A copula-based approach to model serial dependence in financial time series

Fabian Spanhel

A copula-based approach to model serial dependence in financial time series

Fabian Spanhel

Dissertation an der
Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität München

vorgelegt von
Fabian Spanhel

München, den 18. Februar 2015

Erstgutachter: Prof. Stefan Mittnik, PhD

Zweitgutachter: Prof. Dr. Yarema Okhrin

Tag der Disputation: 27. März 2015

Abstract

This thesis is devoted to the development and application of a copula-based approach to model serial dependence in univariate financial time series.

We show that the stationarity and the order of a Markov process can be easily determined by its SD-vine copula, which is a particular regular vine copula, and that this copula has the unique property that only p bivariate (conditional) copulas are required to represent every stationary p-th order Markov process. To tackle the curse of dimensions in practical applications, we propose to model a stationary p-th order Markov process by a simplified $(p+1)$-dimensional SD-vine copula. The resulting finite-order Markov process can be considered as a natural generalization of an autoregressive Gaussian process. Contrary to established methods, where some features of the transition distribution are modeled, the transition distribution of the Markov process is derived from its SD-vine copula and its marginal distribution, giving rise to flexible transition distributions that can not be expressed by location and scale models. We introduce methods that allow for a parsimonious representation of an SD-vine copula-based stationary Markov process with long memory and provide a detailed analysis of its dependence properties. An application to time series of price durations demonstrates a strong superiority of the copula-based approach to the popular class of ACD models.

In order to model financial returns, an understanding of their stylized facts in terms of copulas is required. For this reason, another substantial contribution of this thesis is the development of a theory of copulas that can be used to model volatility clustering. In this regard, we introduce two important dependence properties of a bivariate copula, one of which is sufficient for a positive correlation between squared or absolute symmetric random variables and an increasing conditional variance in the absolute value of the conditioning variable. Moreover, several construction methods of copulas with the desired dependence properties are presented and compared. An application of the copula-based time series model to the returns of three major stock indices and one currency exchange rate documents the competitiveness of our approach with established GARCH models.

To the best of our knowledge, this thesis provides the first evidence for the practical usefulness of a copula-based approach to model time series of price durations and financial returns.

Zusammenfassung

Diese Dissertation widmet sich der Entwicklung und Anwendung eines Copula-basierten Ansatzes, um serielle Abhängigkeit in univariaten Finanzzeitreihen zu modellieren.

Es wird gezeigt, dass die Stationarität und die Ordnung eines Markov-Prozesses sich auf einfache Weise durch seine SD-Vine-Copula feststellen lassen, und dass diese spezielle reguläre Vine-Copula die charakterisierende Eigenschaft aufweist, dass nur p bivariate (bedingte) Copulas benötigt werden, um jeden stationären Markov-Prozess p-ter Ordnung darzustellen. Um den Fluch der Dimensionen in praktischen Anwendungen zu umgehen, schlagen wir vor, einen stationären Markov-Prozess p-ter Ordnung mittels einer vereinfachten $(p+1)$-dimensionalen SD-Vine-Copula zu modellieren. Der resultierende Zeitreihenprozess kann als Verallgemeinerung von autoregressiven Gaußschen Prozessen verstanden werden. Im Gegensatz zu etablierten Methoden, die nur einige Merkmale der Übergangsverteilung abbilden, wird die Übergangsverteilung des Prozesses aus seiner SD-Vine-Copula und seiner Randverteilung hergeleitet, was flexible Übergangsverteilungen ermöglicht, die nicht durch Lokations- und Skalenmodelle dargestellt werden können. Es werden Methoden vorgestellt, die eine sparsame Parametrisierung eines SD-Vine-Copula-basierten Markov-Prozesses mit langem Gedächtnis ermöglichen, und die Abhängigkeitseigenschaften des Prozesses werden ausführlich untersucht. Eine Anwendung unseres Ansatzes auf Zeitreihen von Preisdurationen demonstriert eine starke Überlegenheit gegenüber den populären ACD Modellen.

Für die Modellierung von Finanzrenditen ist ein Verständnis ihrer stilisierten Fakten im Sinne von Copulas erforderlich. Aus diesem Grund ist ein weiterer wesentlicher Beitrag dieser Dissertation die Entwicklung einer Theorie der Copulas, die zur Modellierung von Volatilitätsclustern verwendet werden kann. In diesem Zusammenhang werden zwei wichtige Abhängigkeitseigenschaften von bivariaten Copulas eingeführt, eine davon ist hinreichend für eine positive Korrelation zwischen quadrierten oder absoluten symmetrischen Zufallsvariablen und eine bedingte Varianz welche im Absolutwert der konditionierenden Variable monoton steigend ist. Zudem werden zahlreiche Konstruktionsmethoden von Copulas mit den gewünschten Abhängigkeitseigenschaften präsentiert und verglichen. Eine Anwendung des Copula-basierten Zeitreihenmodells auf die Renditen dreier bedeutender Aktienindizes und einem Währungswechselkurs dokumentiert die Wettbewerbsfähigkeit unseres Ansatzes gegenüber etablierten GARCH Modellen.

Diese Dissertation liefert nach unserem Kenntnisstand die ersten Belege für den praktischen Nutzen einer Copula-basierten Modellierung von Preisdurationen und Finanzrenditen.

Contents

List of Figures

List of Tables

1 Introduction

In the last decade, copulas have been frequently applied to model the cross-sectional dependence between several random variables, i.e., the (conditional) dependence of various random variables at the same data point or time period (see Patton, 2009, 2012; Chen and Fan, 2006a, and the references therein). On the other side, the number of applications that utilize copulas to model serial dependence is still very small, although the basic idea of modeling and characterizing time series with copulas is not new. Darsow et al. (1992) is the first study that characterizes a first-order Markov process by placing conditions on the finite-dimensional copulas of the process, followed by a section in the monograph of Joe (1997), in which several dependence properties of copula-based first-order Markov processes are derived. The most frequently cited study of Chen and Fan (2006b), which focuses on the asymptotic distribution of copula-based first-order Markov processes, has given rise to a number of studies which investigated the theory and application of copula-based time series models. At the moment, the statistical inference for univariate copula-based first-order Markov processes is well understood and, quite recently, regular vine copulas have been proposed to model multivariate copula-based Markov processes. However, so far there is no empirical study which documents that copula-based approaches to time series analysis are competitive with established time series models. In fact, whenever a comparison is drawn, the performance of copula-based time series models turns out to be inferior (see the literature review in Section 3.1.3). Viewed in this light, it appears that Thomas Mikosch, theoretical time series analyst and copula critic, is right in claiming that "copulas completely fail in describing complex space-time dependence structures" and "are not useful for modeling dependence through time" (Mikosch, 2006, p. 18-19).

One reason for the poor performance of copula-based time series models so far is that, until recently, the application of copulas was mainly restricted to the modeling of bivariate dependence, with elliptical copulas constituting the only class that could describe realistic multivariate relations. Bivariate copulas can be used to construct first-order Markov models, but time series processes typically exhibit a more complex and persistent serial dependence. Elliptical copulas can represent fairly flexible cross-sectional dependence structures, but, except for the Gaussian copula, are inappropriate for modeling serial dependence, because there is no finite-order Markov process whose finite-dimensional copulas are elliptical but not Gaussian. Obviously, a copula-based approach to time series analysis stands and falls with a suitable framework of modeling multivariate copulas in a flexible way. The recent introduction of (regular) vine copulas (Bedford and Cooke, 2002; Aas et al., 2009) has been an enormous advance for high-dimensional dependence modeling and greatly extends the number of flexible and useful multivariate copulas. Moreover, simplified vine copula models, which are constructed upon a sequence of bivariate unconditional copulas, offer the means of overcoming the curse of dimension and still allow the modeling of complex dependencies.

This thesis is concerned with the development of a copula-based approach to model serial dependence in univariate time series that is competitive with established methods. To this end, we investigate whether the recent introduction of vine copulas has also opened the doors for a successful application of copula-based time series models. We show that the stationarity and the Markov order of a univariate time series process can be easily characterized by its stationary D-vine copula (SD-vine copula), which is a particular regular vine copula, and that this copula has the unique property that only p bivariate (conditional) copulas are required to represent every stationary p-th order Markov process. To tackle the curse of dimension in practical applications, we propose to model a stationary p-th order Markov process by a simplified $(p+1)$-dimensional SD-vine copula. The resulting finite-order Markov process can be considered as a natural generalization of an autoregressive Gaussian process. However, contrary to established methods, where some features of the transition distribution are modeled, the transition distribution of the Markov process is derived from its SD-vine copula and its marginal distribution, resulting in flexible transition distributions that can not be expressed by location and scale models. We introduce methods that allow for a parsimonious representation of an SD-vine copula-based stationary Markov process with long memory and provide a detailed analysis of its dependence properties. An application to time series of price durations demonstrates the advantages of our framework and reveals a strong superiority of the copula-based approach to the popular class of ACD models in terms of in-sample fit and different out-of-sample criteria.

Although the SD-vine copula gives rise to flexible time series models, we point out that neither martingale difference sequences, nor processes that exhibit volatility clustering, can be modeled if the SD-vine copula is composed of commonly used parametric families of bivariate copulas. We show that a necessary condition for a martingale difference sequence is that all bivariate copulas of the SD-vine copula are not strictly quadrant dependent (QD). Since commonly used parametric copulas are not strictly QD only if they collapse to the product copula, it follows that the only martingale difference sequence that can be constructed is a sequence of iid random variables. We derive sufficient and necessary conditions for a conditionally symmetric martingale difference sequence in terms of copulas. However, martingale difference sequences do not necessarily exhibit volatility clustering. In order to model financial returns, an understanding of their stylized facts in terms of copulas is required. For this reason, another substantial contribution of this thesis is the development of a theory for copulas that can be used to model volatility clustering. In this regard, we introduce two important dependence properties of a bivariate copula that are sufficient for a positive correlation between squared or absolute symmetric random variables and an increasing transition variance in the absolute value of one conditioning variable. Moreover, several construction methods of copulas with the desired dependence properties are presented and compared. An application of the copula-based time series model to returns of three major stock indices and one currency exchange rate documents the competitiveness of our approach with established GARCH models and that copula-based models can successfully model financial returns.

This thesis is structured as follows. Chapter 2 discusses some preliminaries. A short introduction to D-vine copulas is given and some key definitions and results from Spanhel and Kurz (2015), who investigate approximations based on the simplifying assumption, are presented. Moreover, we generalize the concept of quadrant dependence to bivariate conditional distributions and extend Hoeffding's lemma.

In Chapter 3 we introduce a copula-based approach to model serial dependence in time series and elucidate the differences from the classical approach. We review the literature that is concerned with copula-based time series models and point out open research questions that are addressed in this thesis. Stationarity and the Markov property are characterized in terms of copulas. We define the SD-vine copula and establish its unique properties among the class of regular vine copulas, e.g., only p bivariate (conditional) copulas are required to represent every univariate stationary p-th order Markov process. Moreover, the partial autocopula sequence, a non-linear generalization of the partial autocorrelation function that corrects for considerably more dependence caused by intermediate variables, is introduced. Algorithms for specifying and estimating a simplified SD-vine copula-based Markov model are presented and it is illustrated how exploratory data analysis can be utilized to construct a higher-order Markov process by a sequence of bivariate copulas. Since financial time series are often described by models that exhibit a rather large or even an infinite Markov order, we impose a structure on the serial dependence of the copula-based process. We model the SD-vine copula by a lag function, which is parameterized by a low-dimensional parameter vector, in order to obtain a parsimonious representation of a higher-order Markov process. Although a lag function allows for the modeling of a copula-based Markov process with arbitrary memory, it is preferable to choose the smallest possible order in applications, because the number of computations increases quadratically with the order of the process. We propose a new criterion, which is based on the mutual information, to truncate the order and to balance the trade-off between a reasonable computational time and a loss in the goodness-of-fit. In addition, the modeling of the marginal distribution of the copula-based process is addressed and the literature on goodness-of-fit tests for the marginal distribution of dependent data is reviewed. Finally, various dependence properties of SD-vine copula-based higher-order Markov processes, such as monotonicity of the conditional expectation or conditions for a non-negative autocorrelation function, are investigated. Moreover, we comment on the mixing properties of SD-vine copula-based Markov processes.

An application of stationary SD-vine copula-based Markov models to time series of price duration data is conducted in Chapter 4. We use transaction data of five blue chip stocks and evaluate the goodness-of-fit of augmented ACD models in terms of classical measures as well as from a copula perspective. Although the ACD models are successful in removing the autocorrelation structure of price durations, simple transformations of the residuals show that a great amount of non-linear dependence is still present in the filtered durations for all five time series. Copula scatter plots of consecutive durations clearly reveal that the ACD models do not capture the clusters of short durations which are prevalent in the data. We explain how exploratory data analysis can be utilized to set up a simplified SD-vine copula-based Markov model which captures these features. As a result, the copula-based models clearly outperform the ACD models wrt the AIC. Moreover, we find that the transition distributions of the copula-based models exhibit a strongly time-varying dispersion. Since ACD models exhibit a time-constant dispersion, this further explains the superior performance of our approach. The copula-based models also clearly outperform the ACD models in an out-of-sample experiment. The results of the Diebold-Mariano type tests of superior out-of-sample specification of the transition distribution strongly favor the CMP models. An analysis of the logarithmic scores suggest that one reason for this is that the copula-based models perform better in forecasting clusters of short durations.

Chapter 5 is concerned with the development of a theory of copula-based processes that are compatible with well known stylized facts of financial returns. In particular, we investigate to what extent stylized facts of daily asset returns, such as the martingale difference property, volatility clustering and the leverage effect, can be characterized in terms of copulas. By the introduction of vertically symmetric copulas, we obtain necessary and sufficient conditions for a conditionally symmetric martingale difference sequence. These conditions also reveal that the only possible martingale difference sequence that can be generated by commonly used copula families, excluding the Student-t copula with zero correlation parameter, is a sequence of independent random variables. We then focus on the construction of copula-based first-order Markov processes that exhibit volatility clustering. For this purpose, we define two important dependence properties of a bivariate copula. Among other things, the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property implies a non-negative correlation between squared or absolute symmetric random variables. The stronger $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property is sufficient for a first-order Markov process to exhibit a non-negative autocorrelation function of squared and absolute symmetric random variables and an increasing transition variance. We show that the Student-t copula with zero correlation parameter has the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property and that a first-order Markov process with this copula can be represented as a transformed ARCH(1) process. In addition, three different methods to construct parametric copulas with the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property are proposed. The first two methods, the merged and patched X-shaped version of a copula, are based on transformations of established copula families, while the third method extracts the copula of a bivariate distribution that is inspired by the GJR-ARCH(1) process. Finally, we explore the construction of higher-order Markov processes that exhibit volatility clustering. We advocate the use of an SD-vine copula that is composed of copulas that exhibit at least the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property, so that the correlation between squared or absolute random variables, conditional on intermediate random variables, is non-negative.

In Chapter 6, we investigate and compare the fit of the copulas that are proposed in Chapter 5 for modeling clusters of extreme observations. We find that the constructed copulas provide a good fit to financial returns of three major stock indices and one currency exchange rate. In order to evaluate the performance and properties of the copula-based model, we draw a comparison with the GARCH model and illustrate the similarities and differences between both approaches. In terms of log-likelihood and AIC, both model classes perform equally well, with the copula-based model being slightly superior for one stock index. Despite their similar in-sample performance, we find substantial differences in the transition distributions of both model classes during periods of high volatility. While the trajectories of the conditional volatilities of both model classes are very similar in calm market periods, the conditional volatility of the copula-based model is more strongly varying in turbulent market periods. Moreover, the shape of the transition density can be more complex than under the assumption of a location and scale model. Even if the effects of volatility have been accounted for, we detect for two stock indices a great variation in the transition distribution of the copula-based model, giving rise to a bimodal or peaked transition density during turbulent market periods. For the other two time series, the copula-based model supports the assumption that the transition distribution of financial returns is determined by its volatility. The out-of-sample evaluation provides some evidence that the transition distribution of the copula-based model is closer to the true transition distribution in one out of four cases, and otherwise comparable to the GARCH model. Chapter 7 summarizes our main results and discusses directions for future research.

In summary, the main novel contributions of this thesis are as follows. We provide a thorough analysis of univariate stationary higher-order Markov processes in terms of copulas, with a focus on regular vine copulas, and develop a copula-based approach for modeling univariate time series. In order to model Markov processes with long memory, we introduce methods that allow for a parsimonious representation of an SD-vine copula-based stationary Markov processes. In addition, a detailed analysis of the dependence properties of SD-vine copula-based stationary Markov processes is provided, which also contributes to the understanding of dependence in regular vine copula models in general. We develop a theory for copulas that are suited to model volatility clusters and thereby establish the fundament for modeling financial returns with copula-based processes. Two applications show that our proposed framework can be competitive with or may even clearly outperform highly specialized time series models. To the best of our knowledge, this thesis is the first document that demonstrates the practical usefulness of a copula-based approach to model time series of price durations and financial returns.

Software

The computations in this thesis were conducted using the MATLAB programming language (MathWorks, R2014a) and its optimization toolbox. All necessary routines for the estimation of copula-based time series models and ACD models were written by the author. The estimation of GARCH models was done with Kevin Sheppard's MFE Toolbox (`https://www.kevinsheppard.com/MFE_Toolbox`).

2 Preliminary remarks and definitions

In this chapter we give a short introduction to D-vine copulas and present some key definitions and results from Spanhel and Kurz (2015), who investigate approximations based on the simplifying assumption. Moreover, we generalize the concept of quadrant dependence to bivariate conditional distributions and extend Hoeffding's lemma. We assume that the reader is familiar with the concept of copulas, see Joe (1997) or Nelsen (2006) for an excellent introduction to this topic.

Remarks on notation and assumptions

Throughout this thesis, $Y_{1:K} = (Y_1, \ldots, Y_K)$ refers to a K-dimensional random vector with cumulative distribution function (cdf) $F_{1:K}$, if not otherwise indicated. $U_{1:K}$ denotes in general the vector of probability integral transforms, i.e., $U_i = F_i(Y_i)$ for all $i = 1, \ldots, K$, with cdf $C_{1:K} = F_{U_{1:K}} = F^u_{1:K}$. Depending on the context, $F_{1:K}(y_{1:K})$ refers to the K-dimensional list $F_1(y_1), \ldots, F_K(y_K)$ or to the scalar $\mathbb{P}(Y_i \leq y_i, \forall i = 1, \ldots, K)$. For instance, if $F_{1:K}(y_{1:K})$ constitutes the input of a K-dimensional function, then it refers to the K-dimensional list $F_1(y_1), \ldots, F_K(y_K)$, i.e., $C_{1:K}(F_{1:K}(y_{1:K})) := \mathbb{P}(F_i(Y_i) \leq F_i(y_i), \forall i = 1, \ldots, K)$.

For simplicity, we assume that all random vectors are absolutely continuous wrt to the Lebesgue measure and have a strictly positive density so that conditional quantiles are given by the (almost surely) unique inverse of the corresponding conditional distribution function.

2.1 Simplified vine copula approximations and partial copulas

In this section we recall the definition of conditional probability integral transforms and the conditional copula. We also give a short introduction to D-vine copulas and the simplifying assumption. Information on regular vines can be found in Bedford and Cooke (2002) and Kurowicka and Joe (2011). Finally we summarize some key definitions and results from Spanhel and Kurz (2015) who investigate approximations based on the simplifying assumption. In particular, we define partial probability integral transforms and (higher-order) partial copulas which are crucial for understanding the partial autocopula sequence that is introduced in Chapter 3.

Definition 2.1 (Conditional probability integral transform (CPIT))

Let $X_{1:K}$ be an absolutely continuous K-dimensional random vector and $1 \leq j \leq K$. Let $K\backslash j := \{1, \ldots, K\}\backslash j$, i.e., $X_{K\backslash j}$ is the vector of random variables when X_j is being excluded from $X_{1:K}$. We call $F_{X_j|X_{K\backslash j}}(X_j|X_{K\backslash j})$ the conditional probability integral transform of X_j wrt to $X_{K\backslash j}$.

It holds that CPITs are uniformly distributed and $F_{X_j|X_{K\backslash j}}(X_j|X_{K\backslash j}) \perp X_{K\backslash j}$ for all $j = 1, \ldots, K$. Thus, applying the random transformation $F_{X_j|X_{K\backslash j}}(\cdot|X_{K\backslash j})$ to X_j removes possible dependencies between X_j and $X_{K\backslash j}$. The conditional joint distribution of CPITs gives rise to the definition of the conditional copula which has been introduced by Patton (2006).

Definition 2.2 (Bivariate conditional copula – Patton (2006))

Let (Y_1, Y_2) be a bivariate random vector and Z be a K-dimensional random vector with $K \geq 1$. Define the CPITs $U_1(Z) = F_{Y_1|Z}(Y_1|Z)$ and $U_2(Z) = F_{Y_2|Z}(Y_2|Z)$. For each $Z = z$ there exists an (a.s.) unique conditional copula $C_{Y_1,Y_2|Z}(\cdot, \cdot|z)$ for the conditional distribution $F_{Y_1,Y_2|Z}(\cdot, \cdot|z)$, which is defined as the conditional distribution $F_{U_1(Z),U_2(Z)|Z}(\cdot, \cdot|z)$. More precisely, it holds (a.s.) that

$$C_{Y_1,Y_2|Z}(a, b|z) := P(U_1(Z) \leq a, U_2(Z) \leq b|Z = z) = F_{Y_1,Y_2|Z}(F^{-1}_{Y_1|Z}(a|z), F^{-1}_{Y_2|Z}(b|z)|z).$$

Using Patton's definition of a conditional copula one can show that each copula density can be represented by the density of a D-vine copula.

Definition 2.3 (Density of a D-vine copula – Kurowicka and Cooke (2006))

Let $U_{1:K}$, $K \geq 3$, be a random vector with uniform marginals and cdf $F_{U_{1:K}} = F_{1:K} = C_{1:K}$, $(i, j) \in \mathcal{I}_K := \{(i, j)\colon j = 1, \ldots, K-1, i = 1, \ldots, K-j\}$ and $s(i, j) := i+1 : i+j-1$. A D-vine copula decomposes the density of $U_{1:K}$ into $K(K+1)/2$ bivariate conditional copula densities $c_{i,i+j|s(i,j)}$ according to the following factorization:

$$c_{1:K}(u_{1:K}) = \prod_{(i,j)\in\mathcal{I}_K} c_{i,i+j|s(i,j)}\big(F_{i|s(i,j)}(u_i|u_{s(i,j)}), F_{i+j|s(i,j)}(u_{i+j}|u_{s(i,j)})|u_{s(i,j)}\big),$$

where $c_{i,i+j|s(i,j)} = c_{i,i+1}$ for $j = 1$.

From a graph-theoretic point of view, a D-vine copula can be considered as an ordered sequence of trees where j refers to the number of the tree. In general, a bivariate conditional copula $C_{i,i+j|s(i,j)}$ with $j-1$ conditioning arguments must be assigned to each edge of the j-th tree for $j = 1, \ldots, K-1$, in order to represent the density of every K-dimensional copula $C_{1:K}$ by a D-vine copula. However, in some special cases, bivariate unconditional copulas can be assigned to the edges of the tree if $C_{1:K}$ satisfies the simplifying assumption.

Definition 2.4 (The simplifying assumption – Hobæk Haff et al. (2010))

The D-vine copula in Definition 2.3 is of the simplified form if $c_{i,i+j|s(i,j)}(\cdot,\cdot|u_{s(i,j)})$ does not depend on $u_{s(i,j)}$ for all $(i,j)\in\mathcal{I}_K$.

The validity of the simplifying assumption is known to be true for the multivariate Gaussian, Student-t and Clayton copula. If a copula $\tilde{C}_{1:K}$ satisfies the simplifying assumption, its density collapses to a product of $K(K+1)/2$ (bivariate) unconditional copula densities $\tilde{c}_{i,i+j;\,s(i,j)}$, which is then given by

$$\tilde{c}_{1:K}(u_{1:K}) = \prod_{(i,j)\in\mathcal{I}_K} \tilde{c}_{i,i+j;\,s(i,j)}\big(\tilde{F}_{i|s(i,j)}(u_i|u_{s(i,j)}), \tilde{F}_{i+j|s(i,j)}(u_{i+j}|u_{s(i,j)})\big), \tag{2.1.1}$$

where for $k = i, i+j$, $l = i + \mathbb{1}_{\{k=i+j\}}$,

$$\tilde{F}_{k|s(i,j)}(u_k|u_{s(i,j)}) = \partial_{2+i-l}\tilde{C}_{l,l+j-1;\,s(l,j-1)}\big(\tilde{F}_{l|s(l,j-1)}(u_l|u_{s(l,j-1)}), \tilde{F}_{l+j-1|s(l,j-1)}(u_{l+j-1}|u_{s(l,j-1)})\big),$$

i.e., the conditional cdfs can be evaluated using a recursive scheme of partial derivatives of bivariate unconditional copulas.

Due to their general form, D-vine copulas do not give rise to a feasible model framework without further assumptions. Therefore, the simplifying assumption is often used to obtain an approximation of a non-simplified vine copula by means of a simplified vine copula model.

Definition 2.5 (Simplified vine copula approximation)

Let $\mathcal{C}_K$ be the space of K-dimensional absolutely continuous copulas and $\mathfrak{C}_{1:K}$ be the space of K-dimensional absolutely continuous D-vine copulas which satisfy the simplifying assumption. If $\tilde{C}_{1:K} \in \mathfrak{C}_{1:K}$ is used to approximate $C_{1:K} \in \mathcal{C}_K$, we call $\tilde{C}_{1:K}$ a simplified vine copula approximation of $C_{1:K}$.

Although the bivariate unconditional copula $\tilde{C}_{i,i+j;\,s(i,j)}$ in (2.1.1) is only a function of two variables, whereas the bivariate conditional copula $C_{i,i+j|s(i,j)}$ in Definition 2.3 is in general a function of $j+1$ variables, the space of simplified D-vine copula models is still very rich since each unconditional copula $\tilde{C}_{i,i+j;\,s(i,j)}$ can be chosen arbitrarily and the resulting function is always a valid K-dimensional copula. Moreover, a simplified vine copula model does not suffer from the curse of dimension since it is build upon a sequence of bivariate unconditional copulas which renders it very attractive for high-dimensional applications.

Several questions arise if one uses a simplified D-vine copula model given in (2.1.1) to approximate a general D-vine copula stated in Definition 2.3. For instance, it is not obvious what bivariate copulas $\tilde{C}_{i,i+j;\,s(i,j)}$ should be specified or estimated to obtain the best approximation wrt to a certain criterion. Spanhel and Kurz (2015) provide a detailed analysis of this issues. The simplified vine copula approximation that minimizes the KL divergence from the true copula in a tree-by-tree fashion is given by the partial vine copula approximation. The partial vine copula approximation is the simplified vine copula approximation that results if the bivariate conditional copulas in the vine copula of the data generating process are approximated by higher-order partial copulas.

Definition 2.6 (Higher-order partial copulas and partial probability integral transforms)

Assume $K \geq 3$ and let $C_{1:K}$ be the copula of $U_{1:K}$. Define

$$C^{\partial}_{i,i+2;\,i+1}(a,b) = P(U_{i|i+1} \leq a, U_{i+2|i+1} \leq b),$$

for all $i = 1, \ldots, K-1$. Consider the corresponding D-vine copula stated in Definition 2.3. In the first tree, we set for $i = 1, \ldots, K-1$: $C^{\partial 0}_{i,i+1} = C_{i,i+1}$, while in the second tree, we denote for $i = 1, \ldots, K-2, k = i, i+2$: $C^{\partial 1}_{i,i+2;\,i+1} = C^{\partial}_{i,i+2;\,i+1}$ and $U^{\partial 0}_{k|i+1} = U_{k|i+1} = F_{k|i+1}(U_k|U_{i+1})$. In the remaining trees $j = 3, \ldots, K-1$, for $i = 1, \ldots, K-j$, we set $k = i, i+j$, $l = i + \mathbb{1}_{\{k=i+j\}}$, and define

$$U^{\partial j-2}_{k|s(i,j)} = F^{\partial j-2}_{k|s(i,j)}(U_k|U_{s(i,j)}) = \partial_{2+i-l} C^{\partial j-2}_{l,l+j-1;\,s(l,j-1)}(U^{\partial j-3}_{l|s(l,j-1)}, U^{\partial j-3}_{l+j-1|s(l,j-1)}),$$

and

$$C^{\partial j-1}_{i,i+j;\,s(i,j)}(a,b) = \mathbb{P}(U^{\partial j-2}_{i|s(i,j)} \leq a, U^{\partial j-2}_{i+j|s(i,j)} \leq b).$$

For $j = 2, \ldots, K-1, i = 1, \ldots, K-j$, and all $u_{s(i,j)} \in [0,1]^{j-1}$, we call $U^{\partial j-2}_{k|s(i,j)}$ the $(j-2)$-th order **partial probability integral transform (PPIT)** of U_k wrt to $U_{s(i,j)}$ and $C^{\partial j-1}_{i,i+j;\,s(i,j)}$ the $(j-1)$-th order **partial copula** of $F_{i,i+j|s(i,j)}(\cdot,\cdot|u_{s(i,j)})$ that is induced by the D-vine copula given in Definition 2.3.

Proposition 2.1 (Tree-by-tree KL divergence minimization using higher-order partial copulas – Spanhel and Kurz (2015))

Let $K \geq 3$ and $C_{1:K} \in \mathcal{C}_K$ be the true copula and $j, J = 1, \ldots, K-1$. We define $\mathcal{T}^{\partial j-1}_j := (C^{\partial j-1}_{i,i+j;\,s(i,j)})_{i=1,\ldots,K-j}$, so that $\mathcal{T}^{\partial}_{1:J} := \times_{j=1}^{J} T^{\partial j-1}_j$ collects all higher-order partial copulas up to and including the J-th tree. Moreover, let

$$\mathbf{T}_j := \left\{(\tilde{C}_{i,i+j;\,s(i,j)})_{i=1,\ldots,K-j} \colon \tilde{C}_{i,i+j;\,s(i,j)} \in \mathcal{C}_2 \text{ for all } 1 \leq i \leq K-j\right\},$$

so that $\mathbf{T}_{1:J} = \times_{j=1}^{J}\mathbf{T}_j$ represents all possible simplified vine copula models up to and including the J-th tree. Let $\tilde{\mathcal{T}}_j \in \mathbf{T}_j, \tilde{\mathcal{T}}_{1:j-1} \in \mathbf{T}_{1:j-1}$, so that for a simplified vine copula approximation the KL divergence related to the first tree is given by

$$D^{(1)}_{KL}(\tilde{\mathcal{T}}_1(\tilde{\mathcal{T}}_{1:0})) := \sum_{i=1}^{K-1} \mathbb{E}\left[\log \frac{c_{i,i+1}(U_i, U_{i+1})}{\tilde{c}_{i,i+1}(U_i, U_{i+1})}\right],$$

and for the remaining trees $j = 2, \dots, K-1$, the related KL divergence is

$$D^{(j)}_{KL}(\tilde{\mathcal{T}}_j(\tilde{\mathcal{T}}_{1:j-1})) := \sum_{i=1}^{K-j} \mathbb{E}\left[\log \frac{c_{i,i+j|s(i,j)}(F_{i|s(i,j)}(U_i|U_{s(i,j)}), F_{i+j|s(i,j)}(U_{i+j}|U_{s(i,j)})|U_{s(i,j)})}{\tilde{c}_{i,i+j;\, s(i,j)}(\tilde{F}_{i|s(i,j)}(U_i|U_{s(i,j)}), \tilde{F}_{i+j|s(i,j)}(U_{i+j}|U_{s(i,j)}))}\right],$$

where for $j \geq 2, k = i, i+j, l = i + \mathbb{1}_{\{k=i+j\}}$,

$$\tilde{F}_{k|s(i,j)}(U_k|U_{s(i,j)}) := \partial_{2+i-l}\tilde{C}_{l,l+j-1;\, s(l,j-1)}(\tilde{F}_{l|s(l,j-1)}(U_l|U_{s(l,j-1)}), \tilde{F}_{l+j-1|s(l,j-1)}(U_{l+j-1}|U_{s(l,j-1)})).$$

It holds that

$$\forall J = 1, \dots, K-1 \colon \quad \underset{\mathcal{T}_J \in \mathbf{T}_J}{\arg\min}\ D^{(J)}_{KL}(\mathcal{T}_J(\mathcal{T}^{\partial}_{1:J-1})) = \mathcal{T}^{\partial}_J. {}^{J-1}$$

Definition 2.7 (Partial vine copula approximation (PVCA))

We call the simplified vine copula approximation $\mathcal{T}^{\partial}_{1:K-1}$, defined in Proposition 2.1, the partial vine copula approximation.

Thus, if one sequentially minimizes the related KL divergence in each tree, then the optimal simplified vine copula approximation consists of higher-order partial copulas. However, Spanhel and Kurz (2015) also show that the partial vine copula approximation is in general not the global minimizer of the KL divergence from the true distribution since the effect of the J-th tree $\mathcal{T}_J$ on the KL divergences related to higher trees is ignored when the KL divergence in the J-th tree is minimized. It follows that the step-by-step ML estimator (Hobæk Haff, 2012, 2013) of a parametric vine copula model is not a consistent estimator for the parameters that minimize the KL divergence from the true distribution. For more information on simplified vine copula approximations and higher-order partial copulas we refer to Spanhel and Kurz (2015).

2.2 A generalization of quadrant dependence and Hoeffding's lemma

In this section we generalize the concept of quadrant dependence, which is defined for bivariate distributions, to bivariate conditional distributions and investigate its properties. We also generalize Hoeffding's lemma to bivariate conditional distributions that are derived from absolutely continuous transformations. These tools allow us to derive a necessary condition for a martingale difference sequence in terms of copulas (Proposition 5.1) and to establish a sufficient condition for a positive conditional autocorrelation function of absolute or squared random variables of a symmetric martingale difference sequence (Proposition 5.13). In the following let $Y_{1:K} \sim F_{1:K}$.

Definition 2.8 ((Strict) quadrant dependence of a bivariate (conditional) cdf)

Let $K \geq 2$. We say that a bivariate (conditional) cdf $F_{1K|2:K-1}$ or $(Y_1, Y_K)|Y_{2:K-1}$ is positive quadrant dependent (PQD) if for all $(y_1, y_K) \in \mathbb{R}^2$ and F_{K-2}-almost all $y_{2:K-1} \in \mathbb{R}^{K-2}$

$$F_{1K|2:K-1}(y_1, y_K|y_{2:K-1}) \geq F_{1|2:K-1}(y_1|y_{2:K-1})F_{K|2:K-1}(y_K|y_{2:K-1}). \tag{2.2.1}$$

$F_{1K|2:K-1}$ is negative quadrant dependent (NQD) if the inequality in (2.2.1) is reversed. We say that $F_{1K|2:K-1}$ is quadrant dependent (QD) if $F_{1K|2:K-1}$ is PQD or NQD.

For $K = 2$ we obtain the well known definition that a bivariate cdf is PQD or NQD (Joe, 1997; Balakrishnan and Lai, 2009). To the best of our knowledge, the generalization to bivariate conditional cdfs has not been considered in the literature. For bivariate distributions, the QD property is invariant wrt componentwise increasing transformations (Lehmann, 1966) and a property of the copula if the distribution is absolutely continuous. The next lemma establishes a similar invariance of the QD property for bivariate conditional distributions.

Lemma 2.1 (Invariance of the QD property wrt increasing transformations)

For $i = 1, K$, let $G_i \colon \mathbb{R} \to \mathbb{R}$ be an almost-surely increasing function. For $i = 2, \ldots, K-1$, let $G_i \colon \mathbb{R} \to \mathbb{R}$ be an almost-surely strictly increasing function such that $\partial_1 G_i$ exists. Consider the random vector $Z_{1:K}$ whose elements are given by $Z_i = G_i(Y_i)$ for $i = 1, \ldots, K$. Then $(Z_1, Z_K)|Z_{2:K-1}$ is QD if $(Y_1, Y_K)|Y_{2:K-1}$ is QD.

Proof. See Appendix A.1. ■

Lemma 2.2 (QD as a property of the bivariate (conditional) copula)

An absolutely continuous bivariate conditional distribution $F_{1K|2:K-1}$ is PQD (or NQD) if and only if its bivariate conditional copula is PQD (or NQD).

Proof. See Appendix A.2 ■

In general, it is not possible to compute the covariance of $(Y_1, Y_K)|Y_{2:K-1} = y_{2:K-1}$ if only $C_{1K|2:K-1}$ is known. However, if we know that the conditional copula is PQD (or NQD) we can conclude that the conditional covariance is non-negative (or non-positive). For that purpose, we generalize Hoeffding's lemma to bivariate conditional distributions and absolutely continuous transformations.

Lemma 2.3 (Generalized Hoeffding's lemma for conditional covariances)

For $i = 1, K$, let G_i be absolutely continuous (wrt to the Lebesgue measure) with density g_i such that $\mathbb{E}[G_1(Y_1), G_K(Y_K)|Y_{2:K-1}]$ and $\mathbb{E}[G_i(Y_i)|Y_{2:K-1}]$ exist. Then (a.s.)

$$\begin{aligned}
&\mathbb{E}\big[\prod_{i=1,K} G_i(Y_i)\big|Y_{2:K-1}\big] - \prod_{i=1,K}\mathbb{E}\big[G_i(Y_i)|Y_{2:K-1}\big] \\
&= \int_{\mathbb{R}^2}\prod_{i=1,K} g_i(y_i)\Big(F_{1K|2:K-1}(y_1, y_K|Y_{2:K-1}) - \prod_{i=1,K} F_i(y_i|Y_{2:K-1})\Big)\mathrm{d}y_1\mathrm{d}y_k.
\end{aligned}$$

Proof. See Appendix A.3. ■

Mardia and Thompson (1972) prove Lemma 2.3 if $K = 2$ and $G_i(Y_i) = Y_i^r, r \geq 1$. Beare (2009) gives another generalization of Hoeffding's lemma if $K = 2$ and Y_1 and Y_K are bounded random variables. In both studies, multiple Lebesgue-Stieltjes integration by parts is applied to obtain the result. Our proof does not rely on integration by parts and is inspired by the elegant proof of Hoeffding's lemma given in Lehmann (1966).

Lemma 2.4 (QD and covariance)

Let $G_i, i = 1, \ldots, K$, be defined as in Lemma 2.1, $U_{1:K} \sim C_{1:K}$ and $Z_{1:K} = G_{1:K}(U_{1:K})$. If $C_{1K|2:K-1}$ is PQD and $\mathbb{C}\text{ov}\big[Z_1, Z_K|Z_{2:K-1}\big]$ exists for given $G_{1:K}$ then

$$\mathbb{C}\text{ov}\big[Z_1, Z_K|Z_{2:K-1}\big] \geq 0, \quad \text{(a.s.)}, \tag{2.2.2}$$

with equality if and only if $C_{1K|2:K-1} = C^{\perp}$ (a.s.). If $C_{1K|2:K-1}$ is NQD, the inequality sign in (2.2.2) is reversed.

If for all $G_{1:K}$, such that $\mathbb{C}\text{ov}\big[Z_1, Z_K|Z_{2:K-1}\big]$ exists, we have that

$$\mathbb{C}\text{ov}\big[Z_1, Z_K|Z_{2:K-1}\big] \geq 0, \quad \text{(a.s.)}, \tag{2.2.3}$$

then $C_{1K|2:K-1}$ is PQD. If the inequality sign in (2.2.3) is reversed, then $C_{1K|2:K-1}$ is NDQ.

Proof. See Appendix A.4. ■

Lemma 2.4 states that, irrespective of the marginal distributions of the joint distribution, the conditional covariance of a bivariate conditional distribution is non-negative if the underlying bivariate conditional copula has the PQD property. Moreover, the conditional covariance is zero only if the bivariate conditional copula is the product copula. If the bivariate conditional copula has the QD property but is not the product copula, we say that the bivariate conditional copula has the strict QD property.

Definition 2.9 (Strict QD)

$(Y_1, Y_K)|Y_{2:K-1}$ is strictly QD if $(Y_1, Y_K)|Y_{2:K-1}$ is QD and $C_{1K|2:K-1} \neq C^{\perp}$, (a.s.).

2.3 Appendix

A.1 Proof of Lemma 2.1

Let $G[Y_i] := \{z_i \in \mathbb{R} \mid \exists y_i \in \mathbb{R}\colon z_i = G_i(y_i)\}$ be the image of G_i for $i = 1, K$. We have to show that for all $(z_1, z_K) \in G[Y_1] \times G[Y_K]$

$$\mathbb{P}(Z_i \leq z_i, \forall i = 1, K | Z_{2:K-1}) \geq \prod_{i=1,K} \mathbb{P}(Z_i \leq z_i | Z_{2:K-1}) \quad \text{a.s.}$$

For $i = 1, K$, let z_i be fixed and choose $\tilde{y}_i$ such that $G(\tilde{y}_i) = z_i$. Since G_i is bijective and its derivative exists for all $i = 2, \ldots, K-1$, the bivariate conditional distribution of $(Y_1, Y_K)|Z_{2:K-1}$ is given by

$$\mathbb{P}(Y_i \leq y_i, \forall i = 1, K | Z_{2:K-1}) = F_{Y_1, Y_K | Y_{2:K-1}}\big(y_1, y_K | G^{-1}_{2:K-1}(Z_{2:K-1})\big) \quad \text{a.s.}$$

If $(Y_1, Y_K)|Y_{2:K-1}$ is PQD (for NQD reverse the following two inequalities) this implies that

$$\mathbb{P}(Y_i \leq \tilde{y}_i, \forall i = 1, K | Z_{2:K-1}) \geq \prod_{i=1,K} \mathbb{P}(Y_i \leq \tilde{y}_i | Z_{2:K-1}) \quad \text{a.s.}$$

Since G_i is increasing for $i = 1, K$, it follows that

$$\mathbb{P}(G_i(Y_i) \leq z_i, \forall i = 1, K | Z_{2:K-1}) \geq \prod_{i=1,K} \mathbb{P}(G_i(Y_i) \leq z_i | Z_{2:K-1}) \quad \text{a.s.}$$

A.2 Proof of Lemma 2.2

Wlog we only consider PQD. If $F_{1K|2:K-1}$ is PQD, then for $F_{2:K-2}$-almost all $y_{2:K-1} \in \mathbb{R}^{K-2}$ and for all $(y_1, y_K) \in \mathbb{R}^2$ we have that

$$\begin{aligned} F_{1K|2:K-1}(y_1, y_K | y_{2:K-1}) &\geq \prod_{i=1,K} F_{i|2:K-1}(y_i | y_{2:K-1}), \\ \Leftrightarrow C_{1K|2:K-1}(F_{1|2:K-1}(y_1 | y_{2:K-1}), F_{K|2:K-1}(y_K | y_{2:K-1}) | y_{2:K-1}) &\geq \prod_{i=1,K} F_{i|2:K-1}(y_i | y_{2:K-1}). \end{aligned} \tag{2.3.1}$$

Using $y_i = F^{-1}_{i|2:K-1}(u_{i|2:K-1} | y_{2:K-1})$ it follows from (2.3.1) that for $F_{2:K-2}$-almost all $y_{2:K-1} \in \mathbb{R}^{K-2}$ and all $(u_{i|2:K-1}, u_{K|2:K-1}) \in [0,1]^2$ we have

$$C_{1K|2:K-1}(u_{1|2:K-1}, u_{K|2:K-1} | y_{2:K-1}) \geq \prod_{i=1,K} u_{i|2:K-1}. \tag{2.3.2}$$

If $C_{1K|2:K-1}$ is PQD then (2.3.2) holds for $F_{2:K-2}$-almost all $y_{2:K-1} \in \mathbb{R}^{K-2}$ and all $(u_{i|2:K-1}, u_{K|2:K-1}) \in [0,1]^2$. Using $u_{i|2:K-1} = F_{i|2:K-1}(y_i | y_{2:K-1}), i = 1, K$, we see that

(2.3.1) holds for $F_{2:K-2}$-almost all $y_{2:K-1} \in \mathbb{R}^{K-2}$ and for all $(y_1, y_K) \in \mathbb{R}^2$, thus, $F_{1K|2:K-1}$ is PQD.

A.3 Proof of Lemma 2.3

Let $Y_{1:K}$ and $\tilde{Y}_{1:K}$ have the same distribution $F_{1:K}$ and be stochastically independent. Then (a.s.),

$$
\begin{aligned}
&2\Big(\mathbb{E}\Big[\prod_{i=1,K} G_i(Y_i)\Big|Y_{2:K-1}\Big] - \prod_{i=1,K}\mathbb{E}\big[G_i(Y_i)|Y_{2:K-1}\big]\Big)\\
&= \mathbb{E}\big[\prod_{i=1,K} G_i(Y_i) - G_i(\tilde{Y}_i)|Y_{2:K-1}\big]\\
&= \mathbb{E}\big[\prod_{i=1,K}\int_{\tilde{Y}_i}^{Y_i} g_i(t_i)\mathrm{d}t_i|Y_{2:K-1}\big]\\
&= \mathbb{E}\big[\prod_{i=1,K}\int_{\mathbb{R}} \mathbb{1}_{\{\tilde{Y}_i \le t_i \le Y_i\}} g_i(t_i)\mathrm{d}t_i|Y_{2:K-1}\big]\\
&= \mathbb{E}\big[\prod_{i=1,K}\int_{\mathbb{R}} (\mathbb{1}_{\{\tilde{Y}_i \le t_i\}} - \mathbb{1}_{\{Y_i \le t_i\}}) g_i(t_i)\mathrm{d}t_i|Y_{2:K-1}\big]\\
&= \mathbb{E}\big[\int_{\mathbb{R}^2}\Big(\prod_{i=1,K} (\mathbb{1}_{\{\tilde{Y}_i \le t_i\}} - \mathbb{1}_{\{Y_i \le t_i\}}) g_i(t_i)\Big)\mathrm{d}t_1\mathrm{d}t_K|Y_{2:K-1}\big]\\
&= \int_{\mathbb{R}^2} g_1(t_1)g_K(t_K)\Big(\mathbb{E}\big[\mathbb{1}_{\{\tilde{Y}_1 \le t_1\}}\mathbb{1}_{\{\tilde{Y}_K \le t_K\}}|Y_{2:K-1}\big] + \mathbb{E}\big[\mathbb{1}_{\{Y_1 \le t_1\}}\mathbb{1}_{\{Y_K \le t_K\}}|Y_{2:K-1}\big]\\
&\quad - \mathbb{E}\big[\mathbb{1}_{\{\tilde{Y}_1 \le t_1\}}\mathbb{1}_{\{Y_K \le t_K\}}|Y_{2:K-1}\big] - \mathbb{E}\big[\mathbb{1}_{\{Y_1 \le t_1\}}\mathbb{1}_{\{\tilde{Y}_K \le t_K\}}|Y_{2:K-1}\big]\Big)\mathrm{d}t_1\mathrm{d}t_K\\
&= \int_{\mathbb{R}^2} g_1(t_1)g_K(t_K)\Big(2F_{1K|2:K-1}(t_1, t_K|Y_{2:K-1})\\
&\quad - \mathbb{E}\big[\mathbb{1}_{\{\tilde{Y}_1 \le t_1\}}|Y_{2:K-1}\big]\mathbb{E}\big[\mathbb{1}_{\{Y_K \le t_K\}}|Y_{2:K-1}\big]\\
&\quad - \mathbb{E}\big[\mathbb{1}_{\{Y_1 \le t_1\}}|Y_{2:K-1}\big]\mathbb{E}\big[\mathbb{1}_{\{\tilde{Y}_K \le t_K\}}|Y_{2:K-1}\big]\Big)\mathrm{d}t_1\mathrm{d}t_K\\
&= 2\int_{\mathbb{R}^2} g_1(t_1)g_K(t_K)\Big(F_{1K|2:K-1}(t_1, t_K|Y_{2:K-1}) - \prod_{i=1,K} F_i(t_i|Y_{2:K-1})\Big)\mathrm{d}t_1\mathrm{d}t_K.
\end{aligned}
$$

Provided that $\mathbb{E}[G_1(Y_1), G_K(Y_K)|Y_{2:K-1}]$ and $\mathbb{E}[G_i(Y_i)|Y_{2:K-1}], i = 1, K$, exist, we can take the expectation under the integral sign in the third last equality by Fubinis' theorem. In the second last equality we have used the fact that $Y_1 \perp \tilde{Y}_K$ and $\tilde{Y}_1 \perp Y_K$.

