590	0.0368	-0.7730
ψ	2.0104	5.1860	5.7960	6.8370	9.9260	12.4960	14.0023	7.8460	2.6995	0.3622	-0.3304
										# of rejections	
LB test for standardized residuals										0	
LB test for squared standardized residuals										4	
KS test for skew t dist of std. residuals										2	

average estimation of the autoregressive parameter β with around 0.914. Not surprisingly, however, the leverage coefficient δ is statistically significant and positive (0.061 on average). According to this, the volatility is asymmetric, i. e. negative shocks have greater impact on volatility then positive shocks of equal magnitude. The persistence level of volatility (0.958 on average) is less than unity but very close to one.[173]

Next, we discuss the accuracy of the marginal models of stock returns. If the distribution specifications successfully capture the serial correlation in the conditional mean and the conditional volatility, no autocorrelation should be left in the residual series. To validate this, we apply weighted versions of the Ljung-Box test on standardized residuals and squared standardized residuals, respectively. As can be seen from the bottom panel of Table 5.3 Ljung-Box (LB) tests at up to the tenth lag reveal no significant autocorrelation (at the 0.05 level) of the residual series and reject the null of zero autocorrelation for only four of the squared residual series. Therefore, we conclude that the proposed models work quite accurately.

[173] Based on empirical results a common finding in the econometrics literature is that many financial time series exhibit a high degree of persistence (see, e. g., Engle and Bollerslev, 1986, Bollerslev and Engle, 1993, Engle and Patton, 2001).

In a final step, we apply the Kolmogorov-Smirnov (KS) test to check the fit of the skewed t distributions for the standardized residuals. To decide if the samples come from the hypothesized distributions the critical values for the KS test are determined by simulations. Since the null of correct specification is rejected for only two out of the 29 time series, we conclude that the skewed t distribution is well-suited for the underlying data.

We now turn our attention to the estimations of the univariate modeling approach of log-changed *Google Trends* data. As pointed out in section 5.2.3, we find evidence of significant autocorrelation of order five and two distinctive moving average lags, two and five, respectively. For this purpose our model for the conditional mean needs more structure than commonly-used models for daily financial time-series. We recommend using ARMA(5,2) and ARMA(5,5) models, respectively. To account for short- and long-term effects in search queries we use the CS-GARCH(1,1) model of Lee and Engle (1999), which decomposes the conditional variance into a permanent and transitory component. As becomes apparent from Table 5.2 the log-changed *Google Trends* time series also reveal skewness and fat tails. In order to capture these non-normal characteristics, we use the method of Fernandez and Steel (1998) to obtain skewed versions of the skewed t distribution and the skewed *ged* distribution, respectively.

Table 5.4 reports the estimates for the marginal and dependence parameters.

The order of the specific ARMA models are supported by the fact that nearly all estimated coefficients presented in Table 5.4 are statistically significant. For the conditional variance models several findings are noteworthy. First, the permanent component shock term η_2 reveal only mild statistical evidence compared with the persistence of the permanent component autoregressive term η_1 (oscillating between 0.963 and 0.999). Further, the estimates of transitory component terms, α and β, provide no uniform pattern. Regarding the parameters of the marginal distributions, the estimates for both distributions, the skewed student t as well as the generalized error *ged*, indicate fat tails and slight skewness in most cases. Finally, we apply the weighted LB(10) test to the (squared) standardized GARCH residuals. For all residual series,

Table 5.4: **Estimates for the marginal and dependence parameters.**

The table presents summary statistics of the estimated ARMA(5,2)-CS-GARCH(1,1)-Skew $t(\nu;\psi)/ged(\kappa,\xi)$ and ARMA(5,5)-CS-GARCH(1,1)-Skew $t(\nu;\psi)/ged(\kappa,\xi)$ models, respectively, estimated on log-changed *Google Trends* data. For each of the first 29 banks in the sample (see Appendix C.1) the models are estimated for the period from January 2011 to December 2013. The columns present descriptive statistics from the cross-sectional distribution of the parameters listed in the rows. The bottom panel shows the number of rejections (at the 0.05 level) across 29 banks from Ljung-Box tests for serial correlation up to 10 lags. The first row is for standardized residuals of log-changed *Google Trends* data and the second row for squared standardized residuals. The bottom panel shows the number of rejections across 29 banks from the Kolmogorov-Smirnov test of the Skew $t(\nu;\psi)$ and Skew $ged(\kappa,\xi)$ distribution, respectively, used for the standardized residuals.

	Panel A: Parameters estimates of the ARMA models										
	μ	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5	Θ_1	Θ_2	Θ_3	Θ_4	Θ_5
Banco Bradesco	-0.0006	-0.4519	0.3006	0.0920	0.0629	0.0440	-0.1939	-0.7063	-	-	-
Banco Santander	0.0000	-0.2953	-0.3343	-0.3420	-0.2206	0.6132	-0.0084	0.0095	-0.0047	-0.1105	-0.8546
Bank of America	-0.0028	-0.5341	-0.5494	-0.5573	-0.4707	0.4290	0.1364	0.1964	0.2175	-0.0050	-0.6627
Bank of China	0.0003	0.4105	0.7894	-0.0774	-0.8888	0.2601	-1.0382	-0.6594	0.5841	0.9479	-0.7400
Bank of Montreal	-0.0011	-1.8206	-2.2074	-1.4806	-0.6085	0.1548	1.0727	0.7446	-0.3509	-0.7829	-0.8090
Bank of New York Mellon	0.0001	-0.2083	-0.4543	-0.6005	-0.3681	0.1760	-0.5608	0.2225	0.2694	-0.2181	-0.4042
Bank of Nova Scotia	0.0000	0.9488	-0.0091	-0.0737	0.0254	-0.0636	-1.8061	0.8184	-	-	-
Barclays	0.0000	0.4721	-0.0422	0.0047	0.1498	0.0687	-0.7230	-0.1915	-	-	-
BBVA Banco	0.0002	-0.0059	0.3092	0.0730	0.0911	-0.0120	-0.6053	-0.4081	-	-	-
BNP Paribas	-0.0019	0.1078	0.2618	0.0048	0.0936	0.1588	-0.5236	-0.3337	-	-	-
Capital One Financial	0.0001	-0.9013	-0.9045	-0.9159	-0.8665	0.1070	-0.0214	0.0780	0.0533	-0.0675	-0.9229
Charles Schwab	-0.0023	-0.8701	-1.0193	-0.7437	-0.4771	0.2967	0.3066	0.4212	-0.0015	-0.1700	-0.7039
Citigroup	-0.0026	0.6272	-0.0040	-0.0651	0.1191	0.0205	-1.0962	0.1863	-	-	-
Com. Bank of Australia	0.0000	-0.0083	0.1881	0.0707	0.1411	0.1480	-0.6466	-0.2554	-	-	-
Credit Agricole	-0.0000	0.0978	0.6194	0.0503	-0.5897	0.4270	-0.5512	-0.8236	0.1899	0.7957	-0.6272
Deutsche Bank	-0.0025	-0.4109	-0.3918	-0.4119	-0.3474	0.5933	0.0507	0.1032	0.1351	-0.0124	-0.7401
HSBC	0.0003	1.4076	-0.5453	0.0958	0.0106	-0.1108	-1.8303	0.8624	0.5832	-0.4666	-0.5886
Intesa Sanpaolo	0.0000	0.0965	0.1364	0.0990	0.1278	0.1438	-0.8430	-0.1697	-	-	-
Itau Unibanco	-0.0012	0.5369	-0.0874	-0.0022	-0.0059	0.0660	-1.3184	0.3858	-	-	-
JP Morgan Chase	-0.0012	0.5311	-0.0227	-0.0120	0.0251	0.0213	-1.0908	0.1801	-	-	-
Lloyds Banking Group	0.0013	-0.3245	-0.2153	0.2757	0.7722	0.0571	-0.1267	0.0924	-0.4481	-0.7248	0.2564
Morgan Stanley	0.0001	-0.0861	0.1982	0.0439	0.0871	0.1457	-0.4718	-0.4582	-	-	-
National Australia Bank	-0.0001	-0.7192	-0.7310	-0.7060	-0.6346	0.1951	-0.1252	0.0414	0.0054	-0.0422	-0.6495
Nordea Bank	-0.0002	-0.7545	-0.7963	-0.6978	-0.6799	0.1641	-0.1095	0.1358	-0.0916	-0.0022	-0.6687
Royal Bank of Canada	-0.0006	-1.9229	-2.2922	-1.4321	-0.4526	0.2254	1.2456	0.8790	-0.3758	-0.9119	-0.8280
Société Générale	-0.0001	-0.2707	-0.2887	-0.3045	-0.2130	0.6659	-0.0960	0.1000	0.0576	-0.0937	-0.7161
Standard Chartered	-0.0005	-0.3336	-0.3919	-0.3572	-0.2891	0.5759	-0.0782	0.1623	0.0852	-0.0362	-0.6254
Unicredit	-0.0005	0.4701	0.0875	0.0935	0.0646	0.1721	-0.9860	-0.0253	-	-	-
Wells Fargo	0.0011	-0.5943	-0.6022	-0.6231	-0.5686	0.3661	0.0380	0.1112	0.1603	-0.0566	-0.6629

Table 5.4: **Cross-sectional distribution of parameter estimates (continued).**

	Panel B: Parameters estimates of the CS-GARCH and marginal models							
	ω	α	β	η_1	η_2	*distr*	$\nu\|\kappa$	$\psi\|\xi$
Banco Bradesco	0.0004	0.5998	0.2130	0.9965	0.0000	*sstd*	0.7540	2.6043
Banco Santander	0.0000	0.5277	0.0000	0.9995	0.0000	*sstd*	0.8838	3.7210
Bank of America	0.0002	0.7616	0.0613	0.9968	0.0000	*sstd*	0.8952	2.5180
Bank of China	0.0004	0.0606	0.8725	0.9808	0.0000	*sstd*	0.9484	5.3514
Bank of Montreal	0.0001	0.2068	0.4450	0.9973	0.0000	*sstd*	0.7532	3.2622
Bank of New York Mellon	0.0006	0.0538	0.8149	0.9905	0.0000	*sged*	0.8345	2.0003
Bank of Nova Scotia	0.0002	0.1182	0.2619	0.9960	0.0000	*sstd*	0.7218	5.1261
Barclays	0.0000	0.3572	0.2375	0.9958	0.0000	*sstd*	1.1098	3.4835
BBVA Banco	0.0001	0.2741	0.1318	0.9963	0.0000	*sstd*	0.8487	5.5838
BNP Paribas	0.0001	0.4837	0.0574	0.9943	0.0000	*sstd*	0.9148	3.4381
Capital One Financial	0.0012	0.0511	0.0019	0.9799	0.0019	*sged*	0.8101	2.2941
Charles Schwab	0.0002	0.4546	0.2271	0.9971	0.0000	*sstd*	0.7882	2.8257
Citigroup	0.0002	0.3209	0.1755	0.9969	0.0000	*sstd*	1.0148	3.2788
Com. Bank of Australia	0.0003	0.1213	0.4390	0.9943	0.0000	*sged*	1.0868	1.9353
Credit Agricole	0.0000	0.4702	0.0253	0.9972	0.0011	*sstd*	0.8780	3.1315
Deutsche Bank	0.0011	0.3800	0.3422	0.9687	0.1355	*sstd*	0.9936	3.3422
HSBC	0.0000	0.2253	0.0001	0.9934	0.0000	*sstd*	1.2218	3.3586
Intesa Sanpaolo	0.0001	0.1408	0.4611	0.9965	0.0007	*sstd*	0.7734	4.6094
Itau Unibanco	0.0002	0.0227	0.8017	0.9965	0.0000	*sged*	0.9551	2.6818
JP Morgan Chase	0.0016	0.3295	0.1603	0.9633	0.0204	*sstd*	0.8605	3.5131
Lloyds Banking Group	0.0000	0.1148	0.5923	0.9999	0.0038	*sstd*	0.9142	9.5759
Morgan Stanley	0.0000	0.3853	0.0001	0.9982	0.0000	*sstd*	0.9188	3.4493
National Australia Bank	0.0001	0.1147	0.0157	0.9732	0.0157	*sstd*	0.8135	7.2093
Nordea Bank	0.0003	0.0885	0.6293	0.9915	0.0069	*sstd*	0.7836	7.0671
Royal Bank of Canada	0.0002	0.1463	0.4252	0.9926	0.0000	*sstd*	0.8321	7.1366
Société Générale	0.0001	0.6142	0.0000	0.9967	0.0000	*sstd*	0.8162	3.2018
Standard Chartered	0.0001	0.2978	1.8190	0.9945	0.0000	*sstd*	1.0235	5.5138
Unicredit	0.0004	0.8938	0.0253	0.9988	0.0086	*sstd*	1.1138	2.3473
Wells Fargo	0.0006	0.2784	0.4352	0.9808	0.0713	*sstd*	0.8881	2.4831

	# of rejections
LB test for standardized residuals	0
LB test for squared standardized residuals	1
KS test for skewed *t* / skewed *ged* distribution of standardized residuals	0

the null hypothesis of zero autocorrelation cannot be rejected at the 5% level, and the squared standardized residual series reveal significant autocorrelation in only one case, which indicate that serial correlations have been removed. Thus, we conclude our proposed ARMA-CS-GARCH models impressively picked up the first-moment and second-moment dependence, respectively. Finally, we use the KS-test to investigate the fit of the skewed t distribution and the skewed *ged* distribution, respectively, for the standardized residuals, using simulations to obtain critical values that capture the parameter estimation error. The KS-test rejects the null of correct specification in none of the cases.