A.4 Proof of Lemma 2.4.

Wlog we only consider PQD here. If $C_{1K|2:K-1}$ is PQD then $(Z_1, Z_K)|Z_{2:K-1}$ is PQD by Lemma 2.2. Using Hoeffding's lemma (Lemma 2.3) we see that $\mathbb{C}\mathrm{ov}\big[Z_1, Z_K|Z_{2:K-1}\big] \ge 0$, (a.s.). If the conditional covariance is almost-surely zero and $(Z_1, Z_K)|Z_{2:K-1}$ is PQD, it follows from Hoeffding's lemma that $\forall (z_1, z_K) \in \mathbb{R}^2$

$$F_{Z_1,Z_K|Z_{2:K-1}}(z_1, z_K|Z_{2:K-1}) = \prod_{i=1,K} F_{Z_i|Z_{2:K-1}}(z_i|Z_{2:K-1}), \quad \text{(a.s.)},$$

which is equivalent to $C_{1K|2:K-1} = C^{\perp}$ (a.s.). If $K = 2$ we also have that $C_{1K|2:K-1} = C^{\perp}$ because copulas are continuous from the right so that if two copulas agree almost everywhere they also must be identical.

Assume now that $\mathbb{C}\text{ov}\big[Z_1, Z_K|Z_{2:K-1}\big] \geq 0$ (a.s.), for all $G_{1:K}$. For all $(u_1, u_K) \in [0,1]^2$ define $G_i(U_i) = \mathbb{1}_{\{U_i \geq u_i\}}, i = 1, K$. Then

$$\mathbb{C}\text{ov}\big[Z_1, Z_K|Z_{2:K-1}\big] = \mathbb{P}(U_i \geq u_i, \forall i = 1, K|U_{2:K-1}) - \prod_{i=1,K} \mathbb{P}(U_i \geq u_i|U_{2:K-1}) \geq 0 \quad \text{(a.s.)}$$

for all $(u_1, u_K) \in [0,1]^2$. This is equivalent to $(U_1, U_K)|U_{2:K-1}$ being PQD and from Lemma 2.2 we deduce that $C_{1K|2:K-1}$ is PQD.

3 A copula-based approach to model serial dependence in time series

In this chapter, we introduce and investigate a copula-based approach to model serial dependence in time series. Section 3.1 is devoted to a general discussion of a copula-based approach to time series analysis and elucidates the differences from the classical approach to time series analysis. We also review literature that is related to the copula-based modeling of serial dependence and point out open research questions that are addressed in this thesis.

Since the assumption of stationarity imposes constraints on a multivariate copula, we analyze in Section 3.2 which copulas are suited for modeling stationary processes. We find that there is one unique regular vine copula, the so called SD-vine copula, which exhibits desirable properties in this context. In Section 3.3 we discuss copula-based characterizations of the Markov property and demonstrate that the Markov property can be easily characterized for the SD-vine copula. Moreover, we show that the SD-vine copula has the unique property that only p bivariate conditional copulas are required to construct every stationary p-th order Markov process.

In order to model a p-th order Markov process in practice, we propose in Section 3.4. to use a simplified SD-vine copula to model the copula of $p+1$ adjacent random variables. In this way, the conditional autocopula sequence of a Markov process is determined by a sequence of bivariate unconditional copulas. This greatly simplifies the modeling of time series and also allows the effective use of exploratory data analysis to construct a higher-order Markov model. Since financial time series are often modeled by time series models that exhibit a rather large or even an infinite Markov order, we impose in Section 3.5 a structure on the copula sequence of the SD-vine copula. By this means we obtain a parsimonious representation of a higher-order Markov process. Since the modeling of the marginal distribution is a crucial issue in our framework, we review in Section 3.6 the scarce literature on goodness-of-fit tests for the marginal distribution of dependent data and investigate Neyman's smooth test in more detail. In Section 3.7 we analyze various dependence properties of SD-vine copula-based Markov processes. Our results are summarized in Section 3.8.

3.1 Introducing a copula-based approach to model serial dependence in time series

In this section, we introduce a copula-based approach to model serial dependence. First, we provide a review and discussion of the classical approach to time series analysis which is based on the specification of the transition distribution or some of its features. Then, we develop a copula-based approach to time series analysis and investigate its advantages and downsides. In the following, let Y be a (strictly) stationary stochastic process with parameter space $\mathbb{N}$ and state space $\mathcal{S}$ which is a connected subset of $\mathbb{R}$. Moreover, we assume that all finite-dimensional distributions are absolutely continuous wrt to the Lebesgue measure and have a strictly positive density.[1]

3.1.1 A critique of the classical approach to time series analysis

Time series analysis is typically based on the specification of the transition distribution of a stochastic process Y or on the modeling of some features of the transition distribution.[2] To fix ideas, let $F_{t|t-1:1}$ be the transition distribution of Y at time t, i.e.,

$$F_{t|t-1:1} \colon \mathbb{R}^t \to [0,1], \quad F_{t|t-1:1}(y_t|y_{t-1:1}) = \mathbb{P}(Y_t \leq y_t | Y_{t-1:1} = y_{t-1:1}).$$

Moreover, let $\mathcal{F}_{t|t-1:1}$ be the space of transition distributions at point t, and define $G_t \colon \mathcal{F}_{t|t-1:1} \to \bigtimes_{i=1}^{K} F_i$, where F_i is a function space and $K \in \mathbb{N}$. For instance, G_t could be the identity function for all t, i.e., $G_t(F_{t|t-1:1}(y_t|y_{t-1:1})) = F_{t|t-1:1}(y_t|y_{t-1:1})$, so that the complete transition distribution is specified. If only the mean and variance of the transition distribution are of interest then $G_t(F_{t|t-1:1}(y_t|y_{t-1:1})) = (\mathbb{E}[Y_t|Y_{t-1:1} = y_{t-1:1}], \mathbb{V}\mathrm{ar}[Y_t|Y_{t-1:1} = y_{t-1:1}])$ for all t. In the classical approach to time series analysis, a stochastic process Y is then modeled by the sequence $(G_t \circ F_{t|t-1:1})_{t\in\mathbb{N}}$, which we call the transition representation of Y.[3]

Advantages

The underlying rationale for the use of the transition representation is that we are often interested in the transition distribution or some of its features in practical applications. For instance, if the objective is to forecast the next realization, then the mean of the transition distribution is the functional that minimizes the expected value of the squares of the forecast errors. For that purpose, the direct modeling of the transition distribution is convenient. In order to derive the properties of a time series process, the equivalent stochastic representation of the process Y is often used, which is given by

$$Y_t = \mu(Y_{t-1:1}) + \sigma(Y_{t-1:1})\mathcal{E}_t, \quad \mathbb{E}[\mathcal{E}_t|Y_{t-1:1}] = 0, \ \mathbb{V}\mathrm{ar}[\mathcal{E}_t|Y_{t-1:1}] = 1, \tag{3.1.1}$$

[1] As a result, any (conditional) cdf is strictly increasing and the corresponding (conditional) quantile function is given by the usual inverse of the cdf.

[2] The critique in this section focuses on the most popular models of classical time series analysis. In particular, we assume that the transition distribution of the process is a function of the natural filtration that is associated to the process.

[3] Note that if G_t is the identity function for all t, the transition representation also characterizes all finite-dimensional distributions of Y.

so that $\mu(Y_{t-1:1}) = \mathbb{E}[Y_t|Y_{t-1:1}]$, and $\sigma(Y_{t-1:1})^2 = \sigma^2(Y_{t-1:1}) = \mathbb{V}\mathrm{ar}[Y_t|Y_{t-1:1}]$, provided these expressions exist. In this light, classical time series analysis specifies the functions μ, σ, and an error sequence $\mathcal{E} = (\mathcal{E}_t)_{t\in\mathbb{N}}$ to model the transition distribution and thus the stochastic process. If $\mathcal{E} = (\mathcal{E}_t)_{t\in\mathbb{N}}$ is such that $\mathcal{E}_t \perp Y_{t-1:1}$ for all t, then (3.1.1) becomes a location-scale model, i.e., the transition distribution is completely determined by its location μ, scale σ, and the marginal distribution of the error. However, in the general case the dependence structure of the error sequence has to be modeled.

The classical approach of using the transition representation for modeling time series provides many advantages. First of all, it is not necessary to specify the complete transition distribution if the interest lies in some features of the transition distribution. For instance, if we are just interested in the location of the transition distribution we can just set up the model

$$Y_t = \mu(Y_{t-1:1}) + \mathcal{E}_t, \quad \mathbb{E}[\mathcal{E}_t|Y_{t-1:1}] = 0. \tag{3.1.2}$$

In general, $\mathcal{E}$ is not a sequence of independent random variables but a martingale difference sequence wrt to Y. Consequently, we have a loss in estimation efficiency if we do not model the complete transition distribution. This is the price to pay if we treat the dependence structure of $\mathcal{E}$ as a nuisance parameter. In return, the analysis is greatly simplified and, under some assumptions on $\mathcal{E}$, the time series model (3.1.2) provides the means to obtain a consistent estimator for the location of the transition distribution, provided that the functional shape of the location is correctly specified (see Bollerslev and Wooldridge (1992) for an overview on quasi-maximum likelihood estimators). It is also straight forward to impose moment restrictions on the transition distribution. For instance, if we assume that the time series is generated by a martingale difference sequence, then we can simply set $\mu(Y_{t-1:1}) = 0$ in (3.1.1) for all t to satisfy this assumption. Obviously, the location and scale of the transition distribution have closed form expressions whenever we assume closed form expressions for μ and σ. Moreover, the (partial) autocorrelation function of the process is also tractable if the functional shape of μ is simple enough.

Drawbacks

On the other side, one can argue that using the transition representation as model approach has some drawbacks. In the following, we take the position that the classical approach is primarily based on model properties rather than on exploratory data analysis. We also illustrate that the focus on the reproduction of the autocorrelation function is disputable, since its interpretation and importance as a measure of dependence is not clear in the general case. Moreover, we point out that useful properties of the marginal distribution, which are often inspected by kernel density estimators or descriptive statistics, are only partly used to construct the time series model for the data.

The specification of the location and scale functions and the structure of the error sequence is often guided by mathematical convenience to obtain tractable properties of the process but not that much by reference to the actual data at hand. In the majority of the cases, an intrinsic linear specification of the location and scale functions is used. For instance, in the popular class of ARMA-GARCH models the location is a linear function of lagged random variables and the scale is the square root of a linear function of squared lagged random variables. But also in the case of non-linear regime switching time series models, such as autoregressive threshold models (Tong (1996)) or hidden Markov models (Frühwirth-Schnatter (2006)), the specification of location and scale is intrinsically linear in almost all cases. That is, conditional on the regime, the location and scale functions are linear functions of (transformed) lagged variables. Apparently, the major reason for assuming a simple functional shape for the location and scale of the transition distribution is analytical tractability. If the functional shapes are non-linear, conditions that ensure stationarity of the process become noticeably harder to derive and the corresponding autocorrelation function may not be tractable anymore. The modeling of stationary processes with more complex dynamics can get quite cumbersome if one uses the classical approach to time series analysis.

Classical time series analysis often focusses on the modeling of the autocorrelation structure of the process. If the stochastic process is Gaussian, i.e., all finite-dimensional distributions are Gaussian, then the autocorrelation indeed summarizes the dependence structure of the process.[4] However, the meaning of the autocorrelation function is less clear if the time series is not Gaussian which is especially the case in finance. For example, the marginal distribution of duration data or realized volatility has a positive support which is quite different from a Gaussian process. The interpretation of the correlation coefficient is unclear in these cases. To reinforce this argument, we consider the following example.

Example 3.1

Let Y be a time series process which is defined by the difference equation $Y_t = 0.7Y_{t-1} + \sqrt{1-0.7^2}\mathcal{E}_t$, where $\mathcal{E}$ is an iid sequence of standard normal variables. The time series process X is defined by $X_t = F_X^{-1}(\Phi(Y_t))$, where Φ is the cdf of the standard normal distribution and F_X^{-1} is the quantile function of $X = W/\sqrt{\mathbb{V}\text{ar}[W]}$, where W has a log-normal distribution with parameter $(\mu, \sigma) = (0, 2)$.

Note that the marginal distribution of Y is standard normal in Example 3.1. While the autocorrelation function is a sensible dependence measure for Y, Figure 3.1 illustrates that the interpretation of the autocorrelation function can be highly misleading for X. The left panels of Figure 3.1 show a trajectory and the corresponding sample autocorrelation function of Y. In this case, the strong temporal dependence is captured by the autocorrelation function with the first 12 lags being (individually) statistically significant. In the right

[4] This might also explain the ubiquitous use of linear functions in classical time series analysis.

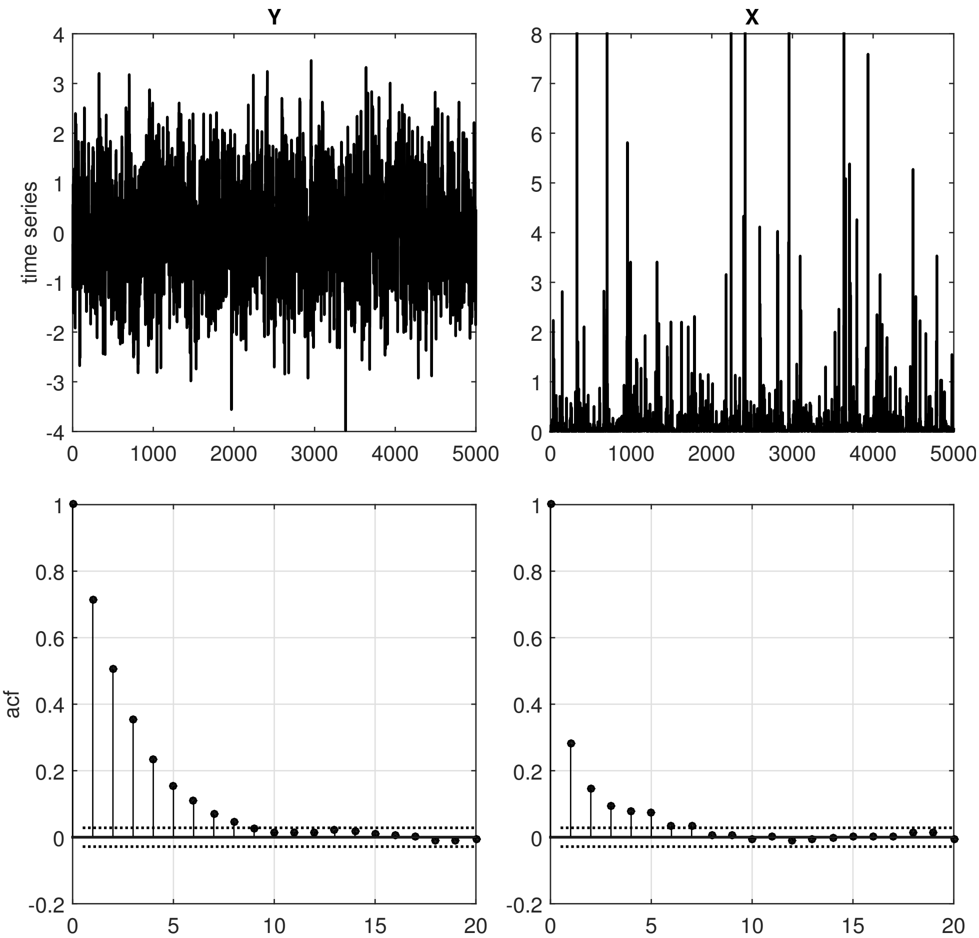


Figure 3.1: Illustration of the trajectory and sample autocorrelation function with 5% Bartlett confidence intervals of two stationary time series processes which only differ in their marginal distribution. Process Y and X are defined in Example 3.1. An element of the trajectory of X was generated by $x_t = F_X^{-1}(\Phi(y_t))$, where y_t is an element of the trajectory of Y. Thus, both empirical autocorrelation functions exhibit the same sampling variability.

panels of Figure 3.1 a trajectory and the corresponding sample autocorrelation function of X is depicted. X has the same finite-dimensional distributions as Y except for the marginal distribution which has been transformed to a log normal distribution. According to the autocorrelation function, the temporal dependence in X is rather weak and only the first two lags are individually statistically significant. Although the two time series processes only differ in their marginal distribution, the autocorrelation functions suggest that these two processes have a quite different dependence structure. Thus, the autocorrelation function might not be a very meaningful dependence measure. From this point of view, the common practice of constructing models that mainly reproduce the autocorrelation function of (non-Gaussian) time series can be called in question.

The qualitative shape of the distribution of the error term $\mathcal{E}_t$ in (3.1.1) is often derived from the shape of the marginal distribution of the process, with the intention to reproduce the marginal distribution of the process. One can also argue that the transition distribution is often constructed as a replica of the marginal distribution but with parameters that depend on the history of the process. However, even if the complete transition distribution is specified, the actual computation of the marginal distribution of the process is not feasible in almost all cases. It is in general not tractable how the functional form of the location and scale function, and the error distribution, determine the marginal distribution of the process. Besides, the relation between the shape of the transition distribution and the shape of the marginal distribution is rather loose. For instance, if the marginal distribution is unimodal, this does not imply that the transition distribution is also unimodal. As a result, the useful information that is contained in the marginal distribution is only partly incorporated in the resulting model. Moreover, the correct specification of the marginal distribution is a necessary condition for the correct specification of the transition distribution. In this regard, it would be beneficial to know if the marginal distribution can be reproduced by the model.

Last but not least, whereas location scale models display the strength of the classical approach, the modeling of the complete transition distribution is less developed in the classical approach and suffers from a fast increase in model parameters and model choices, see Hansen (1994), Haas (2004), and Rigby and Stasinopoulos (2005), for some approaches in this direction.

After having pointed out advantages and drawbacks of the transition-based approach to time series analysis, we now introduce a copula-based approach for modeling time series which fixes some drawbacks of the classical approach at the cost of loosing some of its advantages.

3.1.2 A copula-based approach to model serial dependence in time series

The law of a stochastic process is characterized by its finite-dimensional distributions. Thus, one can in principle model a stationary time series Y by specifying the sequence $(F_{1:t})_{t\in\mathbb{N}}$, where

$$F_{1:t}\colon \mathbb{R}^t \to [0,1], \quad F_{1:t}(y_{1:t}) = \mathbb{P}(Y_i \leq y_i, \forall i = 1, \ldots, t).$$

The major reason why this approach is not popular in time series modeling is that the number of useful multivariate distributions was still mainly restricted to the class of elliptical distributions a couple of years ago. Except for the popular Gaussian process, elliptical distributions do not give rise to a useful modeling framework. For instance, if a process is Markov of order $p \in \mathbb{N}$, then the only stochastic process with finite-dimensional elliptical distributions is a Gaussian process or a sequence of iid random variables.[5] Obviously,

[5] This follows since there is no elliptical distribution which is not the multivariate normal distribution or a vector of independent random variables such that $Y_{t+p+1} \perp Y_t | Y_{t+1:t+p}$, (cf. Cambanis et al., 1981, Section 3).

the direct modeling of the finite-dimensional distribution stands and falls with a suitable framework of modeling multivariate distributions. In this regard, the recent introduction of vine copulas (Bedford and Cooke (2002), Aas et al. (2009)) has greatly extended the number of flexible and useful multivariate distributions. Consequently, the vine copula framework opens the doors to modeling the finite-dimensional distributions of a process and gives rise to a fundamentally different approach to modeling time series. Before we further elaborate on the usefulness of vine copulas to model time series we first want to discuss the general approach of modeling the finite-dimensional distributions of a process with a copula.

While the classical approach aims at modeling the transition representation $(G_t \circ F_{t|t-1:1})_{t\in\mathbb{N}}$, the copula-based approach models the copula representation of a stationary[6] process Y which is given by $(F_Y, (C_{1:t})_{t\in\mathbb{N}})$. F_Y denotes the marginal distribution of Y, i.e., $F_Y(y_i) = F_{Y_i}(y_i) = \mathbb{P}(Y_i \leq y_i)$ for all $i \in \mathbb{N}$, and

$$C_{1:t}\colon [0,1]^t \to [0,1], \quad C_{1:t}(u_{1:t}) = \mathbb{P}(F_Y(Y_i) \leq u_i, \forall i = 1,\ldots,t).$$

In theory, both approaches are equivalent, i.e., we can obtain the transition representation from the copula representation and vice versa. However, from a modeling point of view both approaches are different in several ways as we point out below.

Advantages

If one uses the transition representation, it is, in general, not possible to calculate any finite-dimensional distribution of the process, such as the marginal distribution.[7] The copula representation yields a direct expression for the marginal distribution and, depending on the model for the multivariate copulas, gives access to some of the finite-dimensional distributions of the process. Moreover, the transition distribution of the process can be expressed in terms of a copula and the marginal distribution, i.e.,

$$\begin{aligned} F_{t|t-1:1}(y_t|y_{t-1:1}) &= \partial_{1:t-1}F_{1:t}(y_{1:t})/\partial_{1:t-1}F_{1:t}(y_{1:t-1},\infty) \\ &= \frac{\partial^{t-1}C_{1:t}(F_{1:t}(y_{1:t}))}{\partial y_1\cdots\partial y_{t-1}} \Big/ \frac{\partial^{t-1}C_{1:t}(F_{1:t-1}(y_{1:t-1}),1)}{\partial y_1\cdots\partial y_{t-1}} \\ &= F^u_{t|t-1:1}(F_Y(y_t)|F_{t-1:1}(y_{t-1:1})), \end{aligned}$$

where $\partial_{1:t-1}$ is the mixed partial derivative wrt to the first $t-1$ variables, $F^u_{t|t-1:1} := F_{U_t|U_{t-1:1}}$ and, with a slight abuse of notation, $F_{1:i}(y_{1:i}) := F_Y(y_1),\ldots,F_Y(y_i)$. That is, in the copula-based approach to time series modeling the transition distribution is derived from

[6] The modeling of non-stationary processes is also possible but – also for the sake of notational simplicity – we do not consider it here.

[7] See Chapter 4.2 of Tong (1996) for a discussion of obtaining the marginal distribution using numerical approximations.

a finite-dimensional distribution.

At first sight, this might seem like a redundant detour, which also blows the modeling problem out of proportion if the interest lies primarily in the transition distribution, since now a t-dimensional distribution has to be modeled to obtain a univariate transition distribution at point t. However, the univariate transition distribution at period t is a t-dimensional function which is also true for the t-dimensional copula from which the transition distribution can be derived. Although the copula-based approach requires the modeling of an additional one-dimensional function for the marginal distribution, the implicit modeling of the transition distribution does not really increase the dimensionality of the problem. Moreover, depending on the specific stochastic process, the copula-based approach can result in a more parsimonious model for the transition distribution as we will illustrate later on.

On the other side, one major advantage of the copula-based approach is that we can actually use the information that is contained in the finite-dimensional distribution to obtain a possibly even better model for the transition distribution. For that reason, the small detour might be worthwhile. We now illustrate how exploratory data analysis of the finite-dimensional distribution can successfully be used to develop a model for the transition distribution if sufficient data is available.[8]

In general, modeling data from a copula point of view allows us to separate the modeling of the marginal distribution from the modeling of the dependence structure. In other words, the copula-based approach to time series analysis allows us to obtain a time series process with given marginal distribution. This fact is of great advantage due to the following reasons. By fitting an adequate marginal distribution the resulting transition distribution reproduces the marginal features of the time series in any case, e.g., the implied number of unconditional VaR exceedances should then match the specified probability level, which is not always true if we specify the transition distribution directly as in the classical approach. The importance of a correct specification of the marginal distribution can not be overestimated. As will be illustrated in Chapter 4 and Chapter 6, the marginal distribution contributes the major part to the log-likelihood value of the fitted model. In many cases, the dependence in financial time series is rather weak so that more than 80% of the log-likelihood is accounted for by the marginal distribution.

The modeling of the marginal distribution is also easier than the modeling of the error distribution in the classical approach. The error distribution is not observable but implied by the functional form of the location and scale function of the transition distribution. Thus, the goodness of the error distribution can only be checked after a full model has been fitted. If the model has been judged as inappropriate, then it is often not clear

[8] Of course, the utility of exploratory data analysis increases with the sample size. If the sample size of the time series is too small, then exploring the data can not provide reference points for the construction of a complete model. Fortunately, in financial applications we typically have at least moderate sample sizes, i.e., $T \geq 500$.

whether this is due to a miss-specification of the location and scale function or the error distribution. In any case, one has to fit another full model in order to evaluate if the error distribution is correctly specified. In the copula-based approach, the modeling of the marginal distribution is the first step. We can directly inspect the marginal distribution of a process using descriptive statistics or graphical tools such as kernel density estimation. The goodness of a marginal distribution function for capturing the marginal features of the process can be evaluated at once without the need to specify a full model. Only if the marginal distribution has been correctly specified we should continue with the modeling of the dependence structure. In this regard, the process of modeling a time series is simplified.

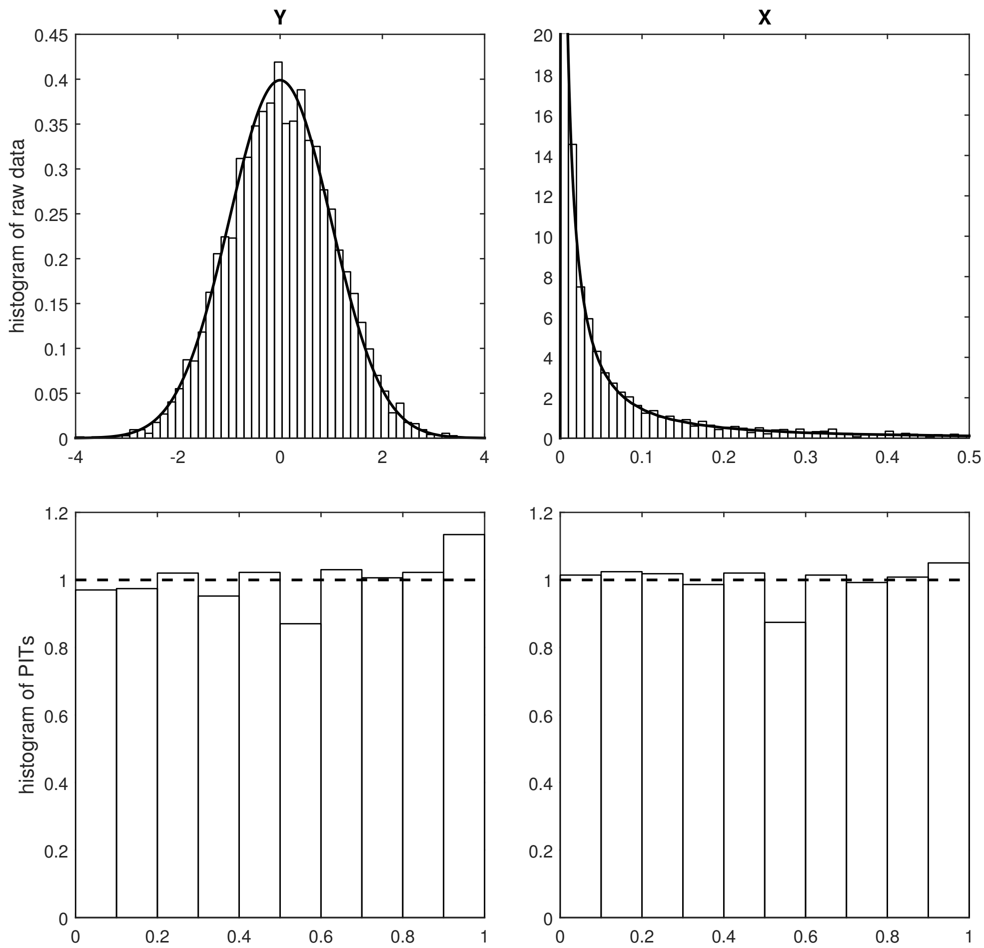


Figure 3.2: Fitting the marginal distribution of a time series process in the copula-based framework. Process Y and X are defined in Example 3.1. The upper panel shows for each process the (normalized) histogram of the marginal distribution and the superimposed density of a fitted normal distribution for Y and a fitted log-normal distribution for X. The lower panel shows the histogram of the estimated probability integral transforms of the marginal distribution, i.e., if $\tilde{F}_Y$ is the estimated marginal distribution of Y then $\tilde{F}_Y(y_t)$ is an observation from the corresponding estimated probability integral transform.

The upper panel of Figure 3.2 shows histograms of the observations from the marginal distribution of the processes Y and X which are defined as in Example 3.1. The superimposed marginal densities indicate that the marginal distribution is correctly specified although the fit of the assumed marginal distribution for X is hard to assess for very small or very large values because of its highly skewed distribution. Therefore, the lower panel also depicts the histograms of the estimated probability integral transforms (PITs) which, independently of the marginal distribution, should resemble the histogram of observations from a standard uniform distribution. In Section 3.6 we also discuss formal goodness-of-fit tests for the marginal distribution, but at the moment the graphical diagnostic checks should indicate that the marginal distribution is correctly specified for both processes.

Besides the marginal distribution, other features of the finite-dimensional distributions can easily be implemented in the copula-based approach. The visual inspection of pairwise dependence within a time series has a long tradition and goes at least back to the famous statistician George Udny Yule (Yule (1927), p. 277), who used scatter plots for the investigation of sunspot numbers. However, scatter plots of, say, consecutive observations of a time series are almost never used for explorative purposes in the classical approach to time series analysis. A possible reason is that the information contained in pairwise dependence can not be used to construct a *direct* model for the transition distribution. Moreover, a scatter plot of consecutive observations is often not very informative if the marginal distribution is not close to being normally distributed. High kurtosis or skewness of the marginal distribution makes it difficult to interpret the dependence structure since the marginal features dominate the visual impression of the scatter plot. The plot in the

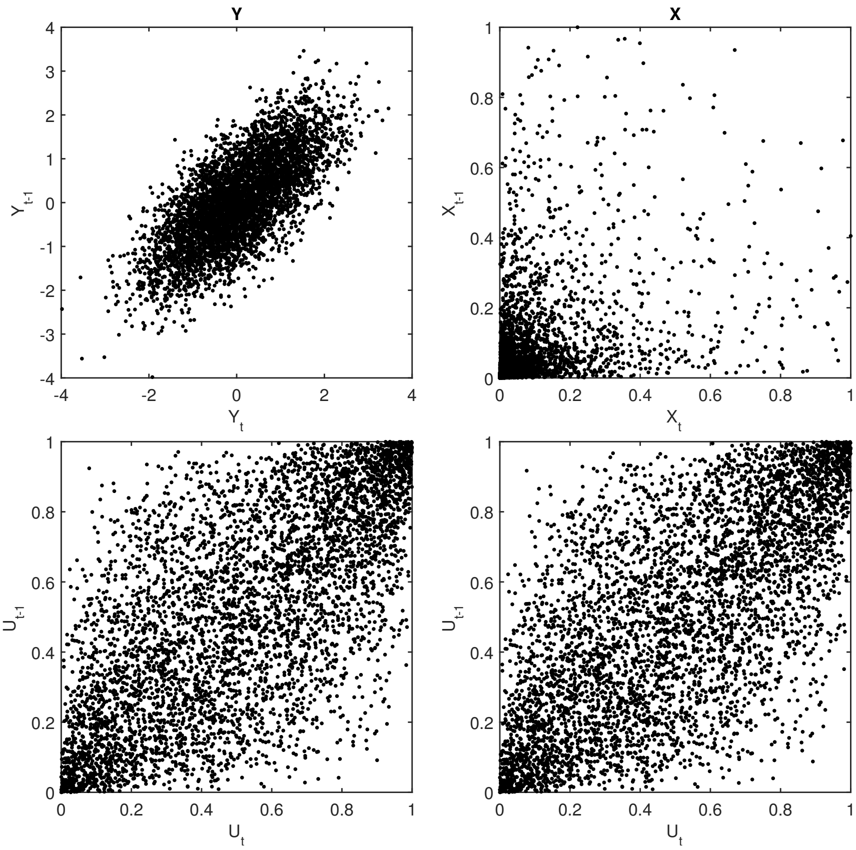


Figure 3.3: Fitting the dependence structure of a time series process in the copula-based framework. Process Y and X are defined in Example 3.1. The upper panel shows for each process the scatter plot of consecutive observations. The lower panel shows for each process the scatter plot of consecutive observations from estimated probability integral transforms, e.g., if F_Y is the estimated marginal distribution of Y then the t-th observation in the scatter plot is given by $(u_t, u_{t-1}) = ((F_Y(y_t), F_Y(y_{t-1}))$.

top right panel of Figure 3.3 illustrates that a scatter plot of consecutive observations from a raw time series is not very revealing if the marginal distribution is highly skewed.

This picture changes if we consider scatter plots of consecutive observations from the time series of estimated probability integral transforms. If the marginal distribution is correctly specified, these scatter plots depict realizations from the copula of consecutive observations. The lower panels in Figure 3.3 strongly suggest that both time series have the same copula for consecutive observations. An important advantage of analyzing data in this way is the conceptual ease of dealing with bivariate distributions that exhibit uniform marginals. In contrast to other distributions, observations from a uniform distribution are evenly scattered on a bounded interval, thus there are no marginal outliers which obscure the dependence structure. Moreover, if we consider observations from both variables that lie within specific quantiles, e.g., observations from both variables are within the respective lower and upper quartiles, this has a clear geometric meaning because these observations are located in a rectangle.

By the use of these copula scatter plots we can investigate whether we have a positive or negative dependence, or a monotone dependence at all, and can easily detect asymmetries in the dependence structure if the sample size is sufficiently large. In this way, expert knowledge can be utilized to specify or at least restrict the qualitative form of the copula that should be used for modeling the dependence between consecutive observations.[9] As a result, the dependence structure of consecutive observations can be an integral part of the model construction in the copula-based approach. This is different from the classical approach, where this dependence structure can only be investigated via simulation. It is also not so obvious how the skewed marginal distribution and the copula of time series X can be reproduced in the classical framework when one directly has to specify the transition distribution. We can not use exploratory data analysis effectively to set up a model for the transition distribution in the traditional approach. Consequently, it may be possible that several models for the transition distribution have to be fitted in the traditional approach until an adequate model is found. And if a model is rejected by the data it may not be clear due to which reasons. In this regard, the copula-based approach seems to be more convenient since it can be based on exploratory data analysis. Obviously, we have not yet addressed the question of how to incorporate information from higher-dimensional distributions in the copula-based model construction. We delay this issue to Section 3.4.2 when we show how (simplified) vine copulas can be used to extend the exploratory analysis of the finite-dimensional distributions.

Finally, it should be noted that copula-based time series models greatly extend the space of available stationary and ergodic time series models. Under mild assumptions on the

[9] Obviously, this is only possible if the sample size is not too low so that a structure in a scatter plot is truly visible. As a rule of thumb we think that at least 500 time series observations are necessary for a reasonable interpretation of a copula scatter plot of consecutive observations.

marginal distribution and the copula (see Section 3.7.2) the resulting time series process is stationary and ergodic. By combining different marginal distributions with different copulas, we can generate a wide variety of processes that exhibit transition distributions which can not be represented by a location scale model. In Figure 3.4 some resulting shapes of the transition distribution of copula-based first order Markov processes are given. We

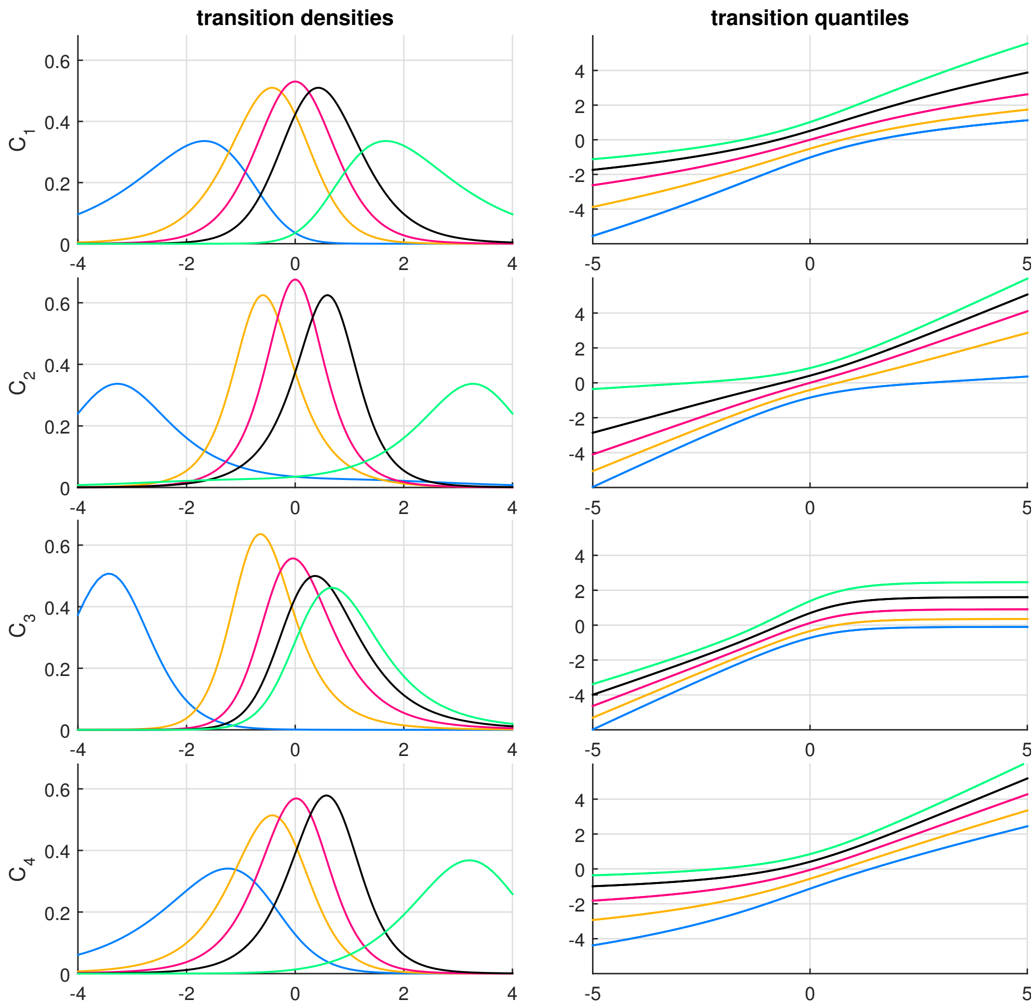


Figure 3.4: Implied densities and quantiles of the transition distribution of a copula-based first order Markov process if the lagged realization y_{t-1} attain its 1%, 25%, 50%, 90%, or 99% quantile. The marginal distribution is a Student-t distribution with 4 dof but the underlying copula is different in any row. C_1 is the Gaussian copula, C_2 the Student-t copula with 2 dof, C_3 is the Clayton copula, C_4 is the Gumbel copula. All copulas have the same Kendall's tau value of 0.5.

observe that, even if the underlying copula is Gaussian and the marginal distribution is symmetric, the transition distribution is in general not symmetric as can be seen in the first row of Figure 3.4. Thus, the shape of the marginal distribution is an unreliable indicator for the shape of the transition distribution. However, the marginal distribution has an effect on the shape of the transition distribution. The copula-based approach combines the marginal distribution together with the copula to obtain the shape of the transition distribution. The third and fourth row of Figure 3.4 also reveal that the marginal distribution can be symmetric around zero even if the transition distribution is always skewed to the same direction.

Drawbacks

In comparison with the classical approach to time series analysis, the copula-based approach also has some disadvantages. First of all, the copula-based approach always models the complete transition distribution. It is not possible to model just some features of the transition distribution like the location and the scale. If the interest lies in some features of the transition distribution then the copula-based approach might not be the first choice. However, even if we are just interested in some features of the transition distribution, the copula-based approach may still result in a better model for these features. In addition, the modeling of the complete transition distribution is appealing since if the transition distribution is correctly specified, "it will be preferred by all forecast users, irrespective of the loss function", as it has been pointed out by Diebold et al. (1998).[10] That is, regardless of the loss function, a correct forecast of the transition distribution is weakly superior to a correct forecast of some features of the transition distribution. Moreover, in financial applications we are often interested in the complete transition distribution. If portfolio optimization is not based on the mean-variance analysis of Harry Markowitz (1952), the joint transition distribution of the assets, and thus the transition distribution of each individual asset, is required in order to compute the expected utility or downside risk measures of a portfolio. Forecast intervals are also only sensible if the complete transition distribution is modeled.

A more severe drawback of the copula-based approach, mainly from a theoretical point of view, is that many properties of the model are not tractable. Moments of the transition distribution are given by one-dimensional integrals of the form

$$\mathbb{E}[Y_t^i | Y_{t-1:1} = y_{t-1:1}] = \int_{\mathbb{R}} \frac{y_t^i c_{1:t}(F_{1:t}(y_{1:t}))}{c_{1:t-1}(F_{1:t-1}(y_{1:t-1}))} f_Y(y_t) \mathrm{d}y_t.$$

In general, these conditional moments have to be approximated using numerical methods. Thus, the copula-based approach is computationally more expensive than the classical approach if an application involves the moments of the transition distribution. Moreover, the implied transition distribution of a copula-based approach is in general not analytically tractable. Let F_Y be the marginal distribution of the process, and $Z_t \sim U(0,1), Z_t \perp Y_{t-1:1}$ for all t. The autoregressive stochastic representation of the process is given by

$$Y_t = F_{t|t-1:1}^{-1}(Z_t | Y_{t-1:1}) = F_Y^{-1} \circ (F_{t|t-1:1}^u)^{-1}(Z_t, Y_{t-1:1}),$$

[10] The loss function of the researcher determines which feature of the transition distribution should be modeled. For instance, the mean squared error is minimized by the mean of the transition distribution whereas expected mean absolute error is minimized by the median of the transition distribution. The squared error is often chosen as loss function due to mathematical convenience but not because it is an appropriate loss function for the data at hand, e.g., the plausibility of the squared error as loss function is questionable if the marginal distribution is skewed.

where $(F^u_{t|t-1:1})^{-1}$ is the inverse of $F^u_{t|t-1:1}$ wrt the first argument. For the process X defined in Example 3.1 this becomes

$$X_t = \frac{F_W^{-1} \circ (F^u_{t|t-1})^{-1}(Z_t, F_W(\sqrt{\mathbb{V}\mathrm{ar}[W]}X_{t-1}))}{\sqrt{\mathbb{V}\mathrm{ar}[W]}}$$
$$= \frac{\exp\Big(2\big[0.7\Phi^{-1}(0.5 + 0.5\mathrm{erf}\big(\log(54.096 \times X_{t-1})/\sqrt{8}\big) + \sqrt{1-0.7^2}\Phi^{-1}(Z_t)\big]\Big)}{54.096},$$

where erf is the error function. Thus, if one changes the marginal distribution of a Gaussian AR(1) process this may already result in a copula-based time series process which is analytically not tractable anymore. In general, the autocorrelation function of a copula-based process has no closed form expression and can only be investigated by simulation. In some cases, one can find conditions for the sign of the autocorrelation function. However, if the finite-dimensional copulas of the time series process are modeled by flexible regular vine copulas, then this is not possible (see Section 3.7).

In the classical approach to time series analysis, the inclusion of exogenous regressors and the modeling of multivariate time series processes is relatively straight forward. The actual implementation of exogenous regressors or the construction of multivariate time series processes is much more complicated in the copula-based framework. Although it is conceptually possible to augment the copula-based framework for these cases, the simplicity of the proposed approach in Section 3.4 for univariate time series is lost because, e.g., there is no longer a unique regular vine copula that exhibits desirable properties.

3.1.3 Literature review of copula-based models for serial dependence

In this section we provide a detailed summary of literature that is related to a copula-based approach to model serial dependence.[11] Patton (2009, 2012), and the monograph of Cherubini (2012), also review some of the literature that is mentioned in the following. At the end of this section, we summarize and discuss the reviewed literature and identify open research questions which form the basis of this thesis.

Stationary copula-based first-order Markov processes

Let C be a bivariate copula and F_Y be a univariate cdf. The properties of stationary copula-based first-order Markov processes of the form

$$Y_t = F_Y^{-1} \circ F^{-1}_{U_1|U_2}(Z_t, F_Y(Y_{t-1})), \quad Z_t \perp Y_{t-1:1} \tag{3.1.3}$$

[11] Consequently, we do not consider the so called copula-GARCH models (Patton, 2006) which use a copula to model cross-sectional dependence.

where $F^{-1}_{U_1|U_2}$ is the inverse of $\partial_2 C_{12}$ wrt to the first argument, have been analyzed by several authors. Darsow et al. (1992) provide a condition equivalent to the Chapman-Kolmogorov equations for first-order Markov processes in terms of copulas and also characterize a first-order Markov process by placing conditions on the finite-dimensional copulas of the process. In Section 8.5 of the monograph of Joe (1997), several dependence properties of copula-based first order Markov processes are derived. Chen and Fan (2004) check the independence and uniformness of the sequence of estimated conditional probability integral transforms $(\tilde{F}_{t|t-1:1}(Y_t|Y_{t-1:1}))_{t\in\mathbb{N}}$, which holds if and only if the transition distributions $\tilde{F}_{t|t-1:1}$ are correctly specified, by modeling this sequence with a copula-based first-order Markov process. Ibragimov and Lentzas (2008) analyze long-memory properties of copula-based first order Markov processes and show via simulations that first-order Markov processes with the Clayton copula exhibit long memory on the level of copulas. Lagerås (2010) builds on the framework of Darsow et al. (1992) and establishes some theoretical results on the dependence of some specific first-order Markov processes. Beare and Seo (2014) propose a test of time reversibility which can be applied to copula-based first order Markov processes.

Mixing properties of processes of the form (3.1.3) have been investigated by several authors. Chen and Fan (2006b) consider a semi-parametric setup where the marginal distribution is estimated by non-parametric methods, such as ranks or kernel density estimators, and the copula is specified by a parametric family. Chen and Fan (2006b) prove the asymptotic normality of their estimators under the assumption that the sequence of β-mixing coefficients decays fast enough. Abegaz and Naik-Nimbalkar (2008a) prove the asymptotic normality of the estimators if the marginal distribution and the copula belong to parametric families and the copula generates a geometrically ergodic process. Chen et al. (2009a) establish that a Markov process generated by the Clayton, Gumbel, or Student-t copula, is geometrically ergodic. Beare (2010) shows that a Markov model generated by a symmetric copula with positive and square integrable copula density is geometrically β-mixing. He also notes that copulas with tail dependence do not have square integrable densities. Moreover, he shows that if the density of a copula is bounded away from zero, the copula generates ρ-mixing Markov processes. Longla and Peligrad (2012) point out that geometric ergodicity and geometric ρ-mixing for symmetric copulas are equivalent. In addition, they show that if the density of a copula is bounded away from zero, it generates ϕ-mixing Markov processes, which strengthens a result of Beare (2010). Beare (2012) derives sufficient conditions for Archimedean copulas to generate geometrically ergodic Markov processes and shows that the Clayton, Gumbel, Frank, and AMH copula, satisfy these conditions. He also shows that Archimedean copulas can be ergodic but not geometrically ergodic, i.e., the decay of the β-mixing coefficients can be sub-exponential.