5.4.2 Multivariate Analysis and Model Comparisons

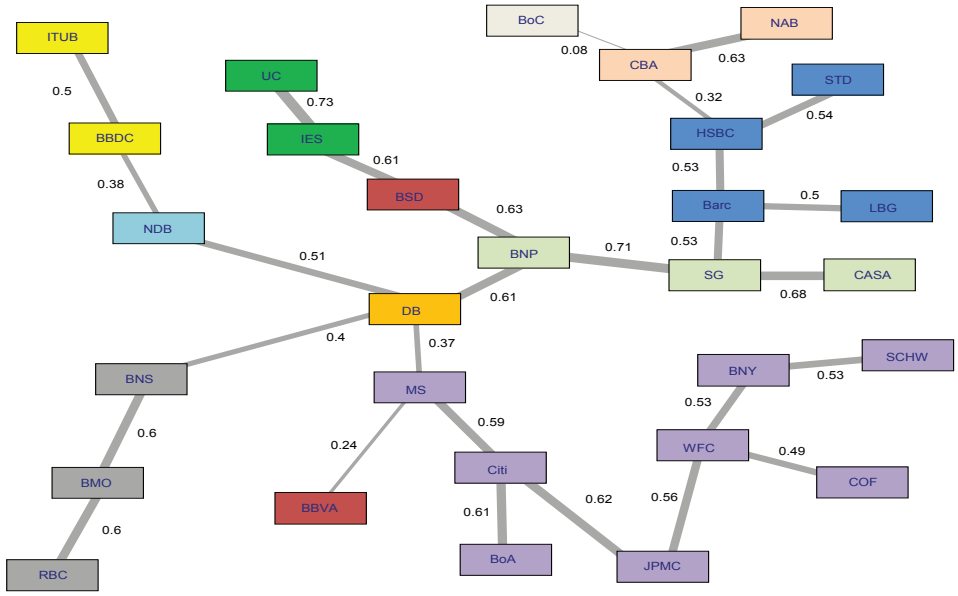
Having selected appropriate ARMA-GARCH models for the univariate margins, we now focus on the analysis of the dependence structure in the underlying data sets. Initially, the resulting standardized residual vectors are converted to uniform pseudo-observations using their empirical distribution function, i. e. $u_i = F_i(\varepsilon_i)$. Then we fit a full R-vine copula to both data samples, the stock returns as well as the log-changed *Google Trends* data, by using the modeling approach presented in section 5.3.2 We employ Kendall's τ to compute the edge weights and choose the pair-copulas from a prespecified range of parametric families by applying AIC. As candidate parametric copula families, we use the Gaussian, Student's t, Clayton, Gumbel, Frank, Joe, Survival Clayton, Survival Gumbel, the rotated Clayton copula (90 and 270 degrees) and the rotated Gumbel copula (90 and 270 degrees).

Figure 5.5 illustrates the first trees of the fitted R-vine copula based on the standardized residuals of stock returns and log-changed *Google Trends* data, respectively. In both Panels, banks located in the same country are highlighted in a uniform color. For increased readability, the correspondence between the acronyms and the bank names are listed in Appendix C.1. Note that our way of vine modeling results in capturing the strongest dependencies in the first tree.

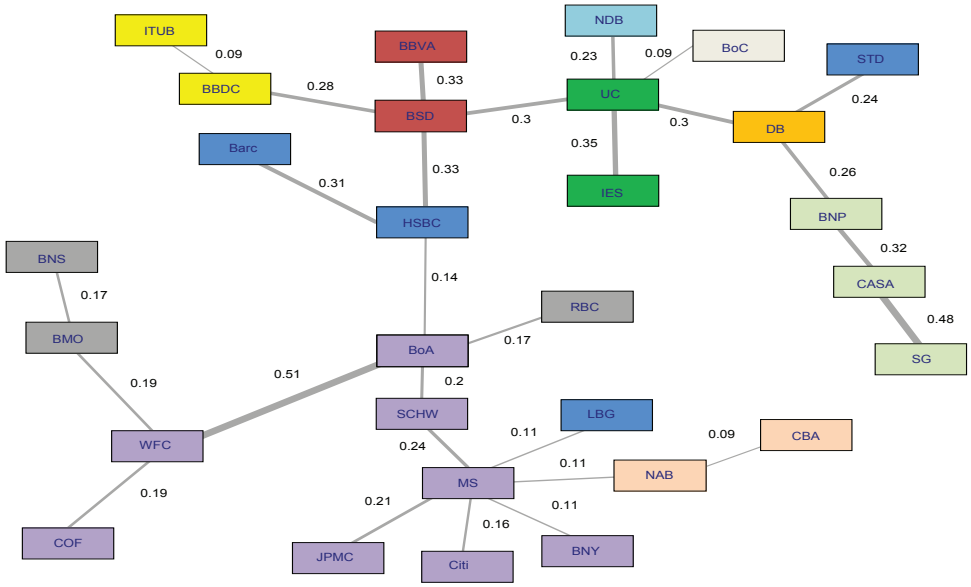
Figure 5.5: **Comparison of country-specific structures arising from the first tree of the fitted R-vines.**

The figure illustrates the first tree of the fitted R-vine copula models for our sample of 29 banks. The models are estimated on the basis of the residuals from AR-GJR-GARCH processes applied to stock returns (Panel (a)) and from ARMA-CS-GARCH processes applied to log-changed *Google Trends* data (Panel (b)), respectively. In both realized dependence structures the thickness of the edges corresponds to the Kendall's τ implied by the respective pair copula parameters. Banks (nodes) residing in the same country are highlighted in the same color.

(a) First tree of the R-vine model fitted to stock return data of 29 banks



(b) First tree of the R-vine model fitted to log-changed *Google Trends* data of 29 banks



The upper Panel in Figure 5.5 highlight several country-specific clusters in the banking network based on stock returns. As expected, the biggest cluster consists of the banks located in the United States, namely: Morgan Stanley, Citigroup, Bank of America, JP Morgan Chase, Wells Fargo, Bank of New York Mellon, Capital One Financial and Charles Schwab. The next biggest cluster consists of the British banks HSBC, Barclays, Lloyds Banking Group and Standard Chartered. Then, two further clusters (Canada and France), each consisting of three banks, can be identified. It is also quite remarkable that the banks located in Spain, Italy, Australia and Brazil are directly linked to each other. If all European clusters are grouped together, a European banking network becomes apparent. With the exception of BBVA Banco, all European banks are directly interconnected. The American and European clusters are linked through Morgan Stanley and Deutsche Bank, respectively. Finally, the plot reveals several central nodes of the banking network. Wells Fargo, Citigroup, Morgan Stanley, Deutsche Bank, BNP Paribas, Société Générale, Barclays and HSBC, respectively, are characterized by at least three direct connections per node indicating their high level of interconnectedness.

Perhaps surprisingly, the first tree of the fitted R-vine copula based on log-changed *Google Trends* data shows a very similar structure. In line with the upper Panel of Figure 5.5 country-specific classifications are clearly evident. First, all banks located in the United States are interconnected and constitute to the biggest country-specific cluster. Moreover, this cluster is only connected to banks of Anglo-Saxon countries: Bank of Montreal (Canada), Royal Bank of Canada (Canada), Lloyds Banking Group (Great Britain) and National Australia Bank (Australia). Another cluster consists of the European banks, in which the banks located in France, Spain and Italy are completely country-linked among each other. Additionally, the European banking network reveals connections to Brazil and China. In contrast to the upper Panel, the American and European clusters are linked through Bank of America and HSBC, respectively. Moreover, the Spanish banks are completely bundled, but, however the British and Canadian banks are not. Comparing the corresponding key nodes of the upper Panel,

we find four common banks via Wells Fargo, Morgan Stanley, Deutsche Bank and HSBC. Additionally, three more central nodes are given by the Bank of America, Unicredit and Banco Santander.

We now turn our attention to pairwise commonalities in the banking network depicted in the first tree. Thus, Figure 5.6 presents the first trees of the fitted R-vine copula models and highlights pairs of identical connections appearing in both first trees.

The two plots in Figure 5.6 reveal nine identical connections: Wells Fargo - Capital One Financial, Citigroup - Morgan Stanley, HSBC - Barclays, Société Générale - Credit Agricole, Unicredit - Intesa Sanpaolo, Bank of Nova Scotia - Bank of Montreal, Com. Bank of Australia - National Australia Bank, Itau Unibanco - Banco Bradesco and BNP Paribas - Deutsche Bank. Except for the last, all connections are country-specific. Regarding the total number of 29 banks in the network, nine identical connections seem to be a remarkable number.

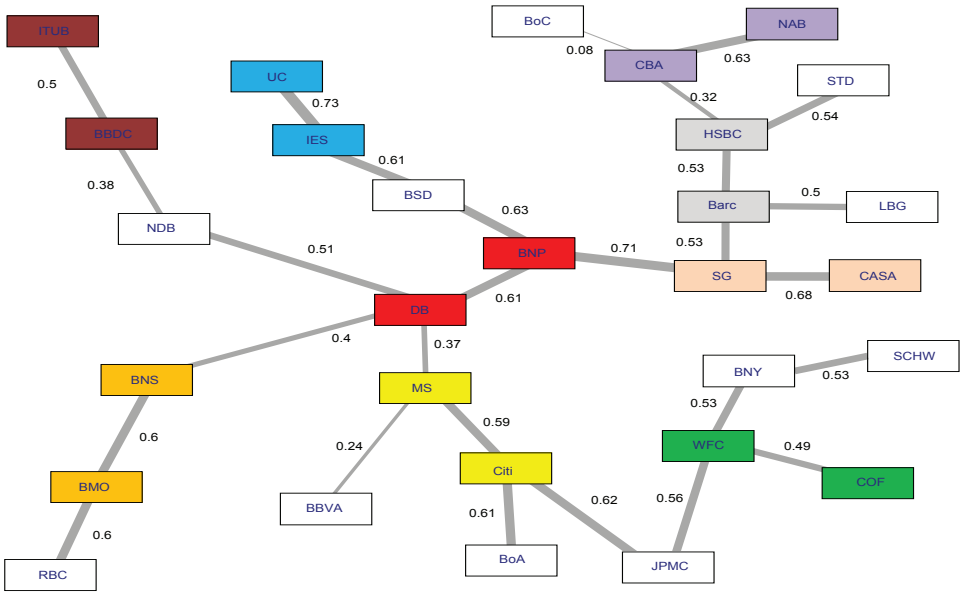
Finally, we focus on the degree of pairwise dependence between the banks, as displayed on the edges. While, the weights in Panel (a) are in the range of 0.08 to 0.73, the corresponding weights in Panel (b) are in the range between 0.09 and 0.51. As a whole, it can be stated that the empirical Kendall's τ values tend to be larger in the first tree of the R-vine model for stock returns. But even in the R-vine model for *Google Trends* data well pronounced dependencies become apparent, particularly for the identical connections emerged from Figure 5.6.

At this point it is natural to ask if *Google Trends* data contains the same information as stock returns. Note that both R-vine models provide strong similarities in the first tree, which captures most of the dependence. We find both the same country-specific characteristics as well as a substantial number of identical pairwise connections. Thus, these findings directly lead us to the discussion about possible causal relations between investor attention and stock returns.

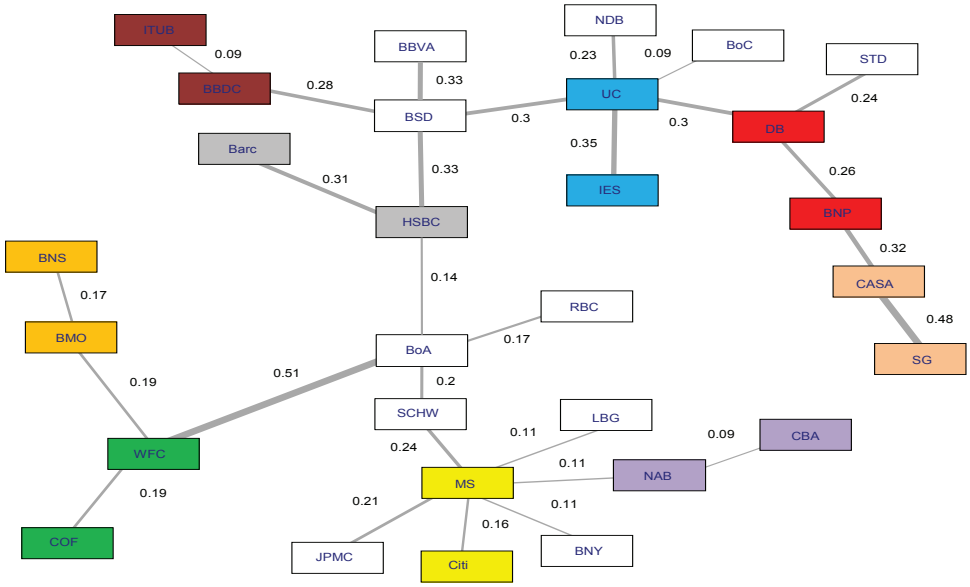
Figure 5.6: **Pairwise identities in the first tree of the fitted R-vines.**

The figure illustrates the first tree of the fitted R-vine copula models for our sample of 29 banks. The models are estimated on the basis of the residuals from AR-GJR-GARCH processes applied to stock returns (Panel (a)) and from ARMA-CS-GARCH processes applied to log-changed *Google Trends* data (Panel (b)), respectively. In both realized dependence structures the thickness of the edges corresponds to the Kendall's τ implied by the respective pair copula parameters. Pairwise connections given in both plots are highlighted in the same color.

(a) First tree of the R-vine model fitted to stock return data of 29 banks



(b) First tree of the R-vine model fitted to log-changed *Google Trends* data of 29 banks



5.4.3 Diagnostics for Dependence

In the analysis so far we have found evidence of commonality in the dependence structure of stock returns and log-changed *Google Trends* data. We now try to tackle possible causal relations between retail investors' attention and stock returns. In particular, we calculate cross-correlations and daily percentages of concordant pairs between the stock returns and log-changed *Google Trends* queries for the corresponding bank name. Moreover, we document the existence of significant tail dependence between both data sets.

Cross Correlations

The cross-correlation function (CCF) is a measure of the similarity between two time series. Analogously to the ACF, the CCF is a function of lag and the value of the corresponding lag indicates which series is leading the other. But additionally, size and position of significant cross-correlations provide information about the strength and direction of the relation between the series. More precisely, regarding two time series y_t and x_t, the series y_t may be related to past lags of the x-series. Then, the sample CCF provides a useful tool for identifying lags of the x-variable that might be appropriate predictors of y_t.

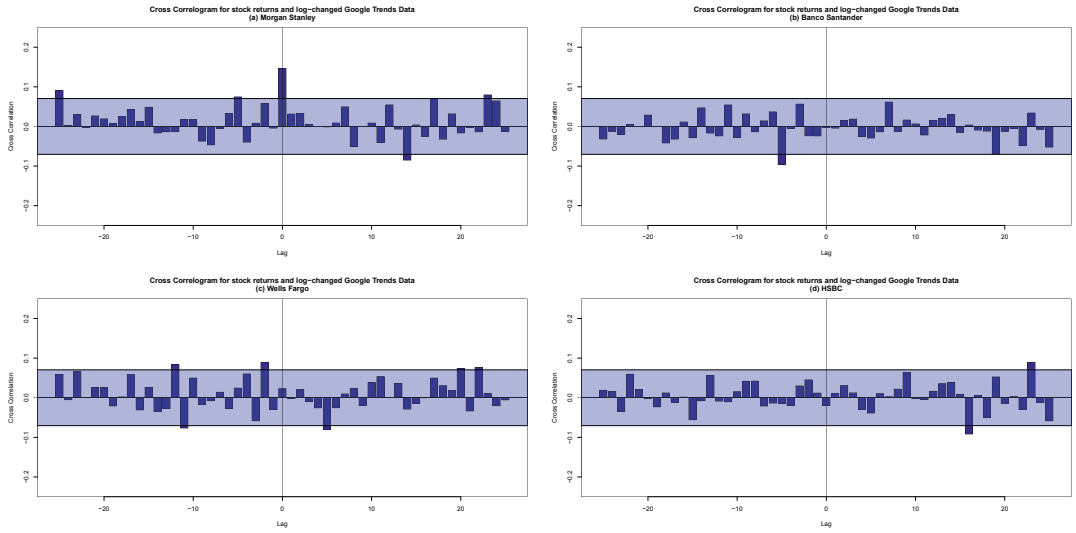
At this point it is natural to mention that time series for which lagged-correlations are of interest are often themselves autocorrelated. The presence of autocorrelation in either series may lead to distortions or misleadings in the estimated cross-correlations. Box and Jenkins (1970) show that the expected standard error for the cross-correlations at various lags depends on the autocorrelation in the individual series. Therefore, our examination is based on white-noise time series resulting from the filtering techniques described in section 5.3.1.

We here only show four examples of cross correlation plots, that represent the distinctive patterns of all other. For the CCF, the presence of significant negative (positive) lags means that *Google Trends* data leads (lags) stock returns. In each plot of Figure

5.7 the CCF is calculated up to lag 25 and the shaded region emphasizes the 95% confidence interval.

Figure 5.7: **Cross Correlation Functions (CCF) for stock returns and log-changed *Google Trends* data.**

The figure shows plots of the cross correlation function (CCF) of the residuals obtained from log-changed *Google Trends* data and stock returns for the full sample period. Panel (a) shows the CCF for Morgan Stanley, Panel (b) for Banco Santander, Panel (c) for Wells Fargo and Panel (d) shows the CCF on the basis of the residuals obtained from HSBC. Negative (positive) lags indicate that *Google Trends* data leads (lags) stock returns. The shaded regions denote the 95% confidence interval around zero.



As can be seen from Panel (a), the CCF for Morgan Stanley reveal statistical significance at various lags. However, the most significant spike is at lag zero, indicating both variables evolve concurrently. The movements in one variable cannot be anticipated by looking at the other and thus, it cannot be stated whether *Google Trends* data leads stocks returns, nor vice versa. Further, the correlation at lag zero is positive, i. e. when one variable increases the other increases as well. The other significant spikes are given at both positive as well as negative lags. Such a pattern is also recognizable for Lloyds Banking Group, Charles Schwab and Standard Chartered. But note, in all these CCFs significant spikes at positive lags occur only at higher orders (from lag 10 upwards), whereas in three out of four cases there is evidence of negative short-term lag cross-correlation.

The upper right Panel (b) represents another particular pattern, here obtained from the CCF of Banco Santander. The plot reveals no significant correlations for positive lags and one statistically significant negative peak at lag -5. Actually the same characteristics are shown by the CCF of BNP Paribas, whereas the CCF of JP Morgan Chase and Bank of Montreal differ only marginally. To be precise, the CCF of JP Morgan Chase exhibits a slightly delayed significant lag at -8 and the latter CCF reveals two significant negative spikes at lags -6 and -11, respectively. Furthermore, the CCFs of Société Générale, Credit Agricole and BBVA Banco generate approximately identically structures. Note, that each of these cross correlation plots shows a significant positive spike at lag -7 and no significant correlations for positive lags. Not surprisingly, there are positive and negative levels of significance. Remember, in case of stock returns, good (bad) news comes in the form of a positive (negative) residual. In case of log-changed *Google Trends* data, on the other hand, bad news as well as good news is associated with a positive residual. Proceeding, significant positive, but higher order spikes (at lags -13 and -14) are also obtained for the CCFs of Bank of China and Commercial Bank of Australia. We thus conclude that these nine CCFs reveal mild statistical evidence for *Google Trends* data might be useful predictors for stocks returns.

The CCF plot of Wells Fargo depicted in the lower left Panel (c) represents another pattern of cross correlations. Apparently, various significant positive as well as negative spikes occur at lags of lower and higher order, respectively. Quite similar types of this structure are obtained by the CCF of Barclays, Deutsche Bank, Capitel One Financial, Bank of New York Mellon, Citigroup, Bank of America, Unicredit, Intesa Sanpaolo and Bank of Nova Scotia. Among each other they differ with regard to the number of significant correlations and the corresponding time lags. Thus, the cross correlation plots of these twelve banks does not permit us to draw any clear conclusions about the lead-lag relation between *Google Trends* data and stock returns.

The CCF plot of HSBC in the lower right Panel (d) illustrates no significant correlations for negative lags and two statistically significant peaks for positive high-order

lags. A similar plot is achieved only for Itau Unibanco; but here statistical significance is given at the low order lag 2. Due to the limited number of samples of this CCF pattern we cannot provide a reliable statement which variable is leading and which is lagging.

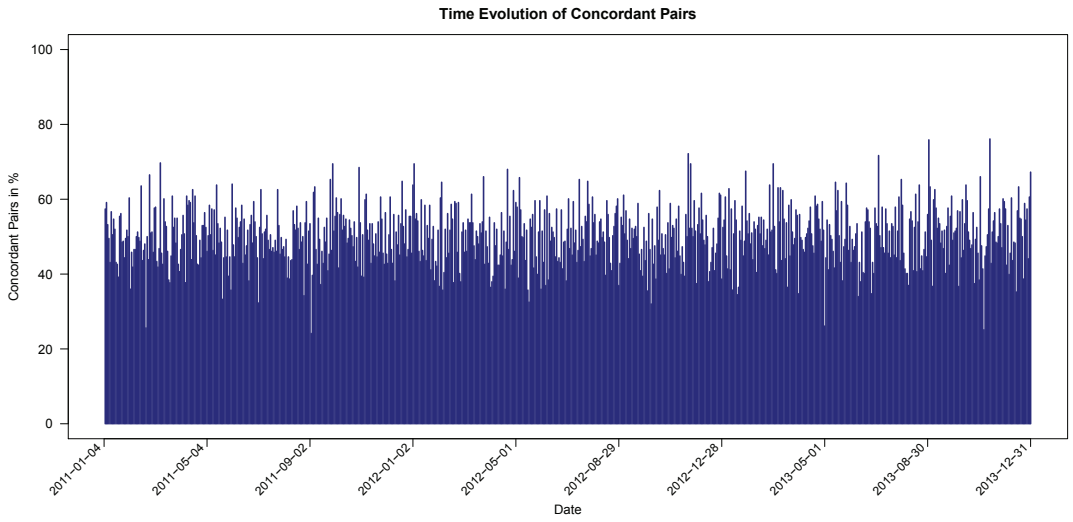
Concordance between *Google Trends* data and stock returns

The main finding from our preliminary analysis is that we have found evidence of commonality in the network structure of stock returns and log-changed *Google Trends* data of the 29 sample banks. But however, our cross correlation investigation so far shows no indications of a unique pattern in the lead-lag relation between the two variables. On the one hand, several CCFs gently suggest *Google Trends* data to be useful predictors for stocks returns. On the other hand and more likely, most CCF plots reveal statistical evidence indicating that both variables evolve concurrently. We now assess the strength of the relation by a measure of association. More precisely, we calculate the average daily percentage of concordant pairs between stock returns and log-changed *Google Trends* data for the corresponding bank names. Figure 5.8 presents the time evolution of the daily percentage of concordant pairs the 29 bivariate time series.

On average, the percentage of concordant pairs is 50.27%, which means that, on average, more than half of both variables evolve in the same direction. This matching is far from being perfect, but indicates a considerably degree of commonality. Over the whole 3-year sample, the percentage has ranged between 24.14% and 76.11%. Note, that the highest average value of a year slightly increases (2011: 69.70%, 2012: 72.17% and 2013: 76.11%), whereas the lowest average value remain relatively stable (2011: 24.14%, 2012: 32.02%, 2013: 25.12%). It is also quite remarkable that the plot provides no specific characteristics over the entire sample period. Even through phases of high volatility as well as relatively calm time intervals in the row data, our measure of association shows no significant effects. From this results we conclude that the rankings produced are stable through time, i. e. we find clear evidence of dependency between both data sets.

Figure 5.8: **Time evolution of the percentage of concordant pairs between stock returns and log-changed *Google Trends* data.**

The figure shows the time evolution of daily percentage of concordant pairs between between stock returns and log-changed *Google Trends* data for the full sample period, January 1, 2011 to December 31, 2013. For each day of the sample period, we calculate the percentage of concordant pairs as the ratio of the number of concordant pairs to the total number of pairs.



Selected parametric pair-copulas in the R-vine model

Both fitted vine copulas show a quite similar structure of the first tree, which captures the strongest dependencies by unconditional pair copulas. We now examine the complete dependence structure of *Google Trends* data in more detail. Consequently, we focus on the fitted bivariate copula families in the R-vine model based on *Google Trends* data, indicating whether or not most of the (un-)conditional dependence in the data is actually linear. The following Table 5.5 presents the corresponding percentages of selected parametric pair-copulas separately by the respective tree level in the estimated R-vine model.[174]

[174] See section 5.4.2 for the candidate parametric copula families from which the pair-copulas of the true vine models are chosen.

Table 5.5: **Treewise selection of parametric pair-copulas (R-vine *Google Trends* data).**

The table reports percentages on the treewise selection of bivariate parametric pair-copulas in our R-vine copula model. The R-vine copula model is estimated on pseudo-observations of log-changed *Google Trends* data for the 29 banks, resulting in 406 ($= 29 \cdot 28/2$) parametric pair-copulas that need to be specified. The 29-dimensional R-vine copula is composed of 28 trees, where copula selection is based on the sequential method as proposed by Dissmann et al. (2013) and conducted using Akaike's Information Criterion (AIC) as the selection criterion to be minimized. Each tree, $i, i = 1, \ldots, 28$, requires the selection and estimation of $29 - i$ bivariate parametric pair-copulas. The results in the table show the number of a particular parametric copula family being selected in the tree i as a percentage of the total number of pair-copulas to be specified in the corresponding tree (that is, $18-i$). As candidate parametric copula families from which the pair-copulas of the vine model are chosen, we use the Gaussian (N), Student's t (t), Clayton (C), Gumbel (G), Frank (F), Joe (J), Survival Clayton (r180 C), Survival Gumbel (r180 G), and rotated versions of the Clayton copula (r90 C and r270 C) and Gumbel copula (r90 G and r270 G), respectively.

	Parametric copula families (in %)											
Tree	N	t	C	G	F	J	r180 C	r180 G	r90 C	r90 G	r270 C	r270 G
1	7.14	28.57	7.14	0.00	10.71	0.00	0.00	46.43	0.00	0.00	0.00	0.00
2	7.41	14.81	11.11	0.00	37.04	0.00	7.41	18.52	0.00	0.00	0.00	3.70
3	15.38	7.69	3.85	3.85	19.23	0.00	11.54	11.54	7.69	3.85	7.69	7.69
4	4.00	12.00	12.00	0.00	24.00	0.00	8.00	16.00	4.00	4.00	8.00	8.00
5	12.50	4.17	4.17	0.00	20.83	0.00	12.50	12.50	0.00	16.67	12.50	4.17
6	13.04	0.00	8.70	0.00	34.78	0.00	13.04	4.35	4.35	8.70	8.70	4.35
7	4.55	18.18	13.64	4.55	9.09	4.55	9.09	9.09	4.55	0.00	13.64	9.09
8	19.05	9.52	9.52	4.76	9.52	0.00	14.29	9.52	4.76	0.00	4.76	14.29
9	15.00	0.00	25.00	0.00	30.00	0.00	15.00	0.00	5.00	0.00	5.00	5.00
10	15.79	5.26	10.53	0.00	21.05	0.00	5.26	0.00	21.05	0.00	15.79	5.26
11	11.11	5.56	16.67	5.56	27.78	0.00	16.67	0.00	0.00	5.56	5.56	5.56
12	0.00	0.00	29.41	0.00	35.29	0.00	5.88	5.88	23.53	0.00	0.00	0.00
13	6.25	6.25	25.00	0.00	12.50	6.25	6.25	12.50	18.75	0.00	0.00	6.25
14	33.33	0.00	20.00	0.00	13.33	0.00	13.33	13.33	0.00	0.00	6.67	0.00
15	7.14	0.00	7.14	0.00	21.43	0.00	7.14	0.00	14.29	7.14	21.43	14.29
16	7.69	0.00	23.08	7.69	23.08	0.00	0.00	7.69	7.69	7.69	15.38	0.00
17	25.00	0.00	0.00	0.00	50.00	8.33	8.33	0.00	8.33	0.00	0.00	0.00
18	9.09	9.09	9.09	0.00	45.45	0.00	0.00	0.00	9.09	0.00	18.18	0.00
19	10.00	0.00	10.00	0.00	50.00	0.00	0.00	0.00	0.00	0.00	20.00	10.00
20	11.11	0.00	22.22	0.00	22.22	0.00	0.00	11.11	0.00	22.22	11.11	0.00
21	12.50	0.00	25.00	25.00	12.50	0.00	0.00	0.00	0.00	0.00	12.50	12.50
22	0.00	0.00	14.29	0.00	28.57	0.00	14.29	14.29	0.00	0.00	0.00	28.57
23	0.00	16.67	16.67	0.00	16.67	0.00	16.67	0.00	33.33	0.00	0.00	0.00
24	20.00	0.00	20.00	0.00	40.00	0.00	0.00	0.00	0.00	0.00	0.00	20.00
25	25.00	0.00	0.00	0.00	50.00	0.00	0.00	25.00	0.00	0.00	0.00	0.00
26	0.00	0.00	0.00	0.00	33.33	0.00	0.00	33.33	0.00	0.00	33.33	0.00
27	0.00	0.00	50.00	0.00	50.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

In the first tree pair-copulas without tail dependence such as the Gaussian copula and Frank copula are chosen for only 7.14% and 10.71%, respectively, while the vast majority of bivariate unconditional data pairs are modeled using tail dependent copulas. In contrast, the percentage for the Gaussian copula and Frank copula increases considerably up in the lower trees. In total, slightly over a third of 406 fitted data pairs are modeled via these both copulas. An essential finding is that in more than 40% of cases

tail dependent copulas are selected.[175] Furthermore, there is a substantial amount of pair-copulas capturing dependencies in the upper left tail and the lower right tail with around 23%.[176] These results so far underline that log-changed *Google Trends* data indeed exhibit strong non-linear dependence.