Other univariate copula-based models of serial dependence

Joe (1997) mentions the construction of stationary p-th order Markov processes using a $(p{+}1)$-dimensional copula for $p{+}1$ consecutive random variables and discusses the necessary and sufficient constraints for the $(p+1)$-dimensional copula so that it is indeed the copula of a stationary p-th order Markov process. Ibragimov (2009) extends the work of Darsow et al. (1992) to higher-order Markov processes.[12] Chen et al. (2009a) employ parametric copula functions to specify non-linear quantile functions of first-order Markov processes. Smith et al. (2010) propose a D-vine copula to model the serial dependence of longitudinal data. Moreover, dynamic copula-based time series models of the form

$$\begin{aligned} Y_t &= F_Y^{-1} \circ F_{U_t|U_{t-1}}^{-1}(Z_t, F_Y(Y_{t-1}); \theta_t), \quad Z_t \perp Y_{t-1:1}, \\ \theta_t &= g(\theta_{t-1:1}, Y_{t-1:1}, V_t), \quad V_t \perp (Z_{t-1:1}, Y_{t-1:1}), \end{aligned} \tag{3.1.4}$$

have been proposed. Note that the dependence parameter θ_t of the copula $C_{t,t-1} = F_{U_t,U_{t-1}}$ generally depends on t in these models. Abegaz and Naik-Nimbalkar (2008b) and Abegaz and Naik-Nimbalkar (2011) develop score tests for testing the constancy of the copula parameter in (3.1.4).

Cherubini et al. (2011) develop a random walk model where the error and the lagged random variable are not independent but their dependence is modeled by a bivariate copula. In the recent monograph of Joe (2014), mixed Markov and q-dependent processes are briefly discussed and copula-based Markov models for count time series are illustrated.

Multivariate copula-based Markov processes

Cherubini et al. (2011) extend the framework of Ibragimov (2009) to bivariate higher-order Markov processes, i.e., they express the Markov property in terms of copulas. Rémillard et al. (2012) generalize the approach of Chen and Fan (2006b) to multivariate first-order Markov processes and discuss parameter estimation and goodness-of-fit tests under high-level assumptions. They also generalize some results of Beare (2010) to the multivariate setting. Smith (2013) employs a D-vine copula to model multivariate higher-order Markov processes. By means of a particular D-vine copula he can express the joint transition distribution in closed form and the order of the process can be easily determined. He also proposes to apply the generalized impulse response function in the sense of Koop et al. (1996) to interpret the model implied dynamics. Brechmann and Czado (2014) focus on the construction of a bivariate higher order Markov process using regular vine copulas and obtain conditions for Granger-causality. Beare and Seo (2014) also employ a regular

[12] Theorem 1 of Ibragimov (2009) also characterizes higher-order Markov processes if the parameter space is continuous. However, the proof uses equivalence formulations of the Markov property which are only valid for a discrete parameter space. As a result, the proof of Theorem 1 is not valid for Markov processes with continuous parameter space.

vine copula to model multivariate higher-order Markov processes. Their framework is very similar to that of Smith (2013) although their vine copula is not a D-vine copula but a so called M-vine copula. They also mention that mixing properties of their copula model are unknown.

Applications of copula-based time series models

Savu and Ng (2005) use the copula process given in (3.1.3) to model order durations of the Deutsche Telekom stock and compare it with the ACD model. Although the copula model provides a better fit for the marginal distribution, it fails to reproduce the autocorrelation structure of the data. In the working paper of Ning et al. (2010) consecutive observations of daily realized volatilities of company stocks, stock indices and foreign exchanges rates, are modeled via a bivariate copula with possibly varying dependence parameter. The static models stated in (3.1.3) perform as well as the dynamic models given in (3.1.4). Following the work of Ning et al. (2010), Sokolinskiy and van Dijk (2011) use in their working paper the copula process given in (3.1.4) to model realized volatilities for S&P500 index futures. They find that the copula-based model outperforms the popular HAR approach (Corsi, 2008) for one-day ahead volatility forecasts in terms of accuracy and that a time-varying copula parameter does not improve the modeling. Vaz Melo Mendes and Accioly (2013) also model realized volatilities from the Brazilian equity market. They use four-dimensional C- and D-vine copulas to construct fourth-order Markov processes of realized volatility. The selection of the lag-order is based on graphical inspection of the autocorrelation function.

Ibragimov and Lentzas (2008) consider in their working paper a process that is similar to an ARCH(1) process but where the dependence between two consecutive squared returns is modeled by a (survival) Clayton copula. Their framework is restricted to first-order Markov processes and only models the variance of the transition distribution. The authors apply their model to returns of the Microsoft stock and find strong evidence that a higher-order Markov process is required. Domma et al. (2009) use the Markov process given in (3.1.3) to model individual series of financial returns using a BB7 copula. Their model implies that there is no considerable serial dependence in the financial returns of four Italian stocks, although they also mention that the examined time series exhibit significant volatility clustering. Mendes and Aíube (2011) apply copula-based first-order Markov processes to daily returns and squared daily returns of 62 US stocks. For returns, the Student-t copula provides the best fit for all series, for squared returns, the Gumbel and survival Clayton copula are preferred. The best fitted parametric copula to the GARCH-filtered data is not significantly different from the product copula. The authors also find that copula-based first-order Markov processes are competitive with Gaussian ARCH(1) processes wrt out-of-sample one-step-ahead VaR forecasts. However, both models do not perform so well in the VaR evaluation. In the working paper of Tinkl and Reichert (2011) the returns of the Commerzbank stock are modeled using copula processes given in (3.1.3). The copula-based first-order Markov processes turn out to be inferior to a standard GARCH model. Modeling the GARCH residuals with a copula also does not improve the fit.

Smith et al. (2010) use Bayesian methods and a D-vine copula to obtain a time-inhomogeneous Markov process of varying order for daily electricity load. They show that the D-vine copula model outperforms a Gaussian copula model. Rakonczai et al. (2012) introduce the concept of empirical autocopulas, which is the sequence given by $(\hat{C}_{1,t})_{t\in\mathbb{N}}$, where $\hat{C}_{1,t}$ is the estimated copula of (Y_1, Y_t), to investigate the dependence structure of time series.

Rémillard et al. (2012), who consider the estimation of copula-based multivariate first-order Markov processes, model the unconditional cross-sectional dependence between the CAN/USD exchange rate and the oil price by means of a bivariate copula, but do not consider the application of a first-order Markov process. Smith (2013) applies a D-vine copula to jointly model a 5-dimensional time series of daily maxima of electricity demand and daily spot prices of electricity in Australia. He applies Bayesian model averaging to determine the "order" of the process. Brechmann and Czado (2014) focus on the construction of a bivariate higher-order Markov process using regular vine copulas. To determine the lag order the authors propose to conduct independence tests or using information criteria. They set up a joint model for inflation, industrial production, stock returns and interest rates, and compare it with a vector autoregressive model. The authors show that their copula-based Markov process is competitive with the VAR model. However, in the working version of the paper it is evident that their copula model is inferior to univariate ARMA-GARCH models with copula-based dependence structure between the residuals (Patton, 2006). They explicitly mention that their copula-based model is not capable of reproducing volatility clusters and that the inclusion of time-varying variances is a possible subject for further research. Beare and Seo (2014), who develop a multivariate higher-order Markov process using regular vine copulas, apply their model to fit a bivariate first-order Markov model to exchange rates.

Summary

The theory on the dependence properties and estimation of univariate stationary copula-based first-order Markov processes is quite exhaustive. However, there is no theory available so far for the dependence properties of univariate or multivariate stationary copula-based higher-order Markov processes, which is also mentioned in the conclusion of Beare and Seo (2014). Although there is some evidence that copula-based time series models might be useful for forecasting realized volatility (Sokolinskiy and van Dijk, 2011), the applications of copula-based models to financial returns have not been convincing due to the following reasons.

First, most applications use a stationary first-order Markov process to model a time series which is probably not Markovian of order one (cf. Savu and Ng, 2005; Ibragimov and

Lentzas, 2008; Domma et al., 2009; Mendes and Aíube, 2011; Tinkl and Reichert, 2011). In the applications that use a copula with a time-varying parameter in the sense of (3.1.4) to allow for a non-Markovian dependence (cf. Ning et al., 2010; Sokolinskiy and van Dijk, 2011), it is often found that the static model performs as well as the dynamic model. This indicates that the approach of specifying a varying dependence of the bivariate copula is not successful in modeling a longer memory of the process.

Second, most parametric copulas have been developed with the aim to model monotone dependencies in a static context. While some static copulas are indeed useful in modeling series of financial data with positive support, such as price durations or realized volatilities, it will become evident in Chapter 5 that static copulas are useless for modeling series of financial returns. Together with an insufficient Markov order, the non-existence of useful parametric copulas for modeling serial dependence of financial returns explains why the models used in Ibragimov and Lentzas (2008); Domma et al. (2009); Mendes and Aíube (2011); Tinkl and Reichert (2011), and Brechmann and Czado (2014), are no match for standard GARCH models.

Recently, vine copulas have been used to model multivariate higher-order Markov processes. While vine copulas definitely allow for the modeling of multivariate time series processes, the introduction of such models in Rémillard et al. (2012); Smith (2013); Brechmann and Czado (2014), and Beare and Seo (2014), seems to be ahead of the times. Obviously, the successful application of a multivariate time series model requires that each individual time series can be adequately modeled by the proposed method. However, to the best of our knowledge, there is no study that investigates the use of vine copulas for constructing univariate higher-order stationary Markov processes in more detail and demonstrates that this approach is competitive with classical approaches. Moreover, the proposed vine copula-based multivariate time series models suffer from the problem that only Markov processes of rather small order can be used in applications. But also for the case of a univariate time series there is no strategy available that allows for a parsimonious modeling of a higher-order Markov process. It is striking that, although the proposed higher-order Markov models can have an arbitrarily large order in theory, the order that is used in practical applications is zero (Rémillard et al., 2012), one (Smith, 2013; Beare and Seo, 2014), or two (Brechmann and Czado, 2014).

Overall, the application of copulas to time series analysis has been considered rather from the perspective of modeling a multivariate distribution but not from a time series perspective. On the contrary, a time series perspective is used in this thesis to construct copula-based higher-order Markov processes. We focus on the modeling of univariate higher-order Markov processes since many important questions need to be addressed here before an effective copula-based joint model of a multivariate time series can be established.

In particular, we have to resolve the following questions in the univariate case:

- How can we obtain flexible stationary copula-based higher-order Markov processes?
- What are suitable copulas to model financial return series?
- How can we model copula-based higher-order Markov processes in a parsimonious way?
- What can we say about the implied dependence properties of the resulting process, e.g., is the autocorrelation function positive?
- Under what conditions do we obtain consistent and asymptotically normally distributed estimators for the parameters of the copula-based model?

The construction of suitable copulas for modeling financial returns is investigated in Chapter 5. The other questions are addressed in the following sections of this chapter.

3.2 Copulas and stationary processes

In order to estimate a time series model from data, one has to impose some structure on the data generating process. A necessary requirement of any statistical analysis of time series data is the existence of some statistical properties which are constant over time or at least over some time intervals. One basic assumption that is often used if one is interested in modeling the transition distribution, is that the underlying stochastic process is (strictly) stationary. A time series process Y is stationary (Tong, 1996, Definition 2.1) if for all $n \in \mathbb{N}$, $i \in \mathbb{N}$, and all $(t_1, \dots, t_n) \in \mathbb{N}^n$,

$$F_{t_1,\dots,t_n} = F_{t_1+i,\dots,t_n+i},$$

where $F_{t_1,\dots,t_n} = P(Y_{t_i} \leq y_{t_i}, \forall i = 1, \dots, n)$. If a process is stationary then all finite-dimensional distributions of the process do not change when shifted in time. Consequently, a time-average of one realization of the process will converge to a unique limit.[13] In this section, we investigate what copulas are suitable for representing the finite-dimensional distributions of a stationary process with a focus on vine copulas.

In virtue of Sklar's theorem stationarity of Y is equivalent to

$$\begin{aligned} &\forall (t,i) \in \mathbb{N}^2 \colon F_t = F_{t+i}, \\ \forall (n,i) \in \mathbb{N}\backslash\{1\} \times \mathbb{N}, (t_1, \dots, t_n) &\in \mathbb{N}^n \colon C_{t_1,\dots,t_n} = C_{t_1+i,\dots,t_n+i}, \end{aligned} \tag{3.2.1}$$

[13] But this does not imply that time-averages from different realizations converge to the same limit. For this to be true we need the stronger notion of ergodicity.

where $C_{t_1,\ldots,t_n}$ is the copula of $(Y_{t_1},\ldots,Y_{t_n})$. Thus, for a process to be stationary the marginal distribution and all finite-dimensional copulas must be invariant under shifts in time.

Clearly, not every multivariate copula is compatible with the constraints given in (3.2.1). Copulas which are based on elliptical distributions require a constraint on the scale while Archimedean copulas automatically satisfy the conditions. Some special structures of hierarchical Archimedean copulas (Joe, 1997; Okhrin et al., 2013), which allow for non-exchangeable dependence structures, are also compatible with the conditions for stationarity.[14] However, the building blocks of hierarchical Archimedean copulas are restricted to Archimedean copulas, e.g., the bivariate marginals are Archimedean, which renders them useless for modeling financial returns (see Chapter 5). Besides hierarchical Archimedean copulas, (regular) vine copulas have been introduced to obtain flexible multivariate copulas.

3.2.1 Regular vine copulas and stationary processes

Just like hierarchical Archimedean copulas, vine copulas (Bedford and Cooke, 2002; Kurowicka and Joe, 2011) are hierarchical structures. Contrary to a hierarchical Archimedean copula, a regular vine copula is constructed upon a sequence of arbitrary bivariate conditional copulas (Patton, 2006) which makes this construction extremely flexible. Moreover, every copula can be expressed as a regular vine copula. In general, the density of a regular vine copula decomposes the density of an absolutely continuous N-dimensional copula, $N \geq 2$, into the following product of $N(N-1)/2$ bivariate conditional copulas,

$$c_{1:N}(u_{1:N}) = \prod_{l=1}^{N-1} \prod_{(i,j,K)\in E_l} c_{i,j|K}(F^u_{i|K}(u_i|u_K), F^u_{j|K}(u_j|u_K)|u_K)$$

where $i,j = 1,\ldots,K,\ i \neq j,\ K \subset \{1,\ldots,K\}\backslash\{i,j\}$, and $F^u_{1:N} = C_{1:N}$ so that $F^u_{l|K} = F_{U_l|U_K}$ for $l = i,j$. From a graph-theoretic point of view, a regular vine copula can be considered as an ordered sequence of $N-1$ trees where bivariate conditional copulas $C_{i,j|K}$ are assigned to the edges (i,j,K). The first two entries of the triple (i,j,K) denote the conditioned set while the last entry K is the conditioning set so that the elements (i,j,K) of the set E_l identify the edges of the l-th tree of the regular vine. For simplicity, we write $C_{i,j|\emptyset} = C_{ij}$, and if $K = \{k_1,\ldots,k_r\}$ then $C_{i,j|K} = C_{ij|k_1,\ldots,k_r}$. The triples (i,j,K) have to satisfy some specific rules to constitute a regular vine which can be found in Bedford and Cooke (2002). As a result, each pair of variables (i,j) occurs exactly once as conditioned set and if two edges have the same conditioning set K, then they are the same edge. Moreover, the cardinality of the set E_l, i.e., the number of edges in the l-th

[14] For instance, if $C(u_{1:3}) = \psi(\psi^{-1} \circ \phi[\phi^{-1}(u_1) + \phi^{-1}(u_3)] + \psi^{-1}(u_2))$, where ϕ and ψ are defined as in equation 4.7 of Joe (1997), then $C_{12} = C_{23}$. However, if $C(u_{1:3}) = \psi(\psi^{-1} \circ \phi[\phi^{-1}(u_1) + \phi^{-1}(u_2)] + \psi^{-1}(u_3))$, then $C_{12} = C_{23}$ does not hold in general. Thus, not all structures of hierarchical Archimedean copulas are compatible with the conditions of stationarity.

tree, is $N-l$. Thus, a N-dimensional vine copula directly specifies $N-1$ of its bivariate marginals whereas the other bivariate marginals have to be obtained via integration. For instance, if $\{(1,2,\emptyset),(1,3,\emptyset)\} \subset E_1$ and $(2,3,\{1\}) \in E_2$ then

$$C_{23}(u_2,u_3) = \int_0^1 C_{23|1}(\partial_1 C_{12}(u_1,u_2), \partial_1 C_{13}(u_1,u_3)|u_1)\mathrm{d}u_1,$$

for which no closed form exists in general. Note that C_{23} is determined by the specification of two bivariate copulas C_{12} and C_{13}, and one conditional copula $C_{23|1}$. Thus, the implied bivariate copula C_{23} is equal to the directly specified copula C_{12} if and only if

$$C_{12}(b,c) = C_{23}(b,c) = \int_0^1 C_{23|1}(\partial_1 C_{12}(a,b), \partial_1 C_{13}(a,c)|a)\mathrm{d}a, \tag{3.2.2}$$

holds for all $(b,c) \in [0,1]^2$. In general, it is not possible to solve this integral equation analytically. Thus, one can not impose conditions on the triple $(C_{12}, C_{13}, C_{23|1})$ such that the implied bivariate copula C_{23} is equal to the directly specified copula C_{12}. It follows that, in general, not every vine copula is compatible with the conditions for stationarity.

If we specify $E_1 = \{(i, i+1, \emptyset) \colon i = 1,\dots,N-1\}$ and set $C_{i,i+1} = C_{12}$ for all $i = 1,\dots,N-1$, then this ensures that the copulas of consecutive random variables $C_{i,i+1}$ are equal. The corresponding tree of these edges is the first tree of a D-vine copula. D-vine copulas are special regular vine copulas since their first tree uniquely determines the remaining trees (Aas et al., 2009). Thus, the edges of the remaining trees are given by $E_l = \{(i, i+l, \{i+1,\dots,i+l-1\}) \colon i = 1\dots,N-l\}, l \geq 2$. The resulting D-vine copula density then reads as follows,

$$c_{1:N}(u_{1:N}) = \prod_{j=1}^{N-1}\prod_{i=1}^{N-j} c_{i,i+j|s(i,j)}\big(F^u_{i|s(i,j)}(u_i|u_{s(i,j)}), F^u_{i+j|s(i,j)}(u_{i+j}|u_{s(i,j)})\big|u_{s(i,j)}\big), \tag{3.2.3}$$

where $s(i,j) = i+1 : i+j-1$. A closer look at (3.2.3) reveals that copulas of the form $C_{i:i+j}$ are determined by the first j trees. If we set $C_{i,i+j|i+1:i+j-1} = C_j$ for $j = 1,\dots,N-1$, where C_j is a bivariate conditional copula with $j-1$ conditioning variables, it is easy to see that the resulting N-dimensional vine copula satisfies the conditions for stationarity given in (3.2.1) for $N \leq n+i$. We call this particular copula the SD-vine copula and its exact definition is given below.

Definition 3.1 (SD-vine copula)

Let $j \in \mathbb{N}$ and C_j be a bivariate (conditional) copula with $j-1$ conditioning variables in the sense of Patton (2006). Denote $F^u_{s(i,j)} = C_{s(i,j)}$ for all $j = 1, \ldots, N-1$, and $i = 1, \ldots, N-j$. Then

$$c_{1:N}(u_{1:N}) = \prod_{j=1}^{N-1} \prod_{i=1}^{N-j} c_j \left(F^u_{i|s(i,j)}(u_i|u_{s(i,j)}), F^u_{i+j|s(i,j)}(u_{i+j}|u_{s(i,j)}) \middle| u_{s(i,j)} \right)$$

is the density of an N-dimensional stationary D-vine copula, or SD-vine copula for short. We denote an N-dimensional SD-vine copula by $\mathbb{C}_{N-1} := (C_j)_{j=1,\ldots,N-1}$.

If and only if $E_1 = \{(i, i+1, \emptyset) \colon i = 1, \ldots, N-1\}$, we can directly specify the copulas of consecutive random variables. For all other regular vine copulas at least one copula of consecutive random variables is given by an integral. Thus, in all other cases we have to satisfy constraints that are given by integral equations in the form of (3.2.2) to ensure that copulas of consecutive random variables are equal. For the particular D-vine given in (3.2.3), we only have to impose the constraint that all copulas in one tree are equal in order to obtain a multivariate copula that is compatible with the conditions for stationarity, which yields the SD-vine copula.

In general, we also have to specify at least two different copulas in the first tree if the regular vine copula is not the SD-vine copula since we can not employ stationarity to specify just one unique copula in the first tree. Obviously, a bivariate conditional copula in one of the remaining trees is in general not identical to another conditional copula of the vine. Thus, in general, we have at least one bivariate conditional copula in each tree that is different from the other specified copulas. It follows that, in general, the number of different copulas that is required to obtain an N-dimensional copula is at least N even if we make use of the conditions for stationarity. Note that, in general, an N-dimensional regular vine copula is specified by $N(N-1)/2$ different copulas so that the SD-vine copula provides an enormous reduction in the number of copulas that we have to specify due to its optimal use of the stationarity conditions. The following proposition summarizes the properties of the SD-vine copula that we have discussed so far.

Proposition 3.1 (Characterization of the SD-vine copula)

For all $N \geq 3$, the N-dimensional SD-vine copula is the unique regular vine copula such that for all stationary processes:

1. The constraints on the bivariate conditional copulas that have to be satisfied in order that the vine copula is the copula of N consecutive random variables of a stationary process are not given by integral equations.

2. The corresponding copula of N adjacent random variables is specified by no more than $N-1$ different bivariate conditional copulas.

Proof. Follows from the previous remarks. ■

If the copula of a multivariate distribution is modeled by a regular vine copula, we call this multivariate distribution a (regular) vine distribution. If $G_{t|t-1:1}$ is the conditional cdf of $G_{1:t}$, which is a regular vine distribution without any constraints for stationarity, then it is possible that the process defined by

$$Y_t = G^{-1}_{t|t-1:1}(Z_t|Y_{t-1:1}), \quad Z_t \sim U(0,1),\ Z_t \perp Y_{t-1:1},$$

is still stationary, see Appendix A.1 for an example and further illustrations. However, the distribution of t consecutive random variables of such a possibly stationary process Y is in general not equal to the distribution of $G_{1:t}$ if no stationarity constraints on $G_{1:t}$ are imposed. Thus, the first statement of Proposition 3.1 does not imply that the SD-vine copula is the only copula that can be used to generate stationary processes.[15] But an implied finite-dimensional distribution of a process with an unique stationary distribution always matches the corresponding vine distribution if it is based on the SD-copula. Regarding the second statement of Proposition 3.1, one should not conclude that the number of different bivariate conditional copulas, that have to be specified for the copula of N adjacent random variables, is always larger than $N+1$ for other regular vine copula. For instance, if we have an iid sequence then all regular vine copulas just consist of product copulas, so we only have to specify one copula for all regular vines.

Finally, the implied transition distribution of the SD-vine copula at time t is given by

$$F_{t|t-1:1}(y_t|y_{t-1:1}) = \partial_2 C_{t,1|2:t-1}\big(F_{t|2:t-1}(y_t|y_{2:t-1}), F_{1|2:t-1}(y_1|y_{2:t-1})\big|y_{2:t-1}\big),$$

Due to the definition of a regular vine, we have closed form expressions for $F_{k|2:t-1}, k = 1, t$, provided that the partial derivatives of all copulas in the vine have closed form expressions. Thus, the implied transition distribution of the SD-vine copula has a convenient representation. However, this property does not characterize the SD-vine copula. If the bivariate conditional copula in the last tree corresponds to the distribution of $(Y_t, Y_1)|Y_{2:t-1}$, i.e., $E_{t-1} = \{(t, 1, \{2 : t-1\})\}$, the implied transition density of a regular vine also has this closed form expression.[16]

[15] This fact seems to be unknown in the literature.
[16] Otherwise we have to compute the transition distribution by

$$F_{t|t-1:1}(y_t|y_{t-1:1}) = \int_{-\infty}^{y_t} \frac{f_{t:1}(s, y_{t-1:1})}{\int_{\mathbb{R}} f_{t:1}(z, y_{t-1:1})\mathrm{d}z}\mathrm{d}s$$

using numerical integration methods.

3.2.2 The partial autocopula sequence and simplified SD-vine copulas

As can be seen from Definition 3.1, an N-dimensional SD-vine copula is determined by $N-1$ bivariate conditional copulas, so that the following characterization of a stationary process can be established.

Definition 3.2 (Conditional autocopula sequence of a stationary process)

Let Y be a stationary process. Then the law of Y is characterized by its marginal distribution F_Y and its conditional autocopula sequence which is given by $(C_{1,1+j|2:j})_{j\in\mathbb{N}}$, where $C_{1,2|2:1} = C_{12}$, and $C_{1,1+j|2:j}$ is the conditional copula of $F_{1,1+j|2:j}$ for all j.

The conditional autocopula function is an equivalent representation of all finite-dimensional copulas of a stationary process. Thus, without further assumptions on the underlying process, the conditional autocopula sequence is not useful for modeling time series processes in practice. A dependence measure of a stationary process that is related to the conditional autocopula sequence is given by the partial autocopula sequence which we now introduce.

Definition 3.3 (Partial autocopula sequence)

The partial autocopula function of a stationary process Y is given by the function sequence $(C^{\partial^{j-1}}_{1,1+j;\,2:j})_{j\in\mathbb{N}}$, where $C^{\partial^0}_{12;\,2:1} = C_{12}$, and $(C^{\partial^{j-1}}_{1,1+j;\,2:j})_{j\in\mathbb{N}}$ is the $(j-1)$-th order partial copula of $F_{1,1+j|2:j}$ that originates from the SD-vine copula (see Definition 2.6).

The partial autocopula sequence can be regarded as a generalization of the partial autocorrelation function.[17] To illustrate this relation, we assume that the marginal distribution of Y has zero mean and unit variance. While the first value of the partial autocorrelation function is the correlation ρ_{12} of the random vector (Y_1, Y_2), the first element of the partial autocopula sequence is the copula of (Y_1, Y_2). If we define $\mathcal{E}_{i|2} = Y_i - \rho_{12}Y_2$ for $i = 1, 3$, then $\mathcal{E}_{i|2}$ does not depend on Y_2 in a linear way, i.e., there is no correlation left, but $\mathcal{E}_{i|2}$ may still be dependent on Y_2. The second value of the partial autocorrelation function is given by $\rho_{13;2} = \mathbb{C}\mathrm{orr}[\mathcal{E}_{3|2}, \mathcal{E}_{1|2}]$, i.e., it is the correlation of (Y_3, Y_1) after each element has been corrected for the linear influence of Y_2. The second element of the partial autocopula sequence is the distribution of $(U_{3|2}, U_{1|2})$. Note that $U_{i|2} := F_{i|2}(Y_i|Y_2) \perp Y_2$, for $i = 1, 3$, i.e., $U_{i|2}$ is a random variable that does not depend on Y_2 (Definition 2.1). Consequently, the second element of the partial autocopula sequence is the copula of (Y_3, Y_1) after each element has been corrected for the influence of Y_2. E.g., if Y_2 has only a linear influence on Y_1 and Y_3, then $C^{\partial}_{31;\,2}$ is the copula of $(Y_3 - \rho_{12}Y_2, Y_1 - \rho_{12}Y_2)$. Define

$$\begin{aligned}\mathcal{E}_{1|23} &= Y_1 - (\rho_{12}(1-\rho_{13;2})Y_2 + \rho_{13;2}Y_3),\\ \mathcal{E}_{4|23} &= Y_4 - (\rho_{34}(1-\rho_{24;3})Y_3 + \rho_{24;3}Y_2).\end{aligned} \tag{3.2.4}$$

[17] We define the first value or lag of the (partial) autocorrelation function as $\mathbb{C}\mathrm{orr}[Y_1, Y_2]$, i.e., we do not consider $\mathbb{C}\mathrm{orr}[Y_1, Y_1]$ as the first value of the (partial) autocorrelation function.

The third value of the partial autocorrelation function is typically expressed as $\rho_{14;23} = \mathbb{C}\mathrm{orr}[\mathcal{E}_{1|23}, \mathcal{E}_{4|23}]$, i.e., it is the correlation of (Y_4, Y_1) after each element has been corrected for the linear influence of (Y_2, Y_3). To establish the relation to the third element of the partial autocopula sequence, define $\mathcal{E}_{i|j} = Y_i - \rho_{ij} Y_j$ for $i = 1, \ldots, 4$, and $j = 2, 3$, and observe that

$$\mathcal{E}_{1|2} = \rho_{13;2}\mathcal{E}_{3|2} + \mathcal{E}_{1|23},$$
$$\mathcal{E}_{4|3} = \rho_{24;3}\mathcal{E}_{2|3} + \mathcal{E}_{4|23}.$$

Thus, we can express the third value of the partial autocorrelation function as

$$\rho_{14;23} = \mathbb{C}\mathrm{orr}[\mathcal{E}_{1|2} - \rho_{13;2}\mathcal{E}_{3|2}, \mathcal{E}_{4|3} - \rho_{24;3}\mathcal{E}_{2|3}],$$

i.e., $\rho_{14;23}$ is the correlation of $(\mathcal{E}_{1|2}, \mathcal{E}_{4|3})$ after $\mathcal{E}_{1|2}$ has been corrected for the linear influence of $\mathcal{E}_{3|2}$ and $\mathcal{E}_{4|3}$ has been corrected for the linear influence of $\mathcal{E}_{2|3}$. If we do not only correct the involved random variables for linear influences but for the complete influence and transform the marginal distribution into the uniform distribution, then $\mathcal{E}_{i|j}$ becomes $F_{i|j}(Y_i|Y_j)$, for $i = 1, \ldots, 4$, and $j = 2, 3$, and $\mathcal{E}_{1|2} - \rho_{13;2}\mathcal{E}_{3|2}$ becomes $F_{U_{1|2}|U_{3|2}}(U_{1|2}, U_{3|2}) =: U^{\partial}_{1|23}$, and $\mathcal{E}_{4|3} - \rho_{24;3}\mathcal{E}_{2|3}$ becomes $F_{U_{4|3}|U_{2|3}}(U_{4|3}, U_{2|3}) = U^{\partial}_{4|23}$, by the property of CPITs. The distribution of $(U^{\partial}_{1|23}, U^{\partial}_{4|23})$ is indeed the third element of the partial autocopula sequence. E.g., if (Y_2, Y_3) has only a linear influence on Y_4 and Y_1, then $C^{\partial^2}_{14;\,23}$ is the copula of $(\mathcal{E}_{1|23}, \mathcal{E}_{4|23})$ which is defined in (3.2.4). It is evident how the remaining elements of the partial autocopula sequence can be interpreted as a generalization of the partial autocorrelation function.

In other words, the partial autocopula sequence at lag j is the copula of Y_1 and Y_{j+1} after each random variable has been adjusted for the influence of the intermediate vector $Y_{2:j}$ that is captured by the $(j{-}2)$-th order partial distribution functions $F^{\partial^{j-2}}_{1|2:j}$ and $F^{\partial^{j-2}}_{j+1|2:j}$. The partial autocopula sequence may be more informative than the partial autocorrelation function since it does not only measure linear association but also corrects for considerably more – but not all – non-linear dependencies that are generated by intermediate random variables. In this light, the partial autocopula sequence may be a helpful tool that provides insights into the dynamical structure of a time series.

Note that although the partial autocopula sequence can be seen as a generalization of the partial autocorrelation function, this does not mean that the partial autocorrelation function can be derived from the partial autocopula sequence and the marginal distribution. If and only if for all $j \geq 3$ there exists $a_{1:j} \in \mathbb{R}^j$, $b_{1:j} \in \mathbb{R}^j$, and distribution functions F_1 and F_{j+1} such that $U^{\partial^{j-2}}_{1|2:j} = F_1(Y_1 - \sum_{i=1}^{j-1} a_i Y_{i+1})$ and $U^{\partial^{j-2}}_{j+1|2:j} = F_{j+1}(Y_{j+1} - \sum_{i=1}^{j-1} b_i Y_{i+1})$, the correlation of $(F_Y^{-1}(U^{\partial^{j-2}}_{1|2:j}), F_Y^{-1}(U^{\partial^{j-2}}_{j+1|2:j}))$ matches the j-th value of the partial autocorrelation function. Thus, the correlation structure that is induced by the partial autocopula sequence is interesting in its own right.

In general, there is no strict connection between the partial autocopula sequence and the conditional autocopula sequence. From Proposition 3.4 in Spanhel and Kurz (2015) we can conclude that if the j-th copula of the conditional autocopula sequence equals the product copula this does not imply that the j-th copula of the partial autocopula sequence is the product copula and vice versa. Although the partial autocopula sequence does not characterize the law of any stationary process, the modeling of the partial autocopula sequence can be seen as a general principle to model complex processes. In classical time series analysis, ARMA processes are often motivated by the fact that they can represent any autocorrelation function. In this sense, ARMA processes reproduce the dependence property of a process which is considered to be the most important feature in classical correlation-based time series analysis. For general non-linear processes the partial autocopula sequence might be a more informative dependence property of the process. Moving away from the focus on the autocorrelation structure towards the modeling of the partial autocopula sequence might provide a better approximation of the data generating mechanism if the process exhibits non-linear dependence structures.

There is also a vast class of stationary processes that is characterized by their marginal distribution and their partial autocopula sequence. If all SD-vine copulas of a stationary process satisfy the simplifying assumption then $C_{1,1+j|2:j} = C^{\partial^{j-1}}_{1,1+j;\,2:j}$ for all $j \in \mathbb{N}$ (Spanhel and Kurz, 2015), i.e., all bivariate conditional copulas are bivariate unconditional copulas, and the partial autocopula sequence characterizes all finite-dimensional copulas of the process. Since the class of simplified vine copulas is very rich, a large class of stationary processes can be modeled by its partial autocopula sequence. We underline this result in the following proposition.

Proposition 3.2 (Stationary processes with partial dependence structure)

Let Y be a stationary process. If and only if all SD-vine copulas of Y satisfy the simplifying assumption then the conditional autocopula sequence and the partial autocopula sequence coincide and Y is characterized by its marginal distribution and the partial autocopula sequence.

It is interesting to investigate which popular time series processes satisfy this constraint on their dependence structure. Unfortunately, an answer to this question is in general not possible. By definition, we need to know the finite-dimensional distributions of a process in order to check if the corresponding SD-vine copula satisfies the simplifying assumption. However, even for linear autoregressive processes with independent innovations the finite-dimensional distributions do not exhibit an analytically tractable form. To the best of our knowledge, only in the case of Gaussian innovations it is known that the finite-dimensional copulas are Gaussian.[18] Since the multivariate Gaussian copula satisfies the simplifying assumption, we obtain the following result for Gaussian processes.

Corollary 3.1 (Partial autocopula sequence of a Gaussian process)

Let Y be a Gaussian process with autocorrelation function given by $(\rho_{1,1+j;2:j})_{j\in\mathbb{N}}$. Then the j-th element of the partial autocopula sequence is given by a Gaussian copula with correlation parameter $\rho_{1,1+j;2:j}$.

In this light, processes with simplified SD-vine copulas, or, equivalently, processes that can be described by the marginal distribution and the partial autocopula sequence, can be regarded as natural generalizations of Gaussian processes with possibly very complex non-linear dependencies.

3.3 Copulas and the Markov property

In this section we introduce higher-order Markov processes and define a condition that allows us to determine the order of a stationary Markov process if we only observe one trajectory of the process which is sufficiently long. We then discuss copula-based characterizations of the Markov property and demonstrate that the Markov property can be easily characterized for the SD-vine copula. Moreover, the SD-vine copula has the unique property that only p bivariate conditional copulas are required to construct every stationary p-th order Markov process.

3.3.1 The Markov property and Markov order identification of degree s

Definition 3.4 (p-th order Markov process)

Y is a Markov process of order $p \in \mathbb{N}$ if p is the smallest number such that for all $t \geq p+2$,

$$Y_t \perp Y_{t-(p+1):1} \,|\, Y_{t-1:t-p}, \tag{3.3.1}$$

or, equivalently,

$$\forall y_t \in \mathbb{R}\colon F_{t|t-1:1}(y_t|Y_{t-1:1}) = F_{t|t-1:t-p}(y_t|Y_{t-1:t-p}) \quad \text{(a.s.)}.$$

If there is no finite p such that the above statements are true, we call Y a Markov process of infinite order.

Note that we explicitly demand that p is the smallest number in Equation 3.3.1 so that the order of the process is unique. A process is Markovian of order p if the information contained in random variables more than p periods ago is redundant in the sense

[18] If the innovation has a stable distribution, it is well known that the marginal distribution of the process is also stable but it seems to be unknown whether higher-dimensional distributions are also stable.

that the transition distribution of the process is already determined by the p most recent random variables. Using properties of conditional probability integral transforms (see Definition 2.1) we can replace (3.3.1) by the equivalent conditions

$$\forall t \geq p+2, \exists \text{ measurable function } g_t\colon Y_t = g_t(Z_t, Y_{t-1:t-p}), \quad Z_t \perp Y_{t-1:1},\ Z_t \sim U(0,1)$$

or

$$Z_t \overset{iid}{\sim} U(0,1), \text{ where } Z_t = F_{t|t-1:t-p}(Y_t|Y_{t-1:t-p}) \text{ for all } t \geq p+2,$$

which is also pointed out in Theorem 3 (b) and (c) in Rüschendorf and Valk (1993). Note that the condition $Z_t \overset{iid}{\sim} U(0,1)$ is closely related to the evaluation of density forecasts along the lines of Diebold et al. (1998), i.e., if and only if the process is Markovian of order p then $(Z_t)_{t\geq p+2}$ should be an iid sequence which means that the transition distribution is correctly specified.

Estimating the order of a Markov process from data is a non-trivial task because, according to Definition 3.4, it involves testing a sequence of conditional independencies which are associated with the t-dimensional random vector $Y_{1:t}$ for each $t \geq p+2$. To be precise, testing the Markov order is only possible when we observe several independent realizations of the process. In many applications, however, we just observe one realization $(y_t)_{t=1,\ldots,T}$ of the process. Consequently, for all $t \geq p+2$, we have only one realization of the random vector $(y_i)_{i=1,\ldots,t}$ to test whether $Y_t \perp Y_{t-(p+1):1} \perp Y_{t-1:t-p}$. Therefore, without additional assumptions, it is not possible to determine the order if we observe just one realization of the process. Even if we assume stationarity of the process, we only have one realization $(y_i)_{t=1,\ldots,t}$ to check whether $Y_t \perp Y_{t-(p+1):1} \perp Y_{t-1:t-p}$ if $t \geq T/2$ and T is the sample size. To render order selection feasible in this case, we may assume the process has order identification of degree s.

Definition 3.5 (Markov order identification of degree s)

A stochastic process has Markov order identification of degree s if

$$\exists p \in \mathbb{N}, \forall t \geq p+s+1\colon (Y_t \perp Y_{t-(p+1):t-(p+s)} \,|\, Y_{t-1:t-p}) \Rightarrow Y_t \perp Y_{t-(p+1):1} \,|\, Y_{t-1:t-p}$$

We call the order identification strict if $s = 1$.

If a process has Markov order identification of degree s, independence of Y_t and $Y_{t-(p+1):t-(p+s)}$ conditional on $Y_{t-1:t-p}$ is sufficient for Y to be a Markov process of order p. In the case of a linear autoregressive process with independent innovations, i.e., $Y_t = \sum_{i=1}^{p} a_i Y_{t-1} + \mathcal{E}_t,\ \mathcal{E}_t \overset{iid}{\sim} F_{\mathcal{E}}$, and if we assume Markov order identification of degree s, the order is chosen as p if the vector $a_{p+1:p+s}$ is zero. When the order identification is strict, the first zero autoregressive coefficient determines the order. If the data generating process has order identification of degree s and is stationary, the Markov order can be obtained from a sufficiently long but single realization $(y_t)_{t=1,\ldots,T}$. To determine the order, we can now use $T-(p+s+1)+1$ observations from $Y_{t:t-(p+s)}$ to check the conditional independence relation $Y_t \perp Y_{t-(p+1):t-(p+s)} \,|\, Y_{t-1:t-p}$.

Clearly, the degree of Markov order identification is crucial for determining the order. A too small degree may result in an underestimation of the order[19] while a too large degree can make the order determination unnecessarily complex. We expect that a Markov order identification with low degree, e.g., $s \leq 3$, is reasonable for stationary financial time series, assuming the time series is adjusted for seasonal or other periodic effects.

3.3.2 Copula-based characterizations of higher-order Markov processes

There are several representations of the Markov property in terms of copulas and some of them are more advantageous than others regarding practical applications. In particular, the characterization of the Markov property in terms of copulas should be such that the order of the process can be easily obtained.

Characterization in the sense of Darsow et al. (1992) and Ibragimov (2009)

We first discuss copula-based characterizations that are related to the work of Darsow et al. (1992) and Ibragimov (2009).

Proposition 3.3

In terms of copula densities, Y is a p-th order Markov process if $p \in \mathbb{N}$ is the smallest number such that for all $t \geq p+2, 1 \leq k \leq t-p-1$,

$$
\begin{aligned}
c_{t:t-(p+k)}\big(u_{t:t-(p+k)}\big) &= \frac{c_{t:t-p}\big(u_{t:t-p}\big)}{c_{t-1:t-p}\big(u_{t-1:t-p}\big)} c_{t-1:t-(p+k)}\big(u_{t-1:t-(p+k)}\big) \\
&= \prod_{i=0}^{k-1} \frac{c_{t-i:t-(i+p)}\big(u_{t-i:t-(i+p)}\big)}{c_{t-(i+1):t-(i+p)}\big(u_{t-(i+1):t-(i+p)}\big)} c_{t-k:t-(p+k)}\big(u_{t-k:t-(p+k)}\big)
\end{aligned}
$$

Proof. See Appendix A.2. ■

To the best of our knowledge, the copula density representation of the Markov property in Proposition 3.3 has not been mentioned before in the literature. However, a strongly related representation in terms of copulas has been derived by Darsow et al. (1992) for the first-order case and by Ibragimov (2009) for the general higher-order case. Integrating the copula densities given in Proposition 3.3 with respect to all arguments and rearranging appropriately, we can reproduce the following characterization of the Markov property in terms of copulas.

[19] For instance, if we have a linear autoregressive process it might happen that $a_1 = 0$ but $a_2 \neq 0$, so using $s = 1$ we would conclude that the order is zero, although the order is larger than one.

Proposition 3.4 (Theorem 1 in Ibragimov (2009))

Y is Markov of order $p \in \mathbb{N}$ if p is the smallest number such that for all $t \geq p+2, 1 \leq k \leq t-p-1$,

$$C_{t:t-(p+k)} = C_{t:t-p} \star^p C_{t-1:t-(p+k)} = \bigotimes_{i=0}^{k} C_{t-i,\ldots,t-(p+i)},$$

where

$$\begin{aligned}
&C_{t:t-p}(u_{t:t-p}) \star^p C_{t-1:t-(p+k)}(u_{t-1:t-(p+k)}) \\
&= \int_0^{u_{t-1}} \cdots \int_0^{u_{t-p}} \frac{\partial^p C_{t:t-p}(u_t, z_{t-1:t-p})}{\partial z_{t-1:t-p}} \times \frac{\partial^p C_{t-1:t-(p+k)}(z_{t-1:t-p}, u_{t-(p+1):t-(p+k)})}{\partial z_{t-1:t-p}} \\
&\quad \times c_{t-1:t-p}(z_{t-1:t-p})^{-1} \mathrm{d}z_{t-1:t-p},
\end{aligned}$$

and

$$\bigotimes_{i=0}^{k} C_{t-i,\ldots,t-(p+i)} = \begin{cases} C_{t:t-p} \star^p C_{t-1:t-(p+1)} & k=1 \\ C_{t:t-p} \star^p \left(\bigotimes_{i=1}^{k} C_{t-i,\ldots,t-(p+i)}\right) & k \geq 2. \end{cases}$$

The central idea behind Proposition 3.4 is that $U_t \perp U_{t-(p+1):t-(p+k)} | U_{t-1:t-p}$ if and only if $C_{t:t-p} \star^p C_{t-1:t-(p+k)} = C_{t:t-(p+k)}$. Theoretically, one can generate higher-dimensional copulas using Proposition 3.4 and by this means construct Markov processes of higher-order. Despite its theoretical appeal, the copula-based representation of Markov processes in Proposition 3.4 is not useful for determining the order of a Markov process in practical applications. Assume that the process has Markov order identification of degree s. Then the order is chosen as p if $p+s$ is the smallest number such that $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$ is the copula of $U_{t:t-(p+s)}$. Testing whether $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$ is the copula of $U_{t:t-(p+s)}$ is in principle possible for small p and s using a goodness-of-fit test for $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$. However, a more serious problem than testing if $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$ is the copula of $U_{t:t-(p+s)}$, is the computation of $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$. Analytical evaluation of $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$ is only possible in very special cases. For instance, if the Markov copula is Gaussian then $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$ is also Gaussian for all $t \geq p+2, 1 \leq s \leq t-p-1$. In general, the copula family is not preserved under the $\star^p$ operation. If the Markov copula belongs to a non-Gaussian elliptical copula or Archimedean copula, which does not collapse to the product copula, then $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$ is not elliptical or Archimedean. This can be seen by recalling that $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$ is the copula of a random vector where $U_t \perp U_{t-(p+1):t-(p+s)} | U_{t-1:t-p}$. However, there is no Archimedean or non-Gaussian elliptical copula which can exhibit this conditional independence relation, except the independence copula. Therefore, analytical computation of $C_{t:t-p} \star^p C_{t-1:t-(p+s)}$, which is an sp-dimensional integral, is not possible, and numerical integration is only feasible for very small p and s.[20]

On the other side, the implied copula densities under the Markov property in Proposition 3.3 do not require integration. However, if the assumed order of the process is p and we assume s for the degree of Markov order identification, we have to test the goodness-of-fit of a $(p+s)$-dimensional density which quickly becomes infeasible for moderate p and s. It also does not seem to be possible to reduce the dimensionality of the testing problem using further assumptions. In contrast, we can impose an additional assumption in the following copula-based characterization which reduces the determination of the Markov order to a two-dimensional problem.

Characterization in terms of bivariate conditional copulas

The copula-based characterization of the Markov property in the sense of Darsow et al. (1992) and Ibragimov (2009) describes the Markov property in terms of some constraints on the multivariate distributions of the process. A more natural interpretation of the Markov property is given in Definition 3.4 which formulates the Markov property as conditional independence relations. The following proposition reformulates Definition 3.4 in terms of bivariate conditional copulas. For that purpose, define in the following, for all $t \geq p+2$, and all $1 \leq k \leq t-p-1$,

$$U_{t|t-1:t-(p+k-1)} = F_{t|t-1:t-(p+k-1)}(Y_t|Y_{t-1:t-(p+k-1)}),$$

and

$$U_{t-(p+k)|t-1:t-(p+k-1)} = F_{t-(p+k)|t-1:t-(p+k-1)}(Y_t|Y_{t-1:t-(p+k-1)}).$$

Proposition 3.5 (The Markov property in terms of bivariate conditional copulas)

Y is a Markov process of order $p \in \mathbb{N}$ if p is the smallest number such that for all $t \geq p+2$, and all $1 \leq k \leq t-p-1$,

$$U_{t|t-1:t-(p+k-1)} \perp U_{t-(p+k)|t-1:t-(p+k-1)} \,|\, Y_{t-1:t-p},$$

[20] Direct integration of the second equality in Proposition 3.3 yields

$$C_{t:t-(p+s)}(u_{t:t-(p+s)}) = \int_0^{u_{t-1}} \cdots \int_0^{u_{t-(p+s-1)}} \frac{\partial^p C_{t:t-p}(u_t, z_{t-1:t-p})}{\partial z_{t-1:t-p}} \times \prod_{j=1}^{s-1} \frac{c_{t-j:t-(j+p)}(u_{t-j:t-(j+p)})}{c_{t-(j+1):t-(j+p)}(u_{t-(j+1):t-(j+p)})}$$
$$\times \frac{\partial^p C_{t-s:t-(p+s)}(z_{t-k:t-(p+s-1)}, u_{t-(p+s)})}{\partial z_{t-k:t-(p+s-1)}} dz_{t-1:t-(p+s-1)},$$

which is only a $(p+s-1)$-dimensional integral and does not require the copula $C_{t-1:t-(p+s)}$. Thus, this integral should be used for numerical integration instead of the integral given in Proposition 3.4.

or, equivalently,

$$C_{t,t-(p+k)|t-1:t-(p+k-1)} = C^{\perp} \quad \text{(a.s.)},$$

where $C_{t,t-(p+k)|t-1:t-(p+k-1)}$ is the conditional copula of $F_{t,t-(p+k))|t-1:t-(p+k-1)}$ in the sense of Patton (2006) and $C^{\perp}$ denotes the bivariate product copula.

Proof. It can be readily verified that (3.3.1) in Definition 3.4 is equivalent to the statement that for all $1 \leq k \leq t-p-1$: $Y_t \perp Y_{t-(p+k)} \,|\, Y_{t-1:t-(p+k-1)}$. The conclusion then follows by the definition of the bivariate conditional copula. ■

The characterization of the Markov property in terms of bivariate conditional copulas naturally leads to the concept of regular vine copulas which are constructed upon a sequence of bivariate conditional copulas. The structure of the regular vine should be chosen such that the corresponding bivariate conditional copulas are directly specified and there is no need for obtaining them by integration. For all $1 \leq k \leq t-p-1$, the t-dimensional D-vine copula with density given by

$$c_{1:t}(u_{1:t}) = \prod_{j=1}^{t-1}\prod_{i=1}^{t-j} c_{i,i+j|s(i,j)}\big(F^u_{i|s(i,j)}(u_i|u_{s(i,j)}), F^u_{i+j|s(i,j)}(u_{i+j}|u_{s(i,j)})\big|u_{s(i,j)}\big), \tag{3.3.2}$$

directly specifies $C_{t-(p+k),t|t-1:t-(p+k-1)}$, which can be seen if we set $i = t-(p+k)$ and $j = (p+k)$ in (3.3.2). By definition, this is also the only regular vine copula with this property. Only in this case, we can directly determine if $C_{t,t-(p+k)|t-1:t-(p+k-1)} = C^{\perp}$ holds for all $1 \leq k \leq t-p+1$, so that the order of the process can be easily determined in theory. Under stationarity, this D-vine copula becomes the SD-vine copula, i.e.,

$$c_{1:t}(u_{1:t}) = \prod_{j=1}^{t-1}\prod_{i=1}^{t-j} c_j\big(F^u_{i|s(i,j)}(u_i|u_{s(i,j)}), F^u_{i+j|s(i,j)}(u_{i+j}|u_{s(i,j)})\big|u_{s(i,j)}\big)$$

so that, if the degree of Markov order identification is s and $t \geq p+s+1$, the process is Markov of order p if $C_j = C^{\perp}$ for all $j = p+1, \ldots, p+s$. This property of the SD-vine copula, together with its convenient properties for modeling stationary processes (Proposition 3.1), renders the SD-vine copula the regular vine copula of choice when it comes to modeling stationary Markov processes. Consequently, we do not have to be concerned about selecting the vine structure if the objective is to model a univariate stationary Markov process. Moreover, if the dependence of a stationary process is characterized by its partial autocopula sequence, we only have to check whether bivariate unconditional copulas are product copulas, which greatly simplifies the determination of the order.

3.4 Modeling stationary Markov processes with simplified SD-vine copulas

In Section 3.2 and Section 3.3 we have shown that SD-vine copulas are suited for modeling stationary processes and that a copula-based characterization of Markov processes from which the order can be easily determined naturally leads to SD-vine copulas. Using Proposition 3.3 or Proposition 3.4, we observe that, if the process is stationary and Markov of order p, all SD-vine copulas with dimension $t \geq p+2$ can be constructed from the sequence $(C_j)_{j=1,\dots,p}$ Thus, we can specify a stationary Markov process of order p by choosing an arbitrary marginal distribution and p arbitrary bivariate conditional copulas C_j.

However, for a reasonable model that can be used in practical applications, we have to impose further assumptions since, in its general form, the modeling of the conditional autocopula sequence is only feasible if the order of the process is rather low. That is because the j-th element of the conditional autocopula sequence is in general a $(j+1)$-dimensional function. While it is in principle possible to model the first few conditional copulas of the conditional autocopula sequence without imposing any constraints, we eventually have to impose conditions on the remaining bivariate conditional copulas to tackle the curse of dimension if the order of the process is rather large.

One very convenient tool of the vine copula framework to tackle the curse of dimension is to use a simplified vine copula model which is in general only an approximation to a non-simplified vine copula, see Spanhel and Kurz (2015) for a detailed analysis of approximations which are based on the simplifying assumption. If we use a simplified SD-vine copula to model the copula of N adjacent random variables, then our model for the conditional autocopula sequence is given by a sequence of bivariate unconditional copulas, which greatly simplifies the modeling of time series. Moreover, if the bivariate unconditional copulas C_j of our model for the conditional autocopula sequence belong to parametric families of copulas which are indexed by a scalar parameter, the determination of the order reduces to checking the value of a scalar parameter for each j, which is much more convenient than checking constraints on a $(j+1)$-dimensional function. In addition, the specification and estimation of bivariate unconditional copulas is much simpler and more worked out than the modeling of bivariate conditional copulas.[21] Finally, we can effectively use exploratory data analysis to construct a higher-order Markov if the conditional autocopula sequence is modeled by a sequence of bivariate copulas. On these grounds, we propose to model the conditional autocopula sequence of a Markov process by a sequence of bivariate unconditional copulas and this section is devoted to a detailed analysis of this framework.[22]

[21] To the best of our knowledge, the modeling of bivariate conditional copulas in the context of vine copulas is currently limited to three dimensions, see Acar et al. (2012).

[22] If the cross-sectional dimension of the process is one and all SD-vine copulas of the data generating process satisfy the simplifying assumption, then the proposed models for stationary multivariate Markov processes of Smith (2013); Brechmann and Czado (2014), and Beare and Seo (2014), which have been

3.4.1 Parametric framework for simplified SD-vine copula-based Markov models

If we specify a parametric cdf for the marginal distributions and assume parametric bivariate unconditional copulas for modeling the conditional autocopula sequence, we obtain the following modeling framework.

Definition 3.6 (Parametric simplified SD-vine copula-based Markov model of order p)

Let $(y_t)_{t=1,\dots,T}$ be an observed time series, $p \leq t-1$, and $K_i \in \mathbb{N}$ for all $i = 0,\dots,p$. Let $F(\tilde{\theta}_0)$ be an absolutely continuous cdf with parameter $\tilde{\theta}_0 \in \Theta_0 \subset \mathbb{R}^{K_0}$ and $(C_i(\tilde{\theta}_i))_{i=1,\dots,p}$ be a sequence of absolutely continuous bivariate unconditional copulas with parameter $\tilde{\theta}_i \in \Theta_i \subset \mathbb{R}^{K_j}$, respectively. Let $u_i^{\tilde{\theta}_0} = F(y_i;\tilde{\theta}_0)$ and set

$$\begin{aligned} u_{i+2|i+1}^{\tilde{\theta}_{0:1}} &= \partial_2 C_1(u_{i+2}^{\tilde{\theta}_0}, u_{i+1}^{\tilde{\theta}_0};\tilde{\theta}_1),\\ u_{i|i+1}^{\tilde{\theta}_{0:1}} &= \partial_1 C_1(u_{i+1}^{\tilde{\theta}_0}, u_{i}^{\tilde{\theta}_0};\tilde{\theta}_1),\\ u_{i+j|i+1:i+j-1}^{\tilde{\theta}_{0:j-1}} &= \partial_2 C_{j-1}(u_{i+j|i+2:i+j-1}^{\tilde{\theta}_{0:j-2}}, u_{i+1|i+2:i+j-1}^{\tilde{\theta}_{0:j-2}};\tilde{\theta}_{j-1}),\\ u_{i|i+1:i+j-1}^{\tilde{\theta}_{0:j-1}} &= \partial_1 C_{j-1}(u_{i+j-1|i+1:i+j-2}^{\tilde{\theta}_{0:j-2}}, u_{i|i+1:i+j-2}^{\tilde{\theta}_{0:j-2}};\tilde{\theta}_{j-1}), \end{aligned} \tag{3.4.1}$$

for $j = 3,\dots,p-1$. Denote the joint log-likelihood function of $F(\tilde{\theta}_0)$ and $(C_i(\tilde{\theta}_i))_{i=1,\dots,p}$ by

$$\mathcal{L}_{0:p}(\tilde{\theta}_{0:p}) = \log f_{1:T}(y_{1:T};\tilde{\theta}_{0:p}) = \sum_{t=1}^{T} \log f(y_t;\tilde{\theta}_0) + \sum_{j=1}^{p}\sum_{i=1}^{T-j} \log c_j(u_{i+j|i+1:i+j-1}^{\tilde{\theta}_{0:j-1}}, u_{i|i+1:i+j-1}^{\tilde{\theta}_{0:j-1}};\tilde{\theta}_j).$$

Define the $\sum_{i=0}^{p} K_i$-dimensional vector $\theta_{0:p}$ by

$$\theta_{0:p} = \underset{\tilde{\theta}_{0:p} \in \times_{i=0}^{p}\Theta_i}{\arg\max}\ \mathcal{L}_{0:p}(\tilde{\theta}_{0:p}), \tag{3.4.2}$$

where we assume that $\theta_{0:p}$ exists and is unique. The parametric simplified SD-vine copula-based Markov model of order p is given by the tuple $(F(\theta_0), \mathbb{C}_p(\theta_{1:p}))$, where $\mathbb{C}_p(\theta_{1:p}) = (C_j(\theta_j))_{j=1,\dots,p}$ denotes the p-dimensional copula sequence of the model.

Since our copula-based Markov model of order p is always based on a simplified SD-vine copula, we also just call it a CMP(p) model. According to Definition 3.6 the transition density of a CMP(p) model is given by

$$f_{t|t-1:1}(y_t|y_{t-1:1}) = f(y_t;\theta_0) \prod_{j=1}^{\min(p,t-1)} c_j(u_{t|t-1:t-(j-1)}^{\theta_{0:j-1}}, u_{t-j|t-1:t-(j-1)}^{\theta_{0:j-1}};\theta_j),$$

developed independently of us, coincide with our model.

where $c_1(u^{\theta_0}_{t|t-1:t}, u^{\theta_0}_{t-1|t-1:t}; \theta_1) = c_1(u^{\theta_0}_t, u^{\theta_0}_{t-1}; \theta_1)$. Under regularity conditions, the copula sequence $\mathbb{C}_p$ is an estimator for the simplified vine copula model which minimizes the KL divergence from the true distribution. If the simplifying assumption holds for all SD-vine copulas of the process, a CMP(p) model estimates the first p elements of the partial autocopula sequence, provided the marginal distribution and the copula families are correctly specified. However, from Proposition 3.2 in Spanhel and Kurz (2015) it follows that, in general, only if we estimate the elements of $\mathbb{C}_p$ successively in a step-wise fashion (see Algorithm 3.2), and not jointly, the parameters of the elements of $\mathbb{C}_p$ converge to the parameters of the partial autocopula sequence, provided that all parametric families are correctly specified. Although the estimated copula sequence $\mathbb{C}_p$ is closely related to the partial autocopula sequence, the statement that the model proposed in Definition 3.6 is an estimator of the partial autocopula sequence is not strictly true in general.

Definition 3.6 defines a CMP(p) model for given copula families and a fixed family for the marginal distribution of the process. In practice, we have to choose the parametric families for the copulas and the marginal distribution, and determine an order for the model. For that purpose, we propose the following steps.

Algorithm 3.1 (Specification, order determination, and estimation of a CMP model)

The specification and estimation of a CMP model proceeds as follows.

1. *Set the degree s of Markov order identification and specify an upper bound P for the order.*
2. *Specify a parametric cdf and perform ML estimation to obtain a model for the marginal distribution of the process.*
3. *Set $j = 0$.*
4. *If $j > P$ go to step 9, else continue.*
5. *Specify the copula families and estimate a CMP($j + s$) model using a step-by-step ML estimation, see Algorithm 3.2.*
6. *Do a joint ML estimation of the CMP($j + s$) model, using the estimates of the step-by-step ML estimation as starting values, see Definition 3.6. This step is optional.*
7. *Check whether the order is j, see Algorithm 3.3.*
8. *Go to the next step if the order is considered to be j, else set $j = j + 1$ and continue with step 4.*
9. *Do a joint ML estimation of the CMP(j) model if $j > 1$ and this has not already been done in step 6, and stop the algorithm.*

If we skip step 6 and 9 in Algorithm 3.1 we obtain an estimate for the partial autocopula sequence of a process. However, if the simplifying assumption is not satisfied, then doing a joint ML estimation in step 5 may not result in a consistent estimator of the partial autocopula sequence. The specification and estimation of a parametric marginal distribution in step 2 is a crucial point in our framework. If the marginal distribution is miss-specified, then it is not possible to obtain correctly specified copulas in the later steps. We discuss the estimation of the marginal distribution in detail in Section 3.6. Besides the marginal distribution, the specification of the copula families in step 5 is an essential step in Algorithm 3.1. For that purpose, we employ the common practice of choosing the copula families in parametric simplified vine copula models by means of the AIC (Dißmann et al., 2013). The following algorithm explains how step-by-step ML estimation of the SD-vine copula and information criteria can be utilized to choose the copula families of a CMP model.

Algorithm 3.2 (Specifying the copula families of a CMP(p) model)

To select the copula families we use step-by-step ML estimation of the SD-vine copula and an information criterion in the following manner.

1. *Use the previously estimated marginal distribution and copula(s) to construct the pseudo-observations* $(u^{\theta_{0:j-1}}_{i+j|i+1:i+j-1}, u^{\theta_{0:j-1}}_{i|i+1:i+j-1})_{i=1,\dots,T-j}$ *for the copula* C_j *which are given by* $u_i^{\tilde{\theta}_0} = F(y_i; \tilde{\theta}_0)$ *if* $j = 1$*, and, for* $j = 2, \dots, p-1$*, the pseudo-observations are given by equation* (3.4.1) *in Definition 3.6.*

2. *Choose parametric copula families for* C_j*. The choice of suitable families can be assisted by analyzing the scatter plot of pseudo-observations* $(u^{\theta_{0:j-1}}_{i+j|i+1:i+j-1}, u^{\theta_{0:j-1}}_{i|i+1:i+j-1})_{i=1,\dots,T-j}$.

3. *Conditional on the previously estimated model* $(F(\theta_0), (C_i(\theta_i))_{i=1,\dots,j-1})$*, use a ML estimation to fit a set of parametric copula families for* C_j*, i.e., the estimated parameter* θ_j *of a copula family* $\tilde{C}_j$ *is given by*

$$\theta_j = \underset{\tilde{\theta}_j \in \Theta_j}{\arg\max} \ \mathcal{L}_j(\tilde{\theta}_j) = \sum_{i=1}^{T-j} \log \tilde{c}_j(u^{\theta_{0:j-1}}_{i+j|i+1:i+j-1}, u^{\theta_{0:j-1}}_{i|i+1:i+j-1}; \tilde{\theta}_j).$$

4. *Choose the fitted copula family* $\tilde{C}_j$ *for* C_j *which minimizes the value of an information criterion.*

5. *Check the goodness-of-fit of the fitted copula. This is an optional step.*

Step 2 of Algorithm 3.2 allows us to use exploratory data analysis to specify suitable copulas for the elements of $\mathbb{C}_p$. In this regard, we can employ exploratory data analysis to construct a model for a higher-order Markov process, see Section 3.4.2 for an illustration. To the best of our knowledge, there is no GoF-test for dependent data available at the moment that can be applied to check the adequateness of the fitted copula family in step 5 of Algorithm 3.2. However, one can visually inspect the goodness of a fitted copula if one compares the scatter plot of the pseudo-observations with a scatter plot of simulated observations from the fitted copula family. By this means, we can also detect what dependence structures a fitted copula family might not capture and can take countermeasures if necessary.

In order to check the order of the CMP model in step 7 of Algorithm 3.1, we can use the following procedure.

Algorithm 3.3 (Determination of the order)

Let s be the degree of Markov order identification. To determine the order of the CMP model we can either:

1. *Use statistical tests, e.g., a likelihood ratio test, to check whether $C_{j+1:j+s} = C^{\perp}_{j+1:j+s}$. If this hypothesis can not be rejected then the order is determined as j.*

2. *Use an information criterion, such as AIC or BIC, to check if the CMP(j) model has a smaller information criterion than the CMP(k) models with order $k = i+1, \ldots, j+s$. If this is true, determine the order as j.*

We strongly favor information criteria to determine the order in Algorithm 3.3 since, without any prior information about the order, we are forced into a multiple hypothesis testing problem if we use statistical tests.

3.4.2 Illustration of simplified SD-vine copula-based Markov models

To illustrate Algorithm 3.1, Algorithm 3.3, and Algorithm 3.3, we examine the following example.

Example 3.2 (Stationary Gaussian AR(2) process)

Consider the stationary Gaussian AR(2) process of the following form,

$$Y_t = 0.6Y_{t-1} - 0.6Y_{t-2} + \mathcal{E}_t, \quad \mathcal{E}_t \overset{iid}{\sim} N(0, \sigma^2_{\mathcal{E}}),$$

with $\sigma^2_{\mathcal{E}} = \frac{(1+a_2)(1-a_1-a_2)(1+a_1-a_2)}{(1-a_2)} = 0.55$, so that $Y_t \sim N(0,1)$ for all $t \in \mathbb{N}$. It holds that $\mathbb{C}\mathrm{orr}[Y_t, Y_{t-1}] = \rho_{t,t-1} = a_1/(1-a_2) = 0.375$, and the partial correlation between Y_t and Y_{t-2} conditional on Y_{t-1} is given by $\rho_{t,t-2;t-1} = a_2 = -0.6$.

We now specify copula families, determine the order and estimate a CMP model for Example 3.2 using Algorithm 3.1. For that purpose, we simulate 1000 observations from the process given in Example 3.2, set $s = 2$ for the degree of Markov order identification, and assume $P = 50$ as an upper bound for the order. For simplicity, we assume that we know that the marginal distribution is normal, in Section 3.6 we discuss the case of an unknown marginal distribution. The fit of a normal distribution is compared with a histogram of the data in the upper plot of the first column in Figure 3.5. Below, the histogram of the estimated probability integral transform is depicted. Both plots suggest that the normal distribution is a reasonable parametric model for the marginal distribution of the data generating process. The estimated parameter $\theta_0 = (\mu, \sigma^2)$ of the fitted normal distribution is given in Table 3.1.

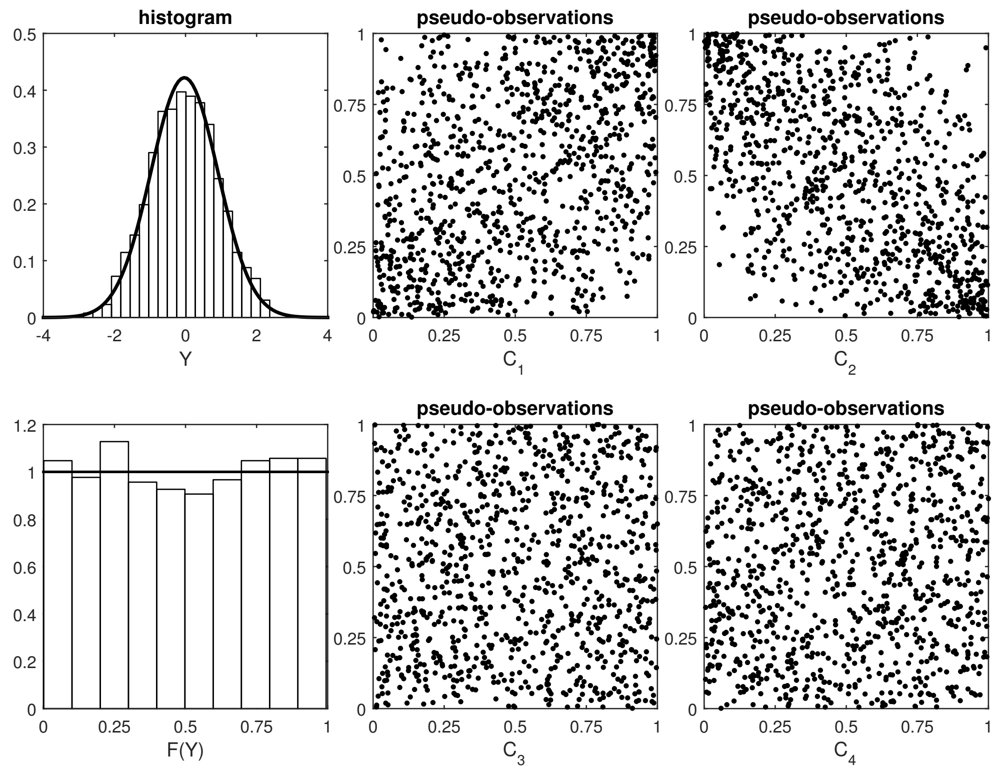


Figure 3.5: Graphical illustration of the fitting procedure of a CMP(2) model for the data generating process given in Example 3.2.

Table 3.1: Results of Algorithm 3.1 if $s = 2, P = 10$, and step 6 and 9 are skipped, which corresponds to the step-by-step ML estimation of the CMP model of Example 3.2.

CMP components	$F(\theta_0)$	$C_1(\theta_1)$	$C_2(\theta_2)$	$C_3(\theta_3)$	$C_4(\theta_4)$
Families and parameters of Example 3.2	$N(0,1)$	$C^{\text{Ga}}(0.375)$	$C^{\text{Ga}}(-0.6)$	$C^{\perp}$	$C^{\perp}$
Estimated CMP(2) model of Example 3.2	$N(0.025, 1.017)$	$C^{\text{Ga}}(0.364)$	$C^{\text{Ga}}(-0.621)$	$C^{\text{Fr}}(-0.06)$	$C^{\text{Cl}}(0.02)$
Average log-likelihoods	-1.436	0.071	0.243	0.00005	0.0002
Decrease in AIC	()	yes	yes	no	no

The possible copula families for C_i are given by the set $\{C^{\text{Ga}}, C^{\text{Fr}}, C^{\text{Cl}}, C^{\text{S-Cl}}, C^{\text{Gu}}, C^{\text{S-Gu}}\}$ for $i = 1, 3, 4$, and by $\{C^{\text{Ga}}, C^{\text{Fr}}, C^{\text{H-Cl}}, C^{\text{V-Cl}}, C^{\text{H-Gu}}, C^{\text{V-Gu}}\}$ for C_2. The AIC is used to select the copula family in Algorithm 3.2. The second row shows the components of the CMP(2) process given in (3.2) and the third row the components of the fitted CMP(4) model. The fourth row displays the average log-likelihood values of each component, i.e., the log-likelihood divided by the sample size $T = 1000$. The last row shows if the additional component of the CMP model results in a decrease of the AIC.

Having estimated the marginal distribution we can now construct the pseudo-observations from the copula of adjacent random variables. The pseudo-observations from this copula are given by $(u_t^{\theta_0}, u_{t-1}^{\theta_0})_{t=2,\ldots,T}$, where $u_t^{\theta_0} = F(y_t; \theta_0)$. A scatter plot of these pseudo-observations is given in the upper picture of the second column in Figure 3.5. The scatter plot of copula realizations indicates a weak positive relationship which seems to be radially symmetric. Consequently, the Gaussian copula, C^{Ga}, or Frank copula, C^{Fr}, seem to be suitable families for our model of C_1. For the purpose of illustration, we also include the Clayton and Gumbel copula as well as their survival versions, which we denote by $C^{\text{Cl}}, C^{\text{S-Cl}}$, and $C^{\text{Gu}}, C^{\text{S-Gu}}$, respectively, in the set of possible families for C_1. We then estimate for each $\tilde{C}_1 \in \{C^{\text{Ga}}, C^{\text{Fr}}, C^{\text{Cl}}, C^{\text{S-Cl}}, C^{\text{Gu}}, C^{\text{S-Gu}}\}$ the parameter θ_1 according to

$$\theta_1 = \underset{\tilde{\theta}_1}{\arg\max}\ \mathcal{L}(\tilde{\theta}_1) = \sum_{t=1}^{T-1} \tilde{c}_1(u_{t+1}^{\theta_0}, u_t^{\theta_0}; \tilde{\theta}_1).$$

and select for C_1 the copula $\tilde{C}_1$ that minimizes the AIC. Table 3.1 shows that a Gaussian copula with $\theta_1 = 0.364$ is selected and that the inclusion of this copula to the CMP model results in a decrease of the AIC in comparison to a CMP(0) model.

We can now construct the pseudo-observations from C_2 which are given by $(u_{t|t-1}^{\theta_{0:1}}, u_{t-2|t-1}^{\theta_{0:1}})_{t=3,\dots,T}$, where $u_{t|t-1} = \partial_2 C_1(u_t^{\theta_0}, u_{t-1}^{\theta_0}; \theta_1)$ and $u_{t-2|t-1} = \partial_1 C_1(u_{t-1}^{\theta_0}, u_{t-2}^{\theta_0}; \theta_1)$. If F and C_1 are correctly specified and the simplifying assumption holds, then these observations are (asymptotically) generated from the distribution of (Y_t, Y_{t-2}) conditional on an arbitrary $Y_{t-1} = y_{t-1}$. If the simplifying assumption does not hold, then these observations are pseudo-observations from $C_{t,t-2;t-1}^{\partial}$ which captures the average dependence between Y_t and Y_{t-2} if the individual influence of Y_{t-1} on Y_t and Y_{t-2} has been eliminated. Even if $C_{t,t-2;t-1}^{\partial}$ does not always completely describe the dependence between Y_t and Y_{t-2} conditional on Y_{t-1}, the scatter plot of the pseudo-observations from $C_{t,t-2;t-1}^{\partial}$ provides valuable information about the underlying bivariate conditional distribution. The upper plot in the third column of Figure 3.5 shows a strong negative relationship between these pseudo-observations and the dependence appears to be radially symmetric. On this basis, the Gaussian and Frank copula should be included in the set of possible models. We also include vertically and horizontally reflected versions of the Clayton and Gumbel copula which allow for negative dependence and which we denote by $C^{\text{H-Cl}}, C^{\text{V-Cl}}$, and $C^{\text{H-Gu}}, C^{\text{V-Gu}}$, respectively. We estimate for each $\tilde{C}_2 \in \{C^{\text{Ga}}, C^{\text{Fr}}, C^{\text{H-Cl}}, C^{\text{V-Cl}}, C^{\text{H-Gu}}, C^{\text{V-Gu}}\}$ the parameter θ_2 according to

$$\theta_2 = \underset{\tilde{\theta}_2}{\arg\max}\ \mathcal{L}(\tilde{\theta}_2) = \sum_{t=1}^{T-2} \tilde{c}_2(u_{t+2|t+1}^{\theta_{0:1}}, u_{t|t+1}^{\theta_{0:1}}; \tilde{\theta}_2),$$

and select for C_2 the copula $\tilde{C}_2$ that minimizes the AIC. Table 3.1 shows that a Gaussian copula with $\theta_2 = -0.621$ is selected and that the inclusion of this copula to the CMP model results in a decrease of the AIC in comparison to a CMP(1) model.

Proceeding in this fashion, we obtain the pseudo-observations from C_3 and C_4 which are depicted in the second and third picture in the second row of Figure 3.5. There is no structure evident in the pseudo-observations from C_3 and C_4. This is also confirmed by the negligible average log-likelihood values of the fitted copulas that are best wrt the AIC. In fact, the AIC of the CMP(3) and CMP(4) model is larger than the AIC of the CMP(2) model. Consequently, we determine the order $p = 2$ since the degree of Markov order identification is $s = 2$. The final parameter estimates of the CMP(2) model are then obtained by a joint ML estimation of the parameters, i.e., by (3.4.2) with $p = 2$. The increase in the log-likelihood and change in estimated parameters is negligible so we do not report the results here. It should be noted that there is a vast difference between the average log-likelihood value for the model of the marginal distribution and the copula models. The value of the log-likelihood of the CMP(2) model is clearly dominated by the marginal distribution. This shows that the modeling of the marginal distribution is a crucial fact.

3.5 Modeling long memory parsimoniously with a lag function

Financial time series are often modeled by time series models that exhibit a rather large or even an infinite Markov order. Regarding our parametric simplified SD-vine copula-based framework this means that we possibly have to estimate a large number of bivariate copulas to reproduce such a long memory behavior. Classical time series analysis imposes a structure on the decay of temporal dependence in order to obtain a parsimonious model with long memory. For instance, a stationary GARCH$(1, 1)$ process is a particular stationary ARCH(∞) process such that the autoregressive parameters of the squared lagged random variables in the variance equation decline exponentially, i.e.,

$$\text{GARCH(1,1)}:\ \sigma_t^2 = \omega + aY_{t-1}^2 + b\sigma_{t-1}^2 \ \Leftrightarrow\ \text{ARCH}(\infty):\ \sigma_t^2 = \omega + \sum_{i=1}^{\infty}(ab^{i-1})Y_{t-i}^2.$$

Using GARCH(p, q) models one can specify a Markov process with infinite order and one also transforms the problem of choosing the order of an ARCH(k) process into the simpler task of choosing the order of an GARCH(p, q) model where p and q are small numbers. We can not directly employ the strategy of GARCH(p, q) models to obtain a parsimonious model with long memory in our copula-based approach because there is no strict analog to the autoregressive coefficients. On the other hand, GARCH(p, q) processes also impose a structure on the partial autocorrelation function of squared random variables,[23] and the partial autocorrelation function corresponds in our copula-based approach to the sequence

[23] Apparently, imposing a structure directly on the partial autocorrelation function of a process has not been considered in the time series literature. It would be interesting to investigate this matter in more detail since, contrary to the sequence of autoregressive coefficients, there seems to be no restriction on the partial autocorrelation function except that it should decay to zero to ensure the stationarity of the process.

$(C_j)_{j\in\mathbb{N}}$. Thus, we propose to impose a structure on the copula sequence $(C_j)_{j\in\mathbb{N}}$ to obtain a parsimonious representation of a higher-order Markov process.

3.5.1 Parameterizing the copula sequence by a lag function

To fix ideas, let $C_j(\theta_j)$ be a bivariate unconditional copula which is parameterized by a scalar parameter θ_j for all $j \in \mathbb{N}$.[24] The parameter θ_j can be the dependence parameter of the copula if all copulas in $(C_j)_{j\in\mathbb{N}}$ belong to the same family. If the copula families in $(C_j)_{j\in\mathbb{N}}$ are different, it is more reasonable to use for θ_j a measure that does not depend on the particular family, e.g., θ_j could be a measure of concordance such as Kendall's tau. We now impose a structure on the sequence $(\theta_j)_{j\in\mathbb{N}}$ by means of a lag function which we define as follows.

Definition 3.7 (Lag function)

The function $g(\cdot;\theta^{\perp},\theta_1,\gamma)$, with parameters $(\theta^{\perp},\theta_1,\gamma) \in \overline{\mathbb{R}} \times \mathbb{R}^{K+1}$, is a lag function if

$$g(\cdot;\theta^{\perp},\theta_1,\gamma)\colon \mathbb{N} \to \Theta \subset \mathbb{R}, \quad g(j;\theta^{\perp},\theta_1,\gamma) =: g_j^{\theta^{\perp}}(\theta_1,\gamma) = \theta_j,$$

so that

$$\lim_{j\to\infty} g_j^{\theta^{\perp}}(\theta_1,\gamma) = \theta^{\perp}.$$

If $\theta^{\perp} = 0$ we set $g_j(\theta_1,\gamma) := g_j^0(\theta_1,\gamma)$.

If $\lim_{\theta_j\to\theta_\perp} C_j(\theta_j) = C^{\perp}$, and $\theta_j \in \Theta$ for all $j \in \mathbb{N}$, we can represent a copula sequence $(C_j)_{j\in\mathbb{N}}$ by the tuple $\big((\mathcal{F}_j)_{j\in\mathbb{N}}, g(\cdot;\theta^{\perp},\theta_1,\gamma)\big)$, where $(\mathcal{F}_j)_{j\in\mathbb{N}}$ denotes the copula families of the copula sequence and g is a lag function with a finite-dimensional vector (θ_1,γ) of small dimension. There are no constraints on the lag function in order that the resulting process is stationary, except that its codomain must be compatible with the values that θ_j can take.[25] However, since we expect that the dependence decreases over time, we assume that the outputs of the lag function eventually decreases to $\theta^{\perp}$ so that the copula sequence converges to the product copula. For instance, if $(C_j)_{j\in\mathbb{N}}$ consists of Student-t copulas with zero correlation parameter and θ_j denotes the degrees of freedom, then $\theta^{\perp} = \infty$. If we further assume that there is a small $k \in \mathbb{N}$ such that the copula families of the sequence $(C_j)_{j\geq k}$ are identical, we can represent the sequence $(C_j)_{j\in\mathbb{N}}$ by a small number of copula families $(\mathcal{F}_j)_{j=1,\dots,k}$, the lag function g and the vector (θ_1,γ). It is also possible to start the lag function at a later lag. For instance, we can model $C_1(\theta_1)$ and $C_2(\theta_2)$ in an unconstrained fashion and then model the remaining copulas $(C_j)_{j\geq 3}$ with a lag function such that $\theta_j = g_{j-2}(\theta_j,\gamma)$ for $j \geq 3$. The following examples present some lag functions that decay to zero, i.e., $\theta^{\perp} = 0$.

[24] It is also possible that θ_j is a vector, but we do not consider this case here.

[25] There definitively would be some constraints on the lag function if it would be imposed on the autoregressive coefficients of a stationary linear autoregressive process.

Example 3.3 (Lag functions that decay to zero)

$$\begin{aligned}
&\text{Geometric/Exponential lag function:} && g_j(\theta_1, \gamma) = \theta_1 \gamma^{j-1}, \quad |\gamma| \leq 1, \\
&\text{Pascal lag function (Solow, 1960):} && g_j(\theta_1, \gamma) = \theta_1 \gamma_1^{j-1} \frac{\binom{j+\gamma_2-1}{j}}{\gamma_2}, \quad |\gamma_1| \leq 1,\ \gamma_2 \geq 0, \\
&\text{Exponential Almon lag function (Ghysels et al., 2007):} && g_j(\theta_1, \gamma) = \theta_1 \frac{\exp(\gamma_1 j + \gamma_2 j^2)}{\exp(\gamma_1 + \gamma_2)}, \quad \gamma_1 \in \mathbb{R},\ \gamma_2 < 0.
\end{aligned}$$

If $\gamma_2 = 1$, then the Pascal lag becomes the geometric lag. Moreover, $g \geq 0$. Note that we have modified the lag functions such that $g_1(\theta_1, \gamma) = \theta_1$. While the geometric lag function is monotonically decreasing to zero, the Pascal and exponential Almon lag function may not be monotonically decreasing and can produce a local maximum that is not attained at the first lag.

The following algorithms summarize the steps for fitting a simplified SD-vine copula-based Markov model which is parameterized by a lag function.

Algorithm 3.4 (Specification and estimation of a CMP model with a lag function)

The specification and estimation of a CMP model that is parameterized by a lag function proceeds as follows. For notational simplicity assume that $\theta^{\perp} = 0$ so that we use the notation $g_j(\theta_1, \gamma)$ for a lag function.

1. *Set an upper bound $P \leq T - 1$ for the order.*
2. *Specify a parametric cdf and use a ML estimation to obtain a model for the marginal distribution of the process.*
3. *Specify the copula families and estimate a CMP(P) model using a step-by-step ML estimation, see Algorithm 3.2.*
4. *Do a joint ML estimation of the specified CMP(P) model, using the estimates of the step-by-step ML estimation as starting values, see Definition 3.6. This step is optional.*
5. *Specify a lag function g such that $g_1(\theta, \gamma) = \theta$, and choose a starting point K for the lag function so that $\theta_j = g_{j-(K-1)}(\theta_j, \gamma)$ for $j = K, \ldots, P$, see Algorithm 3.5.*
6. *Use the estimates from step 4 as starting values for a joint ML estimation of the CMP(P) model so that $\theta_j = g_{j-(K-1)}(\theta_j, \gamma)$ for $j = K, \ldots, P$, i.e.,*

$$(\theta_{0:K}, \gamma) = \underset{(\tilde{\theta}_{0:K}, \tilde{\gamma}) \in \times_{i=0}^{K} \Theta_i \times \Theta_\gamma}{\arg\max} \mathcal{L}_{0:P}(\tilde{\theta}_{0:K-1}, g(\tilde{\theta}_K, \tilde{\gamma})) \tag{3.5.1}$$

where

$$\mathcal{L}_{0:P}(\tilde{\theta}_{0:K-1}, g(\tilde{\theta}_K, \tilde{\gamma})) = \sum_{t=1}^{T} \log f(y_t; \tilde{\theta}_0) + \sum_{j=1}^{K} \sum_{i=1}^{T-j} \log c_j(u_{i+j|i+1:i+j-1}^{\tilde{\theta}_{0:j-1}}, u_{i|i+1:i+j-1}^{\tilde{\theta}_{0:j-1}}; \tilde{\theta}_j)$$
$$+ \sum_{i=1}^{T-(K+1)} \log c_{K+1}(u_{i+j|i+1:i+j-1}^{\tilde{\theta}_{0:K}}, u_{i|i+1:i+j-1}^{\tilde{\theta}_{0:K}}; g_2(\tilde{\theta}_K, \tilde{\gamma}))$$
$$+ \sum_{j=K+2}^{P} \sum_{i=1}^{T-j} \log c_j(u_{i+j|i+1:i+j-1}^{(\tilde{\theta}_{0:K},\tilde{\gamma})}, u_{i|i+1:i+j-1}^{(\tilde{\theta}_{0:K},\tilde{\gamma})}; g_{j-K+1}(\tilde{\theta}_K, \tilde{\gamma})).$$

7. *Obtain the copula sequence with truncated order $p^\star$, see Definition 3.8, and do a joint ML estimation of the truncated CMP model. This step is optional.*

We can also run Algorithm 3.4 for several different lag functions and then choose the truncated CMP that yields the smallest value of an information criteria. In order to select the starting point of the lag function in step 5 of Algorithm 3.4 we apply the following steps.

Algorithm 3.5 (Specification of the lag function)

1. *Determine the earliest possible starting point $k_{\min} \in \mathbb{N}$ for the lag function. E.g., $k_{\min} = \min\{j = 1, \ldots, P \colon \theta_j$ is a scalar$\}$.*

2. *Set the latest possible starting point $k_{\max}$ for the lag function, with $k_{\max} \geq k_{\min}$.*

3. *In order to choose the starting point for the lag function we can either:*

   a) *Investigate the plot of $(\tilde{\theta}_j)_{j=k_{\min},\ldots,k_{\max}}$. The lag function should start at the earliest possible point j such that the variation in the sequence $(\tilde{\theta}_j)_{j=k,\ldots,k_{\max}}$ is smooth.*

   b) *Determine the starting point formally using the following steps.*

      i. *Set $j = k_{\min}$.*

      ii. *If $j = k_{\max}$, set $k = k_{\max}$ and stop, else continue.*

      iii. *Estimate the expected loss in log-likelihood that is generated by using a lag function from the k-th lag on, i.e.,*

$$L_k = \frac{1}{T}\Big(\mathcal{L}_{0:P}(\theta_{0:P}) - \mathcal{L}_{0:P}(\theta_{0:k-1}, g(\theta_k, \gamma))\Big),$$

      where $\mathcal{L}_{0:P}(\theta_{0:P})$ is the log-likelihood function evaluated at the joint ML estimates from (3.4.2) and $\mathcal{L}_{0:P}(\theta_{0:k-1}, g(\theta_k, \gamma))$ is the log-likelihood function evaluated at the joint ML estimates from (3.5.1).

      iv. *If L_k does not exceed a certain threshold δ, set the starting point $k = j$ and end the algorithm, otherwise set $j = j + 1$ and continue with step ii.*

3.5.2 Truncating the order of the copula sequence

By means of a lag function we can, at least in theory, obtain a parsimonious representation of a CMP(p) process with arbitrarily large p. However, there are serious computational reasons why we should not choose the order arbitrarily large in practice. First of all, if we specify a CMP(∞) process and the sample size is T, then we have to evaluate a T-dimensional SD-vine copula to compute the log-likelihood function. In order to evaluate the density of a T-dimensional SD-vine copula, we have to evaluate $(T-1)T/2$ bivariate copula densities and $(T-2)(T-1)$ partial derivatives of bivariate copulas.[26] For instance, if $T = 1000$ then we have to evaluate 499,500 bivariate copula densities and 997,002 partial derivatives of bivariate copulas. The evaluation of the copula densities is rather quick, but the partial derivatives of bivariate copulas are often given by one-dimensional integrals such as in the case of the Gaussian or Student-t copula. Thus, for rather large T the computation of the log-likelihood function is very expensive and ML-estimation can not be accomplished in a reasonable amount of time. Even if the fitting of a CMP(p) process with p being moderately large, say $p = 100$, is feasible in a reasonable amount of time, there are other reasons why we are interested in truncating the order as much as possible, subject to the condition that the truncation does not impair the fit of the model in a significant way.

In applied work, we are often interested in features of the transition distribution, e.g., moments or quantiles. In order to compute moments of the transition distribution we have to approximate the integral

$$\mathbb{E}[Y_t^i|Y_{t-1:t-p}] = \int_{\mathbb{R}} y_t^i f(y_t) \prod_{j=1}^{p} c_j(u_{t|t-1:t-(j-1)}, u_{t-j|t-1:t-(j-1)})\mathrm{d}y_t,$$

where $c_1(u_{t|t-1:t}, u_{t-1|t-1:t}) = c_1(u_t, u_{t-1})$ and $u_{k|t-1:t-(j-1)} = F_{k|t-1:t-(j-1)}(y_k|y_{t-1:t-(j-1)})$ for $k = t, t-j$. Note that $u_{t|t-1:t-(j-1)}$ is a function of y_t and that one evaluation of $u_{t|t-1:t-(j-1)}$ requires the computation of $(j-1)(j-2)+1$ partial derivatives of bivariate copulas for $j = 2, \ldots, p$. Thus, the computation of conditional moments takes considerably more time if the order p is rather large. If the order of the process is p and we use the SD-vine copula for $p+1$ adjacent random variables, one evaluation of the transition distribution requires $1 + (p-1)p$ evaluations of partial derivatives of bivariate copulas. Thus, the number of evaluations that are required for the computation of the transition distribution increases quadratically with the order of the process. This complexity comes into effect when we want to compute functionals of the transition distribution. For instance, the quantile

[26] A D-vine copula is actually the most demanding regular vine copula when it comes to the number of partial derivatives of bivariate copulas that have to be evaluated in order to compute the density. By contrast, a C-vine copula is the cheapest regular vine copula in the sense that it requires only $T(T-1)/2$ evaluations of partial derivatives of bivariate copulas.

function can be expressed as

$$
\begin{aligned}
&F^{-1}_{t|t-1:t-p}(z|y_{t-1:t-p})\\
&= F_Y^{-1} \circ h_1^{-1}\Bigg(h_2^{-1}\Bigg(\dots\big(h_{p-1}^{-1}(h_p^{-1}(z, v_{t-p|t-1}), v_{t-(p-1)|t-1}\big), \dots, v_{t-2|t-1}\Bigg), v_{t-1|t-1}\Bigg),
\end{aligned}
$$

where $v_{t-j|t-1} = u_{t-j|t-1:t-(j-1)}$, and h_j^{-1} is the inverse of $\partial_2 C_j$ wrt to the first argument for $j = 1, \dots, p$. The computation of the quantile function in this fashion requires $(p-1)^2$ evaluations of partial derivatives of bivariate copulas to compute $v_{t-j|t-1}$ for all $j = 1, \dots, p-1$, and p evaluations of inverses of partial derivatives of bivariate copulas to compute h_j^{-1} for all $j = 1, \dots, p$. The evaluation of the quantile function of the transition distribution is also required for simulating trajectories of the process. In particular, for $t \geq p+1$, $(p-1) + \min(p-2, 0)$ evaluations of partial derivatives of bivariate copulas and p evaluations of inverses of partial derivatives of bivariate copulas are required to simulate one observation. Consequently, if the order is large, decreasing the order by one unit results in a substantial decrease of effective calculation time. Simulation is also required for multi-step ahead forecasting or for computing risk measures of a portfolio if each asset is modeled as a CMP. Therefore, the determination of an order that effectively balances the trade-off between a fast computation of the transition distribution and a loss in the goodness of the model fit is of great importance for applied work.

Information criteria can not be utilized to choose the truncated order. The log-likelihood value decreases if the order is reduced, but the number of parameters does not change if a lag function is used, so we should not truncate at all according to any information criteria. There is also no statistical reason why we want to truncate the order of the process, i.e., we do not want to reduce the number of parameters in order to obtain a less complex model wrt the degrees of freedom or to increase the estimation precision, so statistical insignificance of lags is not a good argument to truncate these lags in this case.[27] Instead, the motivation for truncation is a purely deterministic one, namely, to reduce the computational complexity while maintaining a model which is indistinguishable from the non-truncated specification in a deterministic – but not a statistical – sense.

In order to determine the order of the truncated model, we have to define the meaning of equality in this framework. Since we are interested in modeling a transition distribution, it is natural to determine the order of the truncated model by considering its KL divergence from the non-truncated model. Let $D_{KL}(G||H)$ denote the KL divergence of H from G. The specification of a lag function implies that we expect that the "importance" of a lag

[27] For instance, there is no statistical reason to truncate the order of the AR(∞) representation of an ARMA(p,q) process. Statistical reasons are also only a concern when the sample size is small and sampling uncertainty is really an issue. In financial econometrics the sample size of time series is often a four-digit number so lags can be statistically significant but not substantive for the application. It is also not clear what significance level should be chosen, but different significance levels might result in very different truncated orders.

on the transition distribution eventually decreases the further it is located in the past, i.e., $\exists l \in \mathbb{N}, \forall k \geq l\colon D_{KL}(F_{t|t-1:t-k-1}||F_{t|t-1:t-k})$ is decreasing in k. For this reason, we recommend to choose the order of the truncated process by

$$p^\star := \inf\Big\{k = 1, \ldots, P\colon D_{KL}(F_{t|t-1:t-P}||F_{t|t-1:t-k}) \leq \Delta\Big\},$$

where P is the maximal order of the process which has to be specified a priori and Δ is a small positive number so that models with a KL divergence smaller than Δ are considered as equal. Before we address the choice of Δ we note that

$$D_{KL}(F_{t|t-1:t-P}||F_{t|t-1:t-k}) = \sum_{j=k+1}^{P} \mathbb{E}\big[\log c_j(U_{t|t-1:t-j+1}, U_{t-j|t-1:t-j+1})\big].$$

Thus,

$$p^\star := \inf\Big\{k = 1, \ldots, P\colon \sum_{j=k+1}^{P} \mathbb{E}\big[\log c_j(U_{t|t-1:t-j+1}, U_{t-j|t-1:t-j+1})\big] \leq \Delta\Big\},$$

i.e., we choose the order of the truncated model such that the aggregated mutual information of copulas that should be set to product copulas is smaller than Δ. Since the expected log-likelihoods of the copula densities are unknown in practice, we estimate them using the sample means of the respective log-likelihoods.

Obviously, the crucial point in choosing the order of the truncated process is the threshold Δ. Similar to the choice of the significance level of a statistical test, the definite choice of a particular threshold can not be justified by theory and has to be specified by the researcher. However, we can provide some reference point for the choice of Δ. It is common usage to approximate a univariate Student-t distribution with degrees of freedom ν equal to or larger than 30 by a standard normal distribution. If $T(30)$ denotes the Student-t distribution with dof 30 then

$$D_{KL}(T(30)||N(0, 1)) \leq 0.0021.$$

Numerical computations show that if $\nu = 30$, the maximum error in approximating the cdf of a Student-t distribution with a normal distribution is smaller than 0.00525. For many applications of the Student-t distribution this rule of thumb appears to be appropriate. Therefore, we choose $\Delta = 0.0021$ as the threshold for the truncation of the order. In terms of KL divergence, the resulting truncated CMP model is then at least so close to the original CMP model as a standard normal distribution is close to the Student-t distribution with dof 30.

The following definition summarizes our strategy for truncating the order of a simplified SD-vine copula-based Markov process.

Definition 3.8 (Copula sequence with truncated order $p^\star$)

Let $g_j(\theta,\gamma)$ be a lag function and $\big(C_j(g_j(\theta,\gamma))\big)_{j=1,\ldots,P}$ be a copula sequence which is parameterized by g. Let $\Delta \geq 0$ and

$$p^\star := \inf\Big\{k=1,\ldots,P\colon \sum_{j=k+1}^{P} \frac{1}{T-j} \sum_{t=j+1}^{T} \log c_j(u_{t|t-1:t-j+1}, u_{t-j|t-1:t-j+1}) \leq \Delta\Big\},$$

where we have suppressed the dependence of c_j and its arguments on the estimated lag function $g_j(\theta,\gamma)$. The copula sequence with truncated order $p^\star$ is given by $(C_j)_{j=1,\ldots,p^\star}$.