Pairwise comparisons

In the final step of our diagnostics we present a bivariate analysis for some banks. More precisely, we pick up some of the pairwise identities emerged from the first tree of the respective R-vine models (see Figure 5.6) and present a graphical comparison of the underlying data pairs, i. e. the corresponding stocks returns as well as the log-changed *Google Trends* data. In each of the subsequent plots of Figure 5.9, the raw data pairs are illustrated by the panels on the left, whereas on the right side the corresponding pseudo-observations are presented.

We start our comparative study with the connection between Morgan Stanley and Citigroup depicted in Panel (a). As can be seen in the upper left plot actual observations of stock returns display both joint negative and joint positive extreme values. This is confirmed by the scatter plot in the upper right corner, showing clear evidence of tail dependence between the corresponding pseudo-observations. The scatter plot of the actual observations of *Google Trends* data in the lower left corner also reveals evidence of non-linear dependence. Obviously, extreme values (around -0.5 and 0.5) occur simultaneously. This impression is strengthened by the lower right plot of the corresponding pseudo-observations, which depicts lower and upper tail dependence, but however, is somewhat less pronounced compared to pseudo-observations of stock returns.

To study more closely a pairwise identity between two banks of the European cluster, Panel (b) illustrates scatter plots of stock returns and log-changed *Google Trends* data for Intesa Sanpaolo and Unicredit. We find, not surprisingly, a striking similarity

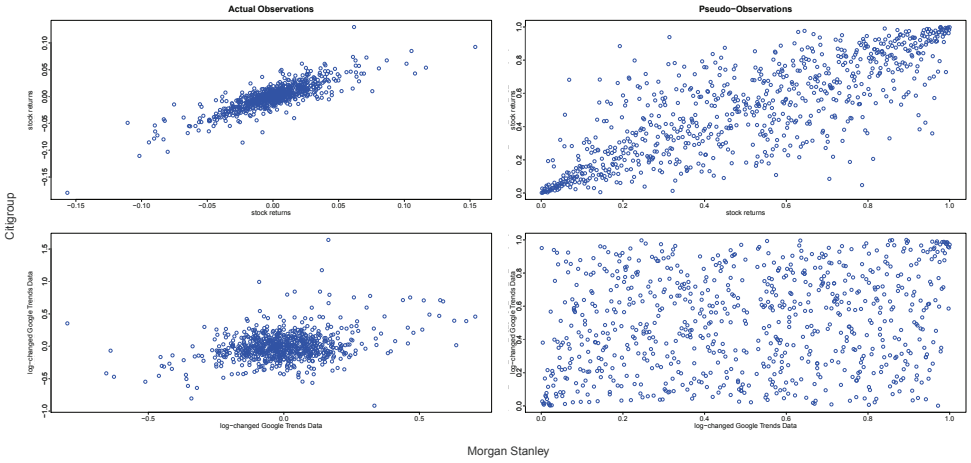
[175] That is, the Student's t, Clayton, Gumbel, Joe, survival Clayton and survival Gumbel copula.

[176] In the following we refer upper left to upper negative tail dependency and lower right to lower negative tail dependency, respectively.

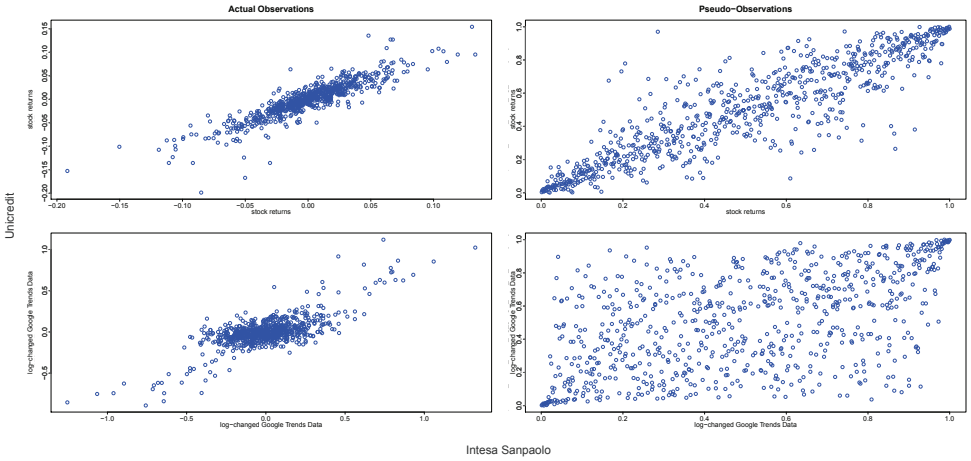
Figure 5.9: **Scatter plots of actual and pseudo observations.**

The figure shows scatter plots based on actual observations as well as pseudo observations of stock returns and log-changed *Google Trends* data, respectively, for three considered pairs of banks. The three bank pairs are Morgan Stanley - Citigroup, Intesa Sanpaolo - Unicredit and Barclays - HSBC. The Panels on the left present plots of the actual observations for both data types whereas the right Panels show the corresponding plots of the pseudo observations. The sample consists of daily observations from January 1, 2011 to December 31, 2013.

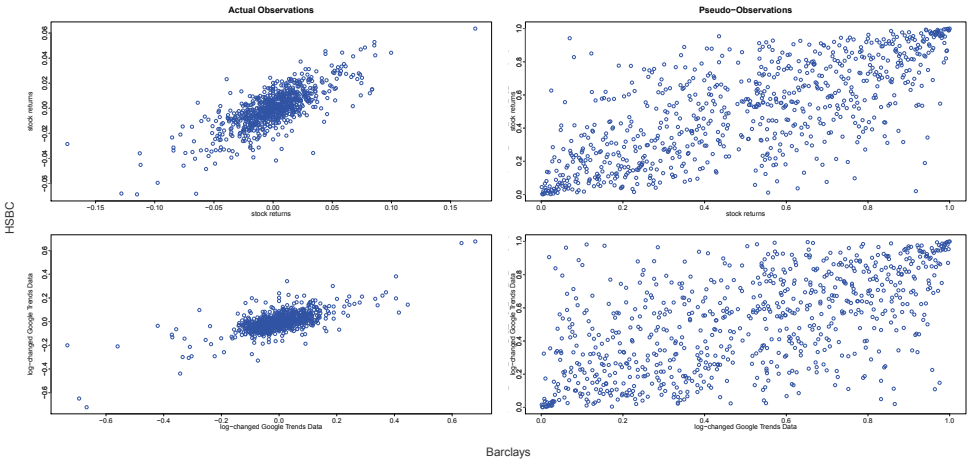
(a) Morgan Stanley vs. Citigroup



(b) Intesa Sanpaolo vs. Unicredit



(c) Barclays vs. HSBC



for the dependence structure in both the actual observations as well as the pseudo-observations of stock returns as highlighted in the upper part of the graphic and those gained from the upper plots of Panel (a), respectively. But more importantly, the dependence structures of log-changed *Google Trends* data depicted in the lower part of Panel (b) seem more or less to coincide with those obtained from the stock returns. Obviously, the presence of nonlinear dependence is more distinctive in the upper plots, but the lower plots furnish clear evidence of the same dependence structure.

This result is also supported by the pairwise comparison of Barclays and HSBC as shown in Panel (c). Again, the plots in the upper part of the graphic reveal strong tail-dependence in both the lower tail and the upper tail. The plot in the lower left corner shows tightly clustered points around zero on the one hand, but also simultaneously occurring extreme values on the other hand. After filtering the time series of actual observations and fitting appropriate univariate marginal models, the resulting scatter plots of pseudo-observations of both stock returns and *Google Trends* data bear a strong resemblance as becomes apparent from the plots on the right side. We thus conclude that non-linearities in the dependence structure of bank stock returns are also manifested in the corresponding log-changed *Google Trends* data.

5.4.4 Extreme dependence between stock returns and *Google Trends* data

In the last step of our analysis we examine the presence of tail dependence between stock returns and log-changed *Google Trends* data for the bank names, with particular focus on asymmetry in the tails. Testing for asymmetric dependence structures is motivated by the fact that for stock returns, bad news comes in the form of a negative residual ($\varepsilon_{i,t} < 0$). For *Google Trends* data, on the other hand, bad news is accompanied by a positive residual ($\varepsilon_{i,t} > 0$) due to increasing retail investor attention. Thus, on the one hand we measure the co-movement in the lower (left) and upper (right)

tail probabilities. On the other hand and more interestingly, we quantify the degree of dependence in the upper left and lower right tail of the bivariate distribution.

In order to assess the simultaneous occurrence of extreme events between stock returns and *Google Trends* data, we initially introduce tail dependence coefficients. For a bivariate copula $C(u_1, u_2)$ the lower (left) tail dependence is measured via the probability limit

$$\lambda_{l,l}(C) = \lim_{\zeta \to 0} \mathbb{P}[X_2 \leq F_2^{\leftarrow}(\zeta) \,|\, X_1 \leq F_1^{\leftarrow}(\zeta)] = \lim_{\zeta \to 0} \frac{C(\zeta, \zeta)}{\zeta}, \tag{5.18}$$

where ζ denotes the tail probability. Analogously, the upper (right) tail dependence is defined by

$$\lambda_{u,u}(C) = \lim_{\zeta \to 0} \mathbb{P}[X_2 > F_2^{\leftarrow}(1-\zeta) \,|\, X_1 > F_1^{\leftarrow}(1-\zeta)] = 2 - \lim_{\zeta \to 0} \frac{1 - C(1-\zeta, 1-\zeta)}{\zeta}. \tag{5.19}$$

Both coefficients, $\lambda_{l,l}$ and $\lambda_{u,u}$, are widely used and well-established in measuring the simultaneously occurrence of extremely small or extremely large outcomes of X_1 and X_2. In our modeling framework, for instance, upper tail dependence quantifies the probability to observe a large bank´s stock return (X_2), assuming a large change of retail investor's attention (X_1) for the corresponding bank. But additionally, the degree of dependence in the upper left and lower right quadrant tail of the considered bivariate distributions is of great interest. More precisely, lower right (upper left) tail dependence captures extremely negative (positive) bank stock returns (X_2), assuming large (low) log-changed *Google Trends* data (X_1). To handle tail dependence in the upper left and lower right quadrant, we define the corresponding dependence coefficients via

$$\lambda_{l,u}(C) = \lim_{\zeta \to 0} \mathbb{P}[X_2 > F_2^{\leftarrow}(1-\zeta) \,|\, X_1 \leq F_1^{\leftarrow}(\zeta)] = 1 - \lim_{\zeta \to 0} \frac{C(\zeta, 1-\zeta)}{\zeta} \tag{5.20}$$

and

$$\lambda_{u,l}(C) = \lim_{\zeta \to 0} \mathbb{P}[X_2 \leq F_2^{\leftarrow}(\zeta) \,|\, X_1 > F_1^{\leftarrow}(1-\zeta)] = 2 - \lim_{\zeta \to 0} \frac{1 - C(1-\zeta, 1-\zeta)}{\zeta}. \tag{5.21}$$

We now aim to plot the functions above, in order to visualize limiting behavior. Therefore, we define

$$LL(\zeta) = \mathbb{P}[X_2 \leq F_2^{\leftarrow}(\zeta) \,|\, X_1 \leq F_1^{\leftarrow}(\zeta)] \tag{5.22}$$

for the lower left tail,

$$UU(\zeta) = \mathbb{P}[X_2 > F_2^{\leftarrow}(1-\zeta) \,|\, X_1 > F_1^{\leftarrow}(1-\zeta)] \tag{5.23}$$

for the upper right tail,

$$LU(\zeta) = \mathbb{P}[X_2 > F_2^{\leftarrow}(1-\zeta) \,|\, X_1 \leq F_1^{\leftarrow}(\zeta)] \tag{5.24}$$

for the upper left tail, and

$$UL(\zeta) = \mathbb{P}[X_2 \leq F_2^{\leftarrow}(\zeta) \,|\, X_1 > F_1^{\leftarrow}(1-\zeta)] \tag{5.25}$$

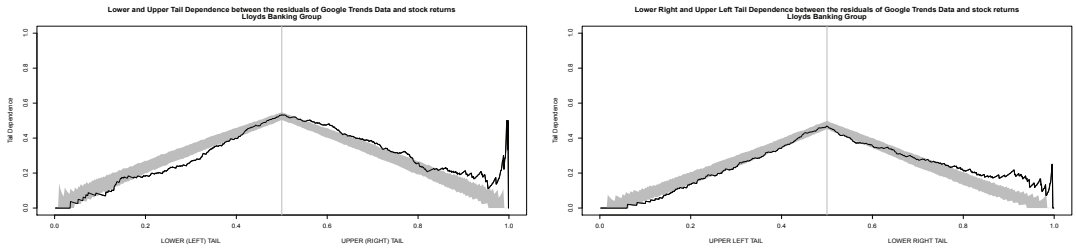
for the lower right tail. It is easily to obtain empirical counterparts of these functions to calculate the limiting behavior for the underlying data pairs. Figure 5.10 depicts evolutions of the tail concentration functions related to the residuals of stock returns and log-changed *Google Trends* data of four considered banks. For each bank two different plots are shown. The plots on the left side present lower (left) and upper (right) tail concentrations, whereas the right panels display the degree of dependence in the upper left and lower right tail. Moreover, we compare the empirical tail concentrations to those implied by the Gaussian copula, represented here by the respective 90% confidence interval.[177]

[177] Comparing empirical tail dependence coefficients to those given by parametric copulas that have the same Kendall's τ, directly highlights commonalities and mismatches in the limiting behavior. Then, a detailed inspection of both tails reveals if the proposed parametric copula accurately captures the limiting behavior. For instance, the Gaussian copula has no tail dependence and therefore, underestimations of joint movements in the tails become directly apparent.