3.5.3 Lag function and truncation: Illustration

Consider the following ARMA(1,1) process,

$$Y_t = 0.99Y_{t-1} - 0.8\mathcal{E}_{t-1} + \mathcal{E}_t, \quad \mathcal{E}_t \overset{iid}{\sim} N(0,1), \tag{3.5.2}$$

which exhibits a rather long memory. We now simulate 1000 observations from this ARMA(1,1) process. The left panel of Figure 3.6 shows that the sample autocorrelation function is slowly decaying. More importantly, the right panel of Figure 3.6, which depicts the sample partial autocorrelation function, reveals that an appropriate order for a Markov model is also rather large.

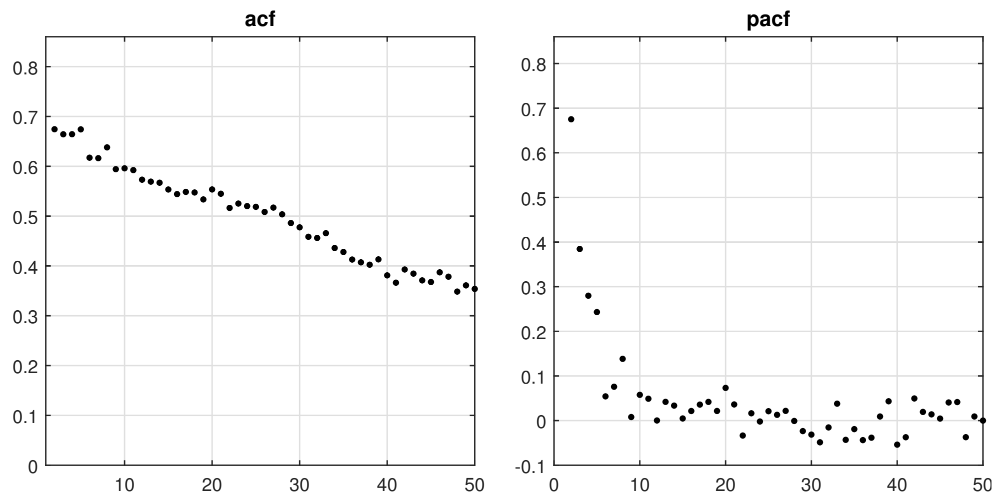


Figure 3.6: Sample (partial) autocorrelation function of 1000 observations from the ARMA(1,1) process given in (3.5.2).

We now specify a normal distribution for F, Gaussian copulas for $(C_j)_{j=1,\ldots,50}$, and estimate a CMP(50) model without using a lag function. The estimated dependence parameters $(\theta_j)_{j=1,\ldots,50}$ of the Gaussian copulas and the estimated values of Kendall's tau $(\tau_j)_{j=1,\ldots,50}$ are depicted as dots in Figure 3.7. We observe that the values of both estimated sequences scatter around zero after the 15th lag. Indeed, we obtain an estimated

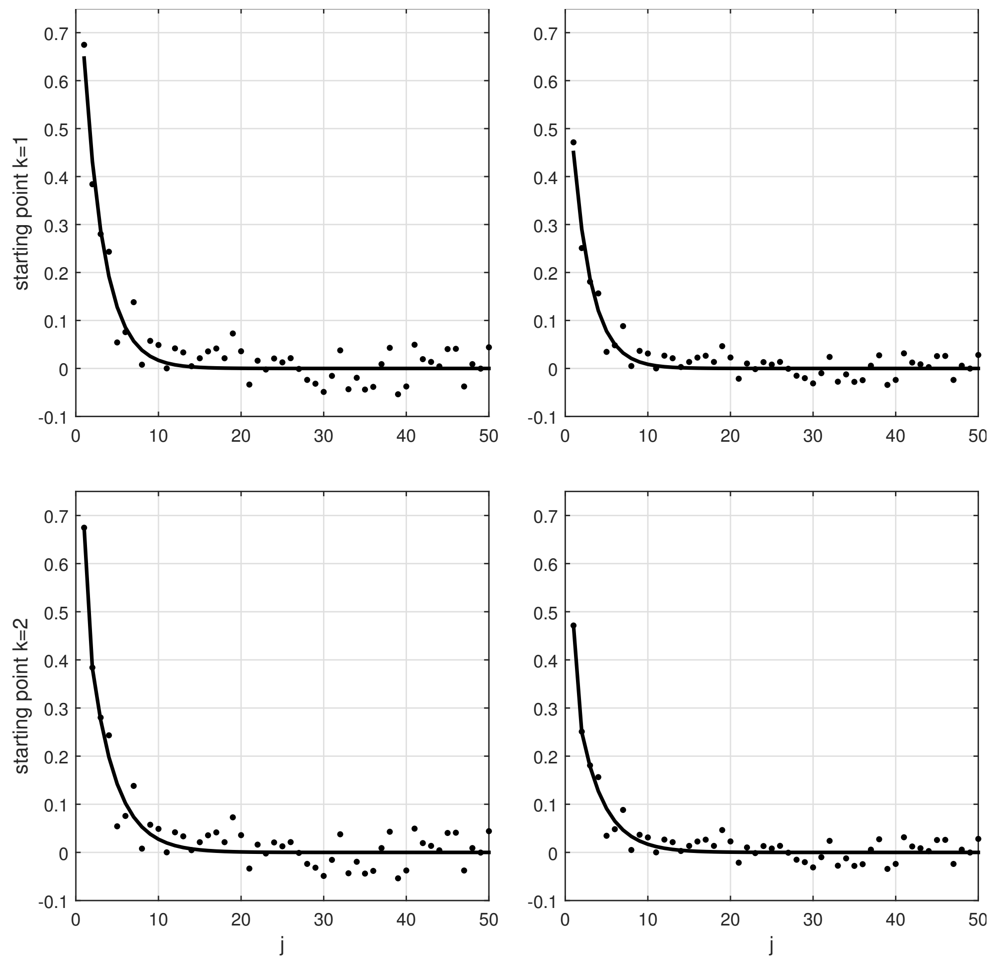


Figure 3.7: Graphical illustration of an unconstrained CMP(50) model and a CMP(50) model which is parameterized by a lag function with starting point $k = 1, 2$. The dots in the left plots show the sequence $(\theta_j)_{j=1,\ldots,50}$, which has been estimated using steps 1-4 in Algorithm 3.4. The dots in the right plots show the corresponding sequence $(\tau_j)_{j=1,\ldots,50}$, where τ_j is the estimated Kendall's tau of the j-th element of $(C_j(\theta_j))_{j=1,\ldots,50}$. The lines in all plots refer to the corresponding $(\theta_j^{g_k})_{j=1,\ldots,50}$ or $(\tau_j^{g_k})_{j=1,\ldots,50}$ sequences which are implied by fitted exponential lag functions using step 5 and 6 of Algorithm 3.4. In the upper plots the lag functions start with the first lag while in the lower plots the lag functions start with the second lag.

CMP(7) model if we assume $s = 4$ for the degree of order identification and use the BIC criterion to select the order.

We now estimate a CMP(50) model using an exponential lag function which starts either with the first or with the second lag. Define $\theta_j^{g_k} = g_{j-(k-1)}(\theta_k, \gamma)$ and $\tau_j^{g_k} = g_{j-(k-1)}(\tau_k, \gamma)$. The lines in Figure 3.7 represent the estimated $(\theta_j^{g_k})_{j=1,\ldots,50}$ and $(\tau_j^{g_k})_{j=1,\ldots,50}$ sequences that are implied if an exponential lag function is used to parameterize the CMP process. In the upper panels, the lag functions start with the first lag, i.e., all elements of $(\theta_j^{g_1})_{j=1,\ldots,50}$ or $(\tau_j^{g_1})_{j=1,\ldots,50}$ are determined by the two parameters of an exponential lag function. In the lower panels, the lag functions start with the second lag, i.e., $(\theta_j^{g_2})_{j=2,\ldots,50}$ or $(\tau_j^{g_2})_{j=2,\ldots,50}$ are determined by two parameters of an exponential lag function but the parameter θ_1 or τ_1 is estimated without any constraints. We observe a small but visible discrepancy between the unconstrained estimated value of θ_1 or τ_1 and the implied value of $\theta_1^{g_1}$ or $\tau_1^{g_1}$ if the lag functions start with the first lag. If the lag functions start with the second lag, there is no visible difference between the first values of the respective sequences. Thus, setting the starting point for the lag function to two seems to be a good idea. If we set the threshold for the truncated copula sequence in Definition 3.8 to $\Delta = 0.0021$, we obtain the truncated order $p^\star = 7$, or $p^\star = 10$, if the lag functions start with the first lag, or with the second lag, respectively.

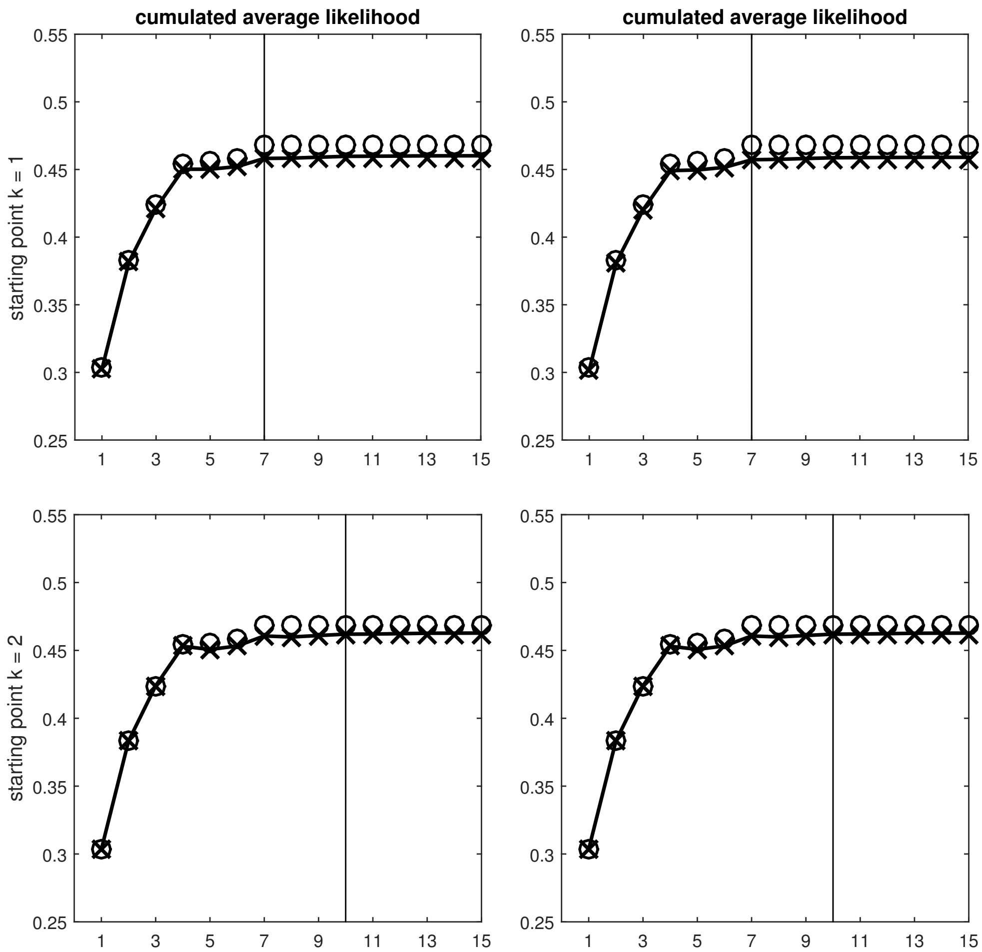


Figure 3.8: The circles display the first 15 cumulated average log-likelihoods of the CMP(7) processes which were estimated without a lag function, i.e., the j-th element is given by $\frac{1}{1000}\sum_{i=1}^{j}\mathcal{L}_i(\psi_i), \psi_i \in \{\theta_i, \tau_i\}$. The solid lines display the first 15 cumulated average likelihoods of the CMP(8) processes with a lag function, i.e., the j-th element is given by $\frac{1}{1000}\sum_{i=1}^{j}\mathcal{L}_i(\psi_i^{g_k}), \psi_i^{g_k} \in \{\theta_i^{g_k}, \tau_i^{g_k}\}$. The marks indicate the first 15 cumulated average likelihoods that result from the truncated copula sequence that is obtained via Definition 3.8 for $\Delta = 0.0021$. The truncation point $p^\star$ is indicated by a vertical line. In the upper panel the decay functions start with the first lag while in the lower panel the decay functions start with the second lag.

The cumulated average log-likelihoods of all fitted copula sequences are displayed in Figure 3.8. If the lag function starts with the first lag, the cumulated average log-likelihoods, that are implied by the lag function, are smaller than the cumulated average log-likelihoods that result from an unconstrained estimation of the CMP model. However, if the lag functions start with the second lag, the differences between the cumulated average log-likelihoods are decreased. Moreover, the CMP$(C_1(\theta_1), (C_j(\theta_j^{g_2})_{j=2,\ldots,10}))$ model, that is given by the lag function and a truncated copula sequence, outperforms the unconstrained CMP$((C_j(\theta_j)_{j=1,\ldots,7})$ model in terms of BIC. While the unconstrained CMP(7) model requires 7 parameters for the copula sequence, the CMP(10) model with a lag function requires only 3 parameters.

3.6 Modeling the marginal distribution

Since the data for a copula is typically not observed in practice, a copula is typically fitted to pseudo-observations which are constructed by applying an estimator of the marginal distribution to the data. If the model of the marginal distribution is misspecified and we provide the "wrong" pseudo-observations for the copula, the estimated copula model will also be misspecified wrt to the joint distribution. Therefore, the specification and estimation of the marginal distribution and the evaluation of its goodness-of-fit is of primary importance.

To circumvent this problem, empirical rank transformations are often considered in theory and practical applications of copulas (Genest and Favre, 2007). In these cases one is solely interested in the dependence structure that is captured by the copula and the marginal distribution is treated as a nuisance parameter. However, in our case the marginal distribution is not a nuisance parameter since we are interested in modeling the transition distribution of a stochastic process which depends on the copula and the marginal distribution. The use of empirical rank transformations results in a non-continuous estimate of the transition distribution. This may be an undesired feature from a practical point of view, e.g., for the computation of extreme quantiles, which are important for risk analysis, a smooth transition distribution may be more appropriate. Therefore, we use continuous parametric models for the marginal distribution to obtain a smooth transition distribution.[28]

The estimation of the marginal distribution of financial time series with parametric models has been conducted by several authors (see Behr and Pötter, 2009, and the references therein). Francq and Zakoïan (2013) provide a rigorous treatment of the asymptotic distribution of the parameter estimators if the dynamics of the data are neglected. While there is a vast literature on goodness-of-fit test for the marginal distribution in the i.i.d. case and for the distributional form of the innovation sequence that drives the stochastic process, goodness-of-fit tests for the marginal distribution of weakly dependent processes have only recently been introduced. So far, L_2-type statistics using kernel estimators (Fan and Ullah, 1999; Neumann and Paparoditis, 2000) or empirical characteristic functions (Leucht, 2012), moment-based tests (Bai and Ng, 2005), wavelets (Bochkina, 2007), and Neyman's smooth test (Ignaccolo, 2004; Munk et al., 2011), have been analyzed in order to check the appropriateness of the parametric marginal distribution of a time series model.

[28] One could also apply kernel or other nonparametric estimators to obtain a smooth estimated marginal distribution. However, a brief analysis indicates that the implied transition density is not smooth enough if the smoothness of the marginal density is only based on the marginal density and not on the transition density. It is an open question how to choose the bandwidth in order that the implied transition density is smooth enough.

The tests based on L_2-distances are computationally demanding as critical values can only be obtained via the bootstrap while the wavelet goodness-of-fit exhibits a multiple testing problem. The moment-based test of Bai and Ng (2005) is computationally fast but can only be applied to check the skewness and kurtosis of the marginal distribution. Moreover, simulations indicate that the test has size problems for a finite number of observations. Neyman's smooth test is simple and the asymptotic distribution of the test statistic only requires the estimation of a long-run covariance matrix. For the i.i.d. case it has also been shown that Neyman's smooth test exhibits substantial better overall performance against a broad range of alternatives than the popular Kolmogorov-Smirnov or Cramer-von Mises tests (Ledwina and Inglot, 1996; Rayner et al., 2009). On these grounds, we apply Neyman's smooth test in Chapter 4 and Chapter 6 in order to evaluate the fit of estimated parametric models of the marginal distribution.

Neyman's smooth test for weakly dependent data

Assume that an ergodic stationary process has marginal distribution F. We want to test the following hypothesis

$$H_0 : Y_t \sim F \text{ vs. } H_1 : Y_t \nsim F.$$

Another way of testing this hypothesis, which can be applied to processes with continuous cdf F, is to utilize the probability integral transform and reduce the problem to testing a distribution on $[0, 1]$. In this sense, Neyman's smooth test considers the following hypothesis

$$H_0\colon\ U_t := F(Y_t) \sim U(0,1) \text{ vs. } H_1\colon\ U_t := F(Y_t) \nsim U(0,1).$$

Note that it is assumed that F is a known cdf, i.e., there is no estimation uncertainty about F.[29] The test statistic of Neyman's smooth test is given by

$$R_k = \sum_{j=1}^{k} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_j(U_t) \right\}^2, \quad k \in \mathbb{N},$$

where $(\phi_j)_{j \in \mathbb{N}}$ is an orthonormal system in $L_2[0,1]$ with $\phi_0(x) = 1$ and k is the number of components. In most cases, the shifted and normalized[30] Legendre polynomials,

$$\phi_j(x) = \sqrt{2j+1}(-1)^j \sum_{k=0}^{j} \binom{j}{k} \binom{n+k}{k} (-x)^k,$$

[29] This might result in an under-rejection of a wrong null hypothesis (McCulloch and Percy, 2013) for smaller sample sizes.

[30] Normalized because these Legendre polynomials have unit length, i.e., $\int_0^1 \phi_j(x)^2 \mathrm{d}x = 1$.

are chosen for $(\phi_j)_{j\in\mathbb{N}}$ and we also consider them in the following. Since $\mathbb{E}[\phi_j(U_t)] = \int_0^1 \phi_j(x_t)\mathrm{d}x_t = \int_0^1 \phi_j(x_t)\phi_0(x_t)\mathrm{d}x_t = 0, j \geq 1$, we observe that $\mathbb{E}[\phi_j(U_t)] = 0, j \geq 1$. Thus, we expect that R_k is not too far away from zero if H_0 is true. Ignaccolo (2004) investigates Neyman's smooth test for strictly-stationary α-mixing data when the number of components is fixed. Under some regularity conditions (Ignaccolo, 2004, Theorem 3.2), the test statistic converges under H_0 to a generalized chi-square distribution with one degree of freedom (Mathai and Provost, 1992) , i.e., if $X_i \overset{iid}{\sim} \chi_1^2$, then

$$R_k \xrightarrow{d} \sum_{i=1}^{k} \lambda_i^{(k)} X_i,$$

where the coefficients $\lambda_i^{(k)}$ are the eigenvalues of the matrix

$$\Sigma^{(k)} = \Sigma_0^{(k)} + \sum_{j=2}^{\infty} \Sigma_j^{(k)} + (\Sigma_j^{(k)})',$$

where

$$\Sigma_j^{(k)} = \mathbb{E}[\Phi_t \Phi_{t-j}'], \quad \Phi_t = (\phi_1(U_t), \ldots, \phi_k(U_t))'.$$

Since the eigenvalues of the matrix $\Sigma^{(k)}$ are unknown, we apply a consistent estimator of Σ which is given by

$$\widehat{\Sigma}^{(k)} = \widehat{\Sigma}_0^{(k)} + \sum_{j=2}^{m} w(j,m)(\widehat{\Sigma}_j^{(k)} + (\widehat{\Sigma}_j^{(k)})')$$

with

$$\widehat{\Sigma}_j^{(k)} = \frac{1}{T-j} \sum_{t=j+1}^{T} \Phi_t \Phi_{t-j}',$$

where $m < T$ and $w(j,m)$ is a weighting function that ensures that $\widehat{\Sigma}^{(k)}$ is positive definite, e.g., $w(j,m) = 1 - j(m+1)$ are the Bartlett weights for the Newey-West estimator (Newey and West, 1987).[31] From $\widehat{\Sigma}^{(k)}$ we obtain consistent estimators for the eigenvalues due to continuity. Note that if $k = 1$ the asymptotic distribution of Neyman's smooth test is gamma distributed with shape parameter 0.5 and scale parameter 2λ. For $k > 1$ we can use the method of Davies (1980) which is based on characteristic functions.

In order to choose the number of components Ledwina (1994) and Kallenberg and Ledwina (1997) establish a consistent data-driven selection using a (modified) Schwarz's selection rule for the case of independent data. Recently, Munk et al. (2011) generalize their approach and establish a data-driven version of Neyman's smooth goodness-of-fit test for weakly dependent data. The number of components $\hat{k}$ is selected by means of a modified

[31] m is selected following Newey and West (1994).

Schwarz's criterion, i.e.,

$$\hat{k} = \min\left\{k : 1 \leq k \leq K_{\max},\ R_k - k\log T \geq R_j - j\log T,\ j = 1,\ldots,K_{\max}\right\},$$

where $K_{\max}$ is the maximal number of components which have to be specified a priori. Munk et al. (2011) show that under H_0 and some regularity conditions

$$\lim_{T\to\infty} P(\hat{k} = 1) = 1$$

also holds in the dependent case, i.e., asymptotically one component is selected with probability one if H_0 is true. It follows that under H_0

$$R_{\hat{k}} \xrightarrow{d} \Gamma(0.5, 2\lambda_1),$$

where Γ denotes the gamma distribution and $\lambda_1 = \sum_{i\in\mathbb{Z}} \mathbb{E}[\phi_1(U_t)\phi_1(U_{t-i})] = 12\sum_{i\in\mathbb{Z}} \mathbb{C}\mathrm{ov}[U_t, U_{t-i}]$.

Thus, under the null hypothesis, the asymptotic distribution of Neyman's smooth test is a gamma distribution. However, if the null hypothesis is false then the number of estimated components $\hat{k}$ will be equal or larger than one in finite samples and also asymptotically. Moreover, if the number of components is estimated by means of a Schwarz's selection rule, the asymptotic distribution of the test statistic is no longer a linear combination of k χ^2 random variables. Consequently, we still have to specify the number of components a priori to obtain a strictly valid testing procedure. Simulations indicate that $k = 4$ is a good choice. In this case, Neyman's smooth test checks constraints that are related to the first four moments of the distribution of the estimated probability integral transform.

3.7 Dependence properties of SD-vine copula-based Markov processes

Financial time series often exhibit a positive autocorrelation function, so it is an interesting question whether we can say something about the correlation structure of our copula-based process. Moreover, in some applications it might be reasonable to expect that the mean of the transition distribution is an increasing function in all lagged values. If this is the case, we say that the mean of the transition distribution is increasing. For first-order CMP processes, Joe (1997, Theorem 8.3) derives sufficient conditions for the copula of adjacent random variables so that the process exhibits a (decaying but) positive autocorrelation function. One can also show that these conditions imply that the mean of the transition distribution is increasing. In the following, we investigate to what extent we can generalize these conditions to higher-order Markov processes. The general case of copula-based Markov processes is discussed first, and then SD-vine copula-based Markov processes are investigated.

3.7.1 Positive autocorrelation function and increasing mean of the transition distribution

A condition that implies for any marginal distribution a positive autocorrelation function of a process is that its copula $C_{1:t}$ is (positively) associated for all $t \in \mathbb{N}$, i.e., $\mathbb{C}\text{ov}[g_1(U_{1:t}), g_2(U_{1:t})] \geq 0$ for all real-valued functions g_1 and g_2 which are increasing in each component and such that the covariance exists (cf. Joe, 1997, Section 2.1.4). In practice, the property of association is not very useful since it is hard to check for a specific copula. But if a copula $C_{t:1}$ or the corresponding random vector $U_{t:1}$ is conditional increasing in sequence (cf. Joe, 1997, Section 2.1.2), which we denote by CIS($U_{t:1}$), the association property for the copula $C_{1:t}$ follows. A copula has the CIS($U_{t:1}$) property if, for all $j = 2, \ldots, T$, we have that SI($U_j|U_{j-1:1}$), i.e, $F_{j|j-1:1}(x_j|x_{j-1:1})$ is decreasing in each element of $x_{j-1:1}$ for all x_j. The following proposition shows that not only the CIS($U_{p+1:1}$) property of $C_{1:p+1}$ is sufficient to obtain a positive autocorrelation function of a p-th order Markov process but also that all bivariate copulas $C_{1,j}$ have the stronger SI($U_j|U_1$) property.

Proposition 3.6 (SI property for bivariate copulas)

Let Y be a stationary p-th order Markov process and let $C_{1:p+1}$ have the CIS($U_{p+1:1}$) property. Then $C_{1,j}$ has the SI($U_j|U_1$) property for all $j \geq 2$.

Proof. See Appendix A.3. ■

Corollary 3.2 (Sufficient condition for a non-negative autocorrelation function)

Let Y be a stationary p-th order Markov process. If $C_{1:p+1}$ has the CIS($U_{p+1:1}$) property, the autocorrelation function of Y is non-negative.

The CIS property is a very strong dependence property (see Müller and Scarsini (2005) for sufficient and necessary conditions for an Archimedean copula to have the CIS property) and also implies that the mean of the transition distribution is increasing as the following lemma demonstrates.

Lemma 3.1 (Increasing conditional expectation)

Let $C_{1:K}$ have the SI($U_1|U_{2:K}$) property, i.e., for all $u_1 \in (0, 1)$ we have that $F_{U_1|U_{2:K}}(u_1|u_{2:K})$ is (a.s.) decreasing in each component of $u_{2:K}$. Then $g(y_{2:K}) := E[Y_1|Y_{2:K} = y_{2:K}]$ is increasing in each variable if $Y_{1:K}$ has the copula $C_{1:K}$.

Proof. See Appendix A.4. ■

Thus, irrespective of the marginal distribution, the mean of the transition distribution of Y is increasing if $C_{1:p+1}$ has the SI($U_{p+1}|U_{p:1}$) property.

The CIS or SI property for SD-vine copula-based Markov processes

We now investigate under what conditions the SD-vine copula $C_{1:p+1}$ can exhibit the CIS$(U_{p+1:1})$ or SI$(U_{p+1}|U_{p:1})$ property. In particular, we want to work out if we can impose any conditions on the bivariate copulas such that the resulting SD-vine copula has CIS$(U_{p+1:1})$ or SI$(U_{p+1}|U_{p:1})$ property. For simplicity, and without loss of generality, we consider the case $p = 2$ and a simplified SD-vine copula. The transition distribution of the second-order Markov process is then given by

$$F_{t|t-1,t-2}(y_t|y_{t-1}, y_{t-2}) = h_2\big(F_{t|t-1}(y_t|y_{t-1}), F_{t-2|t-1}(y_{t-2}|y_{t-1})\big),$$

where $h_2 = \partial_2 C_2$. Since $F_{t-2|t-1}$ is increasing in y_{t-2} it follows that, if and only if C_2 has the SI$(U_{t|t-1}|U_{t-2|t-1})$ property then $F_{t|t-1,t-2}(y_t|y_{t-1}, y_{t-2})$ is decreasing in y_{t-2} for all $(y_t, y_{t-1}) \in \mathbb{R}^2$. Moreover, the proof of Lemma 3.1 reveals that $\mathbb{E}[Y_t|Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-1}] := g(y_{t-1}, y_{t-2})$ is then increasing in y_{t-2} for all $y_{t-1} \in \mathbb{R}$. Thus, we can find simple sufficient and necessary conditions such that the transition distribution is decreasing in the second lag and that the mean of the transition distribution is increasing in the second lag.

Unfortunately, this is not the case for the first lag. Increasing y_{t-1} has an effect on $F_{t|t-1}$ and $F_{t-2|t-1}$ which constitute the arguments for the function h_2. We first consider the effect of y_{t-1} on h_2 through $F_{t|t-1}$. Since $h_2 = F_{U_{t|t-1}|U_{t-2|t-1}}$ is a conditional cdf it is increasing in its first argument in any case. In order to prevent that the first argument $F_{t|t-1}$ of h_2 increases if y_{t-1} increases, we have to assume that C_1 has the SI$(U_t|U_{t-1})$ property. Let us now consider the effect of y_{t-1} on h_2 through $F_{t-2|t-1}$. Recall that the transition distribution is increasing in y_{t-2} if and only if h_2 has the SI$(U_{t|t-1}|U_{t-2|t-1})$ property. Consequently, if h_2 has the SI$(U_{t|t-1}|U_{t-2|t-1})$ property, we have to assume that C_1 has the SD$(U_{t-1}|U_t)$ property, i.e., $F_{t-2|t-1}(\cdot|y_{t-1})$ is increasing in y_{t-1}, to ensure that the second argument $F_{t-2|t-1}$ of h_2 increases if y_{t-1} increases. Otherwise, h_2 might be increasing if y_{t-1} increases. However, C_1 can only exhibit both the SI$(U_t|U_{t-1})$ and the SD$(U_{t-1}|U_t)$ property if and only if $C_1 = C^{\perp}$. This follows because the SI property implies a non-negative correlation and the SD property implies a non-positive correlation.

As a result, we can only assume that C_1 either has the SI$(U_t|U_{t-1})$ or the SD$(U_{t-1}|U_t)$ property. In both cases, it is not clear if the transition distribution is decreasing in y_{t-1} since we observe two counteracting effects. For instance, if C_1 has the SI$(U_t|U_{t-1})$ property, an increase in y_{t-1} decreases the first argument $F_{t|t-1}$ of h_2, so that h_2 decreases on the one hand. But we can not ensure that the second argument $F_{t-2|t-1}$ of h_2 also increases, which would only be true if $C_{t,t-1}$ has the SD$(U_{t-1}|U_t)$ property. Thus, an increase in y_{t-1} can decrease the second argument $F_{t-2|t-1}$ of h_2, so that h_2 can be increasing on the other hand if it has the SI$(U_{t|t-1}|U_{t-2|t-1})$ property. Therefore, the answer to the question whether the transition distribution is decreasing in the first lag y_{t-1} depends on whether the decrease in $F_{t|t-1}$ is sufficiently large enough relative to the decrease in $F_{t-1|t-2}$.

If C_1 and C_2 are Gaussian copulas with positive correlation parameter, then this is obviously true since the resulting copula $C_{1:3}$ has the CIS($U_{3:1}$) property. If $C_1(\theta_1)$ and $C_2(\theta_2)$ belong to the family of FGM copulas then elementary computations show that

$$F_{U_3|U_2,U_1}(u_3|u_2,u_1) = u_3\big\{\theta_2[u_3(\theta_1(1-2u_2)(1-u_3)+1)-1] \\ \times [u_1(\theta_1(1-2u_1)(1-u_1)+1)-1]+1\big\}(\theta_1(1-2u_2)(1-u_3)+1),$$

which is decreasing in u_2 and u_1 for all $u_3 \in [0,1]$ if θ_1 and θ_2 are non-negative, which is equivalent to C_1 having the SI($U_t|U_{t-1}$) and C_2 having the SI($U_{t|t-1}|U_{t-2|t-1}$) property (cf. Joe (1997), p. 148). However, in general, $C_{t:t-2}$ does not exhibit the CIS($U_{t:t-2}$) property if C_1 has the SI($U_t|U_{t-1}$) and C_2 has the SI($U_{t|t-1}|U_{t-2|t-1}$) property as it is illustrated in Figure 3.9.

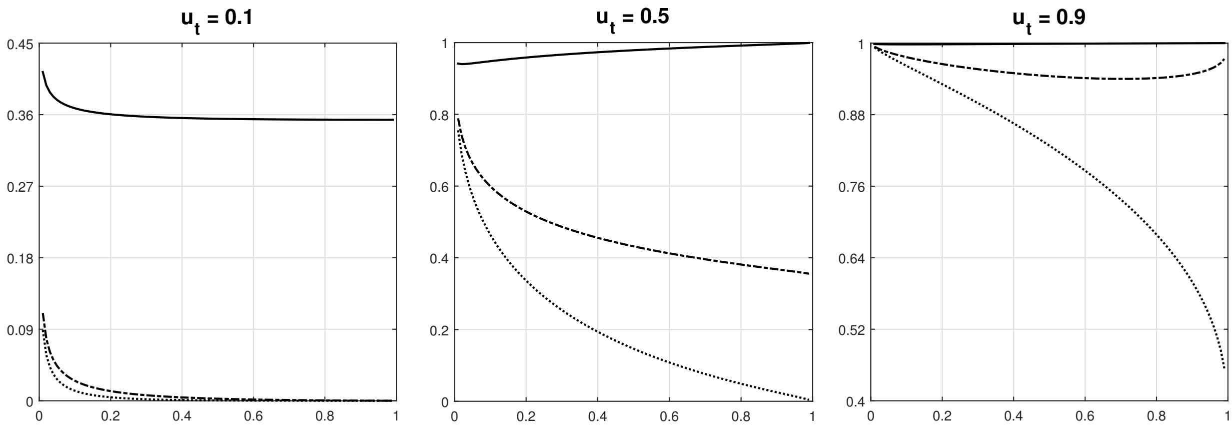


Figure 3.9: Implied transition distribution for a CMP(2) if $C_1 = C^{\mathrm{Ga}}(0.5)$ and $C_2 = C^{\mathrm{Cl}}(2)$ as a function of u_{t-1} for different values of u_t and u_{t-1}. The solid, dashed, and dotted lines represent the transition distribution if $u_{t-2} = 0.1, 0.5, 0.9$, respectively.

Although the Gaussian copula C_1 with correlation parameter 0.5 has the SI($U_t|U_{t-1}$) property and the Clayton copula C_2 with parameter 2 has the SI($U_{t|t-1}|U_{t-1|t-2}$) property (cf. Joe, 1997, pp. 140-141), Figure 3.9 clearly shows that the transition distribution is not decreasing in u_{t-1} if $(u_t, u_{t-2}) = (0.5, 0.9)$ or if $(u_t, u_{t-2}) = (0.9, 0.5)$. Note that the Gaussian and Clayton copula both have a TP2 density in this case (cf. Joe, 1997, pp. 140-141), so even the strongest positive dependence property is not sufficient for a transition distribution that is decreasing in each argument. Moreover, Figure 3.10 shows that $\mathbb{E}[U_t|U_{t-1} = U_{t-1}, U_{t-2} = u_{t-2}] = g(u_{t-1}, u_{t-2})$ is not decreasing in u_{t-1} for all $u_{t-2} \in (0,1)$, i.e., it might happen that the mean of the transition distribution is not decreasing in the first lagged variable.

Similar effects for the transition distribution can be observed if we replace C_1 and C_2 with other copulas that have the respective SI property or a TP2 density. For instance, setting C_1 and C_2 to the Frank copula also shows that the transition distribution is not always decreasing in the first lagged variable. In general, the likelihood of transition distributions

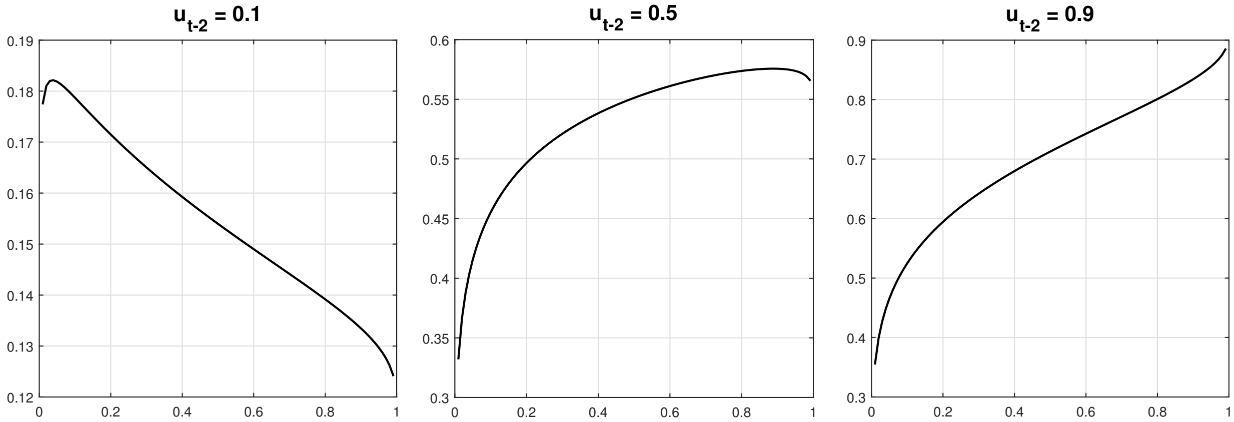


Figure 3.10: Implied mean of a transition distribution for a CMP(2) if $C_1 = C^{\mathrm{Ga}}(0.5)$ and $C_2 = C^{\mathrm{Cl}}(2)$ as a function of u_{t-1} for different values of u_{t-2}.

which are not always decreasing in the first lagged variable depends on the copula families of C_1 and C_2, and on the value of Kendall's tau. In many cases, the decrease in $F_{t|t-1}$ is not sufficiently large enough to overcompensate the decrease in $F_{t-1|t-2}$ if y_{t-1} increases so that the transition distribution is not decreasing in y_{t-1} for all combinations of y_t and y_{t-2}.

These opposing effects also do not vanish if we consider a non-simplified SD-vine copula or if $p > 2$. In general, it is not possible to verify the $\mathrm{SI}(U_{p+1}|U_{p:1})$ property of $C_{1:p+1}$ if we just consider the building blocks of the SD-vine copula. Instead we have to investigate the joint dependence of $C_{1:p+1}$ to check whether it has the $\mathrm{SI}(U_{p+1}|U_{p:1})$ property. Thus, for each set of bivariate (conditional) copulas C_j we have to analyze whether the resulting SD-vine copula has the $\mathrm{SI}(U_{p+1}|U_{p:1})$ property. Unfortunately, the implied dependence structure of a regular vine copula is not accessible due to its complex structure which consists of bivariate (conditional) copulas and partial derivatives of bivariate (conditional) copulas.[32] Consequently, checking whether $C_{1:p+1}$ has the $\mathrm{SI}(U_{p+1}|U_{p:1})$ property is a non-trivial task.

An exception is the case if all bivariate conditional copulas are bivariate unconditional Gaussian or Student-t copulas such that the joint distribution is a multivariate Gaussian or Student-t copula. The resulting copula $C_{1:p+1}$ then obviously has the $\mathrm{SI}(U_{p+1}|U_{p:1})$

[32] Assuming that all partial derivatives exist, we obtain for the SD-vine copula

$$\begin{aligned}\partial_2 F^u_{t|t-1,t-2}(u_t|u_{t-1},u_{t-2}) &= c_2(F^u_{t|t-1}(u_t|u_{t-1}), F^u_{t-1|t}(u_{t-2}|u_{t-1})|u_{t-1})\partial_2 F^u_{t|t-1}(u_t|u_{t-1})\\ &\quad + \partial_2 h_2(F^u_{t|t-1}(u_t|u_{t-1}), F^u_{t-1|t}(u_{t-2}|u_{t-1})|u_{t-1})\partial_2 F^u_{t-1|t}(u_{t-2}|u_{t-1})\\ &\quad + \partial_3 h_2(F^u_{t|t-1}(u_t|u_{t-1}), F^u_{t-1|t}(u_{t-2}|u_{t-1})|u_{t-1}).\end{aligned}$$

Thus, even under the simplifying assumption we have to check whether

$$\begin{aligned}&c_2(F^u_{t|t-1}(u_t|u_{t-1}), F^u_{t-1|t}(u_{t-2}|u_{t-1}))\partial_2 F^u_{t|t-1}(u_t|u_{t-1})\\ &+\partial_2 h_2(F^u_{t|t-1}(u_t|u_{t-1}), F^u_{t-1|t}(u_{t-2}|u_{t-1}))\partial_2 F^u_{t-1|t}(u_{t-2}|u_{t-1})\end{aligned}$$

is negative for all $(u_t, u_{t-1}, u_{t-2}) \in (0,1)^3$ to show that $F^u_{t|t-1,t-2}$ is decreasing in the first lagged variable. Note that $c_2 \geq 0$ and $\partial_2 h_2 \leq 0$, since C_2 must have the $\mathrm{SI}(U_{t|t-1}|U_{t-2|t-1})$ property so that $F^u_{t|t-1,t-2}$ is decreasing in the second lagged variable. Thus, it is in general a very difficult problem to check whether $\partial_2 F^u_{t|t-1,t-2}$ is non-positive.

property. However, in many cases it can happen that the transition distribution is not decreasing in each lagged value and it is difficult to analyze this behavior analytically. We summarize these findings in the following proposition.

Proposition 3.7 (The SI property for SD-vine copulas in terms of its building blocks)

In general, there are no sufficient conditions that we can impose on the bivariate conditional copulas $(C_j)_{j=1,\dots,p}$ of an SD-vine copula $C_{1:p+1}$ such that $C_{1:p+1}$ has the $\mathrm{SI}(U_{p+1}|U_{p:1})$ property. Instead, the functional form of the implied transition distribution $F_{p+1|p:1}$ has to be investigated. If and only if the partial copula $C^{\partial}_{1,p+1|2:p}$ has the $\mathrm{SI}(U_{p+1|p:2}|U_{1|p:2})$ property then $F_{p+1|p:1}$ is decreasing in y_1 and $g(y_{p:1}) = \mathbb{E}[Y_{p+1}|Y_{p:1} = y_{p:1}]$ is increasing in y_1.

Proof. Directly follows from the previous remarks if $p = 2$. For $p > 2$ similar arguments apply since a necessary condition for $C_{1:p+1}$ to have the $\mathrm{SI}(U_{p+1}|U_{p:1})$ property is that $C^{\partial}_{1,p+1|2:p}$ has the $\mathrm{SI}(U_{p+1|p:2}|U_{1|p:2})$ property. We leave the details to the reader, as no new ideas are required. ■

The consequence of Proposition 3.7 is that we can not apply Corollary 3.2 to ensure a positive autocorrelation function of an SD-vine copula-based Markov process. Moreover, there are no sufficient conditions for the bivariate conditional copulas such that $\mathbb{E}[Y_{p+1}|Y_{p:1} = y_{p:1}] = g(y_{p:1})$ is increasing in each conditioning variable. However, it might still be the case that we can find other sufficient conditions such that the autocorrelation function is positive.

Unfortunately, it turns out that sufficient conditions for a positive autocorrelation function are hard to establish. This is not a particular problem of the SD-vine copula but a general problem of regular vine copulas. We illustrate this in the following. The autocorrelation function is positive if and only if $C_{1,t}$ has the PQD property for all $t \in \mathbb{N}$ (see Section 2.2 for details on the PQD property). $C_{1,t}$ is given by

$$C_{1,t}(u_1, u_t) = \int_{[0,1]^{t-2}} C_{1,t|2:t-1}\big(F_{U_1|U_{2:t-1}}(u_1|u_{2:t-1}), F_{U_t|U_{2:t-1}}(u_t|u_{2:t-1})\big|u_{2:t-1}\big)\mathrm{d}C_{2:t-1}(u_{2:t-1}),$$

i.e., $C_{1,t}$ is determined by $C_{1,t|2:t-1}$ and the conditional cdfs $F_{U_1|U_{2:t-1}}(u_1|u_{2:t-1})$ and $F_{U_t|U_{2:t-1}}(u_t|u_{2:t-1})$. Thus, we have to find conditions for $C_{1,t|2:t-1}$ and the conditional cdfs such that $C_{1,t}(u_1, u_t) \geq u_1 u_t$ for all $(u_1, u_t) \in [0, 1]^2$. We first note that the PQD property of the bivariate conditional copula $C_{1,t|2:t-1}$ does not necessarily imply that the associated copula $C_{1,t}$ has the PQD property. For instance, if $Y_{1:3}$ has a normal distribution with $(\rho_{12}, \rho_{23}, \rho_{13;2}) = (-0.5, 0.5, 0.2)$, then $\rho_{13} = -0.1$. However, in this case, $Y_{1:3}$ is not a random vector of a stationary process. For a stationary Gaussian process the PQD property of $C_{13|2}$ is a sufficient and necessary condition so that $\mathbb{C}\mathrm{ov}[Y_1, Y_3]$ is non-negative.[33] Thus, it seems reasonable to assume that $C_{1,t|2:t-1}$ should have the PQD property in order that $\mathbb{C}\text{ov}[Y_1, Y_t]$ is non-negative. Even with this assumption, the PQD property of $C_{1,t}$ is hard to verify as the following lemma demonstrates.

Lemma 3.2

Let $t \geq 3$ and $C_{1,t|2:t-1}$ be PQD. Then $C_{1,t}$ has the PQD property if and only if

$$C_{1,t}(u_1, u_t)^{C_{1,t|2:t-1}=C^{\perp}} = \int_{[0,1]^{t-2}} \prod_{i=1,t} F_{U_i|U_{2:t-1}}(u_i|u_{2:t-1}) \mathrm{d}C_{2:t-1}(u_{2:t-1}) \geq u_1 u_t. \quad (3.7.1)$$

Proof. See Appendix A.5. ■

The copula $C_{1,t}(u_1, u_t)^{C_{1,t|2:t-1}=C^{\perp}}$ is the copula of (Y_1, Y_t) that arises if $C_{1,t|2:t-1} = C^{\perp}$. Thus, Lemma 3.2 states that if $C_{1,t|2:t-1}$ is PQD, then $C_{1,t}$ is PQD if and only if the bivariate copula of (Y_1, Y_t) that would arise if $C_{1,t|2:t-1} = C^{\perp}$ is PQD. Unfortunately, (3.7.1) is difficult to check because there is no analytical expression for the integral in general. But for $t = 3$ we can use the following lemma.

Lemma 3.3

If $C_{1:3}$ has the $\text{SI}(U_1|U_2)$ and the $\text{SI}(U_3|U_2)$ property, and $C_{13|2}$ is PQD, then C_{13} is PQD.

Proof. Due to Lemma 3.2 it is sufficient to consider the case when $C_{13|2} = C^{\perp}$. Since $C_{1:3}$ has the $\text{SI}(U_1|U_2)$ and the $\text{SI}(U_3|U_2)$ property we obtain the stochastic representation

$$U_1 = F^{-1}_{U_1|U_2}(Z_1|U_2),$$
$$U_3 = F^{-1}_{U_3|U_2}(Z_3|U_2),$$

where $Z_1 \perp Z_3$ since $C_{13|2} = C^{\perp}$. An application of Theorem 1 in Lehmann (1966) shows that C_{13} is PQD. ■

Unfortunately, it does not seem to be possible to generalize Lemma 3.3 for $C_{1,t}$ if $t \geq 4$.[34] Consequently, it may be possible that (3.7.1) has to be verified for every bivariate copula of the SD-vine copula. We have investigated several other approaches to show that $C_{1,t}$ is PQD, but all approaches just seem to work if $t = 3$. It appears that, in general, we have to

[33] It would be interesting to investigate whether $\mathbb{C}\text{ov}[Y_1, Y_t]$ is also non-negative for $t > 3$ if Y is a Gaussian process and $C_{1,t|2:t-1}$ has the PQD property.

[34] The proof of Lemma 3.3 is based on Theorem 1 in Lehmann (1966), which says that if (X_i, Y_i) is PQD for all $i = 1, \ldots, N$, and $(X_i, Y_i) \perp (X_j, Y_j)$ for all $i \neq j$, then (A, B) is PQD if $A = g(X_{1:N})$, $B = h(Y_{1:N})$, and g and h are both increasing (or decreasing) in the i-th coordinate for $i = 1, \ldots, N$. If $C_{14|23} = C^{\perp}$, and the simplifying assumption holds, then $Y_1 = F^{-1}_{Y_1|U_2,U_{3|2}}(Z_1|U_2, U_{3|2})$, $Y_4 = F^{-1}_{Y_4|U_3,U_{2|3}}(Z_4|U_3, U_{2|3})$, where $Z_1 \perp Z_4$, and $F^{-1}_{Y_1|U_2,U_{3|2}}$ and $F^{-1}_{Y_4|U_3,U_{2|3}}$ are both increasing in each component of the conditioning variables. However, simulations show that $(U_2, U_3) \perp (U_{2|3}, U_{3|2})$ is not true in general, so we can not apply Theorem 1 in Lehmann (1966).

simulate the autocorrelation function of a SD-vine copula-based Markov process in order to investigate its functional form.