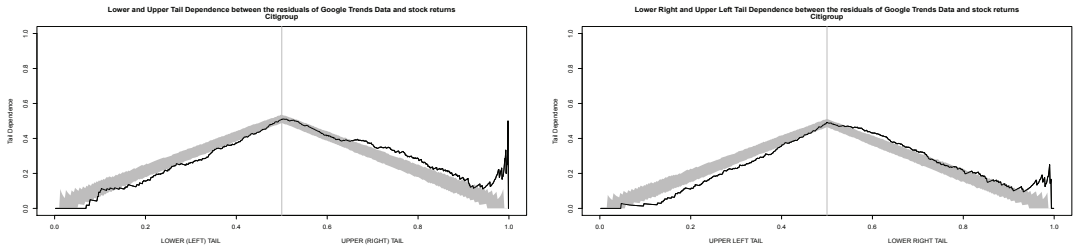
Figure 5.10: **Empirical tail concentration between stock returns and *Google Trends* data.**

The Figure shows evolutions of the empirical tail concentration related to the residuals of stock returns and log-changed *Google Trends* data for four selected banks subject to various threshold levels. The four banks are Lloyds Banking Group, Citigroup, Morgan Stanley and Charles Schwab. The Panels on the left present lower (left) and upper (right) tail concentrations, whereas the right Panels display the degree of dependence in the upper left and lower right tail. The shaded regions denote the 90% confidence interval given by the bivariate Gaussian copula with the corresponding Kendall's τ.

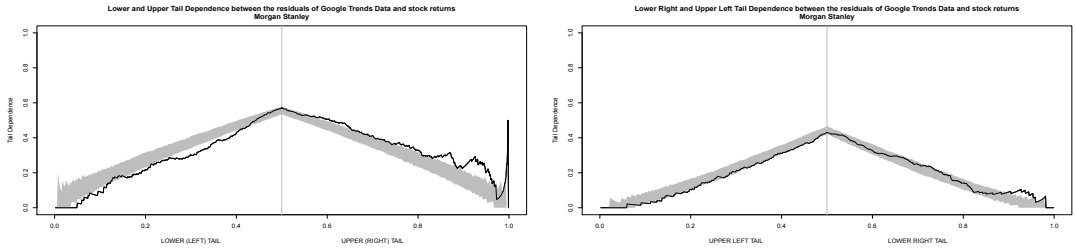
(a) Lloyds Banking Group



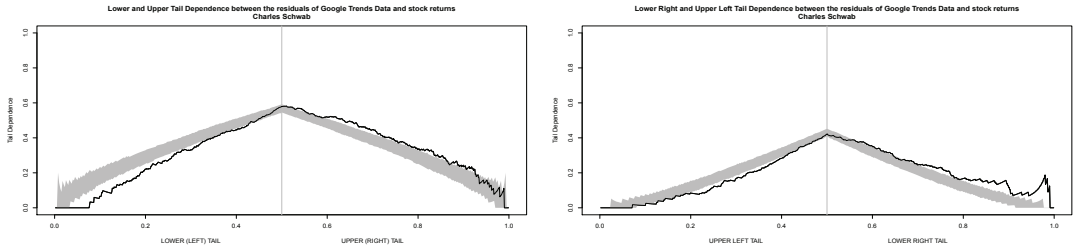
(b) Citigroup



(c) Morgan Stanley



(d) Charles Schwab



At the top of Figure 5.10 the empirical tail concentrations for Lloyds Banking Group are depicted. The left plot of Panel (a) shows that the residuals of stock returns and log-changed *Google Trends* data are lower tail independent. In contrast, the upper tail function indicates well pronounced tail dependence with around $\hat{\lambda}_{u,u} = 0.2$. Perhaps

surprisingly, the corresponding graph for the upper left and lower right tail concentrations follows a very similar pattern. As can be seen from the right plot of Panel (a) the data is upper left tail independent, but reveals clear evidence of lower right tail dependence with approximately $\hat{\lambda}_{l,u} = 0.2$. The only difference is the more distinctive peak at the right margin of the plots. Panel (b) of Figure 5.10 reveals a similar limiting behavior for the tail dependence coefficients of Citigroup. Again, the stock returns and log-changed *Google Trends* data are lower as well as upper left tail independent, but clear evidence of tail dependence in upper and lower right quadrants become apparent.

Panels (c) and (d) of Figure 5.10 present the empirical tail concentrations for Morgan Stanley and Charles Schwab. The left half of all four plots confirm our previous finding that there are no substantial tail concentrations in the left quadrants. Additionally, the degree of dependence in the left tails of Morgan Stanley and Charles Schwab is weaker than Gaussian. To put it more precisely, we do not find any statistical evidence for extremely high and low stock returns, respectively, occurring simultaneously with a significant decrease in retail investors attention for the corresponding bank. Furthermore, the tail concentrations in the right quadrants possess different characteristics. On the one hand, the upper right tail dependence coefficient of Morgan Stanley is clearly significant and reveals a sharp peak at the limit. But however, the right plot of Panel (c) reveals only mild lower right tail concentration. As can be seen from Panel (d), on the other hand, exactly the contrary is true for the empirical tail dependence concentration for Charles Schwab.[178] Our results so far emphasize that the bivariate dependence structure inherent in residual pairs of stock returns and log-changed *Google Trends* data for the bank name cannot be adequately modeled using tail independent Gaussian copulas. The data sample is characterized by significant tail concentrations in the upper right and lower right quadrants. More importantly, it can be easily demonstrated that most bivariate copulas cannot capture the simultaneous existence of both the upper right and the lower right tail concentrations. Thus, in the last part of our diagnostics,

[178] In unreported results, empirical estimates of the tail dependence for other banks show quite similar results.

we address the question how to capture asymmetric dependence structures in the sense described above. In order to tackle this issue, we follow Wu (2014) and employ asymmetric copulas. The proposed method of constructing asymmetric copulas differs from other existing methods due to the additional flexibility to handle tail dependence in each of the four quadrants.

The key idea behind asymmetric copulas is to build a copula in which variables are not exchangeable with other variables. Focusing on the bivariate case a given Copula $C(u_1, u_2)$ ($u_1, u_2 \in [0, 1]$) is called symmetric, if $C(u_1, u_2) = C(u_2, u_1)$, i. e. u_1 and u_2 are exchangeable. In the literature most of the proposed parametric copulas are symmetric in nature.[179] Note, Wu (2014) proves for symmetric copulas the implication $\lambda_{l,u}(C) = \lambda_{u,l}(C)$. Also note, that this relation poses the problem that symmetric copulas cannot fit data exhibiting unbalanced tail concentrations in the upper left and lower right quadrant of a bivariate distribution. To overcome this drawback, we define

$$\breve{C}_1(u_1, u_2) = u_2 - C(1 - u_1, u_2), \tag{5.26}$$

$$\breve{C}_2(u_1, u_2) = u_1 - C(u_1, 1 - u_2), \tag{5.27}$$

and it can easily be proven that $\breve{C}_1$ and $\breve{C}_2$ are copulas.[180] Then, for the lower tail dependence coefficient of $\breve{C}_1$ and $\breve{C}_2$ the following relation is given

$$\lambda_{l,l}(\breve{C}_1) = \lambda_{l,u}(C), \tag{5.28}$$

$$\lambda_{l,l}(\breve{C}_2) = \lambda_{u,l}(C). \tag{5.29}$$

By using a convex combination of $C, \breve{C}_1$ and $\breve{C}_2$ we obtain a asymmetric bivariate copula. With the mixing proportions p_k ($k = 0, 1, 2$) the corresponding asymmetric copula is given by

$$\bar{C}(u_1, u_2) = p_0 C(u_1, u_2) + p_1 \breve{C}_1(u_1, u_2) + p_2 \breve{C}_2(u_1, u_2), \tag{5.30}$$

[179] According to this, for instance, both the Clayton as well as the Gumbel copula are symmetric.
[180] The copulas $\breve{C}_1$ and $\breve{C}_2$ are also discussed in the monograph by Nelsen (2006) for a different purpose.

where the nonnegative quantities p_k sum to one. As the components $\breve{C}_1(u_1, u_2)$ and $\breve{C}_2(u_1, u_2)$ are copulas, it easily follows that (5.30) again defines a bivariate copula.

The four tail dependence coefficients of the copula $\bar{C}(u_1, u_2)$ are then calculated via

$$\lambda_{l,l}(\bar{C}) = p_0\lambda_{l,l}(C) + p_1\lambda_{u,l}(C) + p_2\lambda_{l,u}(C), \tag{5.31}$$

$$\lambda_{l,u}(\bar{C}) = p_0\lambda_{l,u}(C) + p_1\lambda_{u,u}(C) + p_2\lambda_{l,l}(C), \tag{5.32}$$

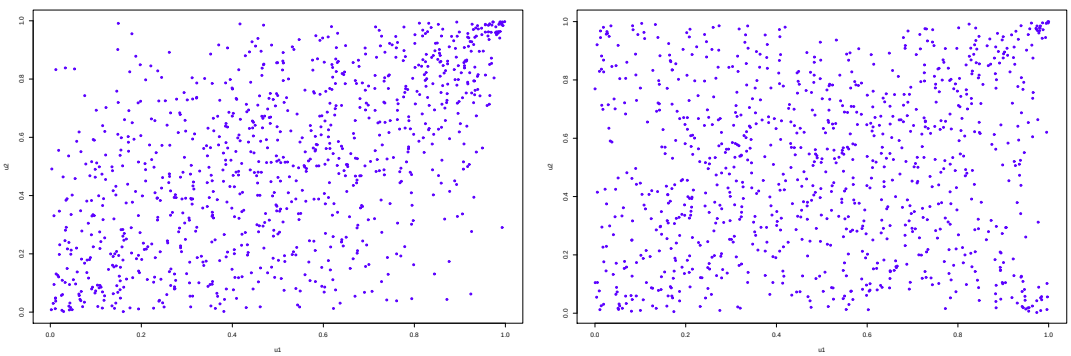
$$\lambda_{u,l}(\bar{C}) = p_0\lambda_{u,l}(C) + p_1\lambda_{l,l}(C) + p_2\lambda_{u,u}(C), \tag{5.33}$$

$$\lambda_{u,u}(\bar{C}) = p_0\lambda_{u,u}(C) + p_1\lambda_{l,u}(C) + p_2\lambda_{u,l}(C). \tag{5.34}$$

This representation highlights the way constructing asymmetric copulas with $\lambda_{l,u}(\bar{C}) \neq \lambda_{u,l}(\bar{C})$ and/or $\lambda_{l,l}(\bar{C}) \neq \lambda_{u,u}(\bar{C})$. For example, the upper right tail dependent Gumbel copula has $\lambda_{u,u}(C) = 2 - 2^{1/\theta}$, where $0 < \theta \leq 1$ is the parameter controlling the dependence, and $\lambda_{l,l}(C) = \lambda_{l,u}(C) = \lambda_{u,l}(C) = 0$. But using (5.30) the corresponding asymmetric Gumbel copula exhibits $\lambda_{l,l}(\bar{C}) = 0$ and $\lambda_{l,u}(\bar{C}) = (2 - 2^{1/\theta})p_1$, $\lambda_{u,l}(\bar{C}) = (2 - 2^{1/\theta})p_2$ as well as $\lambda_{u,u}(\bar{C}) = (2 - 2^{1/\theta})p_0$. Figure 5.11 highlights the differences between a Gumbel copula and an asymmetric Gumbel copula. It particularly shows the presence of significant tail concentrations in the upper right quadrant as well as in the lower right quadrant.

Figure 5.11: **Scatter plots of a Gumbel copula and an asymmetric Gumbel copula.**

The figure presents plots of $N = 1000$ observations simulated from two bivariate Gumbel copulas. Panel (a) shows the plot of simulated observations from the Gumbel copula $C_G(u_1, u_2)$ with $\theta = 1.75$. Panel (b) shows the plot of simulated observations from the asymmetric Gumbel copula $\bar{C}_G(u_1, u_2) = p_0C_G(u_1, u_2) + p_2\breve{C}_{G_2}(u_1, u_2)$ with $\theta = 1.75$, $p_0 = 0.6$ and $p_2 = 0.4$.



At this point it is natural to ask if simple copula models are able to capture asymmetry in the dependence between stock returns and *Google Trends* data for the respective bank name. For this purpose, we draw a comparison between simple copula models and the proposed asymmetric copulas with respect to the fitting accuracy. To be precise, we fit the stock returns and the *Google Trends* data of each underlying bank to a prespecified set of copula families. Then, we calculate the AIC for each copula family considered and construct via (5.30) corresponding asymmetric copulas for those with the minimum AIC value. The results of this comparison are presented in Table 5.6. In each row the minimum AIC value of the simple copulas are highlighted in red.

The results of Table 5.6 support our previous findings regarding the presence of significant tail concentrations in the upper right and lower right quadrant tails between the stock returns and log-changed *Google Trends* data. Focusing on simple copulas, in 19 of 29 cases an upper right tail dependent copula is chosen, particularly 17 times the survival Clayton copula, and once a Gumbel copula as well as Joe copula. In nine out of the 29 cases lower right tail dependent copulas are selected, i. e. five times the rotated Clayton copula (90 degrees) and four times the rotated Gumbel copula (270 degrees). In contrast, only in one case the AIC criterion indicates to chose an upper left tail dependent copula and lower left tail dependent copulas are totally ignored. The last column of Table 5.6 presents the AIC values of the corresponding asymmetric copulas. Based on AIC, asymmetric copulas lead to considerable improvements on the fitting accuracy of the underlying dependence structure in 27 out of 29 cases. Together, these results strongly recommend using asymmetric copulas to capture the (asymmetric) tail dependence that exists between the stock returns and log-changed *Google Trends* data for the respective bank name.

Table 5.6: **AIC-values of different bivariate copulas.**

The table presents the performance of different bivariate copulas fitted to the (pseudo-)observations of stock returns and log-changed *Google Trends* data for each of the 29 banks. Based on AIC, in 27 of 29 cases an asymmetric copula is best and survival Clayton is second best. Sensitivity analysis, for copulas that use more asymmetric tail dependence than likelihood-based estimates can substantially improve the model fit. (N=Gauss, t=Student t, C=Clayton, G=Gumbel, F=Frank, J=Joe, A=Asymmetric)

	AIC-values												
Bank	N	t	C	G	F	J	r180 C	r180 G	r90 C	r90 G	r270 C	r270 G	A
Banco Bradesco	0.72	4.47	2.01	2.02	1.15	2.01	2.01	2.02	0.00	2.01	1.77	-1.04	-1.05
Banco Santander	1.73	7.63	2.01	2.02	1.77	2.02	2.01	2.02	1.38	2.02	2.00	2.01	1.38
Bank of America	0.75	7.94	1.96	2.00	0.40	2.01	1.32	2.00	2.01	2.02	2.01	2.03	0.39
Bank of China	1.67	7.38	2.00	2.01	1.12	2.01	2.01	2.01	2.00	2.01	1.84	2.01	0.75
Bank of Montreal	1.97	7.33	2.00	1.90	1.79	1.90	1.99	2.01	1.29	2.02	2.01	1.94	1.24
Bank of New York Mellon	1.97	6.86	2.01	2.01	1.81	2.01	1.72	2.01	1.99	2.02	2.01	1.73	1.71
Bank of Nova Scotia	2.00	8.95	2.01	2.02	1.97	2.02	1.70	2.02	2.00	2.02	2.01	2.01	1.67
Barclays	0.54	7.71	2.02	2.01	0.60	2.00	2.00	2.04	-0.55	2.01	2.01	0.16	-1.06
BBVA Banco	2.00	8.60	2.00	2.02	1.99	2.02	2.01	2.01	2.00	2.02	2.01	1.64	0.84
BNP Paribas	0.06	6.63	1.25	2.00	0.79	2.01	0.95	1.56	2.01	2.04	2.02	2.03	0.06
Capital One Financial	-0.34	4.46	2.02	2.02	1.05	2.02	2.01	2.03	-0.16	1.87	1.30	-0.01	-0.45
Charles Schwab	-22.99	-20.74	-2.14	-25.62	-31.81	-21.78	-23.61	-7.29	2.02	2.09	2.06	2.02	-33.87
Citigroup	0.07	-4.09	2.01	-7.26	1.19	-9.23	-12.28	2.01	0.19	2.04	2.03	-0.75	-15.74
Com. Bank of Australia	1.49	9.52	2.00	2.01	1.62	2.01	1.85	2.01	2.01	2.02	2.01	2.02	1.49
Credit Agricole	1.97	9.49	2.00	2.00	1.79	2.00	2.00	2.01	2.00	2.02	2.01	2.01	1.66
Deutsche Bank	2.00	7.42	2.00	2.02	1.64	2.02	1.70	2.01	2.00	2.01	2.01	2.01	1.38
HSBC	1.99	-23.54	2.02	-32.64	0.20	-34.06	-20.85	2.02	2.00	1.91	1.70	1.60	-35.95
Intesa Sanpaolo	2.00	1.85	2.01	0.78	1.02	0.43	-1.14	2.01	0.23	2.02	2.02	0.02	-2.88
Itau Unibanco	0.30	2.81	2.02	2.00	-1.69	1.68	1.06	2.04	0.70	1.92	1.91	2.00	-3.51
JP Morgan Chase	1.89	4.12	2.01	1.36	1.99	0.96	-0.72	2.01	0.58	2.03	2.02	-1.08	-1.72
Lloyds Banking Group	-6.39	-14.17	2.00	-14.99	-2.56	-16.19	-22.32	0.75	1.88	2.05	2.04	0.58	-25.02
Morgan Stanley	-22.42	-21.79	0.22	-29.42	-21.31	-29.10	-31.54	-4.92	2.02	2.08	2.05	2.06	-32.63
National Australia Bank	1.50	10.45	2.00	2.00	1.98	2.00	1.59	2.01	2.01	2.03	2.01	2.03	1.28
Nordea Bank	0.92	5.08	1.36	1.75	1.69	1.88	0.82	1.49	2.01	2.02	2.01	2.02	0.70
Royal Bank of Canada	1.85	8.75	2.01	1.88	1.90	1.81	1.17	2.02	1.91	2.03	2.02	2.00	1.11
Société Générale	-1.48	2.36	1.31	0.19	0.32	0.59	-2.74	1.05	2.02	2.04	2.02	2.03	-2.77
Standard Chartered	-1.35	2.09	2.01	2.04	0.22	2.03	2.02	2.04	-5.45	2.00	2.00	-3.55	-5.81
Unicredit	1.76	6.38	2.01	1.83	1.99	1.74	-0.27	2.01	2.00	2.02	2.01	2.01	-0.31
Wells Fargo	1.44	6.74	2.01	1.45	1.92	1.30	-0.35	2.01	2.01	2.03	2.01	2.02	-0.42

5.5 Conclusion

Motivated by the growing interest in measuring retail investors' attention by daily internet search data, this paper starts with a comprehensive time series analysis to extract meaningful statistics and other characteristics of Google search data. We find both autoregressive dynamics and conditional heteroskedasticity structures inherent in the time-series data. Our analysis shows that the first- and second-moment dependencies are well captured by asymmetric ARMA-CS-GARCH models with the skewed t and skewed *ged* distribution for the Google search data residuals.

To model the joint distribution of Google search data we propose R-vine copulas. The great flexibility to capture dependencies of extreme values made them popular in finance. By using vine copulas we are able to document the existence of strong non-linear and asymmetric dependence in the Google search data. While non-linear dependence has been shown to exist in financial time series like stock returns and credit risk, this paper is the first that demonstrates that investor attention measured by Google search data is characterized by extreme dependence as well. Further, we find a striking similarity between the joint distribution of a multivariate bank stock portfolio and the corresponding portfolio of Google search queries. Consistent with this result, we retrieve tail dependence between stock returns in the respective search query pairs.

Several studies document that the co-movement between investor attention and stock returns is bi-directional. However, the methods used so far only take into account linear and symmetric dependencies. Exploiting our modeling approach, we provide first empirical evidence of significant tail dependence between investor attention and stock returns. Moreover, we explore possible causal relations between stock returns and log-changed *Google Trends* search queries by calculating cross-correlations and daily concordant pairs. We show the movements in one variable cannot be anticipated by looking at the other and thus, it cannot be stated whether stock returns lead *Google Trends* data, nor vice versa. More precisely, stock returns and *Google* search data seem to evolve concurrently in real time. This result suggests that investor attention

measured by internet search data and stock returns reveal almost the same information. Consequently, future research should investigate the existence of external factors which drive the level of investor attention and stock returns in the same way and nearly to the same extent.

Appendix A

Supplementary Material for Chapter 2

A.1 Hilfssatz für den Beweis von Satz 2.2.5

Lemma A.1.1. *Es gilt:*

$$cov\left[\frac{dC}{C}, \frac{df}{f}\right] = \frac{\sigma_f^2}{f}\left[\left(\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf} - h^*\right)\frac{C_\Phi}{C}\Phi + \frac{C_f}{C}f + \frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\frac{C_s}{C}s\right].$$

Beweis: Analog zu dem Beweis von Satz 2.2.4 müssen nur die Terme $C_\Phi\, d\Phi, C_s\, ds$ und $C_f\, df$ betrachtet werden. Daraus ergibt sich:

$$\begin{aligned}\operatorname{cov}\left[\frac{dC}{C}, \frac{df}{f}\right] &= \operatorname{cov}\left[\frac{C_\Phi}{C}d\Phi + \frac{C_s}{C}ds + \frac{C_f}{C}df, \frac{df}{f}\right] \\ &= \frac{C_\Phi}{C}\Phi\operatorname{cov}\left[\frac{d\Phi}{\Phi}, \frac{df}{f}\right] + \frac{C_s}{C}s\operatorname{cov}\left[\frac{ds}{s}, \frac{df}{f}\right] + \frac{C_f}{C}f\operatorname{var}\left[\frac{df}{f}\right].\end{aligned} \tag{A.1}$$

Für die Kovarianzterme und dem Varianzterm auf der rechten Seite von Gleichung (A.1) ergibt sich unter Berücksichtigung der optimalen Hedge-Rate h^*:

$$\operatorname{cov}\left[\frac{d\Phi}{\Phi}, \frac{df}{f}\right] = \frac{\sigma_f^2}{f}\left(\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf} - h^*\right) \tag{A.2}$$

sowie

$$\operatorname{cov}\left[\frac{ds}{s}, \frac{df}{f}\right] = \frac{\sigma_f^2}{f}\left(\frac{\sigma_s\sqrt{f}}{\sigma_f\sqrt{s}}\rho_{sf}\right) \tag{A.3}$$

bzw.

$$\operatorname{var}\left[\frac{df}{f}\right] = \frac{\sigma_f^2}{f}. \tag{A.4}$$

Durch Einsetzen der Gleichungen (A.2), (A.3) und (A.4) in die Gleichung (A.1) und anschließendes Ausklammern des Terms $\frac{\sigma_f^2}{f}$ ergibt sich unmittelbar die Behauptung. □

Appendix B

Supplementary Material for Chapter 3

B.1 Simulation algorithm for a canonical vine

Algorithm 3: Simulation algorithm for a canonical vine

Input: N, d, fam, $\boldsymbol{\theta}$

Let N be the number of simulations and d the dimension of the vine. If the underlying vine is parametrically specified fam denotes the set of $d(d-1)/2$ bivariate parametric copula families defining the PCC structure and $\boldsymbol{\theta}$ the vector of the related parameters. In case of a Bernstein PCC fam denotes $d(d-1)/2$ nonparametric Bernstein copulas and $\boldsymbol{\theta}$ contains the corresponding contingency tables.

Output: x

x is an array of simulations of size $N \times d$ from the specified C-vine.

begin
  for $n = 1$ *to* N **do**
    Sample $w_1, \ldots, w_d$, independent uniform on $[0, 1]$;
    Set $x_{n,1} = v_{1,1} = w_1$;
    Sample $d - 1$ dependent variables for the nth simulation via;
    for $i = 2$ *to* d **do**
      Set $v_{i,1} = w_i$;
      for $k = i - 1$ *to* 1 **do**
        Set $v_{i,1} = h^{-1}(v_{i,1}, v_{k,k}, \theta_{k,i-k})$
      end
      $x_{n,i} = v_{i,1}$;
      if $i = d$ **then**
        STOP
      end
      Compute the required conditional distribution functions needed for sampling the next $(i + 1)$th variable by using the h-function recursively with the previously determined conditional distribution functions $v_{i,j} = F(x_{n,i}|x_{n,1}, \ldots, x_{n,j-1})$ with respect to the set of parameters $\theta_{j,i}$ of the corresponding copula density $c_{j,j+i|1,\ldots,j-1}(\cdot)$;
      for $j = 1$ *to* $i - 1$ **do**
        Set $v_{i,j+1} = h(v_{i,j}, v_{j,j}, \theta_{j,i-j})$
      end
    end
  end
end

B.2 Simulation algorithm for a D-vine

Algorithm 4: Simulation algorithm for a D-vine

Input: N, d, fam, $\boldsymbol{\theta}$

Let N be the number of simulations and d the dimension of the vine. If the underlying vine is parametrically specified fam denotes the set of $d(d-1)/2$ bivariate parametric copula families defining the PCC structure and $\boldsymbol{\theta}$ the vector of the related parameters. In case of a Bernstein PCC fam denotes $d(d-1)/2$ nonparametric Bernstein copulas and $\boldsymbol{\theta}$ contains the corresponding contingency tables.

Output: x

x is an array of simulations of size $N \times d$ from the specified D-vine.

begin
- **for** $n = 1$ *to* N **do**
  - Sample $w_1, \ldots, w_d$, independent uniform on $[0, 1]$;
  - Set $x_{n,1} = v_{1,1} = w_1$;
  - Set $x_{n,2} = v_{2,1} = h^{-1}(w_2, v_{1,1}, \theta_{1,1})$;
  - Set $v_{2,2} = h^{-1}(v_{1,1}, v_{2,1}, \theta_{1,1})$;
  - Sample $d-1$ dependent variables for the nth simulation via;
  - **for** $i = 3$ *to* d **do**
    - Set $v_{i,1} = w_i$;
    - **for** $k = i-1$ *to* 2 **do**
      - Set $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,2k-2}, \theta_{k,i-k})$
    - **end**
    - Set $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,1}, \theta_{1,i-1})$;
    - Set $x_{n,i} = v_{i,1}$;
    - **if** $i = d$ **then**
      - STOP
    - **end**
    - Compute the required conditional distribution functions needed for sampling the next $(i+1)$th variable by using the h-function recursively with the previously determined conditional distribution functions $v_{i,j} = F(x_{n,i}|x_{n,1}, \ldots, x_{n,j-1})$ with respect to the set of parameters $\theta_{j,i}$ of the corresponding copula density $c_{i,i+j|i+1,\ldots,i+j-1}(\cdot)$;
    - Set $v_{i,2} = h(v_{i-1,1}, v_{i,1}, \theta_{1,i-1})$;
    - Set $v_{i,3} = h(v_{i,1}, v_{i-1,1}, \theta_{1,i-1})$;
    - **if** $i > 3$ **then**
      - **for** $j = 2$ *to* $i-2$ **do**
        - Set $v_{i,2j} = h(v_{i-1,2j-2}, v_{i,2j-1}, \theta_{j,i-j})$ Set $v_{i,2j+1} = h(v_{i,2j-1}, v_{i-1,2j-2}, \theta_{j,i-j})$
      - **end**
    - **end**
    - Set $v_{i,2i-2} = h(v_{i-1,2i-4}, v_{i,2i-3}, \theta_{i-1,1})$
  - **end**
- **end**

end

B.3 Fitting a canonical vine copula

Algorithm 5: Fitting a canonical vine copula

Input: $class$, $data$, fam

Let $class$ be a binary coded variable indicating either the model should be fit parametrically (= 1) or nonparametrically (= 0). $data$ is a $n \times d$ sample being the basis for performing the model fit. If $class = 1$, fam denotes the set of candidate bivariate parametric copula families. For $class = 0$ fam denotes the nonparametric Bernstein copula.

Output: PCC

PCC is an array of $d(d-1)/2$ bivariate fitted (parametric or nonparametric) copulas and a set $\boldsymbol{\theta}$ of related parameters ($class = 0$) or contingency tables ($class = 1$).

```
begin
    k = 1;
    Start at the first tree;
    for edge i = 1 to d − 1 do
        Set y = [data[, 1], data[, i + 1]];
        Obtain the best fitting bivariate copula PCC_k for edge i by inserting y in bifit(·);
        k = k + 1;
        Set v_{1,i} := h(data[, i], data[, 1], θ_{1,i});
    end
    for trees j = 2 to d − 1 do
        for edge i = 1 to d − j do
            Set y = [v_{j−1,1}, v_{j−1,i+1}];
            Obtain the best fitting bivariate copula PCC_k by inserting y in bifit(·);
            k = k + 1;
            if j < d − 1 then
                Set v_{j,i} := h(v_{j−1,i+1}, v_{j−1,1}, θ_{j,i})
            end
        end
    end
end
```

B.4 Fitting a D-vine copula

Algorithm 6: Fitting a D-vine copula

Input: $class$, $data$, fam

Let $class$ be a binary coded variable indicating either the model should be fit parametrically (= 1) or nonparametrically (= 0). $data$ is a $n \times d$ sample being the basis for performing the model fit. If $class = 1$, fam denotes the set of candidate bivariate parametric copula families. For $class = 0$ fam denotes the nonparametric Bernstein copula.