Since the autocorrelation function can be obtained from the partial autocorrelation function one might conjecture that one can use this relation in order to find sufficient conditions for a positive autocorrelation function. For instance, it holds for a stationary process that $\rho_{13} = \rho_{13;2}(1-\rho_{12}^2) + \rho_{12}^2$. Thus, if we assume that C_1 is PQD then $\rho_{12} \geq 0$, and it only remains to show that $\rho_{13;2} \geq 0$. $C_{13|2}$ (together with $F_{1|2}$ and $F_{3|2}$) represents the bivariate conditional distribution $F_{13|2}$ which can be used to compute the conditional correlation $\rho_{13|2}$. However, in general, $F_{13|2}$ does not give access to the partial correlation $\rho_{13;2}$. Moreover, the relation between conditional correlation and partial correlation is rather loose, e.g., the conditional covariance can always be positive but the partial correlation can be negative.[35] Therefore, in general, we can not show that $\rho_{13;2} \geq 0$ if we specify an SD-vine copula and we can not use the implied partial autocorrelation function to show that the autocorrelation function of an SD-vine copula-based Markov process has a specific functional form.

3.7.2 Ergodicity and mixing of SD-vine copula-based Markov processes

Although SD-vine copula-based Markov processes are stationary processes by construction, that does not necessarily imply that they are ergodic. For instance, if

$$C_{12}(u_1,u_2) = \begin{cases} 0.5 + 2(u_1-0.5)(u_2-0.5) & (u_1,u_2) \in (0.5,1]^2 \\ 2\min(0.5,u_2)\min(0.5,u_1) & \text{else} \end{cases},$$

then the process U generated by

$$U_t = F^{-1}_{U_1|U_2}(Z_t|U_{t-1}), \quad Z_t \perp U_{t-1:1}$$

is stationary but not ergodic since if $U_1 \in (0,0.5]$ the process is a sequence of iid random variables with marginal distribution $U(0,0.5)$, and if $U_1 \in (0.5,1)$ the process is a sequence of iid random variables with marginal distribution $U(0.5,1)$. The reason for this behavior is that $P(U_1 \geq 0.5|U_2 \leq 0.5) = P(U_1 \leq 0.5|U_2 \geq 0.5) = 0$, i.e., some states of the process are not reachable from every state of the process. This reducibility of the process is obviously implied by the fact that the density of C_{12} is only strictly positive on $[0,0.5]^2 \cup (0.5,1]^2$ in this case. A simple condition that ensures beta mixing of an SD-vine copula-based

[35] For instance, Let $Y_2 \sim N(0,1)$, $\mathcal{E} \sim N(0,1)$, $Y_2 \perp \mathcal{E}$. If we define $Y_1 = 0.5Y_2 + Y_2^2 + \mathcal{E}$, and $Y_3 = 0.5Y_2 - Y_2^2 + \mathcal{E}$, then $\mathbb{C}\mathrm{ov}[Y_1,Y_3|Y_2=y_2] = 1$, but $\mathbb{C}\mathrm{ov}[Y_2^2+\mathcal{E}, -Y_2^2+\mathcal{E}] = -\mathbb{V}\mathrm{ar}[Y_2^2] + \mathbb{V}\mathrm{ar}[\mathcal{E}] = -1$, so that $\rho_{13;2} = \frac{-1}{\mathbb{V}\mathrm{ar}[Y_1-0.5Y_2]}$ is negative. Moreover, it can be shown that ρ_{13} is also negative in this case. However, scatter plots suggest that neither C_{12} nor C_{23} have the PQD property in this case.

Markov processes of order p is that the densities of $C_{1:p+1}$ and the marginal distribution are strictly positive. This follows because the corresponding stationary first-order Markov process $\mathcal{Y} := (\mathcal{Y}_t)_{t\in\mathbb{N}}$, where $\mathcal{Y}_t = Y_{t+p:t}$, is irreducible and aperiodic, which implies that $\mathcal{Y}$ and thus Y is beta mixing (cf. Longla and Peligrad (2012, Proposition 2) or Chen and Fan (2006b, p. 314)). We also know that beta mixing implies strong mixing, and strong mixing implies ergodicity (Doukhan, 1994). Thus, under typical regularity conditions that are assumed in the iid setup, the ML-estimator of an SD-vine copula-based Markov process of order p is strongly consistent if the marginal density and the density of $C_{1:p+1}$ are strictly positive. It is straightforward to show that the density of the SD-vine copula $C_{1:p+1}$ is strictly positive if and only if the density of $C_i \in \mathbb{C}_p$ is strictly positive for all $i = 1, \ldots, p$.

To establish the asymptotic distribution of the ML estimator one needs to know more about the dependence structure of an SD-vine copula-based Markov process, which is usually characterized in terms of mixing coefficients. For the case of first-order copula-based Markov processes we can distinguish between three different approaches that have been proposed to determine the rate of mixing. We now briefly discuss to what extent these approaches could be generalized to higher-order SD-vine copula-based Markov processes.

If C_1 is exchangeable and its density is square-integrable and strictly positive, the sequences of β or ρ-mixing coefficients decay exponentially fast (see Beare, 2010, Theorem 3.1, Theorem 3.2, and Theorem 4.1), which also implies geometric ergodicity. It is not obvious how this result can be extended to higher-order copula-based Markov processes. The proof relies on the fact that the density of a bivariate exchangeable distribution admits a mean square convergent series representation if it is square-integrable. These representations have been extensively studied by several authors for bivariate distributions (see Beare, 2010, supplement), but, to the best of our knowledge, these representations have not been considered for multivariate distributions. Even if strictly positive densities of multivariate distributions can be represented by a (useful) mean square convergent series expansion, a possible generalization of these sufficient conditions for geometric β- or ρ-mixing would not apply to many interesting SD-vine copula-based Markov processes. In the bivariate case, these sufficient conditions rule out copulas with tail dependence because their density is not square-integrable (Beare, 2010, Theorem 3.3) and we strongly conjecture that these sufficient conditions also rule out multivariate copulas with the property that at least one bivariate margin has tail dependence. To ensure that the density of the SD-vine copula $C_{1:p+1}$ is square-integrable we can assume that the density of $C_i \in \mathbb{C}$ is bounded for all $i = 1, \ldots, p$. However, almost all commonly used bivariate copula families do not have a bounded density.

A high-level condition that ensures geometrical ρ-mixing of a first-order Markov process is that the density of C_1 is bounded away from zero, i.e., the density is not only strictly positive but there is a constant $d > 0$ such that $c_1 \geq d$ (Beare, 2010, Theorem 4.1 and 4.2). We conjecture that a similar result might also hold for the higher-order case, i.e., if the density of $C_{1:p+1}$ is bounded away from zero then the resulting p-th order Markov process might be geometrical ρ-mixing. Obviously, the density of $C_{1:p+1}$ is bounded away from zero if the density of $C_i \in \mathbb{C}_p$ is bounded away from zero for all $i = 1, \ldots, p$. Even if this would be a sufficient condition for geometrical ρ-mixing, the possible scope of this result would be limited. For many popular bivariate copula families the density is not bounded away from zero, such as for the Student-t, Clayton, or Gumbel copula. Consequently, if this generalization of sufficient conditions is possible, it would only apply to very specific SD-vine copula–based Markov processes.

Foster-Lyapunov drift conditions (cf. Meyn and Tweedie, 2009) have been used for specific copula families to verify geometric ergodicity of the resulting first-order Markov processes. This approach is used by Chen et al. (2009b) to show that the Clayton, Gumbel, and Student-t copula, generate geometrically ergodic first-order Markov processes, and also by Beare (2012) to prove that many Archimedean copulas generate first-order Markov processes which are geometrically ergodic. The drift condition is a constraint on the transition distribution of the Markov process. In general, these drift conditions might also be applied to prove geometric ergodicity of higher-order Markov processes. However, the implied transition distribution of an SD-vine copula-based Markov process has a very complicated structure since it is a composition of various partial derivatives of bivariate (conditional) copulas. In the first-order case, the transition distribution is given by one partial derivative of a bivariate copula which often results in a nice analytical expression of the transition distribution. But already in this case, the application of the Foster-Lyapunov drift conditions is already quite sophisticated (Chen et al., 2009b, Appendix A). Moreover, in light of our results regarding the $\mathrm{SI}(U_{p+1}|U_{p:1})$ property of an SD-vine copula $C_{1:p+1}$, it may be possible that there are no sufficient conditions that we can impose on the bivariate (conditional) copulas of $\mathbb{C}_p$ in order that the SD-vine copula $C_{1:p+1}$ generates geometrically ergodic Markov processes. It may be possible that the decay rates of mixing coefficients have to be investigated for each particular SD-vine copula. Altogether, it is an open question how fast SD-vine copula-based Markov process are mixing. Simulations of some of the fitted processes in Chapter 4 and Chapter 6 suggest that in theses cases the ML estimator is asymptotically normally distributed.

3.8 Conclusion

In this chapter, we conducted a thorough investigation of univariate stationary higher-order Markov processes in terms of copulas, with a focus on regular vine copulas, and developed a copula-based approach for modeling univariate time series. We argued that, contrary to classical approaches, exploratory data analysis can be utilized to set up a copula-based time series model which captures central features of the dependence structure. Moreover, the transition distribution of the Markov process is derived from its copulas and its marginal distribution, resulting in flexible transition distributions that can not be expressed by location and scale models. On the other side, it is not possible to model only some features of the transition distribution. If the interest lies in some features of the transition distribution then the copula-based approach might not be the first choice. In addition, properties of a copula-based process are often not tractable and the implementation of exogenous regressors or the construction of multivariate time series processes is much more complicated in the copula-based framework.

The literature review documented that the theory on the dependence properties and estimation of univariate stationary copula-based first-order Markov processes is quite exhaustive. Nevertheless, only few studies consider the application of copulas to model serial dependence. Except for the study of Sokolinskiy and van Dijk (2011), which indicates that copula-based time series models successfully forecast realized volatility, there is no empirical evidence that copula-based approaches might be useful for modeling financial time series. We argued that one reason for that is that most applications use a stationary first-order Markov process to model a time series which is probably not Markovian of order one. Moreover, we pointed out that most copulas have been developed for static applications and that there are no useful parametric copulas for modeling time series of financial returns.

In order to model time series with copulas, we examined the characterization of stationarity and the Markov property in terms of copulas. In this regard, we introduced the SD-vine copula and established its unique properties among the class of regular vine copulas. Moreover, the partial autocopula sequence, a non-linear generalization of the partial autocorrelation function that corrects for considerably more dependence caused by intermediate variables, was introduced. We developed algorithms for the specification and estimation of simplified SD-vine copula-based Markov models and illustrated how exploratory data analysis can be utilized to construct a higher-order Markov process by a sequence of bivariate unconditional copulas. In order to obtain a parsimonious representation of a higher-order Markov process, we proposed to model the SD-vine copula by a lag function, which is parameterized by a low-dimensional parameter vector. Since the number of computations increases quadratically with the Markov order of the process, we suggested to truncate the order of the process such that the loss in average log-likelihood does not exceed a certain threshold.

We derived sufficient conditions for a non-negative autocorrelation function of copula-based processes and that the mean of the transition distribution is increasing in all conditioning variables. However, it appears that these sufficient conditions are often not satisfied by SD-vine copula-based Markov processes. In addition, we showed that we can not impose conditions on the bivariate (conditional) copulas of the SD-vine copula so that these sufficient conditions hold. Consequently, the verification of a non-negative autocorrelation function or an increasing conditional mean is generally not possible because the joint dependence structure of regular vine copulas is in general not tractable. We also discussed the mixing rates of SD-vine copula-based higher-order Markov processes. Except for a few cases which are not interesting for applications, it is not possible to generalize the results for the mixing rates of first-order Markov processes to SD-vine copula-based higher-order Markov processes. It is an open question whether one can find conditions for a sufficiently fast mixing rate of SD-vine copula-based Markov processes, or regular vine copula-based Markov processes in general.

3.9 Appendix

A.1 Illustration that other regular vine copulas than the SD-copula can generate stationary processes

It is well known that a trivariate normal distribution with standard normal margins can be represented by a vine distribution where all copulas of the vine are bivariate Gaussian copulas and the marginal distributions are standard normal. That is, the vine density given by

$$f_{1:3}(y_{1:3})=\prod_{i=2,3} c_{1i}\big(\Phi(y_1),\Phi(y_i);\rho_{1i}\big)c_{23;1}\big(F_{2|1}(\Phi(y_2)|\Phi(y_1)),F_{3|1}(\Phi(y_3)|\Phi(y_1));\rho_{23;1}\big)\prod_{i=1}^{3}\phi(y_i), \tag{3.9.1}$$

where c_{12}, c_{13}, and $c_{23;1}$ are bivariate Gaussian copulas with correlation parameter ρ_{12}, ρ_{13} and $\rho_{23;1}$, respectively, is the density of a trivariate normal distribution with standard normal marginals. However, the density given in (3.9.1) is not compatible with the conditions for stationarity unless the constraint $\rho_{23;1} = \frac{\rho_{12}(1-\rho_{13})}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13}^2}}$ is satisfied so that $\rho_{12} = \rho_{23}$. Nevertheless, the transition distribution that is implied by (3.9.1) gives rise to a stationary process. Let $F_{1|2:3}$ be the conditional cdf of (3.9.1). Using well known properties of the multivariate normal distribution, one can show that the transition distribution of the stationary process defined by

$$Y_t = \left(\rho_{12} - \rho_{23}\rho_{13;2}\frac{\sqrt{1-\rho_{12}^2}}{\sqrt{1-\rho_{23}^2}}\right)Y_{t-1} + \rho_{13;2}\left(\frac{\sqrt{1-\rho_{12}^2}}{\sqrt{1-\rho_{23}^2}}\right)Y_{t-2} + \sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13;2}^2}\mathcal{E}, \tag{3.9.2}$$

where

$$\mathcal{E} \sim N(0,1), \quad \mathcal{E} \perp (Y_{t-1}, Y_{t-2}), \quad \rho_{23} = \rho_{23;1}\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13}^2} + \rho_{12}\rho_{13},$$

is equal to $F_{1|2:3}$. To put it differently, if we simulate a stochastic process using

$$Y_t = F_{1|2:3}^{-1}(Z_t|Y_{t-1:2}), \quad Z_t \sim U(0,1),\ Z_t \perp Y_{t-1:1},$$

where $F_{1|2:3}^{-1}$ is the quantile function of $F_{1|2:3}$, then the resulting process is a stationary Gaussian autoregressive process of order two with a representation given in (3.9.2).

However, $\rho_{12} \neq \mathbb{C}\mathrm{ov}[Y_t, Y_{t-1}]$ unless $\rho_{23|1} = \frac{\rho_{12}(1-\rho_{13})}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13}^2}}$, i.e., the density given in (3.9.1) is not the stationary density, and the process is only asymptotically stationary if it is not started from its stationary distribution.

A.2 Proof of Proposition 3.3

For simplicity, we drop the arguments of the functions in the following. The Markov property states that

$$\forall t \geq p+2, \forall 1 \leq k \leq t-p-1\colon\ f_{t|t-1:t-p} = f_{t|t-1:t-(p+k)},$$

which, due to Sklar's theorem, is equivalent to

$$\forall t \geq p+2, \forall 1 \leq k \leq t-p-1\colon\ \frac{c_{t:t-p}}{c_{t-1:t-p}} f_t = \frac{c_{t:t-(p+k)}}{c_{t-1:t-(p+k)}} f_t,$$

so that

$$\forall t \geq p+2, \forall 1 \leq k \leq t-p-1\colon\ c_{t:t-(p+k)} = \frac{c_{t:t-p}}{c_{t-1:t-p}} c_{t-1:t-(p+k)},$$

which proves the first equality in Proposition 3.3.

Assume that the following inductive hypothesis holds

$$c_{s+k-1:s-p} = \prod_{i=0}^{k-2} \frac{c_{s+1+i:s+1+i-p}}{c_{s+i:s+1+i+p}} c_{s:s-p}.$$

From the first equality in Proposition 3.3 it follows that for all $s \geq p+1$ and all $k \in \mathbb{N}$ we have that

$$c_{s+k:s-p} = \frac{c_{s+k:s+k-p}}{c_{s+k-1:s+k-p}} c_{s+k-1:s-p},$$

and, as a special case,

$$c_{s+1:s-p} = \frac{c_{s+1:s+1-p}}{c_{s:s+1-p}} c_{s:s-p},$$

so that the basis for the induction is valid. Thus,

$$\begin{aligned} c_{s+k:s-p} &= \frac{c_{s+k:s+k-p}}{c_{s+k-1:s+k-p}} c_{s+k-1:s-p} \\ &= \frac{c_{s+k:s+k-p}}{c_{s+k-1:s+k-p}} \prod_{i=0}^{k-2} \frac{c_{s+1+i:s+1+i-p}}{c_{s+i:s+1+i+p}} c_{s:s-p} \\ &= \prod_{i=0}^{k-1} \frac{c_{s+1+i:s+1+i-p}}{c_{s+i:s+1+i+p}} c_{s:s-p}. \end{aligned}$$

Setting $s+k=t, s-p=t-(p+k)$, we obtain that

$$c_{t:t-(p+k)} = \prod_{i=0}^{k-1} \frac{c_{t-k+1+i:t-k+1+i-p}}{c_{t-k+i:t-k+1+i-p}} c_{t-k:t-(p+k)}.$$

Noting that

$$\prod_{i=0}^{k-1} \frac{c_{t-k+1+i:t-k+1+i-p}}{c_{t-k+i:t-k+1+i-p}} = \prod_{i=0}^{k-1} \frac{c_{t-i:t-(i+p)}}{c_{t-(i+1):t-(i+p)}},$$

the second equality in Proposition 3.3 is proved.

A.3 Proof of Proposition 3.6

Since $\mathbb{P}(U_j \leq a|U_{j-1:1} = u_{j-1:1}) = \mathbb{P}(U_j \leq a|U_{j-1:j-p} = u_{j-1:j-p})$ for $j \geq p+1$, it follows that $U_{1:j}$ has the CIS($U_{j:1}$) property for all $j \in \mathbb{N}$. We now show that C_{13} has the SI($U_3|U_1$) property. For that purpose, note that $(U_1, U_2) \stackrel{d}{=} (U_1, F_{2|1}^{-1}(Z_2|U_1))$, where $F_{2|1} = F_{U_2|U_1}$ and $Z_2 \perp U_1$. Moreover, $(U_1, U_3) \stackrel{d}{=} (U_1, F_{3|21}^{-1}(Z_3|U_2, U_1))$, where $F_{3|21} = F_{U_3|U_{2:1}}$ and $Z_3 \perp U_{2:1}$. Thus,

$$\begin{aligned}
\mathbb{P}(U_3 \leq a|U_1 = b) &= \mathbb{P}(Z_3 \leq F_{3|21}(a|U_2, b)|U_1 = b) \\
&= \mathbb{P}(Z_3 \leq F_{3|21}(a|F_{2|1}^{-1}(Z_2|b), b)|U_1 = b) \\
&= \int_{[0,1]} \mathbb{P}(Z_3 \leq F_{3|21}(a|F_{2|1}^{-1}(z_2|b), b)|U_1 = b, Z_2 = z_2) f_{Z_2}(z_2) \mathrm{d}z_2 \\
&= \int_{[0,1]} F_{3|21}\big(a|F_{2|1}^{-1}(z_2|b), b\big) \mathrm{d}z_2,
\end{aligned}$$

which is decreasing in b if $F_{3|21}(\cdot|u_2, u_1)$ is decreasing in u_2 and u_1, and $F_{2|1}^{-1}(\cdot|u_1)$ is increasing in u_1. This is true if $U_{1:3}$ has the CIS($U_{3:1}$) property.

To show that C_{1j} has the SI($U_j|U_1$) property for all $j \geq 3$ we use induction. Assume that $C_{1,k}$ has the SI($U_k|U_1$) property for all $k = 2, \ldots, j$, i.e., $F_{k|1}^{-1}(\cdot|b)$ is increasing in b for all $k = 2, \ldots, j$, with $F_{k|1} = F_{U_k|U_1}$. It follows that

$$\begin{aligned}
\mathbb{P}(U_{j+1} \leq a|U_1 = b) &= \mathbb{P}(Z_j \leq F_{j+1|j:1}(a|U_{j:2}, b)|U_1 = b) \\
&= \mathbb{P}(Z_j \leq \mathbb{P}(U_{j+1} \leq a|U_1 = b, U_k = F_{k|1}^{-1}(Z_k|b), \forall k = 2, \ldots, j)|U_1 = b) \\
&= \int_{[0,1]^{j-1}} \mathbb{P}\big(U_{j+1} \leq a|U_1 = b, U_k = F_{k|1}^{-1}(z_k|b)\big) \mathrm{d}z_{2:j}
\end{aligned}$$

is decreasing in b because $U_{j+1:1}$ has the CIS($U_{j+1:1}$) property. The last equality follows since $Z_{2:j}$ is a vector of jointly independent uniform random variables.

A.4 Proof of Lemma 3.1

Since $Y_1 = \int_{\mathbb{R}}(\mathbb{1}_{\{t\geq 0\}} - \mathbb{1}_{\{Y_1\leq t\}})\mathrm{d}t$ and the conditional expectation exists, we obtain by Fubini's theorem that

$$\begin{aligned}\mathbb{E}[Y_1|Y_{2:K}] &= \mathbb{E}\left[\int_{\mathbb{R}}(\mathbb{1}_{\{t\geq 0\}} - \mathbb{1}_{\{Y_1\leq t\}})\mathrm{d}t|Y_{2:K}\right]\\ &= \int_{\mathbb{R}}\big(\mathbb{1}_{\{t\geq 0\}} - \mathbb{E}\big[\mathbb{1}_{\{Y_1\leq t\}})|Y_{2:K}\big]\big)\mathrm{d}t\\ &= \int_{\mathbb{R}}\big(\mathbb{1}_{\{y_1\geq 0\}} - F_{1|2:K}(y_1|y_{2:K})\big)\mathrm{d}y_1\\ &= \int_{\mathbb{R}}\Big(\mathbb{1}_{\{y_1\geq 0\}} - F_{U_1|U_{2:K}}\big(F_1(y_1)|F_{2:K}(y_{2:K})\big)\Big)\mathrm{d}y_1. \end{aligned} \tag{3.9.3}$$

The last equality follows because

$$\begin{aligned}F_{1|2:K}(y_1|y_{2:K}) &= \frac{\partial^{K-1}}{\partial y_2\ldots\partial y_K}\frac{F_{1:K}(y_{1:K})}{f_{2:K}(y_{2:K})} = \frac{\partial^{K-1}}{\partial y_2\ldots\partial y_K}\frac{C_{1:K}(F_{1:K}(y_{1:K}))}{c_{2:K}(F_{2:K}(y_{2:K}))\prod_{i=2}^K f_i(y_i)}\\ &= \frac{\partial^{K-1}}{\partial F_2(y_2)\ldots\partial F_K(y_K)}\frac{C_{1:K}(F_{1:K}(y_{1:K}))\prod_{i=2}^K f_i(y_i)}{c_{2:K}(F_{2:K}(y_{2:K}))\prod_{i=2}^K f_i(y_i)}\\ &= F_{U_1|U_{2:K}}(F_1(y_1)|F_{2:K}(y_{2:K})).\end{aligned}$$

Obviously, if $C_{1:K}$ has the SI$(U_1|U_{2:K})$ property then the integrand in (3.9.3) is increasing in each $y_i, i = 2,\ldots,K$, so that $E[Y_1|Y_{2:K} = y_{2:k}]$ is increasing in each variable.

A.5 Proof of Lemma 3.2

$$\begin{aligned}&C_{1,t}(u_1,u_t) - u_1u_t\\ &= \mathbb{E}\Big[\mathbb{C}\mathrm{ov}[\mathbb{1}_{\{U_1\leq u_1\}},\mathbb{1}_{\{U_t\leq u_t\}}|U_{2:t-1}]\Big] + \mathbb{C}\mathrm{ov}\Big[\mathbb{E}[\mathbb{1}_{\{U_1\leq u_1\}}|U_{2:t-1}],\mathbb{E}[\mathbb{1}_{\{U_t\leq u_t\}}|U_{2:t-1}]\Big]\\ &= \int_{[0,1]^{t-2}}\Big(F_{U_1,U_t|U_{2:t-1}}(u_1,u_t|u_{2:t-1}) - \prod_{i=1,t}F_{U_i|U_{2:t-1}}(u_i|u_{2:t-1})\Big)\mathrm{d}C_{2:t-1}(u_{2:t-1})\\ &\quad + \int_{[0,1]^{t-2}}\prod_{i=1,t}F_{U_i|U_{2:t-1}}(u_i|u_{2:t-1})\mathrm{d}C_{2:t-1}(u_{2:t-1}) - u_1u_t\\ &\geq \int_{[0,1]^{t-2}}\prod_{i=1,t}F_{U_i|U_{2:t-1}}(u_i|u_{2:t-1})\mathrm{d}C_{2:t-1}(u_{2:t-1}) - u_1u_t,\end{aligned}$$

where the last inequality follows if $C_{1,t|2:t-1}$ is PQD.

4 Modeling price durations with copula-based processes

4.1 Introduction

The modeling of price durations allows for the quantification of the risk for a given price change within a particular time interval. Moreover, the conditional hazard function of a price duration process is closely related to the instantaneous return volatility. Thus, duration-based volatility estimators account for time structures in the price process which are not considered by classical volatility models and offer new ways of estimating volatility. The autoregressive conditional duration (ACD) model (Engle and Russell, 1998) and its many varieties are the de facto standard for modeling financial price duration data. In this chapter, we use price duration data of five blue-chip stocks to compare the in- and out-of-sample performances of simplified SD-vine copula-based Markov models with the popular class of ACD models.

The class of ACD models focuses on the mean of the transition distribution but there are many reasons why the modeling of the complete transition distribution of price durations is of interest. First of all, the modeling of the transition distribution is required in order to derive a model for the conditional hazard function[1] which can be used to estimate the integrated conditional variance over an interval. Recently, Tse and Yang (2012) use this approach to estimate the intraday volatility of a stock by integrating the instantaneous conditional volatility obtained from the ACD model. They show via simulations that the resulting duration-based daily volatility estimator has a lower mean squared error than popular realized volatility estimators in almost all cases. The basic assumption for the application of the ACD model is that the transition distribution of price durations and the conditional hazard function is just a function of the conditional mean and the error distribution of the ACD model. Weakening this restriction and setting up a model for the complete transition distribution of price durations can improve the modeling of the implied conditional hazard function and result in better estimates of the integrated conditional variance over an interval.

Another benefit of the copula-based approach in the context of financial durations is that the conditional mean and the conditional variance are not necessarily connected as in the

[1] The implied conditional hazard function of a price duration process with underlying cumulative absolute price change dp can be expressed as

$$\lambda^{dp}(t-t_i; y_{i:1}) = \frac{f_{i+1|i:1}(t-t_i|y_{i:1})}{1-F_{i+1|i:1}(t-t_i|y_{i:1})},$$

where $F_{i+1|i:1}$ is the distribution of the $(i+1)$-th price duration conditional on all past durations.

class of ACD models.[2] As it has been pointed out by Ghysels et al. (2004), a model that allows for distinct dynamic patterns of the conditional mean and the conditional variance is important for the modeling of liquidity risk in financial markets. Moreover, risk measures for liquidity risk, such as the Time at Risk (TaR) (Ghysels et al., 2004), can not be derived from the mean or variance of the transition distribution and the ACD-implied Times at Risk are constrained to be parallel for all forecast horizons. In this sense, the modeling of the transition distribution of price durations is crucial if one is concerned about liquidity risk.

This chapter is structured as follows. Section 4.2 describes the data and provides a descriptive analysis of the data. In Section 4.3 we explain the family of augmented ACD models which are employed for the empirical application. Moreover, the goodness-of-fit of estimated ACD models is evaluated in terms of classical measures as well as from a copula perspective. The estimation of simplified SD-vine copula-based Markov models is addressed in Section 4.4. We show how exploratory data analysis can be used to set up the simplified SD-vine copula-based Markov model and investigate the goodness-of-fit of the copula-based models. The in- and out-of-sample performances of the ACD and CMP models are examined in Section 4.5. We also analyze to what extent a CMP model implies a time-varying conditional dispersion and contrast this feature with the time-constant conditional dispersion of an ACD model. Section 4.6 summarizes our results.

4.2 Data description

We consider price durations from transaction data of the Apple, Cisco, Hewlett-Packard, Intel and Microsoft stock from July 1 to September 31, 2009.[3] We use the period from July 1 to August 31 for in-sample evaluation and the month September for an out-of-sample forecasting experiment. In order to avoid periods of lower frequency trading we only consider trades from 10:00 to 16:00 and remove overnight spells. Let P_t denote the midquote between the best ask and bid prices at time t for a particular stock. We compute for some starting value P_{t_0} the time series of observed price durations $(y_i^\star)_{i=1,\ldots,N}$ as follows

$$y_i^\star := \inf\{t : |P_{t_i+t} - P_{t_i}| \geq dp\}, \quad \text{where } t_i = t_0 + \sum_{j=1}^{i} y_j^\star.$$

The size of the underlying cumulative absolute price change dp is chosen such that it corresponds to 30 times the average absolute trade-to-trade midquote change. In this

[2] This follows because a CMP(1) model can represent any stationary first-order Markov process.
[3] I thank Andreas Fuest who provided the data and also implemented the necessary steps in order to obtain the diurnally adjusted durations.

way, the aggregation levels of the different stocks are approximately comparable.[4] This yields an aggregation level dp of $\$0.261, \$0.03, \$0.06, \$0.027$ and $\$0.032$, for Apple, Cisco, Hewlett-Packard, Intel and Microsoft, respectively.

For each day, price durations typically exhibit a "diurnal pattern" with small durations at the beginning and the end of the day and rather large durations during the lunch break in the middle of the day. To account for intraday seasonality, we first estimate the expectation of the raw durations conditional on the time of the day. For that purpose, we average the unadjusted durations over thirty-minute intervals and then fit a cubic smoothing spline through these nodes to obtain a smooth time-of-day function f that should capture the diurnal pattern (cf. Bauwens and Hautsch, 2006; Hautsch, 2012). The smoothing parameter of the smoothing spline is chosen by generalized cross validation. A diurnally adjusted duration y_i is then given by $y_i := y_i^\star / f_i$, where f_i is the corresponding value of the estimated spline for the raw duration $y_i^\star$.[5] Finally, we remove zero price durations. For the respective stocks we obtain 2737, 9480, 9018, 12299, and 9065 price durations for the in-sample period and 1610, 4867, 3626, 6169, and 4103 observations for the out-of-sample evaluation. Thus, the number of price durations of the Cisco, Hewlett-Packard, Intel and Microsoft stock are comparable whereas the number of price durations of the Apple stock is considerably smaller. A first look at some of the price durations is given in Figure 4.1.[6] The kernel density estimates suggest that the actual probability of small durations is much larger than under the assumption of an exponentially distributed marginal distribution. Moreover, it appears that all marginal distributions have a local mode near 0.2. The high probability of small durations is also confirmed by the histograms of the probability integral transforms which are derived from the fitted exponential distributions. For all stocks, the actual number of price durations that are smaller than the implied 10% quantile of the fitted exponential distribution is approximately twice as large.

4.3 Modeling price durations with ACD models

4.3.1 The (A)ACD model

Let $Y = (Y_i)_{i \in \mathbb{N}}$ be a stationary series of (price) durations, i.e., the series has been corrected for diurnal effects, and $\mathcal{I}_{i-1} = \sigma(Y_j : j \leq i-1)$ be the information set available at point $i-1$. To account for the temporal dependence that is typically present in financial duration

[4] We also investigated price durations with a size of the underlying cumulative absolute price change dp that corresponds to 20 and 40 times the average absolute trade-to-trade midquote change. The results are qualitatively comparable.

[5] These diurnally adjusted durations should not exhibit an intraday periodicity in the mean. However, it may be possible that the adjusted durations still exhibit other diurnal patterns, e.g., the variance of an adjusted duration may still be dependent on the time of the day. If the interest lies in the complete transition distribution, it may be worthwhile to adjust the durations for all diurnal patterns. We leave this open for further research.

[6] The price durations of the Cisco and the Intel stock display the same features.

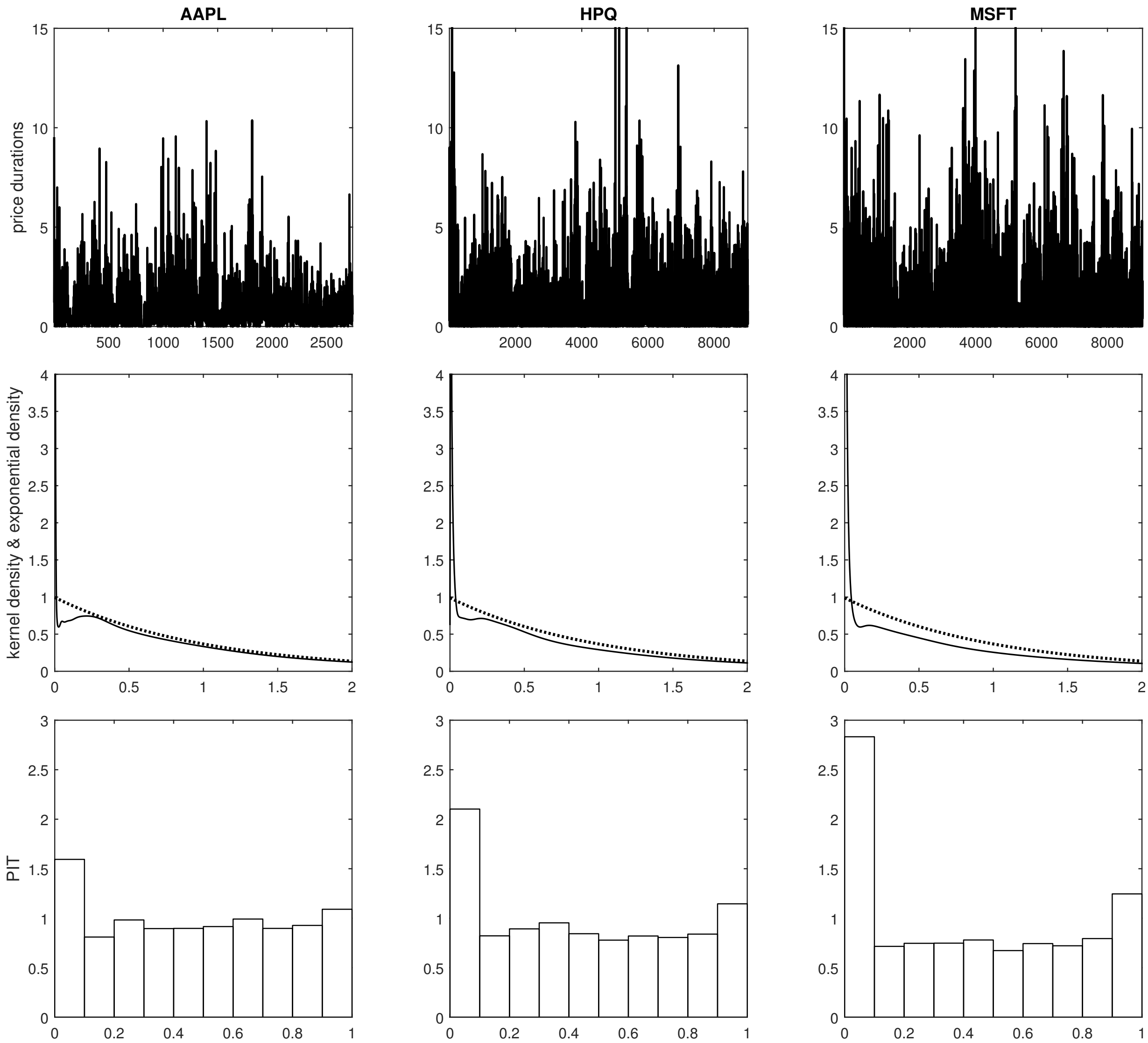


Figure 4.1: Price durations of the Apple, Hewlett-Packard and Microsoft stock. The first row shows the series of price durations from July 1 to August 31. The second row shows the kernel density estimates (solid line) and the densities of fitted exponential distributions (dotted line) for the marginal distribution of price durations. The third row shows the histograms of $(F(y_i))_{i=1,\ldots,N}$, where F is the cdf of the fitted exponential distribution and y_i is the i-th price duration.

data Engle and Russell (1998) introduced the autoregressive conditional duration (ACD) model. For an exhaustive survey of the ACD model and its applications, Pacurar (2008) is an excellent reference. In its general form, the ACD model is given by

$$Y_i = \Psi_i \mathcal{E}_i, \quad \text{where } \mathcal{E}_i \overset{iid}{\sim} F_{\mathcal{E}},\ \mathcal{E}_i \geq 0\ (a.s.),\ \text{and } \mathbb{E}[\mathcal{E}_i] = 1. \tag{4.3.1}$$

The ACD model assumes that the serial dependence within a duration process is completely characterized by the conditional duration $\Psi_i = \mathbb{E}[Y_i|\mathcal{I}_{i-1}]$.[7] Therefore, all features of the

[7] It is possible to relax the assumption that $\mathcal{E} = (\mathcal{E}_i)_{i\in\mathbb{N}}$ is an iid sequence. For instance, Drost and Werker (2004) assume that $E[\mathcal{E}_i|\mathcal{I}_{i-1}] = 1$ for all $i \in \mathbb{N}$. However, we maintain the assumption of iid innovations since we are interested in models that describe the transition distribution and not only some of its features.

transition distribution such as higher-order moments are a function of the conditional duration Ψ_i. This also implies that the conditional dispersion,

$$\frac{\mathbb{V}\mathrm{ar}[Y_i|\mathcal{I}_{i-1}]}{\mathbb{E}[Y_i|\mathcal{I}_{i-1}]^2} = \frac{\Psi_i^2 \mathbb{V}\mathrm{ar}[\mathcal{E}_i]}{\Psi_i^2} = \mathbb{V}\mathrm{ar}[\mathcal{E}_i],$$

is constant and does not depend on the information set. This fundamental property of the ACD model has has been criticized in the literature and evidence for the existence of a varying dynamic pattern for the conditional dispersion has been reported by Ghysels et al. (2004) and Giot (1999).

Several instances of the ACD model are obtained by choosing different specifications for the conditional expected duration Ψ_i and different distributions for the innovation sequence $\mathcal{E}$ in (4.3.1). In the linear ACD(p,q) model, that has been proposed by Engle and Russell (1998), the conditional expected duration Ψ_i depends linearly on the p most recent durations and the q most recent conditional expected durations, i.e.,

$$Y_i = \omega + \sum_{j=1}^{p} \alpha_j Y_{i-j} + \sum_{j=1}^{q} \beta_j \Psi_{i-j}. \tag{4.3.2}$$

The linear ACD(p,q) model can also be expressed as an ARMA(max(p,q),q) model for durations Y_i, i.e.,

$$Y_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) Y_{i-j} - \sum_{j=1}^{q} \beta_j \eta_{i-j} + \eta_i,$$

where $(\eta_i)_{i\in\mathbb{N}}$ is a martingale difference sequence. From this perspective, it is clear that the basic ACD model essentially aims at de-correlating price durations. However, the interpretation of correlation is not clear if the data is generated by a highly non-elliptical distribution as it is the case for price durations. In the empirical study we show that low correlation does not always indicate low dependence and that getting the correlation right is not enough to adequately capture the temporal dependence of the data. Apart from that issue, there is empirical evidence that the basic ACD model overpredicts conditional durations after very short or very long durations (Pacurar, 2008).

To address the latter problem, several alternative specifications of the conditional expected duration have been proposed. A very popular extension of the ACD model, the augmented ACD (AACD) model, was introduced by Fernandes and Grammig (2006). The AACD model originates from applying a Box-Cox transformation to the conditional duration process Ψ_i, combined with a nonlinear function of $\mathcal{E}_i$ to capture asymmetric responses of the conditional expected duration to small and large innovations. The conditional expected duration of the AACD(1, 1) model is given by

$$\Psi_i = \left(\omega + \alpha_1 \Psi_{i-1}^{\lambda}[|\mathcal{E}_{i-1} - b| + c(\mathcal{E}_{i-1} - b)]^{\nu} + \beta_1 \Psi_{i-1}^{\lambda}\right)^{(1/\lambda)}.$$

The parameter vector (λ, ν, b, c) determines the response to innovations and whether the shock impact curve is rotated or shifted. Several ACD models can be recovered by imposing restrictions on the parameter vector (λ, ν, b, c). For instance, the basic linear ACD model is recovered by setting $(\lambda, \nu, b, c) = (1, 1, 0, 0)$.

In the following empirical application we are especially interested in one subclass of the AACD model, namely, the Box-Cox ACD (BCACD) model of Dufour and Engle (2000) which is obtained by letting $\lambda \to 0$ and setting $b = c = 0$ so that

$$\log \Psi_i = \omega + \alpha_1 \mathcal{E}_{i-1}^{\nu} + \beta_1 \log \Psi_{i-1}. \tag{4.3.3}$$

The BCACD model does not suffer so much from identifiability issues and local maxima of the likelihood function as the general AACD model. At the same time it also seems to be the superior member of the AACD family with respect to the Akaike information criterion, see Fernandes and Grammig (2006, table 3-5), which is also confirmed for our data set.[8] We define the conditional expected duration of the BCACD(p, q) model as

$$\log \Psi_i = \omega + \sum_{j=1}^{p} \alpha_j \mathcal{E}_{i-j}^{\nu_j} + \sum_{j=1}^{q} \beta_j \log \Psi_{i-j}.$$

Besides the parameterization of the conditional expected duration another crucial point is the specification of the innovation distribution which can be any distribution defined on a positive support. Commonly used distributions are the exponential and Weibull distribution (Engle and Russell, 1998), and the Burr distribution (Grammig and Maurer, 2000). We also advocate the use of Generalized Beta Distribution of the Second Kind (GB2) which was introduced by McDonald (1984). To the best of our knowledge, the GB2 distribution has not been used in this context before. Yet, the GB2 distribution itself includes a variety of distributions (exponential, Weibull, log-normal, gamma, Burr, generalized Pareto) that have been used before for modeling the distribution of the innovation in the ACD model.

[8] The study of Fernandes and Grammig (2006) reports in table 3 and 4 normalized AIC values for different subclasses of the AACD model for IBM price durations. The normalized AIC is given by $\mathrm{AIC}^{\mathrm{norm}} = -2(\log \mathcal{L} - k)/T$, where $\log \mathcal{L}$ is the value of the maximized log-likelihood, k the number of estimated parameters and T the number of observations. According to the computed normalized AIC values by the authors ($\mathrm{AIC}^{\mathrm{norm}}_{\mathrm{BCACD}} = 1.6729, \mathrm{AIC}^{\mathrm{norm}}_{\mathrm{AACD}} = 1.6721$) the AACD model is preferred to the BCACD model . However, when we try to replicate the normalized AIC values we arrive at the converse solution. According to the authors, the sample size is $T = 4484$, $\log \mathcal{L}_{\mathrm{BCACD}} = -4920.5, k_{\mathrm{BCACD}} = 6$, and $\log \mathcal{L}_{\mathrm{AACD}} = -4918.4, k_{\mathrm{BCACD}} = 9$. Thus, $\mathrm{AIC}^{\mathrm{norm}}_{BCACD} = 2(6 - 4920.5)/4484 = 2.1974, \mathrm{AIC}^{\mathrm{norm}}_{AACD} = 2(11 - 4918.4)/4484 = 2.1987$, and we conclude that the BCACD is preferred to the AACD wrt AIC. Table 5 also indicates that the BCACD model is also the best model or among the best models wrt AIC for price durations of Boeing, Coke, Disney, and Exxon.

The density of the GB2 distribution is given by

$$f(y; a, b, p, q) = \frac{a(y/b)^{ap-1}}{bB(p, q)(1 + (y/b)^a)^{p+q}},$$

where $B(p, q)$ is the beta function and all parameters a, b, p, and q, are positive. Setting

$$b = (\Gamma(p)\Gamma(q))/(\Gamma(p + 1/a)\Gamma(q - 1/a))$$

ensures that the mean of the innovation is one.[9]

However, it may be possible that even the very flexible GB2 distribution is not sufficient to adequately model the error distribution since the marginal distribution of each price duration shows a bimodal structure. Therefore, we also investigate the modeling of the innovation distribution by means of a finite mixture distribution (McLachlan and Peel, 2000; Frühwirth-Schnatter, 2006). For a mixture of two exponential distributions, this has been proposed before by De Luca and Zuccolotto (2003) and De Luca and Gallo (2004, 2008). In order to estimate an ACD model with a mixture distribution for the innovation we use the following algorithm which is inspired by Engle and González-Rivera (1991).

Algorithm 4.1 (Estimation of an ACD model with mixture distribution for the innovation)

1. *Fit an ACD model with a flexible parametric distribution for the innovation, i.e., the GB2 distribution, and obtain the residuals* $\widehat{\epsilon}_i = y_i/\widehat{\psi}_i$, *where* $\widehat{\psi}_i$ *is the current estimate of* ψ_i.

2. *Check if the residuals have unit mean, if not scale them appropriately.*

3. *Estimate the distribution of the residuals using a mixture distribution.*

4. *Re-estimate the parameters of the conditional duration while keeping the parameters of the mixture distribution fixed.*

5. *Iterate steps (2-4) until the increase in the log-likelihood value is negligible, e.g., stop if* $|\mathcal{L}_{i+1} - \mathcal{L}_i| < 10^{-6}$, *where* $\mathcal{L}_i$ *is the log-likelihood value of the* i*-th iteration.*

Note that we do not have to impose constraints on the mixture distribution such that the error has a unit mean since the residuals are appropriately scaled during the fitting procedure.

4.3.2 Estimation of the ACD models

Table 4.1 and Table 4.2 show the estimation results for the ACD(1,1), BCACD(1,1), ACD(2,2), and BCACD(2,2) models of Microsoft price durations. The estimation results

[9] For this purpose, we also add the constraint that the mean of the GB2 distribution exists, which is true if $aq \geq 1$.

for the price durations of the other four stocks are given in Appendix A.1. The estimation results for the five different stocks are qualitatively comparable, so that the following interpretation of the estimation results for Microsoft price durations is representative for all considered price durations.

Table 4.1 shows for Microsoft price durations the estimation results of the ACD(1,1) and BCACD(1,1) models with different error distributions. In terms of log-likelihood and AIC, the fit of the models is greatly improved if we allow for a flexible marginal distribution of the error $\mathcal{E}$. The use of the GB2 distribution for the distribution of $\mathcal{E}$ increases the log-likelihood of the fitted models to a large extent and the in-sample-fit can be further improved if we use a mixture of the GB2 and the log-Normal distribution. By means of the mixture distribution we can model the high probability of very small durations and also the local mode of the marginal density that is visible in Figure 4.1. Applying Neyman's smooth test for the error distribution reveals that a mixture distribution is necessary for a correct specification of the marginal distribution of $\mathcal{E}$. The more complex specification of the conditional expectation duration in the BCACD(1,1) model also improves the in-sample-fit. However, in comparison with the improvement due to a better specification of the distribution of $\mathcal{E}$ this improvement is almost negligible.[10] Even the basic ACD(1,1) model with a GB2 error distribution greatly outperforms the BCACD(1,1) model with an exponential error distribution.

Table 4.2 reports the estimation results of the ACD(2,2) and BCACD(2,2) models with different error distributions. We have included (BC)ACD(2,2) models in our study because the first two lags of the sample autocorrelation functions of the residuals of the (BC)ACD(1,1) models still exhibit a non-negligible correlation. A closer look at Table 4.1 and Table 4.2 shows that the improvement in the log-likelihood is larger when we move from an ACD(1,1) model to an ACD(2,2) model than when we move from an ACD(1,1) model to a BCACD(1,1) model. The best ACD model in terms of AIC, the BCACD(2,2)-GB2-LOGN model with 14 parameters, is not very parsimonious, however such a complex model is required to guarantee that classical goodness-of-fit tests do not reject the model.

4.3.3 Goodness-of-fit of the ACD models

Performance of the ACD model wrt classical goodness-of-fit criteria

The first row of Figure 4.2 visualizes classical goodness-of-fit tests for the BCACD(2,2)-GB2-LOGN model of Microsoft price durations. The results for the other price durations are qualitatively comparable and given in Appendix A.2. The first panel in the first row of Figure 4.2 shows the histogram of estimated probability integral transforms of the error $\mathcal{E}$.

[10] We have also investigated other specifications of the AACD family such as the augmented ACD or the asymmetric logarithmic ACD (Fernandes and Grammig, 2006). However, the BCACD(1,1) model turns out to be the best member wrt to the AIC in almost any case.

Table 4.1: Estimation results of the ACD(1,1) and BCACD(1,1) models with different error distributions for price durations of Microsoft.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.025	-0.136	0.023	-0.096	0.031	-0.076
α	0.089	0.158	0.078	0.106	0.069	0.080
β	0.889	0.941	0.901	0.967	0.899	0.962
ν	-	0.663	-	0.799	-	0.868
a	-	-	3.069	3.101	0.526	0.491
b	-	-	2.532	2.532	30.132	31.456
p	-	-	0.133	0.132	4.023	4.677
q	-	-	0.908	0.869	28.300	29.178
μ	-	-	-	-	-5.043	-5.006
σ^2	-	-	-	-	1.722	1.727
w	-	-	-	-	0.243	0.245
$\log \mathcal{L}$	-8492.787	-8468.038	-5808.859	-5798.014	-5308.072	-5300.928
AIC	16991.573	16944.076	11629.717	11610.028	10636.144	10623.856

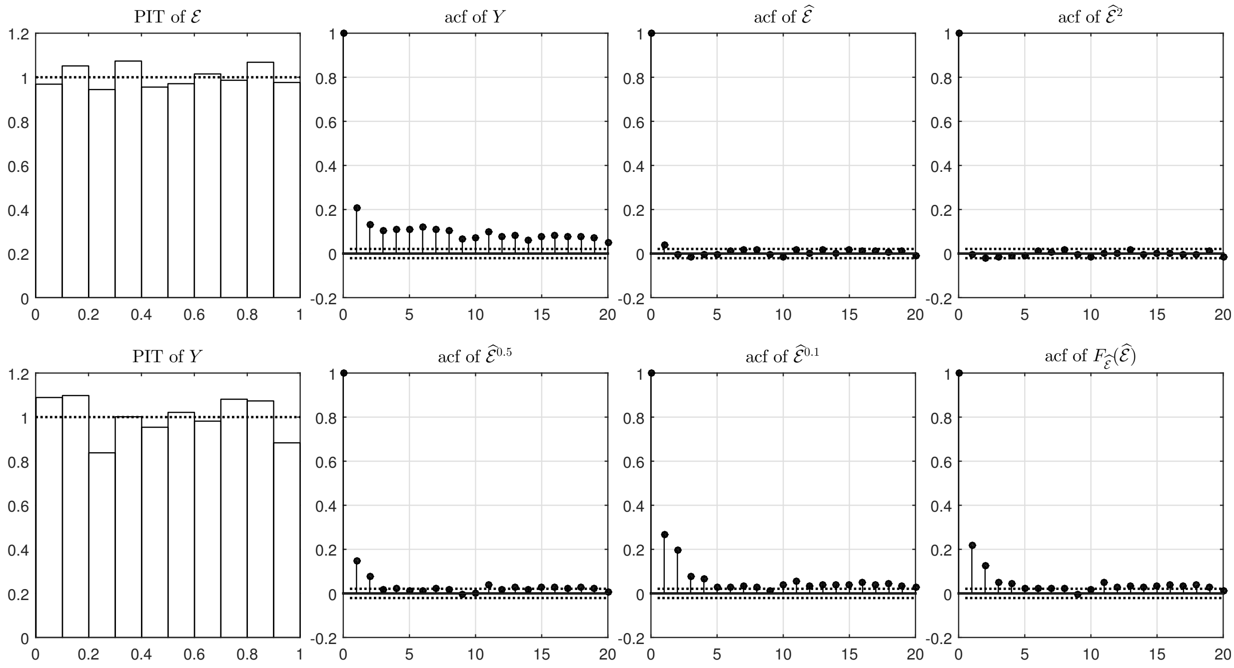
Table 4.2: Estimation results of the ACD(2,2) and BCACD(2,2) models with different error distributions for price durations of Microsoft.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.035	-0.790	0.032	-0.652	0.043	0.065
α_1	0.133	0.219	0.117	0.164	0.100	0.391
α_2	0.000	0.603	0.000	0.507	0.000	-0.455
β_1	0.386	0.509	0.400	0.515	0.435	1.434
β_2	0.450	0.414	0.454	0.439	0.423	-0.443
ν_1	-	0.648	-	0.758	-	0.323
ν_2	-	0.000	-	0.000	-	0.209
a	-	-	2.491	2.491	0.522	0.479
b	-	-	3.103	3.134	31.333	32.274
p	-	-	0.132	0.132	4.088	4.934
q	-	-	0.884	0.842	28.343	29.515
μ	-	-	-	-	-5.028	-4.981
σ^2	-	-	-	-	1.719	1.724
w	-	-	-	-	0.243	0.245
$\log \mathcal{L}$	-8459.646	-8431.008	-5793.348	-5779.569	-5297.989	-5234.043
AIC	16929.292	16876.015	11602.697	11579.138	10619.977	10496.086

$(\omega, \alpha, \beta, \nu)$ are the parameters of the conditional expected duration, see (4.3.2) and (4.3.3). (a, b, p, q) are the parameters of the estimated GB2 distribution of $\mathcal{E}_i$, see (4.3.1). In the last two columns a mixture of the GB2 distribution and the log-normal distribution is estimated for the marginal distribution of $\mathcal{E}$. w denotes the weight for the log-normal distribution with parameters (μ, σ^2). $\log \mathcal{L}$ reports the value of the maximized log-likelihood function and AIC reports the Akaike information criterion.

If and only if the distribution of $\mathcal{E}$ is correctly specified, the histogram of the probability integral transforms corresponds to the histogram of realizations from the standard uniform distribution. We observe that the heights of the bins are close to one which suggests that the distribution of $\mathcal{E}$ is correctly specified. Applying Neyman's smooth test confirms this expectation, the p-value is close to one and given in Table 4.3. The next question we address is how well the ACD model captures the temporal dependence of price durations. The second panel in the first row of Figure 4.2 shows the sample autocorrelation function of the price durations, which is slowly decaying. If N is the sample size, $\widehat{\Psi}_i$ denotes the estimated conditional expected duration at point i and $\widehat{\mathcal{E}}_i = Y_i/\widehat{\Psi}_i$, then the residual series $\widehat{\mathcal{E}} = (\widehat{\mathcal{E}}_i)_{i=1,\dots,N}$ of the ACD model should not exhibit any autocorrelation. The third and fourth panels in the first row of Figure 4.2 show the autocorrelation functions of the ACD-implied residuals and squared residuals. We observe that the estimated autocorrelation functions are close to zero. Thus, the BCACD(2,2)-GB2-LOGN model is successful in removing the autocorrelation structure of price durations.

Figure 4.2: Goodness-of-fit diagnostics for the BCACD(2,2)-GB2-LOGN model of Microsoft price durations.



The left panel shows the histogram of the estimated probability integral transform (PIT) of the error $\widehat{\mathcal{E}}$ and the price duration Y. The estimated marginal distribution of the ACD model is obtained via simulation. The other plots show the empirical autocorrelation of the duration process Y and various transformation of the error sequence $\widehat{\mathcal{E}}$ with approximate 95% confidence intervals. For $q = 2, 0.5, 0.1$, $\widehat{\mathcal{E}}^q$ is the sequence given by $\widehat{\mathcal{E}}^q = (\widehat{\mathcal{E}}_i^q)_{i=1,\dots,N}$ and $F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}) = (F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}_i))_{i=1,\dots,N}$ is the sequence of estimated PITs.

Nevertheless, there may be important non-linear dependencies which the ACD model fails to capture. If the ACD model describes the transition distribution of the duration series, then $\widehat{\mathcal{E}}$ should not only be a sequence of uncorrelated but of independent random variables. As a result, applying a transformation to the residuals should not induce correlation among the transformed residuals. The second and third panel in the second row of Figure 4.2 show the sample autocorrelation functions of the transformed residuals $\widehat{\mathcal{E}}^{0.5} = (\widehat{\mathcal{E}}_i^{0.5})_{i=1,\ldots,N}$ and $\widehat{\mathcal{E}}^{0.1} = (\widehat{\mathcal{E}}_i^{0.1})_{i=1,\ldots,N}$. It is evident that the first lags of these autocorrelation function are not close to zero, implying that the ACD model does not capture all (non-linear) dependencies which are present in the price durations of the Microsoft stock.

This interesting finding demonstrates the inherent problem of using Pearson's product-moment correlation coefficient to measure the dependence in a series of positive random variables. As it is well known, the classical correlation coefficient is not invariant under nonlinear strictly increasing transformations (McNeil et al., 2005) and its interpretation as a measure of dependence is not clear for non-elliptical distributions. The remaining non-linear dependence structure in the highly skewed residual series $\widehat{\mathcal{E}}$ becomes more pronounced in linear terms if we take the residuals to the power of 0.1. That is because the marginal distribution of $\widehat{\mathcal{E}}^{0.1}$ is more symmetric. As a result, the joint distribution of $(\widehat{\mathcal{E}}_i^{0.1}, \widehat{\mathcal{E}}_j^{0.1})$ is closer to an elliptical distribution, so that $\widehat{\mathcal{E}}_i^{0.1}$ and $\widehat{\mathcal{E}}_j^{0.1}$ are more linearly related than $\widehat{\mathcal{E}}_i$ and $\widehat{\mathcal{E}}_j$ for $i, j \in \mathbb{N}, i \neq j$.

A more appropriate dependence measure is Spearman's correlation coefficient (Kruskal, 1958), which is a measure of concordance (Scarsini, 1984) and invariant under (almost surely) monotone transformations of the margins. The last panel in the second row of Figure 4.2 shows the autocorrelation function in terms of Spearman's correlation coefficient and illustrates that there is a moderate concordance between the residuals of the ACD model. Once again, this rejects the assertion that the ACD model is an adequate model to represent the transition distribution of price durations.

Finally, we also investigate whether the implied marginal distribution of the ACD model is correctly specified.[11] Since it is impossible to compute the marginal distribution of the ACD model directly, we approximate the implied marginal distribution by applying a kernel density estimator to simulated observations from the fitted ACD model. For that purpose, we simulate $N = 1,000,000$ observations from the fitted ACD model and keep each tenth observation, so that we have a total number of $100,000$ observations from the marginal distribution. We then apply a kernel density estimator to log-durations and transform the estimator back to plain durations in order to avoid boundary effects due to the non-negativeness of durations. The first panel in the second row of Figure 4.2 shows the histogram of the PIT which is derived from the estimated marginal distribution of the ACD model. Moreover, Table 4.3 shows the p-value of Neyman's smooth test for the ACD-implied marginal distribution.[12]

[11] A correctly specified distribution of the error does not imply that the marginal distribution is correctly specified and vice versa.

[12] The resulting estimate of the ACD-implied marginal distribution seems to be robust, i.e., the p-value of Neyman's smooth test for the ACD-implied marginal distribution in Table 4.3 only varies slightly in the second decimal place when running the simulation again.

Table 4.3: p-values of Neyman's smooth test with four components for the BCACD(2,2)-GB2-LOGN models.

	Apple	Cisco	Hewlett-Packard	Intel	Microsoft
Marginal distribution of $\mathcal{E}$	0.940	0.957	0.982	0.992	0.958
Marginal distribution of Y	0.294	0.006	0.086	0.420	0.018

Despite the fact that the distribution of the error appears to be correctly specified, we find that the implied marginal distribution of the ACD model of Microsoft price durations is rejected by Neyman's smooth test at a 5% significance level. However, the computation of the p-value assumes that we have a direct expression for the marginal distribution of the ACD model and it also does not account for the uncertainty in parameter estimates.[13] Nevertheless, this result might also indicate that the ACD model does not take all (non-linear) dependencies into consideration. Table 4.3 also shows that Neyman's smooth test rejects the ACD-implied marginal distribution of the price durations of Hewlett-Packard at a 10% significance level and the fit of the ACD-implied marginal distribution of the price durations of Cisco is rejected at a 1% significance level.

Performance of the ACD model in terms of copulas

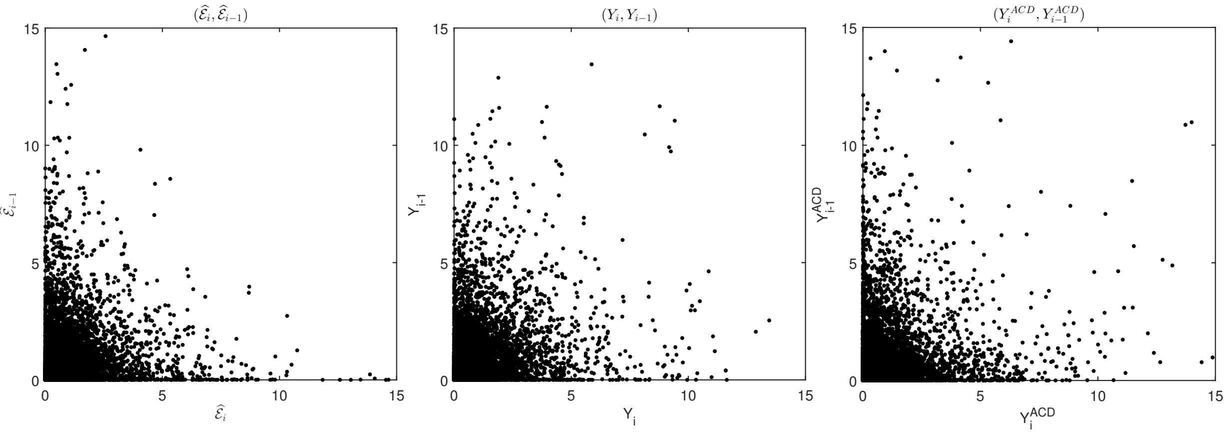
In order to shed light on the reasons why the ACD model does not capture all (non-linear) dynamics, we apply graphical methods of exploratory data analysis. As a first step, we analyze the dependence between two consecutive observations from the price duration data and the ACD-implied dependence between two consecutive durations by means of simple scatter plots. The panels in Figure 4.3 display the scatter plots of consecutive residuals which are obtained from the ACD model, consecutive observations from the data of price durations, and consecutive price durations which are simulated from the fitted ACD model. Because of the very skewed marginal distributions, the scatter plots of plain durations in Figure 4.3 are not very informative and obscure the underlying dependence structures.

In order to analyze the dependence structure, we use copula scatter plots which are the scatter plots of normalized ranks. The normalized ranks of $(y_i)_{i=1,\dots,N}$ are given by the values of the rescaled empirical distribution function which is defined as $F_Y^R(y_i) = \frac{1}{N+1}\sum_{k=1}^{N} \mathbb{1}_{\{y_k \leq y_i\}}$. The normalized ranks are non-parametric estimates of the probability integral transforms and a scatter plot of consecutive normalized ranks corresponds to a scatter plot of consecutive pseudo-observations from the underlying copula. If two time series have different marginal distributions but exhibit the identical dependence structure, then the copula scatter plots of consecutive observations should look the same.

The first row of Figure 4.4 shows copula scatter plots of consecutive residuals of the fitted ACD model, the actual price durations, and simulated price durations from the

[13] Regarding the sample size of price durations, the estimation uncertainty should be negligible, and the error of our simulation setup should also be rather small.

Figure 4.3: Microsoft: Residuals of the BCACD(2,2)-GB2-LOGN model, dependence of price durations, and BCACD-GB2-LOGN implied dependence of price durations.



The left plot shows the scatter plot of consecutive residuals of the fitted ACD model. The middle plot shows the scatter plot of consecutive price durations. The right plot shows the scatter plot of simulated consecutive price durations from the fitted ACD model.

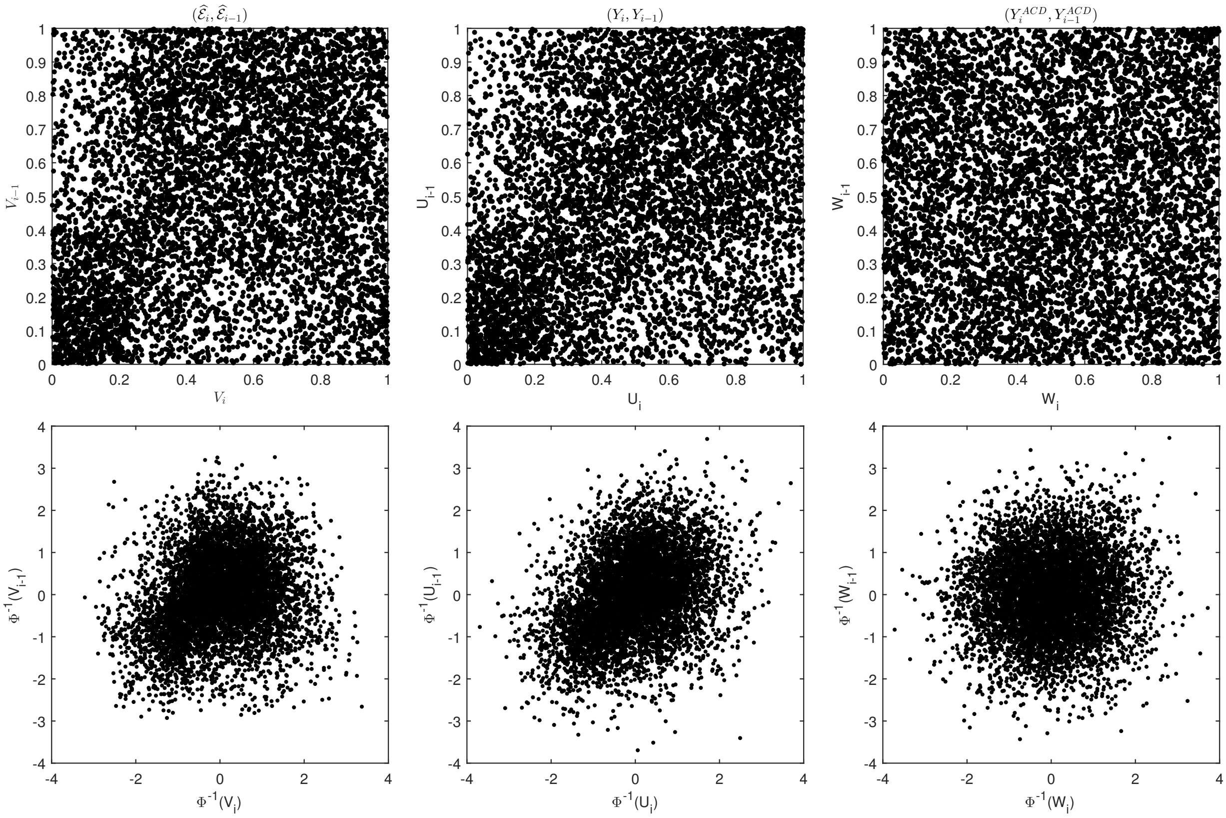
fitted ACD model.[14] The second row of Figure 4.4 shows the corresponding scatter plots if the quantile function of the standard normal distribution is applied to the normalized ranks. The figures for the other price durations are qualitatively comparable and given in Appendix A.3.

The left panel in the first row of Figure 4.4 shows the copula scatter plot of consecutive ACD-implied residuals. We observe that this scatter plot is very similar to the copula scatter plot of consecutive price durations which is shown in the middle panel of the first row of Figure 4.4. That is, although the fitted ACD model removes the autocorrelation of price durations, the largest part of the dependence between consecutive price durations is still present. This is also confirmed by the third panel in the first row of Figure 4.4 which shows the copula scatter plot of consecutive durations that are simulated from the fitted ACD model. The copula scatter plot of simulated price durations shows no clear pattern. It rather looks like a scatter plot of realizations that are drawn from the product copula.

Contrary to the ACD-implied dependence, there is a moderate positive dependence in the copula scatter plot of consecutive observations from the price duration data. Moreover, the dependence structure is asymmetric wrt the side diagonal given by the set $\{(u_i, u_{i-1})\colon u_{i-1} = 1 - u_i\}$. Consider the area given by $\{(u_i, u_{i-1})\colon (u_i, u_{i-1}) \in (0,1) \times (0, 0.2)\}$ which corresponds to consecutive normalized ranks of durations where the previous duration was below its 20% quantile. The next durations strongly cluster in the area $\{(u_i, u_{i-1})\colon (u_i, u_{i-1}) \in (0, 0.2) \times (0, 0.2)\}$. As a result, we expect a small duration if the former duration was small. On the other hand, the clustering of large durations

[14] We simulate $N = 1,000,000$ observations from the fitted ACD model and keep each tenth observation, so that we have a total number of 100,000 observations from the marginal distribution. Checks indicate that the resulting scatter plot is representative and robust against sampling variation.

Figure 4.4: Microsoft: Dependence of the residuals of the BCACD-GB2-LOGN model, dependence of price durations, and BCACD-GB2-LOGN-implied dependence of price durations.



The first row shows copula scatter plots of consecutive residuals $\widehat{\mathcal{E}}$ of the BCACD(2,2)-GB2-LOGN model, consecutive price durations Y, and consecutive observations Y^{ACD} that are simulated from the BCACD(2,2)-GB2-LOGN model. The second row shows the scatter plots of the corresponding pseudo-$N(0,1)$ observations, i.e., the scatter plots that arise if the quantile function of the $N(0,1)$ distribution is applied to the normalized ranks. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data if the number of scatter plot observations exceeds 7000 observations.

seems to be less pronounced. If we consider the area given by $\{(u_i, u_{i-1})\colon (u_i, u_{i-1}) \in (0,1) \times (0.8,1)\}$, we see that large durations are rather followed by large durations, but there are less observations in the area $\{(u_i, u_{i-1})\colon (u_i, u_{i-1}) \in (0.8,1) \times (0.8,1)\}$ than in the area $\{(u_i, u_{i-1})\colon (u_i, u_{i-1}) \in (0,0.2) \times (0,0.2)\}$, indicating a weaker positive dependence between large durations. It is also interesting to see that the normalized ranks of consecutive durations appear to be symmetrically scattered about the main diagonal $\{(u_i, u_{i-1})\colon u_{i-1} = u_i\}$. This indicates a symmetric distribution between plain durations and thus a time-reversible process.

4.4 Modeling price durations with CMP models

The normalized ranks of price durations in the middle panel of Figure 4.4 represent pseudo-observations from the corresponding copula of consecutive durations, which captures the complete dependence structure between consecutive durations. We now apply an SD-vine copula-based Markov process to obtain a time series model which captures the dependence of consecutive price durations.

4.4.1 Estimation of the marginal distribution

First of all, we have to set up a model for the marginal distribution of the series of price durations. Neyman's smooth test shows that a mixture of the GB2 distribution and the log-Normal distribution provides an adequate fit of the marginal distribution. The corresponding estimation results for all five price durations and the p-vales of Neyman's smooth test are given in Table 4.4. For Apple, Cisco, and Microsoft, the marginal models provide, in terms of log-likelihood and AIC, a better fit to the data than the (BC)ACD(1,1) or (BC)ACD(2,2) models with a GB2 distribution. In this regard, the adequate modeling of the marginal distribution is a crucial point in the modeling of the price duration data.

Table 4.4: Estimation results for the GB2-LOGN-mixture model of the marginal distribution of price durations.

	Apple	Cisco	Hewlett-Packard	Intel	Microsoft
a	0.797	0.788	0.629	0.632	0.528
b	22.448	4.813	20.006	8.816	18.198
p	1.743	1.922	2.596	3.002	3.944
q	20.373	7.053	18.718	12.535	18.864
μ	-7.043	-4.959	-5.193	-4.331	-5.226
σ^2	1.548	1.673	1.607	1.877	1.743
w	0.090	0.172	0.147	0.186	0.243
$\log \mathcal{L}$	-2406.241	-7858.997	-7178.171	-10763.800	-5633.211
AIC	4826.481	15731.994	14370.343	21541.600	11280.422
GoF	0.994	0.902	0.937	0.839	0.858

A mixture of the GB2 distribution and the log-normal distribution is estimated for the marginal distribution of price durations. (a, b, p, q) are the parameters of the estimated GB2 distribution, see (4.3.1), w denotes the weight for the log-normal distribution with parameters (μ, σ^2). $\log \mathcal{L}$ reports the value of the maximized log-likelihood function and AIC reports the Akaike information criterion. GoF displays the p-values of Neyman's smooth test with four components.

4.4.2 Specification and estimation of the copula sequence

The exploratory data analysis of consecutive normalized ranks of Microsoft price durations (see the middle panels of Figure 4.4 or the first panel in the first column of Figure 4.5) has shown that the clustering of short durations is more pronounced than the clustering of large durations. Therefore, bivariate copulas that place more probability mass in the lower left corner of the unit cube than in its upper right corner should be specified to model the dependence of consecutive durations. Simple copula families with a scalar parameter that reproduce these features are the Clayton copula and the survival version of the Gumbel copula. However, for moderate positive dependence, these copula families induce a very strong clustering of observations in the lower left corner of the unit cube, which is not present in the scatter plot of consecutive normalized ranks of Microsoft price durations. In order to obtain a copula with moderate positive dependence and less strongly asymmetric dependence structure, we also consider the BB1 copula and mixtures of the Clayton or survival Gumbel copula with radially symmetric copulas, like the Gaussian or Frank copula.

Table 4.5: Specification of the copula family of consecutive Microsoft price durations.

Family	(w, θ) or θ	$\log \mathcal{L}$	AIC
rGU-FR	$(0.735, 0.264, 1.182, 5.486)$	663.94	-1321.9
CL-FR	$(0.359, 0.640, 0.927, 1.820)$	656.56	-1307.1
CL-rGU	$(0.713, 0.286, 0.273, 1.962)$	651.91	-1297.8
rGU-GA	$(0.487, 0.512, 1.579, 0.164)$	641.52	-1277.0
CL-GA	$(0.329, 0.670, 1.300, 0.235)$	640.21	-1274.4
BB1	$(0.371, 1.079)$	631.05	-1258.1
rGU	1.283	627.62	-1253.2
FR	2.259	590.67	-1179.3
CL	0.489	590.25	-1178.5
GA	0.340	565.18	-1128.4

GU, FR, GA are the Gumbel, Frank and Gaussian copula. r refers to the survival version and - indicates a mixture, i.e., rGU-FR is a mixture of the rotated Gumbel and the Frank copula. The second column contains the weights $w = (w_1, w_2)$ and parameters $\theta = (\theta_1, \theta_2)$ of the mixture copula or the scalar parameter θ of the copula. $\log \mathcal{L}$ and AIC are the log-likelihood or AIC value of the corresponding copula.

Table 4.5 reports the estimation results for various copula families of consecutive Microsoft price durations. Among the simple copula families that are parameterized by a scalar parameter, we find that the survival version of the Gumbel copula gives the best fit, which is in line with the exploratory data analysis. Contrary to the Frank and Gaussian copula, the survival version of the Gumbel copula places more probability mass in the lower left corner of the unit cube than in its upper right corner. Moreover, the induced asymmetric dependence structure of the survival version of the Gumbel copula is less pronounced than it is for the Clayton copula. Table 4.5 also shows that more complex copula specifications clearly outperform simple copula families with only one scalar parameter. The best specification is given by a mixture of the survival version of the Gumbel copula and the Frank copula. Thus, we choose this mixture copula as the first copula in the simplified SD-vine copula-based Markov model.

We now set $P = 40$ for the maximal order of the Markov process. In order to select the remaining copula families in the copula sequence $(C_i)_{i=1,\ldots,40}$ we apply Algorithm 3.1. That is, for $i = 2, \ldots, 40$, we use C_{i-1} to compute the pseudo-observations of C_i and then use the AIC to choose the copula family for C_i. For each $i = 2, \ldots, 40$, we specify the same copula families as in Table 4.5 as possible candidates for the copula family. For $i = 1, 2, 3, 4$, Figure 4.5 shows the corresponding scatter plot of pseudo-observations from C_i. It is striking that the dependence in the scatter plots quickly decreases with i. Already the scatter plot of the pseudo-observations from C_2 does not indicate a clear dependence pattern. After having selected the copula families of the CMP(40) model, we

Figure 4.5: Microsoft: Pseudo-observations from $C_i \in \mathbb{C}_p$ for $i = 1, \ldots, 4$. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data.

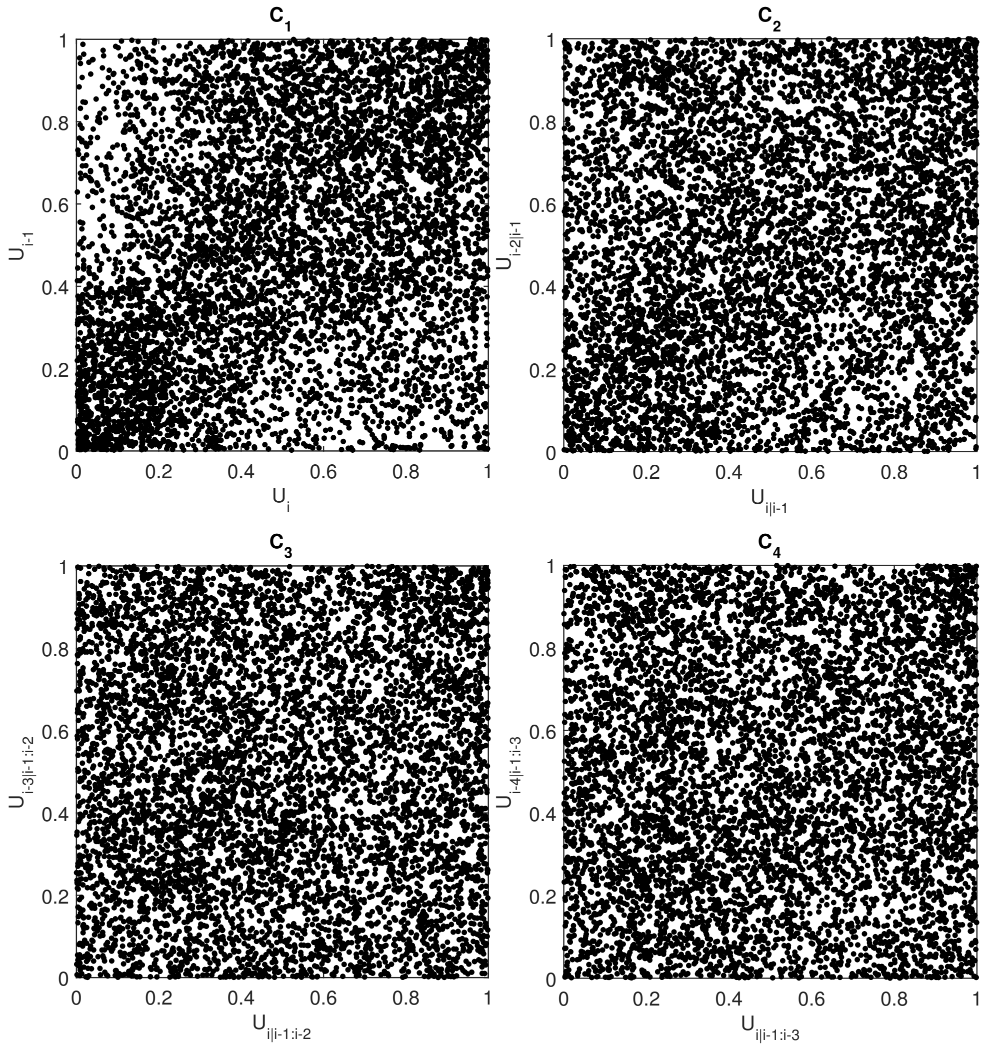
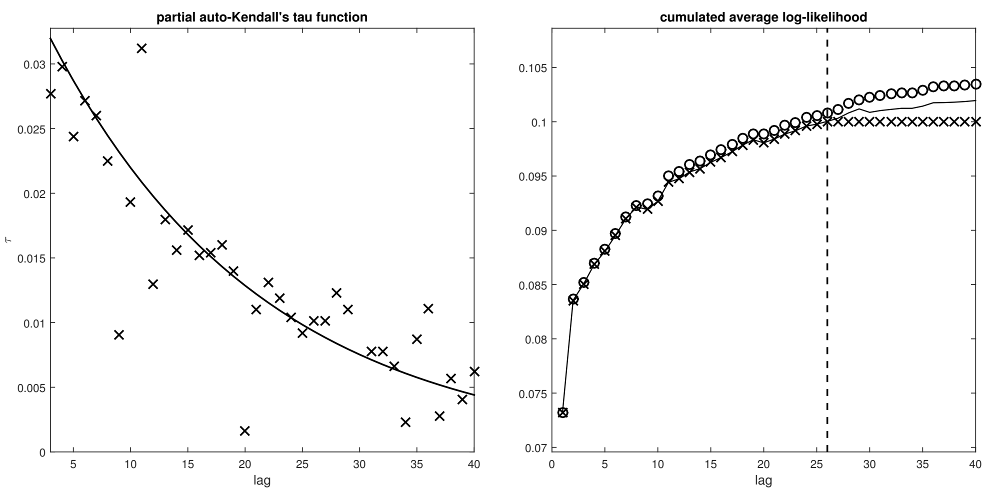


Figure 4.6: Microsoft: Graphical illustration of the CMP model which is parameterized by an exponential lag function g_3 with starting point $k = 3$.



Left panel: The marks show the sequence $(\tau_j)_{j=3,\ldots,40}$, where τ_j is the estimated Kendall's tau of C_j using the first four steps in Algorithm 3.4. The line refers to the corresponding $(\tau_j^{g_3})_{j=3,\ldots,40}$ sequence which is implied by an exponential lag function g_3 using step 5 and 6 of Algorithm 3.4. Right panel: The circles display the cumulated average log-likelihoods $\frac{1}{T}\sum_{i=1}^{j}\mathcal{L}_i(\tau_i)$ that correspond to the dots in the left panel. The line displays the cumulated average log-likelihoods $\frac{1}{T}\sum_{i=1}^{j}\mathcal{L}_i(\tau_i^{g_3})$ that are associated with the line in the left panel. The marks indicate the cumulated average log-likelihoods that result from the truncated copula sequence that is obtained via Definition 3.8 if $\Delta = 0.0021$. The truncation point $p^\star$ is indicated by the vertical line.

apply Algorithm 3.4 in order to estimate a parsimonious CMP model with a lag function. We specify an exponential lag function that starts with the third lag and truncate the copula sequence so that the loss in average log-likelihood does not exceed $\Delta = 0.0021$.[15]

Figure 4.6 illustrates the differences between the CMP(40) model that is estimated without any constraints and the CMP(40) model that is parameterized by a lag function that starts with the third lag. The order of the truncated CMP model is 26 and Table 4.6 shows the corresponding estimation results. According to the log-likelihood of the fitted copula C_i, the temporal dependence in the price durations quickly decreases from the first to the third lag and then slowly decays. This slow decay is successfully modeled by the exponential lag function as Figure 4.6 demonstrates. The second parameter of the exponential lag function is close to one, which confirms the slow decay of temporal dependence from the third copula on. The estimation results for the other four price durations are qualitatively comparable and are given in Appendix A.4 and Appendix A.5.

4.4.3 Goodness-of-fit of the CMP models

Performance of the CMP model wrt classical goodness-of-fit criteria

We now evaluate the goodness-of-fit of the estimated CMP models and begin with classical goodness-of-fit criteria where the ACD models perform well. These classical goodness-of-fit criteria are visualized for the CMP model of Microsoft price durations in the first row of Figure 4.7. The third and fourth panel in the first row of Figure 4.7 show the autocorrelation functions of the residuals and squared residuals of the CMP model under the assumption that the residuals have the same marginal distribution as the residuals of the ACD model.[16] Similar to the ACD model, the CMP model is successful in removing the autocorrelation structure of price durations and the estimated autocorrelation function of the squared residuals is also close to zero. However, contrary to the ACD model, the CMP model also removes much more non-linear dependencies. The autocorrelation functions of the transformed residuals $\widehat{\mathcal{E}}^{0.5} = (\widehat{\mathcal{E}}_i^{0.5})_{i=1,\dots,N}$ and $\widehat{\mathcal{E}}^{0.1} = (\widehat{\mathcal{E}}_i^{0.1})_{i=1,\dots,N}$, which exhibit a non-negligible pattern for the ACD model (see Figure 4.2), are close to zero for the CMP model as can be seen from the second and third panel in the second row of Figure 4.7. Moreover, the last panel in the second row of Figure 4.7 shows that there is no concordance between the residuals of the CMP model. According to these goodness-of-fit measures, the CMP model seems to be an adequate model of the transition distribution of price durations.

[15] The use of a Pascal lag or an exponential Almon lag function does not improve the fit. The choice that the exponential lag function starts with the third lag is based on the scatter plot of pseudo-observations from C_i, for $i = 1, 2, 3, 4$.

[16] The canonical residuals $(\widehat{z}_i)_{i=1,\dots,N}$ of a CMP model are given by the evaluated estimated transition distributions, i.e., $\widehat{z}_i = \widehat{F}_{i|i-1:i-p}(y_i|y_{i-1:i-p})$. In order to make the autocorrelation function of the residuals of the CMP model comparable with the autocorrelation function of the residuals of the ACD model we apply the quantile function G^{ACD} of the ACD-implied residuals to the residuals of the CMP model, i.e., $\widehat{\epsilon}_i = G^{\mathrm{ACD}}(\widehat{z}_i)$ for $i = 1, \dots, N$.

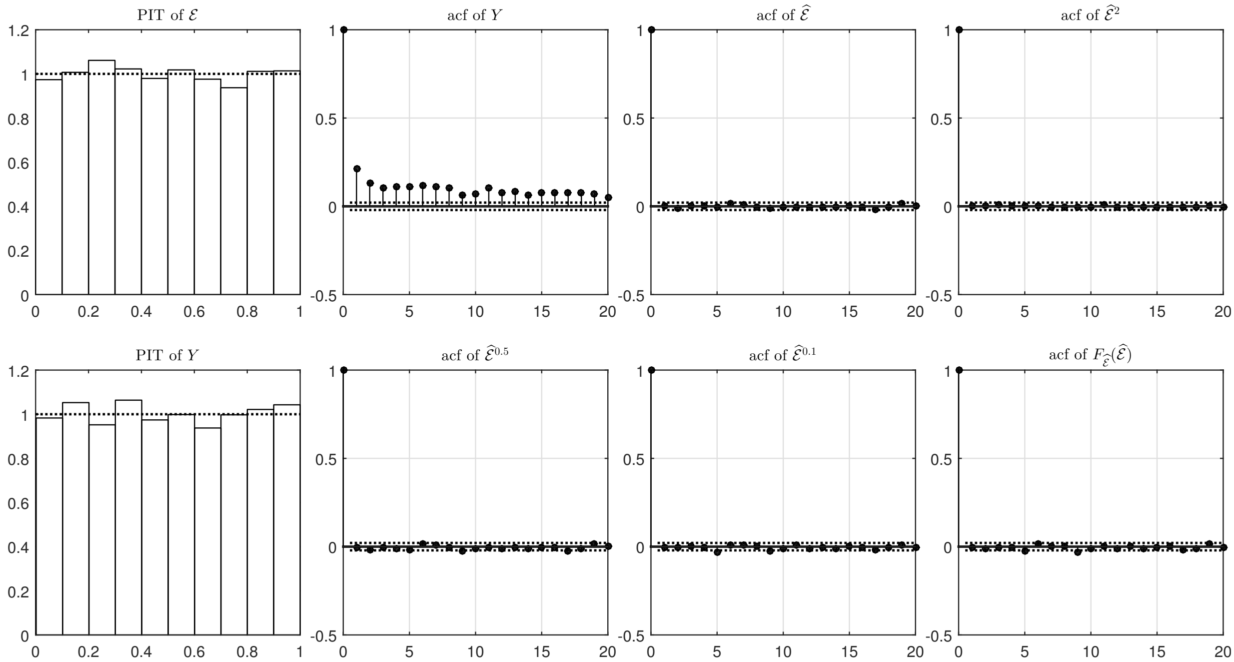
Table 4.6: Microsoft: Estimation results for the truncated CMP(26) model with a lag function. Kendall's τ of $C_{3:26}$ is modeled by an exponential lag decay function, i.e., $\tau_i = \tau_3\gamma^{i-3}$ for $i = 3, \dots, 26$.

	Family	(w, θ) or θ	Kendall's τ	$\log \mathcal{L}$	AIC
C_1	rGU-FR	(0.714, 0.286, 1.184, 5.247)	-	663.651	-1321.302
C_2	rGU-GA	(0.083, 0.917, 2.198, 0.091)	-	93.663	-181.325
C_3	rCL	0.066	0.032	13.854	-25.900
C_4	GU	1.031	0.030	16.240	-30.430
C_5	rCL	0.059	0.029	10.782	-20.275
C_6	rCL	0.056	0.027	13.402	-24.755
C_7	GU	1.026	0.026	14.162	-26.048
C_8	rCL	0.050	0.024	9.593	-16.724
C_9	rCL	0.047	0.023	-2.043	-1.046
C_{10}	CL	0.045	0.022	6.826	-11.395
C_{11}	rCL	0.043	0.021	15.780	-32.018
C_{12}	GU	1.020	0.020	2.956	-4.779
C_{13}	rCL	0.038	0.019	5.713	-9.617
C_{14}	GA	0.028	0.018	2.806	-3.527
C_{15}	rCL	0.034	0.017	5.231	-8.501
C_{16}	CL	0.032	0.016	4.163	-6.330
C_{17}	GU	1.015	0.015	4.621	-7.168
C_{18}	rCL	0.029	0.014	5.557	-7.743
C_{19}	GU	1.014	0.014	4.479	-6.140
C_{20}	rCL	0.026	0.013	-2.520	1.901
C_{21}	GU	1.012	0.012	3.162	-3.400
C_{22}	rGU	1.012	0.012	4.228	-5.931
C_{23}	rCL	0.022	0.011	2.679	-3.481
C_{24}	GU	1.010	0.010	3.901	-5.956
C_{25}	rCL	0.020	0.010	1.599	-1.109
C_{26}	GU	1.009	0.009	2.366	-2.830
lag-decay	Exponential	(0.032, 0.947)	-	-	-

The second column displays the copula family that is best wrt the AIC. GU, FR, GA are the Gumbel, Frank and Gaussian copula. r refers to the survival version of a copula and - indicates a mixture, i.e., rGU-FR is a mixture of the survival Gumbel and the Frank copula. The first 26 rows of the third column contain the weights $w = (w_1, w_2)$ and parameters $\theta = (\theta_1, \theta_2)$ of the mixture copula or the scalar parameter θ of the copula. The fourth column shows the values of Kendall's τ of the copulas $C_{3:26}$ which are implied by the fitted exponential lag function. $\log \mathcal{L}$ and AIC represent the log-likelihood or AIC value of the copula C_i. The third column in the second last row contains the parameters (τ_3, γ) of the fitted lag function.

The first panel in the first row of Figure 4.7 shows the histogram of the estimated probability integral transforms of the innovation sequence of the CMP model. Although the heights of the bins are close to one, Neyman's smooth test rejects the correct specification of the marginal distribution of the innovation at a 5% level, see Table 4.7. This is also true for the Cisco stock. In general, the p-values for Neyman's smooth test are much smaller for the marginal distribution of the innovation $\mathcal{E}$ than for the marginal distribution of price durations Y (see Table 4.4). This is possibly due to the fact that the temporal dependence of the CMP-implied innovation sequence $\mathcal{E}$ is much weaker than the temporal dependence of the price durations, which might improve the power of the test. Moreover, the sample sizes of the price durations of Cisco, Hewlett-Packard, Intel, and Microsoft, are around 10,000 observations. For the Apple stock, we have less than 3000 observations and here the estimated marginal distribution of the innovation sequence is not rejected. In this light, the estimated marginal distributions of the innovation sequences of the CMP

Figure 4.7: Goodness-of-fit diagnostics for the CMP model of Microsoft price durations.



The first column shows the histogram of the estimated PIT of the error $\mathcal{E}$ and the price duration Y. The other plots show the empirical autocorrelation function of Y and various transformations of $\widehat{\mathcal{E}}$ with approximate 95% confidence intervals. For $p = 2, 0.5, 0.1$, $\widehat{\mathcal{E}}^q$ is the sequence given by $\widehat{\mathcal{E}}^q = (\widehat{\mathcal{E}}_i^q)_{i=1,\ldots,N}$ and $F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}) = (F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}_i))_{i=1,\ldots,N}$.

models seem to be adequate. Finally, the second panel in the second row of Figure 4.7

Table 4.7: p-values of Neyman's smooth test with four components for the CMP models.

	Apple	Cisco	Hewlett-Packard	Intel	Microsoft
Marginal distribution of $\mathcal{E}$	0.417	0.020	0.089	0.099	0.017

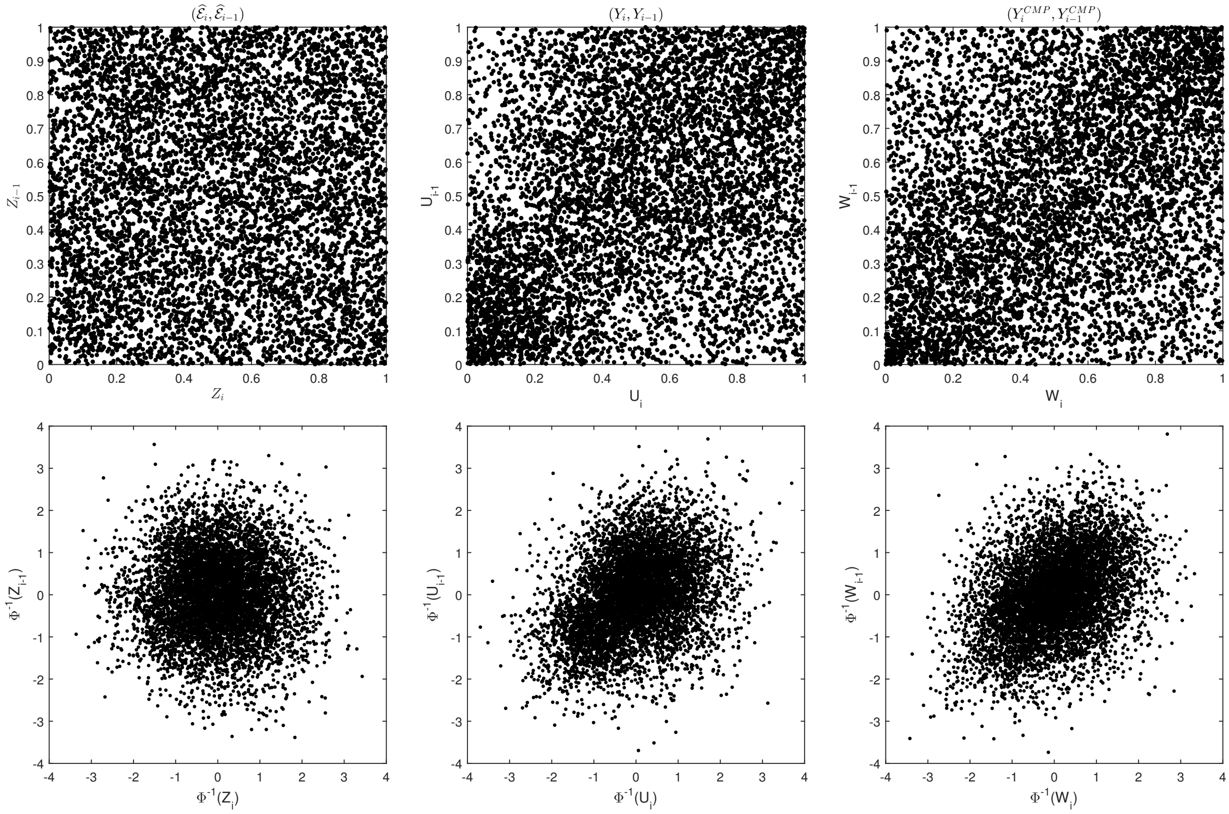
shows the histogram of estimated probability integral transforms of price durations which are based on the marginal distribution of the CMP model. By construction, the marginal distribution of the price durations is well reproduced.[17] In Appendix A.6 it is illustrated that the CMP models for the other price durations also provide a good fit wrt to classical goodness-of-fit criteria.