Output: PCC

PCC is an array of $d(d-1)/2$ bivariate fitted (parametric or nonparametric) copulas and a set $\boldsymbol{\theta}$ of related parameters ($class = 0$) or contingency tables ($class = 1$).

begin
- $k = 1$;
- Start at the first tree;
- **for** $i = 1$ *to* $d - 1$ **do**
  - Set $\mathbf{y} = [data[, i], data[, i + 1]]$;
  - Obtain the best fitting bivariate copula PCC_k for edge i by inserting $\mathbf{y}$ in $bifit(\cdot)$;
  - $k = k + 1$;
  - Set $v_{1,1} := h(data[, 1], data[, 2], \theta_{1,1})$;Set $v_{1,2d-4} := h(data[, d], data[, d - 1], \theta_{1,d-1})$;
  - **if** $d > 3$ **then**
    - Set $v_{1,2i} := h(data[, i + 2], data[, i + 1], \theta_{1,i+1})$;
    - Set $v_{1,2i+1} := h(data[, i + 1], data[, i + 2], \theta_{1,i+1})$;
  - **end**
- **end**
- **for** *trees* $j = 2$ *to* $d - 1$ **do**
  - **for** $k = 1$ *to* $d - j$ **do**
    - Set $\mathbf{y} = [v_{j-1,2i-1}, v_{j-1,2i}]$;
    - Obtain the best fitting bivariate copula PCC_k by inserting $\mathbf{y}$ in $bifit(\cdot)$;
    - $k = k + 1$;
    - Set $v_{j,1} := h(v_{j-1,1}, v_{j-1,2}, \theta_{j,1})$;
    - Set $v_{j,2d-2j-2} := h(v_{j-1,2d-2j}, v_{j-1,2d-2j-1}, \theta_{j,d-j})$;
    - **if** $d > 4$ *and* $d - j - 2 > 0$ **then**
      - **for** $i = 1$ *to* $d - j - 2$ **do** /* ... */
        - Set $v_{j,2i} := h(v_{j-1,2i+2}, v_{j-1,2i+1}, \theta_{j,i+1})$;
        - Set $v_{j,2i+1} := h(v_{j-1,2i+1}, v_{j-1,2i+2}, \theta_{j,i+1})$;
      - **end**
    - **end**
  - **end**
- **end**

end

Appendix C

Supplementary Material for Chapter 5

C.1 Sample Banking Firms

Table A: **Identifiers of analyzed banks.**

The table presents the banks investigated in the course of our study as well as the corresponding acronyms and the Google data frequency (d=daily, w=weekly, na=not available), respectively.

#	ID	Bank	Country	Google data
1	WFC	Wells Fargo	USA	d
2	JPMC	JP Morgan Chase	USA	d
3	HSBC	HSBC	GBR	d
4	BoA	Bank of America	USA	d
5	Citi	Citigroup	USA	d
6	BoC	Bank of China	CHN	d
7	CBA	Com. Bank of Australia	AUS	d
8	BSD	Banco Santander	ESP	d
9	BNP	BNP Paribas	FRA	d
10	RBC	Royal Bank of Canada	CAN	d
11	LBG	Lloyds Banking Group	GBR	d
12	NAB	National Australia Bank	AUS	d
13	BNS	Bank of Nova Scotia	CAN	d
14	ITUB	Itau Unibanco	BRA	d
15	BBVA	BBVA Banco	ESP	d
16	Barc	Barclays	GBR	d
17	MS	Morgan Stanley	USA	d
18	BBDC	Banco Bradesco	BRA	d
19	NDB	Nordea Bank	SWE	d
20	UC	Unicredit	ITA	d
21	STD	Standard Chartered	GBR	d
22	IES	Intesa Sanpaolo	ITA	d
23	SG	Société Générale	FRA	d
24	DB	Deutsche Bank	DEU	d
25	COF	Capital One Financial	USA	d
26	BMO	Bank of Montreal	CAN	d
27	BNY	Bank of New York Mellon	USA	d
28	CASA	Credit Agricole	FRA	d
29	SCHW	Charles Schwab	USA	d

Table A: **Identifiers of analyzed banks (continued).**

#	ID	Bank	Country	Google data
30	ICBC	Industrial and Commercial Bank of China	CHN	w
31	CCB	China Construction Bank	CHN	w
32	ABC	Agricultural Bank of China	CHN	w
33	AIB	Allied Irish Banks	ROI	w
34	WBC	Westpac Banking Corporation	AUS	w
35	TD	Toronto-Dominion Bank	CAN	na
36	ANZ	Australia and New Zealand Banking Group	AUS	na
37	MUFG	Mitsubishi UFJ Financial Group	JPN	na
38	USB	US Bancorp	USA	w
39	UBS	United Bank of Switzerland	CHE	na
40	GS	Goldman Sachs Group	USA	w
41	SMFG	Sumitomo Mitsui Financial Group	JPN	na
42	RBS	Royal Bank of Scotland Group	GBR	w
43	CS	Credit Suisse Group	CHE	w
44	SBER	Sberbank of Russia	RUS	na
45	MHFG	Mizuho Financial Group	JPN	na
46	PNC	PNC Financial Services Group	USA	na
47	BoCom	Bank of Communications	CHN	w
48	CMB	China Merchants Bank	CHN	w
49	CMBC	China Minsheng Banking	CHN	na
50	CCB	China Citic Bank	CHN	w

Bibliography

Aas, K. and D. Berg (2009): "Models for construction of multivariate dependence - a comparison study," *The European Journal of Finance*, 15, 639–659.

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009): "Pair-Copula Constructions of Multiple Dependence," *Insurance: Mathematics and Economics*, 44, 182–198.

Albers, W., ed. (1981): *Handwörterbuch der Wirtschaftswissenschaft*, vol. 6, Stuttgart: Fischer.

Albrecht, P. and R. Maurer (2008): *Investment- und Risikomanagement*, Stuttgart: Schäffer-Pöschel, 3rd ed.

Arrow, K. J. (1965): *Aspects of the Theory of Risk-Bearing*, Helsinki: Yrjo Jahnsson Saatio.

——— (1971): *Essays in the Theory of Risk-Bearing*, Amsterdam: North-Holland.

Ausín, M. C. and H. F. Lopes (2010): "Time-varying joint distribution through copulas," *Computational Statistics & Data Analysis*, 54, 2383–2399.

Awartani, B. M. A. and V. Corradi (2010): "Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries," *International Journal of Forecasting*, 21, 167–183.

Baglioni, A. and U. Cherubini (2013): "Within and between systemic country risk. Theory and evidence from the sovereign crisis in Europe," *Journal of Economic Dynamics and Control*, 37, 1581–1597.

Barber, B. M., T. Odean, and N. Zhu (2009): "Do Retail Trades Move Markets?" *Review of Financial Studies*, 22, 151–186.

Basel Committee on Banking Supervision (1996): *Supervisory framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements*, Tech. rep.: Bank of International Settlements.

Bedford, T. and R. M. Cooke (2001): "Probability density decomposition for conditionally dependent random variables modeled by vines," *Annals of Mathematics and Artificial Intelligence*, 32, 245–268.

——— (2002): "Vines: A New Graphical Model for Dependent Random Variables," *The Annals of Statistics*, 30, 1031–1068.

Bellman, R. (1957): *Dynamic Programming*, Princeton (New Jersey): Princeton University Press.

Benninga, S., R. Eldor, and I. Zilcha (1983): "Optimal hedging in the futures market under price uncertainty," *Economics Letters*, 13, 141–145.

Berkowitz, J., P. Christoffersen, and D. Pelletier (2011): "Evaluating Value-at-Risk models with desk-level data," *Management Science*, 57, 2213–2227.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T. and R. F. Engle (1993): "Common Persistence in Conditional Variances," *Econometrica*, 61, 167–186.

Bollerslev, T. and V. Todorov (2011): "Tails, Fears, and Risk Premia," *The Journal of Finance*, 66, 2165–2211.

Box, G. E. P. and G. M. Jenkins (1970): *Time Series Analysis - Forecasting and Control*, Holden Day.

Brailsford, T. J. and R. W. Faff (1996): "An evaluation of volatility forecasting techniques," *Journal of Banking & Finance*, 20, 419–438.

Brechmann, E. C. and C. Czado (2013): "Risk management with high-dimensional vine copulas: An analysis of the Euro Stoxx 50," *Statistics & Risk Modeling*, 30, 307–342.

Brechmann, E. C., C. Czado, and K. Aas (2012): "Truncated Regular Vines in High Dimensions with Application to Financial Data," *The Canadian Journal of Statistics*, 40, 68–85.

Brechmann, E. C. and U. Schepsmeier (2013): "Modeling Dependence with C- and D-Vine Copulas," *Journal of Statistical Software*, 52, 1–27.

Breeden, D. T. (1984): "Futures Markets and Commodity Options: Hedging and Optimality in Incomplete Markets," *Journal of Economic Theory*, 32, 275–300.

Breymann, W., A. Dias, and P. Embrechts (2003): "Dependence structures for multivariate high-frequency data in finance," *Quantitative Finance*, 3, 1–14.

Briys, E., M. Crouhy, and H. Schlesinger (1990): "Optimal Hedging under Intertemporally Dependent Preferences," *The Journal of Finance*, 45, 1315–1324.

——— (1993): "Optimal hedging in a futures market with background noise and basis risk," *European Economic Review*, 37, 949–960.

Briys, E. and H. Schlesinger (1993): "Optimal Hedging When Preferences Are State Dependent," *The Journal of Futures Markets*, 13, 441–451.

Briys, E. and B. Solnik (1992): "Optimal currency hedge ratios and interest rate risk," *Journal of International Money and Finance*, 11, 431–445.

Broll, U., K. W. Chow, and K. P. Wong (2001): "Hedging and nonlinear risk exposure," *Oxford Economic Papers*, 53, 281–296.

Broll, U., E. Clark, and E. Lukas (2010): "Hedging mean-reverting commodities," *IMA Journal of Management Mathematics*, 21, 19–26.

Broll, U. and J. E. Wahl (2012): *Risikomanagement im Unternehmen*, Wiesbaden: Springer Gabler.

Broll, U. and K. P. Wong (2002): "Optimal full-hedging under state-dependent preferences," *The Quarterly Review of Economics and Finance*, 42, 937–943.

Brown, S., W. Goetzmann, B. Liang, and C. Schwarz (2009): "Estimating Operational Risk for Hedge Funds: The ω-Score," *Financial Analysts Journal*, 65, 43–53.

Chan, J. C. C. and D. P. Kroese (2010): "Efficient estimation of large portfolio loss probabilities in t-copula models," *European Journal of Operational Research*, 205, 361–367.

Chang, F.-R. (2004): *Stochastic optimization in continuous time*, Cambridge: Cambridge University Press.

CHERUBINI, U., E. LUCIANO, AND W. VECCHIATO (2004): *Copula Methods in Finance*, Wiley.

CHIANG, A. C. (1984): *Fundamental Methods of Mathematical Economics*, New York: McGraw-Hill, 3rd ed.

——— (1992): *Elements of Dynamic Optimization*, New York: McGraw-Hill.

CHOLLETE, L., A. HEINEN, AND A. VALDESOGO (2009): "Modeling International Financial Returns with a Multivariate Regime-Switching Copula," *Journal of Financial Econometrics*, 7, 437–480.

CHRISTIAANS, T. (2004): *Neoklassische Wachstumstheorie*, Nordersted: Books on Demand.

CHRISTIE, A. A. (1982): "The stochastic behavior of common stock variance: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 10, 407–432.

CHRISTOFFERSEN, P. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

CHRISTOFFERSEN, P., V. R. ERRUNZA, K. JACOBS, AND H. LANGLOIS (2012): "Is the Potential for International Diversification Disappearing? A Dynamic Copula Approach," *Review of Financial Studies*, 25, 3711–3751.

CHRISTOFFERSEN, P., K. JACOBS, X. JIN, AND H. LANGLOIS (2013): "Dynamic Dependence in Corporate Credit," Working Paper.

CHRISTOFFERSEN, P. AND D. PELLETIER (2004): "Backtesting Value-at-Risk: A Duration-Based Approach," *Journal of Financial Econometrics*, 2, 84–108.

COX, J., J. INGERSOLL, AND R. S. (1985): "A theory of the term structure of interest rates," *Econometrica*, 53, 385–467.

CZADO, C. (2010): "Pair-copula constructions of multivariate copulas," in *Copula Theory and its Applications*, Springer, 93–109.

CZADO, C., S. JESKE, AND M. HOFMANN (2013): "Selection strategies for regular vine copulae," *Journal of the French Statistical Society*, 154, 174–191.

DA, Z., J. ENGELBERG, AND P. GAO (2011): "In Search of Attention," *The Journal of Finance*, 66, 1461–1499.

——— (2015): "The Sum of All FEARS: Investor Sentiment and Asset Prices," *Review of Financial Studies*, 28, 1–32.

Deheuvels, P. (1979): "La fonction de dépendance empirique et ses propriétés - Un test non paramétrique d'indépendance," *Académie Royale de Belgique - Bulletin de la Classe des Sciences - 5e Série*, 65, 274–292.

——— (1981): "An asymptotic decomposition for multivariate distribution-free tests of independence," *Journal of Multivariate Analysis*, 11, 102–113.

Delatte, A.-L. and C. Lopez (2013): "Commodity and equity markets: Some stylized facts from a copula approach," *Journal of Banking & Finance*, 37, 5346–5356.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum likelihood from incomplete via the EM algorithm," *Journal of the Royal Statistical Society. Series B*, 39, 1–38.

Dias, A. and P. Embrechts (2009): "Testing for structural changes in exchange rates' dependence beyond linear correlation," *The European Journal of Finance*, 15, 619–637.

Diers, D., M. Eling, and S. D. Marek (2012): "Dependence modeling in non-life insurance using the Bernstein copula," *Insurance: Mathematics and Economics*, 50, 430–436.

Dimpfl, T. and S. Jank (forthcoming): "Can Internet search Queries help to predict stock market volatility?" *European Financial Management*.

Dissmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013): "Selecting and Estimating Regular Vine Copulae and Application to Financial Returns," *Computational Statistics & Data Analysis*, 59, 52–69.

Dixit, A. K. (1990): *Optimization In Economic Theory*, Oxford: Oxford University Press, 2nd ed.

Dixit, A. K. and R. S. Pindyck (1994): *Investment under Uncertainty*, Princeton et al.: Princeton University Press.

Duffie, D. (1989): *Futures Markets*, Englewood Cliffs (New Jersey): Prentice-Hall.

Eeckhoudt, L., C. Gollier, and H. Schlesinger (2005): *Economic and financial decisions under risk*, Princeton: Princeton University Press.

EMBRECHTS, P. (2009): "Copulas: A Personal View," *The Journal of Risk and Insurance*, 76, 639–650.

EMBRECHTS, P., A. J. MCNEIL, AND D. STRAUMANN (2002): "Correlation and dependence in risk management: properties and pitfalls," in *Risk Management: Value at Risk and Beyond*, ed. by M. A. H. Dempster, Cambridge University Press, 176–223.

ENGLE, R. F. AND T. BOLLERSLEV (1986): "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

ENGLE, R. F. AND V. K. NG (1993): "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

ENGLE, R. F. AND A. J. PATTON (2001): "What good is a volatility model?" *Quantitative Finance*, 1, 237–245.

ENGLE, R. F. AND J. R. RUSSELL (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

ESCANCIANO, J. C. AND P. PEI (2012): "Pitfalls in backtesting Historical Simulation VaR models," *Journal of Banking & Finance*, 36, 2233–2244.

FAMA, E. F. (1965): "The Behavior of Stock-Market Prices," *Journal of Business*, 38, 34–105.

FANTAZZINI, D. (2009): "The Effects of Misspecified Marginals and Copulas on Computing the Value at Risk: A Monte Carlo Study," *Computational Statistics and Data Analysis*, 53, 2168–2188.

FARIA, J. R. AND P. MCADAM (2013): "Anticipation of Future Consumption: A Monetary Perspective," *Journal of Money, Credit and Banking*, 45, 423–446.

FERNANDEZ, C. AND M. F. J. STEEL (1998): "On Bayesian Modeling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93, 359–371.

FISCHER, M., C. KLÖCK, S. SCHLÜTER, AND F. WEIGERT (2009): "An empirical analysis of multivariate copula models," *Quantitative Finance*, 9, 839–854.

FRANKE, G. AND H. HAX (2009): *Finanzwirtschaft des Unternehmens und Kapitalmarkt*, Berlin: Springer.

FRENKEL, M. AND H.-R. HEMMER (1999): *Grundlagen der Wachstumstheorie*, München: Vahlen.

Genest, C., M. Gendron, and M. Bourdeau-Brien (2009a): "The advent of copulas in finance," *The European Journal of Finance*, 15, 609–618.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995): "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, 82, 543–552.

Genest, C., B. Rémillard, and D. Beaudoin (2009b): "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and Economics*, 44, 199–213.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Goldfarb, D. and A. Idnani (1982): "Dual and Primal-Dual Methods for Solving Strictly Convex Quadratic Programs," in *Numerical Analysis: Proceedings of the third IIMAS Workshop Held at Cocoyoc, Mexico, January 1981*, ed. by J. P. Hennard, Berlin: Springer, 226–239.

Gollier, C. (2001): *The economics of risk and time*, Cambridge: MIT Press.

Grønneberg, S. and N. L. Hjort (2014): "The Copula Information Criteria," *Scandinavian Journal of Statistics*, 41, 436–459.

Grundke, P. and S. Polle (2012): "Crisis and risk dependencies," *European Journal of Operational Research*, 223, 518–528.

Hafner, C. M. and O. Reznikova (2010): "Efficient estimation of a semiparametric dynamic copula model," *Computational Statistics & Data Analysis*, 54, 2609–2627.

Hamid, A. and M. Heiden (2014): "Forecasting Volatility with Empirical Similarity and Google Trends," Working Paper.

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

Heinen, A. and A. Valdesogo (2009): "Asymmetric CAPM dependence for large dimensions: the Canonical Vine Autoregressive Model," CORE Discussion Papers 2009/69.

Ho, T. S. Y. (1984): "Intertemporal Commodity Futures Hedging and the Production Decision," *The Journal of Finance*, 39, 351 – 376.

Hobæk-Haff, I. (2013): "Parameter estimation for pair-copula constructions," *Bernoulli*, 19, 462–491.

Hobæk-Haff, I., K. Aas, and A. Frigessi (2010): "On the simplified pair-copula construction - Simply useful or too simplistic?" *Journal of Multivariate Analysis*, 101, 1296–1310.

Hobæk-Haff, I. and J. Segers (2012): "Nonparametric estimation of pair-copula constructions with the empirical pair-copula," Working Paper.

Hu, L. (2006): "Dependence patterns across financial markets: A mixed copula approach," *Applied Financial Economics*, 717–729.

Huang, C.-F. and R. H. Litzenberger (1988): *Foundations for Financial Economics*, New York: Prentice-Hall.

Ince, O. S. and R. B. Porter (2006): "Individual Equity Return Data From Thomson Datastream: Handle With Care!" *Journal of Financial Research*, 29, 463–479.

Ingersoll, J. E. (1987): *Theory of Financial Decision Making*, Savage: Rowman & Littlefield.

Itô, K. (1951): "On stochastic differential equations," *Memoirs of American Mathematical Society*, 4, 1–51.

Joe, H. (1996): "Families of m-variate distributions with given margins and m(m-1)/2 bi-variate dependence parameters," in *Distributions with Fixed Marginals and Related Topics*, ed. by L. Rüschendorf, B. Schweizer, and M. Taylor, Institute of Mathematical Statistics, 120–141.

——— (1997): *Multivariate Models and Dependence Concepts*, Chapman & Hall.

Johnson, L. L. (1960): "The Theory of Hedging and Speculation in Commodity Futures," *The Review of Economic Studies*, 27, 139–151.

Jondeau, E. and M. Rockinger (2006): "The Copula-GARCH model of conditional dependencies: An international stock market application," *Journal of International Money and Finance*, 25, 827–853.

Keynes, J. M. (1930): *A Treatise on Money, Vol. II: The Applied Theory of Money*, London: Macmillan.

Kim, D., J.-M. Kim, S.-M. Liao, and Y.-S. Jung (2013): "Mixture of D-vine copulas for modeling dependence," *Computational Statistics and Data Analysis*, 64, 1–19.

Kim, G., M. J. Silvapulle, and P. Silvapulle (2007): "Comparison of semiparametric and parametric methods for estimating copulas," *Computational Statistics & Data Analysis*, 51, 2836–2850.

Kim, S.-H. and D. Kim (2014): "Investor sentiment from internet message postings and the predictability of stock returns," *Journal of Economic Behavior & Organization*, 107, 708–729.

Kole, E., K. Koedijk, and M. Verbeek (2007): "Selecting copulas for risk management," *Journal of Banking & Finance*, 31, 405–423.

Kulpa, T. (1999): "On approximation of copulas," *International Journal of Mathematics and Mathematical Sciences*, 22, 259–269.

Kurowicka, D. (2011): "Optimal truncation of vines," in *Dependence Modeling: Vine Copula Handbook*, ed. by D. Kurowicka and H. Joe, New Jersey et al.: World Scientific Publishing Co., 233–248.

Lee, G. J. and R. F. Engle (1999): "A permanent and transitory component model of stock return volatility," in *Cointegration Causality and Forecasting A Festschrift in Honor of Clive W. J. Granger*, Oxford University Press, 475–497.

Li, D. X. (2000): "On Default Correlation: A Copula Function Approach," *The Journal of Fixed Income*, 9, 43–54.

Li, Q., J. B. Brown, H. Huang, and P. J. Bickel (2011): "Measuring Reproducibility of High-Throughput Experiments," *The Annals of Applied Statistics*, 5, 1752–1779.

Liu, Y. and R. Luger (2009): "Efficient estimation of copula-GARCH models," *Computational Statistics and Data Analysis*, 53, 2284–2297.

Mandelbrot, B. B. (1963): "The Variation of Certain Speculative Prices," *The Journal of Business*, 36, 394–419.

Manner, H. (2007): "Estimation and model selection of copulas with an application to exchange rates," *METEOR research memorandum, Maastricht University*, 07/056.

Markowitz, H. M. (1991): *Portfolio selection efficient diversification of investments*, Cambridge et al.: Blackwell.

McLachlan, G. and D. Peel (2000): *Finite Mixture Models*, Wiley.

Mendes, B. V. d. M., M. M. Semeraro, and R. P. C. Leal (2010): "Pair-copulas modeling in finance," *Financial Markets and Portfolio Management*, 24, 193–213.

Merton, R. C. (1969): "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *The Review of Economics and Statistics*, 51, 247–257.

——— (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–887.

Min, A. and C. Czado (2010): "Bayesian Inference for Multivariate Copulas using Pair-copula Constructions," *Journal of Financial Econometrics*, 8, 511–546.

——— (2011): "Bayesian model selection for D-vine pair-copula constructions," *Canadian Journal of Statistics*, 39, 239–258.

Mondria, J. and T. Wu (2011): "Asymmetric Attention and Stock Returns," *AFA 2012 Chicago Meetings Paper*.

Morales-Napoles, O., R. M. Cooke, and D. Kurowicka (2010): "About the number of vines and regular vines on n nodes," Working Paper.

Munk, C. (2013): *Financial Asset Pricing Theory*, Oxford: Oxford University Press.

Nelsen, R. B. (2006): *An Introduction to Copulas*, Springer, 2nd ed.

Nelson, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Nikoloulopoulos, A. K., H. Joe, and H. Li (2012): "Vine copulas with asymmetric tail dependence and applications to financial return data," *Computational Statistics & Data Analysis*, 56, 3659–3673.

Ning, C. (2010): "Dependence structure between the equity market and the foreign exchange market – A copula approach," *Journal of International Money & Finance*, 29, 743–759.

Oh, D. H. and A. J. Patton (2013): "Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads," Economic Research Initiatives at Duke (ERID) Working Paper No. 167.

Pfeifer, D., D. Strassburger, and J. Philipps (2009): "Modelling and simulation of dependence structures in nonlife insurance with Bernstein copulas," Working Paper, Carl von Ossietzky University, Oldenburg.

Pindyck, R. S. and D. L. Rubinfeld (1998): *Econometric models and economic forecasts*, vol. 4, Boston: Irwin McGraw-Hill.

——— (2011): *Mikroökonomie*, vol. 7 (Nachdruck), München: Pearson Studium.

Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko (1962): *The Mathematical Theory of Optimal Processes*, New York: John Wiley & Sons.

Poon, S.-H., M. Rockinger, and J. Tawn (2004): "Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications," *The Review of Financial Studies*, 17, 581–610.

Pratt, J. W. (1964): "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 122–136.

Ramsey, F. P. (1928): "A Mathematical Theory of Saving," *The Economic Journal*, 38, 543–559.

Rodriguez, J. C. (2007): "Measuring Financial Contagion: A Copula Approach," *Journal of Empirical Finance*, 14, 401–423.

Ross, S. A. (1981): "Some Stronger Measures of Risk Aversion in the Small and the Large with Applications," *Econometrica*, 49, 621–638.

Rüschendorf, L. (2013): *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios*, Heidelberg: Springer.

Ruenzi, S. and F. Weigert (2013): "Crash Sensitivity and the Cross-Section of Expected Stock Returns," Working Paper.

Samuelson, P. A. (1969): "Lifetime Portfolio Selection By Dynamic Stochastic Programming," *The Review of Economics and Statistics*, 51, 239–246.

Sancetta, A. and S. Satchell (2004): "The Bernstein copula and its applications to modeling and approximations of multivariate distributions," *Econometric Theory*, 20, 535–562.

Saunders, A. (2002): *Financial Institutions Management*, Boston: Irwin/McGraw-Hill.

Sharpe, W. F. (1966): "Mutual Fund Performance," *The Journal of Business*, 39, 119–138.

——— (1975): "Adjusting for Risk in Portfolio Performance Measurement," *The Journal of Portfolio Management*, 1, 29–34.

——— (1994): "The Sharpe Ratio," *Journal of Portfolio Management*, 21, 49–58.

Shen, X., Y. Zhu, and L. Song (2008): "Linear B-spline copulas with applications to nonparametric estimation of copulas," *Computational Statistics & Data Analysis*, 52, 3806–3819.

Siganos, A., E. Vagenas-Nanos, and P. Verwijmeren (2014): "Facebook's daily sentiment and international stock markets," *Journal of Economic Behavior & Organization*, 107, 730–743.

Stulz, R. M. (1984): "Optimal Hedging Policies," *Journal of Financial and Quantitative Analysis*, 19, 127–140.

Tomann, H. (2005): *Volkswirtschaftslehre*, Heidelberg: Physica-Verlag.

Varian, H. R. (2010): *Intermediate Microeconomics*, New York: Norton.

Vlastakis, N. and R. N. Markellos (2012): "Information demand and stock market volatility," *Journal of Banking & Finance*, 36, 1808–1821.

Vozlyublennaia, N. (2014): "Investor attention, index performance, and return predictability," *Journal of Banking & Finance*, 41, 17–35.

Wahl, J. E. and U. Broll (2003): "Hedging-Effektivität und Risikogestaltung," *Controlling*, 12, 689–693.

Weiss, G. (2011): "Are Copula-GoF-tests of any practical use? Empirical evidence for stocks, commodities and FX futures," *The Quarterly Review of Economics and Finance*, 51, 173–188.

Weiss, G. N. F. (2013): "Copula-GARCH vs. Dynamic Conditional Correlation - an empirical study on VaR and ES forecasting accuracy," *Review of Quantitative Finance and Accounting*, 41, 179–202.

Weiss, G. N. F. and M. Scheffer (2012): "Smooth Nonparametric Bernstein Vine Copulas," Working Paper.

——— (2015): "Mixture Pair-Copula-Constructions," *Journal of Banking & Finance*, 54, 175–191.

Weiss, G. N. F. and H. Supper (2013): "Forecasting Liquidity-Adjusted Intraday Value-at-Risk with Vine Copulas," *Journal of Banking & Finance*, 37, 3334–3350.

Whelan, N. (2004): "Sampling from Archimedean copulas," *Quantitative Finance*, 4, 339–352.

Working, H. (1953): "Futures Trading and Hedging," *The American Economic Review*, 43, 314–343.

Wu, S. (2014): "Construction of asymmetric copulas and its application in two-dimensional reliability modelling," *European Journal of Operational Research*, 238, 476–485.

Ye, W., X. Liu, and B. Miao (2012): "Measuring the subprime crisis contagion: Evidence of change point analysis of copula functions," *European Journal of Operational Research*, 222, 96–103.

Eidesstattliche Versicherung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig verfasst habe und mich ausschließlich der angegebenen Hilfsmittel bedient habe. Die Dissertation ist nicht bereits Gegenstand eines erfolgreich abgeschlossenen Promotions- oder sonstigen Prüfungsverfahrens gewesen.

Ratingen, 23. Juli 2015