Performance of the ACD model in terms of copulas

Figure 4.7 and the corresponding figures in Appendix A.6 illustrate that the CMP models provide a good fit wrt to classical goodness-of-fit criteria. We now examine if this is also true for the copula of consecutive price durations. The left panel in the first row of Figure 4.8 shows the copula scatter plot of consecutive CMP-implied residuals. Contrary to the ACD

[17] Due to the joint estimation of the marginal distribution and the copula sequence, the parameters of the estimated marginal distribution of the CMP model are not equal to the parameters of the estimated marginal distribution in Table 4.4. However, the difference between these two marginal models is negligible and the p-values of Neyman's smooth test are identical up to three decimal places.

Figure 4.8: Microsoft: Residuals of the CMP model, dependence of price durations, and CMP-implied dependence of price durations.



The first row shows copula scatter plots of consecutive residuals $\widehat{\mathcal{E}}$ of the CMP model, consecutive price durations Y, and consecutive observations Y^{CMP} that are simulated from the CMP model. The second row shows the scatter plots of the corresponding pseudo-$N(0,1)$ observations, i.e., the scatter plots that arise if the quantile function of the $N(0,1)$ distribution is applied to the data. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data if the number of scatter plot observations exceeds 7000 observations.

model, the scatter plot of consecutive normalized ranks of the CMP-implied residuals shows no clear dependence structure but rather indicates realizations from the product copula. Thus, the CMP model does not only remove the linear dependence between consecutive price durations but also the non-linear dependencies which the ACD model fails to capture.

A simulation of the CMP model reproduces the moderate positive and slightly asymmetric dependence of consecutive price durations. The CMP-implied dependence structure of consecutive price durations in the right panels of Figure 4.8 strongly resembles the actual dependence of consecutive price durations which is visualized in the middle panels of Figure 4.8. According to Figure 4.5, the dependence of consecutive price durations seems to be the central driving factor in the dynamics of price durations, so that the reproduction of this feature is a strong argument for the CMP model. The copula diagnostics for the other four price durations are given in Appendix A.7. Also for the other four series of price durations, realizations from the copulas of consecutive CMP-implied residuals look like realizations from the product copula. Moreover, for each of the four stocks, the moderate positive and slightly asymmetric dependence structure of consecutive durations is reproduced by a simulation of the fitted CMP model.

4.5 Comparison of ACD and CMP models

Table 4.8 reports the AIC values of the estimated ACD and CMP models for all five stocks. Obviously, the in-sample-fit of the CMP models in terms of AIC is much better than the in-sample-fit of the ACD models. Since the ACD models have 14 parameters and the CMP models have 15 parameters, the differences in AIC values are strongly related to the differences in the log-likelihoods of the models.

Table 4.8: In-sample AIC values of the ACD and the CMP models.

Model/Stock	Apple	Cisco	Hewlett-Packard	Intel	Microsoft
BCACD(2,2)	4512.123	15143.296	13477.839	20717.787	10482.086
CMP	4418.602	14280.378	12748.354	19554.053	9450.019

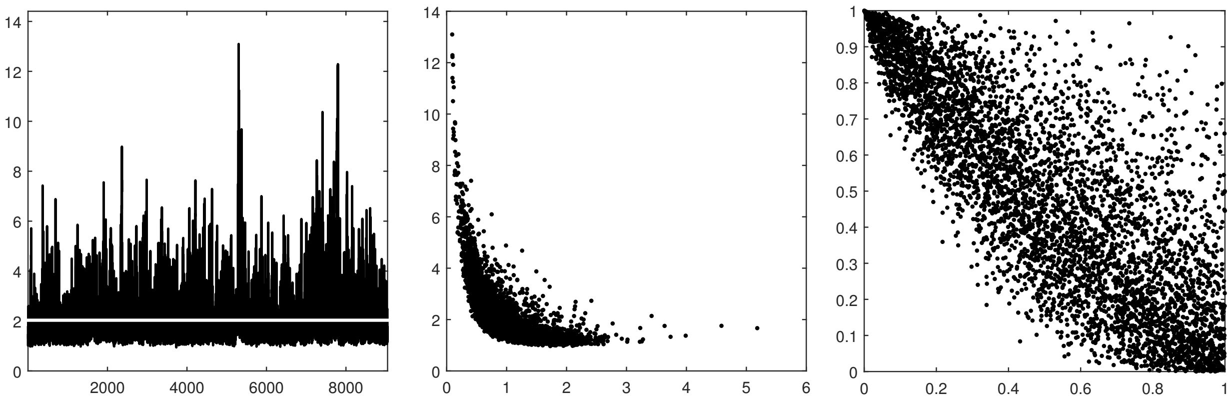
The superior performances of the CMP models can be explained by the fact that, contrary to the ACD models, the CMP models capture the complete dependence of consecutive price durations. In order to elucidate this statement, we consider the AIC values for Microsoft price durations in more detail. Assume that both the ACD and CMP model specify the marginal distribution of Microsoft price durations equally well and that the AIC value of the marginal model is given by 11280.422, which is the AIC value of the marginal model of price durations in Table 4.4. The remaining AIC value of the ACD model which is not accounted for by the marginal distribution is then -798.336 and represents the AIC value of the copula dependence structure of the ACD model. Under the stated assumptions, the AIC of the copula dependence structure of the CMP model has a value of -1830.403. Moreover, the AIC value of the first copula C_1 is -1321.302, so that already the first copula of the copula sequence contributes 72.18% to the overall AIC value of the copula sequence of the CMP model. Consequently, the better modeling of the dependence of consecutive price durations seems to be the main reason for the outstanding in-sample performance of the CMP model.

Time-varying dispersion of the CMP model

The class of ACD models implies that the dispersion of the transition distribution, i.e., σ_i^2/μ_i^2, with $\mu_i = \mathbb{E}[Y_i|\mathcal{I}_{i-1}]$ and $\sigma_i^2 = \mathbb{V}\mathrm{ar}[Y_i|\mathcal{I}_{i-1}]$, is constant and does not depend on the past of the process. Although the CMP-implied dispersion of the transition distribution is in general not tractable, CMP models do in general not impose such a restriction, since for $p = 1$ any duration process can be modeled by our framework. In order to evaluate the CMP-implied dispersion of the transition distribution, we use numerical integration to compute the sequence of conditional dispersions $(\sigma_i^2/\mu_i^2)_{i=1,\ldots,N}$.

The left panel of Figure 4.9 shows the ACD-implied constant dispersion and the CMP-implied time-varying dispersion for the price durations of the Microsoft stock. The ACD-implied constant dispersion has a value of 2.031, which is quite different from the CMP-implied time-varying dispersion which displays a large fluctuation around this value. The 25% and 75% quantiles of the conditional dispersion are given by 1.405 and 2.581, respectively, indicating a remarkable variation of the conditional dispersion over time. Indeed, the constraint of a time-constant conditional dispersion might explain why the ACD model is no match for the CMP model in terms of in-sample-fit.

Figure 4.9: Microsoft: Time-varying dispersion of the CMP model.



The left plot shows the CMP-implied time-varying dispersion σ_i^2/μ_i^2 for the in-sample period. The horizontal white line is the constant dispersion of the BCACD(2,2)-GB2-LOGN model. The middle plot shows the relation between the mean (x-axis) and the dispersion (y-axis) of the transition distribution. The right plot shows the relation between the mean and the dispersion of the transition distribution on the copula scale. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data if the number of scatter plot observations exceeds 7000 observations.

The scatter plot of the mean and the dispersion of the transition distribution in the middle panel of Figure 4.9 indicates a rather strong negative relation between the mean and the dispersion which is highly non-linear. We see that the dispersion of the transition distribution increases enormously if its mean approaches zero. This is also confirmed by the last panel of Figure 4.9 which visualizes the relation between the mean and the dispersion of the transition distribution on the copula scale. The results for the other price durations in Appendix A.8 also confirm a great variation of the CMP-implied time-varying dispersion and a negative relation between the mean and the dispersion of the transition distribution.

Out-of-sample evaluation

Although the CMP models demonstrate an impressive in-sample-fit, this does not imply that their out-of-sample forecast performance is also superior to the ACD models. However, the forecast performance is often the most important criterion for a successful time series model. We now forecast the price durations in September 2009 on the basis of the models that have been fitted to the price durations from July 1 to August 31. Since we are interested in modeling the complete transition distribution, we compare the out-of-sample performances of the CMP and ACD models using the difference in logarithmic scores (Giacomini and White, 2006; Diks et al., 2011). The logarithmic score of a time series model M at observation y_i is given by the logarithm of the evaluated transition density, i.e., $S_i^M(y_i) = \log f^M_{i|i-1:1}(y_i|y_{i-1:1})$. If the value of the transition density evaluated at y_i is large, we obtain a high score, whereas the score is low if the transition density evaluated at y_i is small. The expected logarithmic score is strongly related to the KL divergence of the model-implied transition distribution from the true transition distribution. Clearly, a higher expected value of the logarithmic score implies a lower value of the KL divergence. Consequently, the use of the logarithmic score results in a proper scoring rule (Gneiting and Raftery, 2007) which means that a incorrect density forecast does not receive a better average score than the true transition density. Moreover, the model that yields the highest expected logarithmic score is the superior model in the sense that it is the best approximation of the true transition distribution.

The difference in logarithmic scores

$$D_i = S_i^{CMP}(Y_i) - S_i^{ACD}(Y_i),$$

can be used to set up the null hypothesis that the transition distributions of the ACD model are superior to the CMP model in terms of KL divergence, i.e.,

$$H_0 : \mathbb{E}[D_i] \leq 0 \quad \text{v.s.} \quad H_1 : \mathbb{E}[D_i] > 0.$$

In order to test the null hypothesis, we use the following Diebold-Mariano (Diebold and Mariano, 1995) type test statistic

$$\tau = \frac{\bar{D}_i}{\sqrt{\widehat{\sigma^2_{\bar{D}_i}}/T}}, \tag{4.5.1}$$

where T is the number of out-of-sample observations, $\bar{D}_i := \frac{1}{T}\sum_{i=1}^{T} D_i$, and $\widehat{\sigma^2_{\bar{D}_i}}$ is a heteroscedasticity and autocorrelation-consistent estimator of the long-run variance $\lim_{T\to\infty} T\mathbb{V}\text{ar}[\bar{D}_i]$.[18] Under weak regularity conditions (see Diks et al., 2011, Theorem 1), the test statistic τ converges to the standard normal distribution if T goes to infinity.

The first three rows of Table 4.9 report the average logarithmic scores of three different models for all five price duration series. For every future event time, the naive forecast of the transition distribution is the estimated marginal distribution in Table 4.4. The naive forecast provides a benchmark against which more sophisticated models can be compared.

[18] We use the Newey-West (Newey and West, 1987) heteroscedasticity and autocorrelation-consistent estimator of the long-run variance.

Table 4.9: Results of the log-score test for superior out-of-sample specification of the transition distribution. Naive refers to the naive forecast method, which uses the estimated marginal distribution given in Table 4.4 as forecast. The p-value refers to the null hypothesis that the ACD model produces larger log-scores.

	Apple	Cisco	Hewlett-Packard	Intel	Microsoft
average log-score(Naive)	-0.648	-0.634	-0.806	-0.725	-0.518
average log-score(ACD)	-0.592	-0.590	-0.744	-0.699	-0.477
average log-score(CMP)	-0.542	-0.505	-0.665	-0.628	-0.396
test statistic (4.5.1)	3.335	5.878	5.970	6.532	6.555
p-value	0.000	0.000	0.000	0.000	0.000

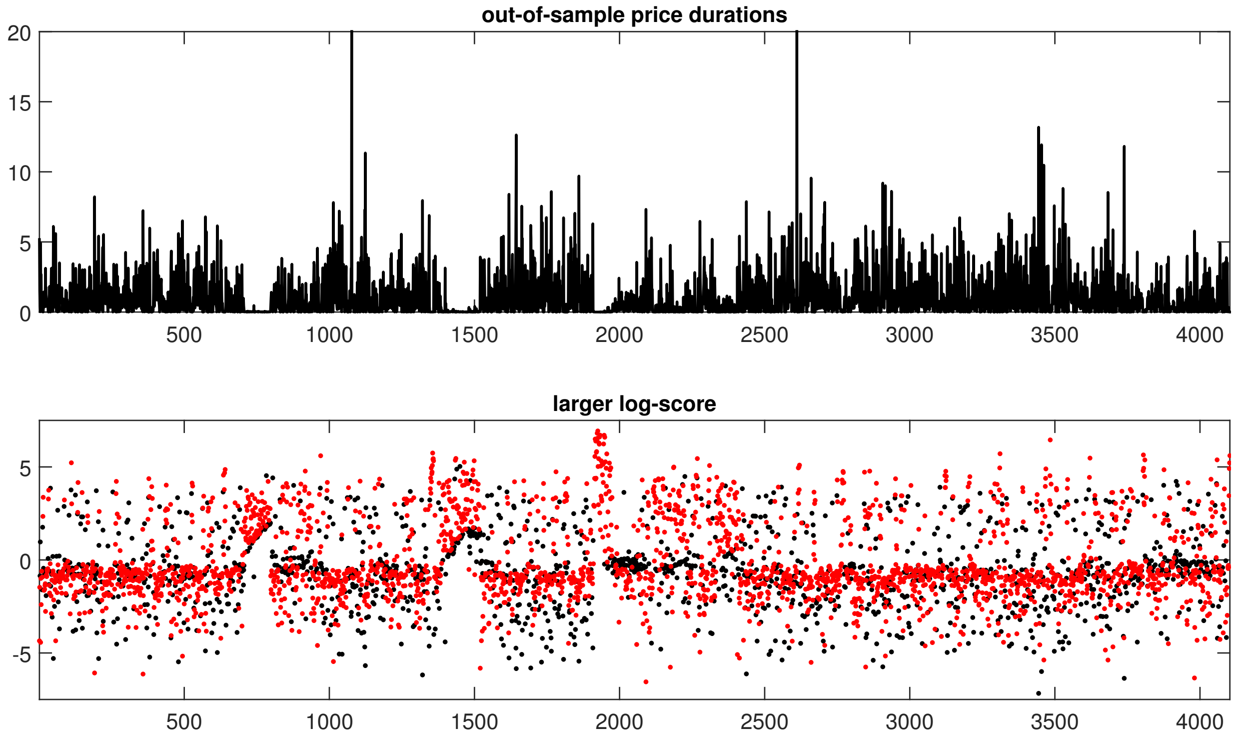
We see that both the ACD and the CMP models exhibit higher average logarithmic scores than the naive forecast. The null hypothesis that the naive forecast produces a larger average logarithmic score than the ACD or CMP model is clearly rejected at a 1% significance level for all five price duration series using the test statistic given in (4.5.1). Regarding the out-of-sample performance of the ACD and CMP models, the average logarithmic scores of the CMP models are in every case larger than those of the ACD models. The p-values of the Diebold-Mariano type tests of superior out-of-sample specification of the transition distribution strongly favor the CMP models. Except for the Apple stock, the values of the test statistics are so large that the corresponding p-values are practically zero. Thus, the superior performance of the CMP models is also confirmed for the out-of-sample experiment.

In order to obtain some insights in the differences of the model-implied transition distributions, we plot the out-of-sample Microsoft price durations and the corresponding larger logarithmic score in Figure 4.10. We see that the CMP model performs notably better for clusters of short durations. This is in line with our exploratory analysis of the in-sample-fit of the ACD and CMP model. The ACD model fails to reproduce clusters of short durations and this shortcoming is also visible in the out-of-sample performance. The CMP model successfully models clusters of short durations which also results in a better performance for the out-of-sample period.

Finally, we investigate the competitiveness of the CMP models wrt to forecasting the mean of the transition distribution. For that purpose, we use the same Diebold-Mariano type test statistic as before with the difference that the score of a model M is now given by the squared forecast error $S_i^M(Y_i) = (Y_i - \mathbb{E}[Y_i|Y_{i-1:1}])^2$. Since the class of ACD models focuses on the mean of the transition distribution it may be possible that the ACD models perform better in this regard. Table 4.10 shows that for three out of five price durations, the ACD models exhibit a lower average squared forecast error than the CMP models. However, in each of these cases, the null hypothesis that the squared forecast errors of the ACD and CMP models are equal can not be rejected at a 10% level. For the Apple and Intel stock, the average squared forecast errors of the CMP models are smaller than those of the ACD models. Moreover, the null hypothesis that the squared forecast errors of the

ACD and CMP models are equal can be rejected for the Cisco stock at a 1% significance level.

Figure 4.10: Microsoft: Out-of-sample price durations and log-scores.



The upper panel shows the price durations of the out-of-sample period. The lower panel shows for each event time the larger logarithmic score. The red dots and black crosses refer to the logarithmic scores of the CMP and ACD model, respectively.

Table 4.10: Results of the Diebold-Mariano test for the transition mean. Naive refers to the naive forecast method, which uses the estimated unconditional mean as forecast. The p-value refers to the null hypothesis that the ACD model produces smaller squared forecast errors.

	Apple	Cisco	Hewlett-Packard	Intel	Microsoft
MSFE(Naive)	1.666	2.671	3.428	2.594	3.207
MSFE(ACD)	0.849	1.644	2.010	1.593	2.013
MSFE(CMP)	0.848	1.651	2.024	1.575	2.021
test statistic	0.270	-1.022	-1.134	2.520	-0.869
p-value	0.394	0.847	0.872	0.006	0.808

4.6 Summary

In this chapter, we applied simplified SD-vine copula-based Markov models to price duration data of five blue-chip stocks and compared their performances with the class of the popular ACD models. We demonstrated that common goodness-of-fit tests, such as the autocorrelation function of the residuals, are inadequate when it comes to the goodness-of-fit of the transition distribution. Although the considered ACD models capture the autocorrelation structure of price durations adequately, simple transformations of the residuals show that a great amount of non-linear dependence is still present in the filtered price durations. Exploratory data analysis based on scatter plots of normalized ranks of consecutive durations revealed that the ACD models fail to capture the clusters of short durations which are prevalent in the data.

We showed how exploratory data analysis can be used to specify the copula of consecutive price durations in the CMP model. As a result, the fitted CMP models reproduce the dependence of consecutive price durations, which seems to be the central factor in the dynamics of price durations. Moreover, we found that the CMP models imply a strongly time-varying dispersion of price durations and that the relation between the mean and the dispersion of the transition distribution is negative and highly non-linear. Since ACD models exhibit a time-constant dispersion, this further explains the superior performance of the CMP models in terms of the AIC. The strong evidence of a strongly varying conditional dispersion also shows that there is no member of the class of ACD models with iid errors that can represent the transition distribution of the considered price duration data. In order to model the transition distribution, one has to model the dependence of the error sequence of the ACD model. For instance, one could specify for the conditional distribution of the errors a gamma distribution with unit mean and where the shape parameter depends on the past. However, it is unclear if this improvement can be realized in practice and whether it is sufficient to obtain ACD models that are competitive with the CMP models.

The out-of-sample evaluation of the ACD and CMP models showed that the CMP models clearly outperform the ACD models. Although the numbers of out-of-sample observations are pretty large, it is impressive that the p-value of the Diebold-Mariano test for a superior transition distribution of the ACD model is practically zero for all five price duration series. An analysis of the logarithmic scores for the models of Microsoft price durations confirmed that one reason for the better out-of-sample performance is that the CMP models are successful in forecasting clusters of short durations.

4.7 Appendix

A.1 Estimation results for the ACD and BCACD model of price durations of Apple, Cisco, Hewlett-Packard, and Intel

Table 4.11: Apple: Estimation results for the ACD(1,1) and BCACD(1,1) model with different error distributions.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.028	-0.153	0.033	-0.131	0.035	-0.111
α	0.116	0.175	0.100	0.144	0.096	0.124
β	0.860	0.956	0.867	0.954	0.872	0.959
ν	-	0.635	-	0.668	-	0.720
a	-	-	3.966	3.994	0.860	0.899
b	-	-	2.180	2.180	14.653	10.534
p	-	-	0.136	0.135	1.698	1.600
q	-	-	0.810	0.793	17.537	13.665
μ	-	-	-	-	-6.963	-6.964
σ^2	-	-	-	-	1.508	1.501
w	-	-	-	-	0.089	0.088
$\log \mathcal{L}$	-2781.342	-2774.740	-2469.505	-2465.310	-2257.176	-2253.742
AIC	5568.685	5557.480	4951.010	4944.620	4534.353	4529.484

Table 4.12: Apple: Estimation results for the ACD(2,2) and BCACD(2,2) model with different error distributions.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.044	-0.234	0.054	-1.514	0.057	-0.180
α_1	0.162	0.265	0.130	0.180	0.121	0.175
α_2	0.036	0.001	0.044	1.355	0.043	0.025
β_1	0.131	0.348	0.116	0.306	0.135	0.285
β_2	0.635	0.591	0.655	0.624	0.648	0.650
ν_1	-	0.604	-	0.749	-	0.687
ν_2	-	2.764	-	0.007	-	0.954
a	-	-	2.194	2.194	0.800	0.848
b	-	-	3.944	4.034	28.338	16.423
p	-	-	0.137	0.134	1.878	1.732
q	-	-	0.822	0.787	27.990	19.090
μ	-	-	-	-	-6.945	-6.946
σ^2	-	-	-	-	1.514	1.505
w	-	-	-	-	0.089	0.088
$\log \mathcal{L}$	-2774.784	-2765.874	-2465.788	-2459.918	-2254.113	-2249.062
AIC	5559.569	5545.749	4947.576	4939.835	4532.226	4526.123

$(\omega, \alpha, \beta, \nu)$ are the parameters of the conditional expected duration, see (4.3.2) and (4.3.3). (a, b, p, q) are the parameters of the estimated GB2 distribution of $\mathcal{E}_i$, see (4.3.1). In the last two columns a mixture of the GB2 distribution and the log-normal distribution is estimated for the marginal distribution of $\mathcal{E}$. w denotes the weight for the log-normal distribution with parameters (μ, σ^2). $\log \mathcal{L}$ reports the value of the maximized log-likelihood function and AIC reports the Akaike information criterion.

Table 4.13: Cisco: Estimation results for the ACD(1,1) and BCACD(1,1) model with different error distributions.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.048	-0.065	0.033	-0.092	0.026	-0.066
α	0.079	0.075	0.076	0.108	0.061	0.076
β	0.876	0.971	0.894	0.966	0.916	0.975
ν	-	0.720	-	0.629	-	0.715
a	-	-	2.653	2.600	0.575	0.630
b	-	-	2.293	2.293	16.046	8.857
p	-	-	0.186	0.191	3.396	2.969
q	-	-	1.060	1.085	17.957	12.719
μ	-	-	-	-	-4.862	-4.849
σ^2	-	-	-	-	1.644	1.631
w	-	-	-	-	0.175	0.174
$\log \mathcal{L}$	-9483.523	-9462.515	-7925.367	-7905.093	-7620.013	-7601.094
AIC	18973.046	18933.030	15862.733	15824.187	15260.025	15224.187

Table 4.14: Cisco: Estimation results for the ACD(2,2) and BCACD(2,2) model with different error distributions.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.056	-0.129	0.038	-0.128	0.031	0.043
α_1	0.099	0.071	0.094	0.152	0.077	0.388
α_2	0.000	0.077	0.000	0.000	0.000	-0.426
β_1	0.545	0.004	0.574	0.609	0.589	1.475
β_2	0.304	0.939	0.297	0.346	0.306	-0.483
ν_1	-	0.731	-	0.606	-	0.287
ν_2	-	0.712	-	1.106	-	0.200
a	-	-	2.282	2.282	0.572	0.622
b	-	-	2.652	2.580	15.831	6.717
p	-	-	0.186	0.193	3.436	3.151
q	-	-	1.056	1.090	17.915	11.325
μ	-	-	-	-	-4.853	-4.712
σ^2	-	-	-	-	1.644	1.635
w	-	-	-	-	0.176	0.177
$\log \mathcal{L}$	-9474.301	-9461.072	-7920.200	-7897.020	-7615.611	-7564.648
AIC	18958.603	18936.143	15856.400	15814.041	15255.222	15157.296

$(\omega, \alpha, \beta, \nu)$ are the parameters of the conditional expected duration, see (4.3.2) and (4.3.3). (a, b, p, q) are the parameters of the estimated GB2 distribution of $\mathcal{E}_i$, see (4.3.1). In the last two columns a mixture of the GB2 distribution and the log-normal distribution is estimated for the marginal distribution of $\mathcal{E}$. w denotes the weight for the log-normal distribution with parameters (μ, σ^2). $\log \mathcal{L}$ reports the value of the maximized log-likelihood function and AIC reports the Akaike information criterion.

Table 4.15: Hewlett-Packard: Estimation results for the ACD(1,1) and BCACD(1,1) model with different error distributions.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.041	-0.101	0.023	-0.096	0.026	-0.075
α	0.120	0.109	0.091	0.101	0.082	0.075
β	0.841	0.969	0.887	0.971	0.892	0.972
ν	-	0.771	-	0.810	-	0.923
a	-	-	2.643	2.575	0.584	0.587
b	-	-	2.396	2.396	24.736	25.743
p	-	-	0.199	0.205	3.437	3.261
q	-	-	1.193	1.274	23.638	29.797
μ	-	-	-	-	-4.765	-4.812
σ^2	-	-	-	-	1.724	1.691
w	-	-	-	-	0.161	0.156
$\log \mathcal{L}$	-8173.723	-8113.634	-7074.324	-7053.425	-6796.195	-6769.912
AIC	16353.446	16235.268	14160.648	14120.849	13612.391	13561.825

Table 4.16: Hewlett-Packard: Estimation results for the ACD(2,2) and BCACD(2,2) model with different error distributions.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.032	0.033	0.028	-2.302	0.030	0.051
α_1	0.132	0.439	0.120	0.146	0.101	0.265
α_2	0.000	-0.469	0.000	2.166	0.000	-0.316
β_1	0.431	1.369	0.463	0.580	0.532	1.442
β_2	0.407	-0.379	0.390	0.381	0.337	-0.452
ν_1	-	0.359	-	0.776	-	0.417
ν_2	-	0.247	-	0.000	-	0.226
a	-	-	2.519	2.519	0.595	0.593
b	-	-	2.551	2.551	26.544	27.543
p	-	-	0.206	0.207	3.170	3.179
q	-	-	1.305	1.293	29.701	29.657
μ	-	-	-	-	-4.811	-4.745
σ^2	-	-	-	-	1.691	1.689
w	-	-	-	-	0.156	0.156
$\log \mathcal{L}$	-8115.704	-8025.545	-7053.956	-7042.222	-6774.690	-6731.919
AIC	16241.408	16065.091	14123.911	14104.445	13573.380	13491.839

$(\omega, \alpha, \beta, \nu)$ are the parameters of the conditional expected duration, see (4.3.2) and (4.3.3). (a, b, p, q) are the parameters of the estimated GB2 distribution of $\mathcal{E}_i$, see (4.3.1). In the last two columns a mixture of the GB2 distribution and the log-normal distribution is estimated for the marginal distribution of $\mathcal{E}$. w denotes the weight for the log-normal distribution with parameters (μ, σ^2). $\log \mathcal{L}$ reports the value of the maximized log-likelihood function and AIC reports the Akaike information criterion.

Table 4.17: Intel: Estimation results for the ACD(1,1) and BCACD(1,1) model with different error distributions.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.042	-0.089	0.029	-0.073	0.021	-0.051
α	0.087	0.097	0.073	0.077	0.058	0.051
β	0.873	0.954	0.899	0.969	0.922	0.979
ν	-	0.829	-	0.885	-	0.996
a	-	-	2.408	2.412	0.551	0.549
b	-	-	2.432	2.432	26.444	25.721
p	-	-	0.221	0.220	3.707	3.752
q	-	-	1.301	1.299	29.603	29.709
μ	-	-	-	-	-4.318	-4.307
σ^2	-	-	-	-	1.831	1.834
w	-	-	-	-	0.182	0.182
$\log \mathcal{L}$	-12092.278	-12085.234	-10654.611	-10649.000	-10382.046	-10374.662
AIC	24190.556	24178.468	21321.222	21312.001	20784.093	20771.325

Table 4.18: Intel: Estimation results for the ACD(2,2) and BCACD(2,2) model with different error distributions.

	ACD-EXP	BCACD-EXP	ACD-GB2	BCACD-GB2	ACD-GB2-LOGN	BCACD-GB2-LOGN
ω	0.050	0.021	0.052	-0.104	0.039	6.567
α_1	0.112	0.450	0.065	0.112	0.050	0.103
α_2	0.000	-0.470	0.069	0.000	0.058	-6.669
β_1	0.512	1.513	0.000	0.640	0.000	0.610
β_2	0.329	-0.518	0.817	0.319	0.855	0.365
ν_1	-	0.336	-	0.831	-	0.782
ν_2	-	0.277	-	1.635	-	0.001
a	-	-	2.431	2.431	0.555	0.539
b	-	-	2.420	2.391	27.742	26.432
p	-	-	0.219	0.223	3.667	3.910
q	-	-	1.295	1.303	29.608	29.844
μ	-	-	-	-	-4.323	-4.278
σ^2	-	-	-	-	1.831	1.836
w	-	-	-	-	0.181	0.184
$\log \mathcal{L}$	-12070.533	-11923.348	-10649.247	-10637.916	-10375.652	-10351.893
AIC	24151.065	23860.695	21314.493	21295.832	20775.303	20731.787

$(\omega, \alpha, \beta, \nu)$ are the parameters of the conditional expected duration, see (4.3.2) and (4.3.3). (a, b, p, q) are the parameters of the estimated GB2 distribution of $\mathcal{E}_i$, see (4.3.1). In the last two columns a mixture of the GB2 distribution and the log-normal distribution is estimated for the marginal distribution of $\mathcal{E}$. w denotes the weight for the log-normal distribution with parameters (μ, σ^2). $\log \mathcal{L}$ reports the value of the maximized log-likelihood function and AIC reports the Akaike information criterion.

A.2 Goodness-of-fit of the BCACD(2,2)-GB2-LOGN model

Figure 4.11: Apple

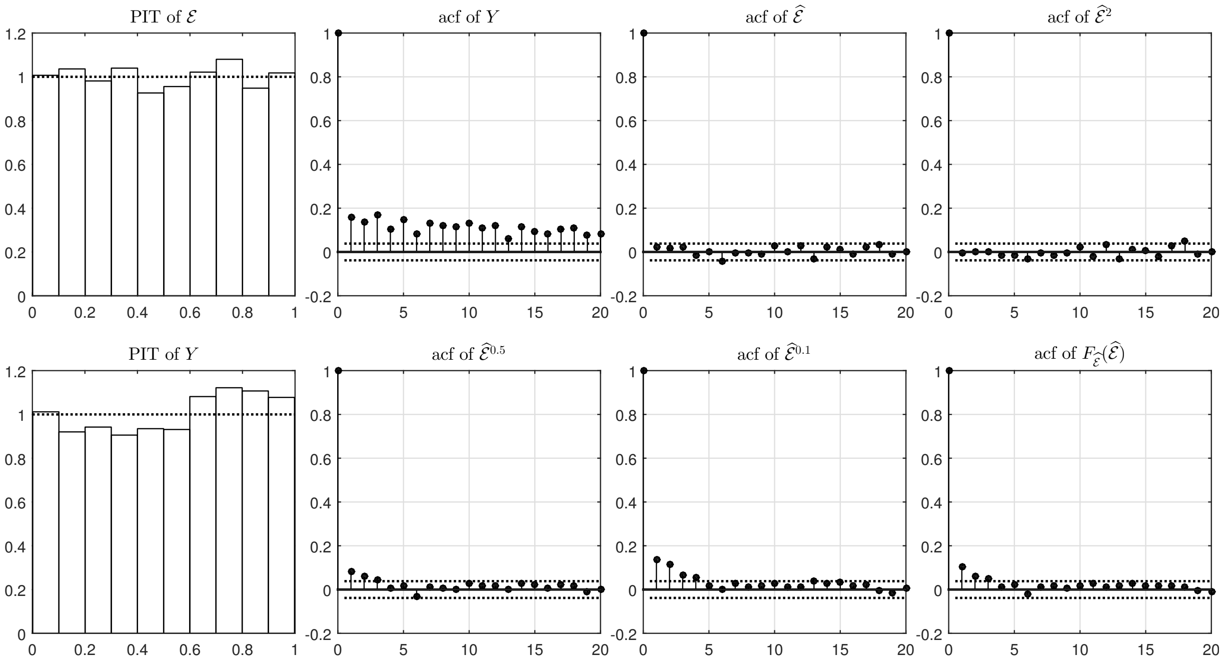
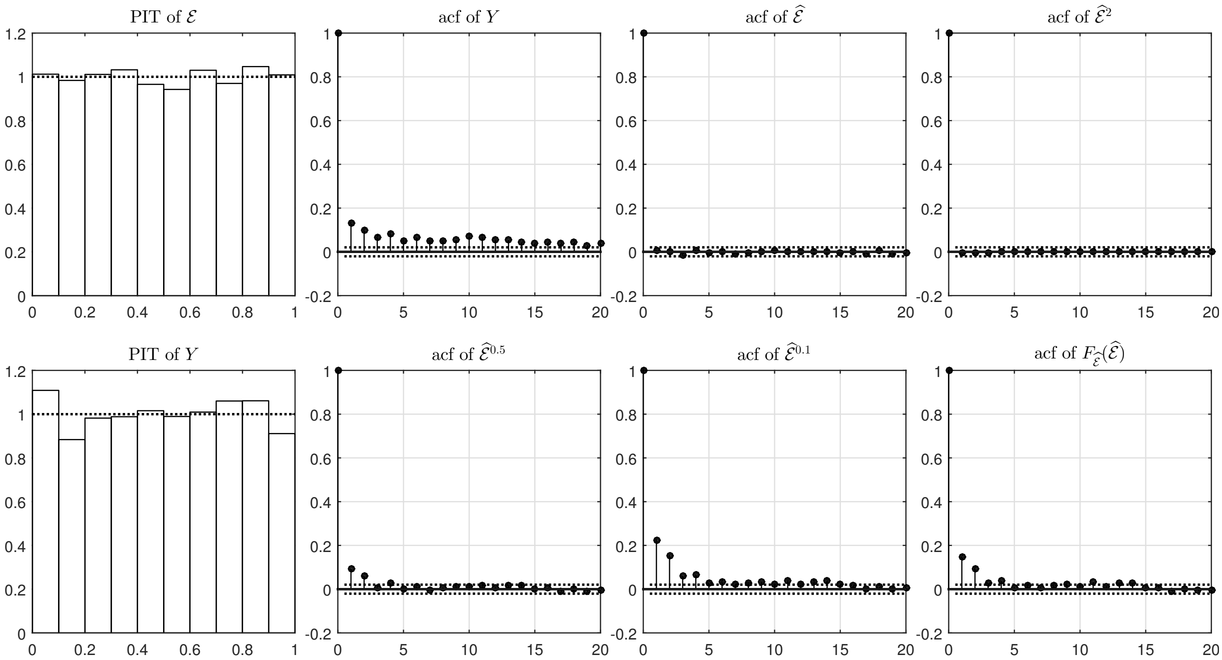


Figure 4.12: Cisco



The left panel in each figure shows the histogram of the estimated PIT of the error $\widehat{\mathcal{E}}$ and the price duration Y. The other plots show the empirical autocorrelation function of Y and various transformation of $\widehat{\mathcal{E}}$ with approximate 95% confidence intervals. For $q = 2, 0.5, 0.1$, $\widehat{\mathcal{E}}^q$ is the sequence given by $\widehat{\mathcal{E}}^q = (\widehat{\mathcal{E}}_i^q)_{i\in\mathbb{N}}$ and $F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}) = (F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}_i))_{i\in\mathbb{N}}$. The implied marginal distribution of the ACD model was obtained via simulation.

Figure 4.13: Hewlett-Packard

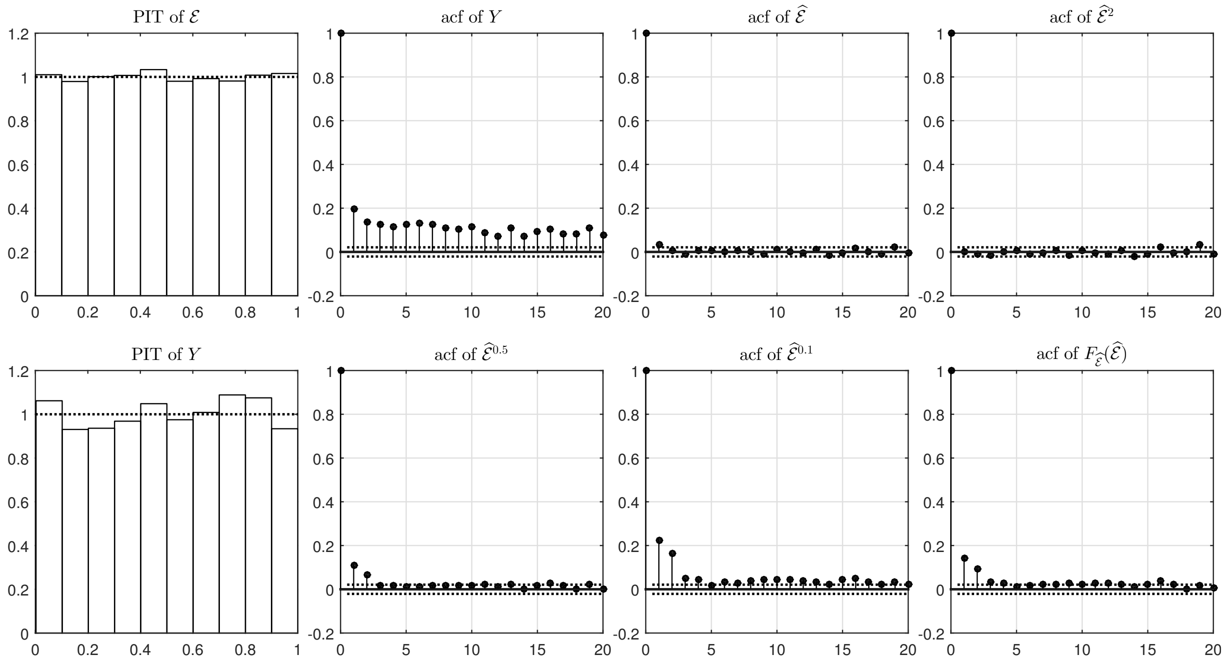
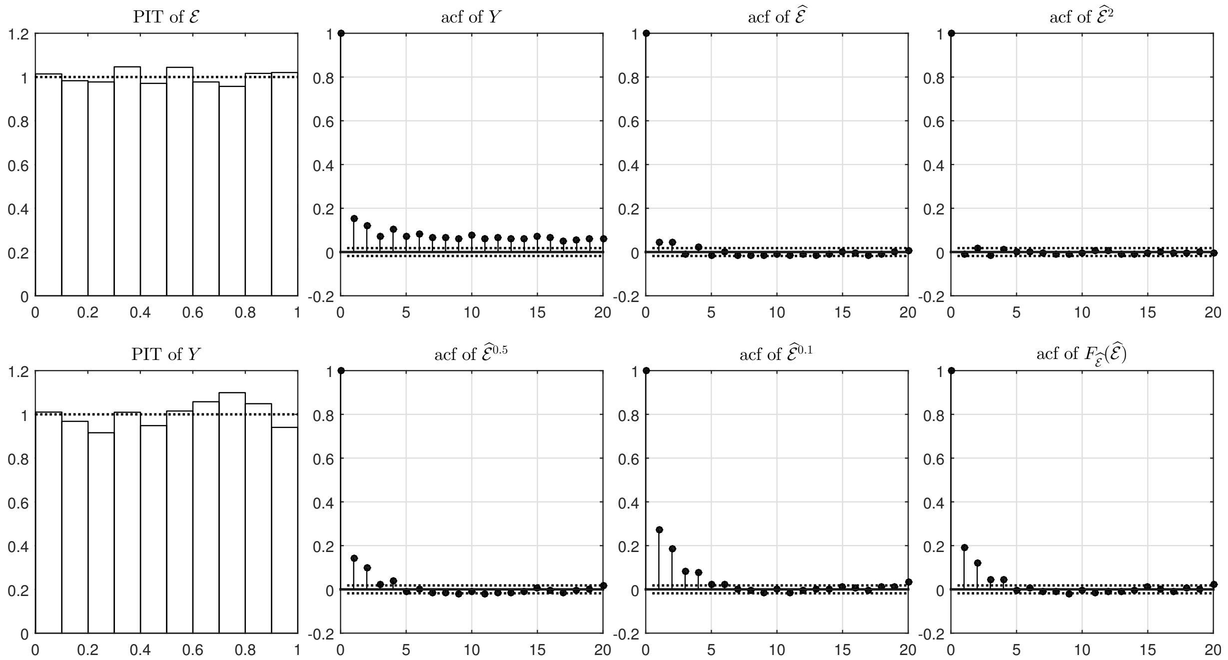


Figure 4.14: Intel



The left panel in each figure shows the histogram of the estimated PIT of the error $\widehat{\mathcal{E}}$ and the price duration Y. The other plots show the empirical autocorrelation function of Y and various transformation of $\widehat{\mathcal{E}}$ with approximate 95% confidence intervals. For $q = 2, 0.5, 0.1$, $\widehat{\mathcal{E}}^q$ is the sequence given by $\widehat{\mathcal{E}}^q = (\widehat{\mathcal{E}}_i^q)_{i\in\mathbb{N}}$ and $F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}) = (F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}_i))_{i\in\mathbb{N}}$. The implied marginal distribution of the ACD model was obtained via simulation.

A.3 ACD-implied and true dependence of price durations

Figure 4.15: Apple

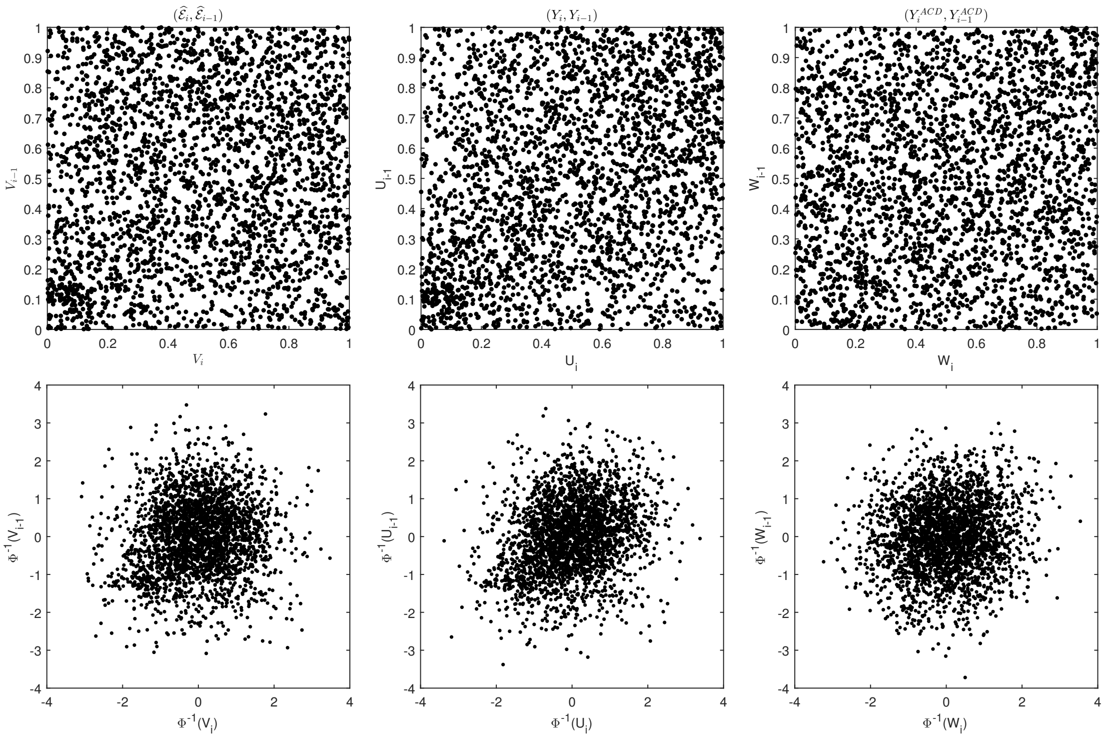
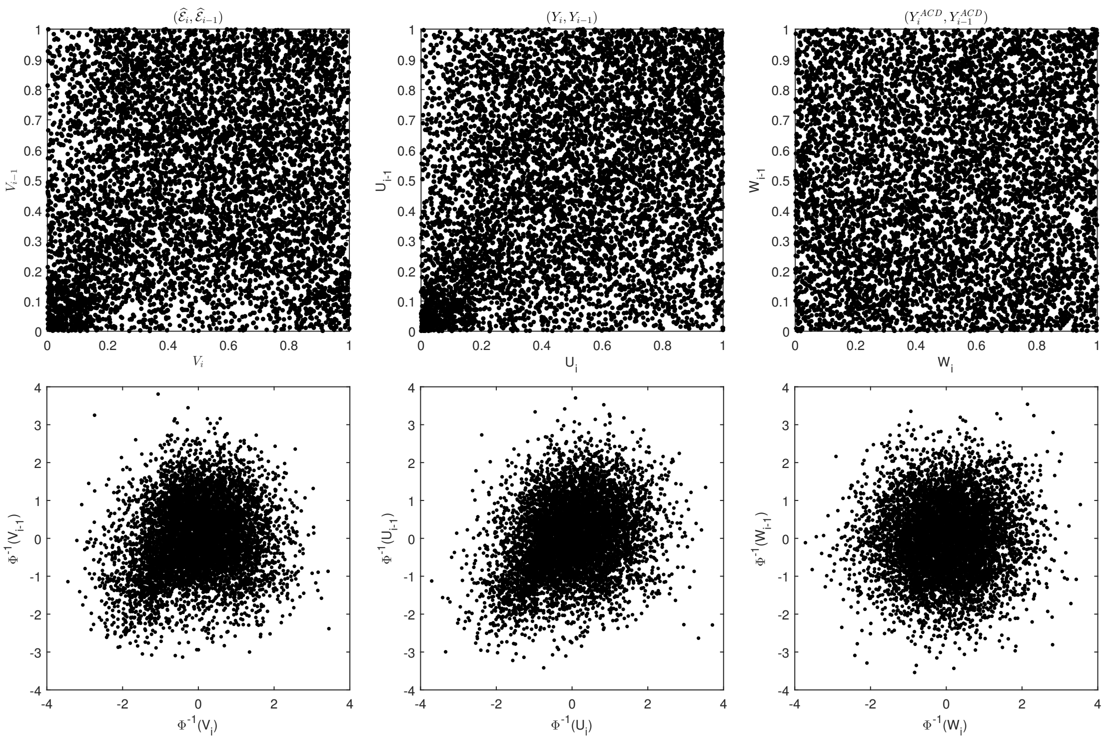


Figure 4.16: Cisco



The first row shows copula scatter plots of consecutive residuals $\widehat{\mathcal{E}}$ of the BCACD(2,2)-GB2-LOGN model, consecutive price durations Y, and consecutive observations Y^{ACD} that are simulated from the BCACD(2,2)-GB2-LOGN model. The second row shows the scatter plots of the corresponding pseudo-$N(0,1)$ observations, i.e., the scatter plots that arise if the quantile function of the $N(0,1)$ distribution is applied to the normalized ranks. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data if the number of scatter plot observations exceeds 7000 observations.

Figure 4.17: Hewlett-Packard

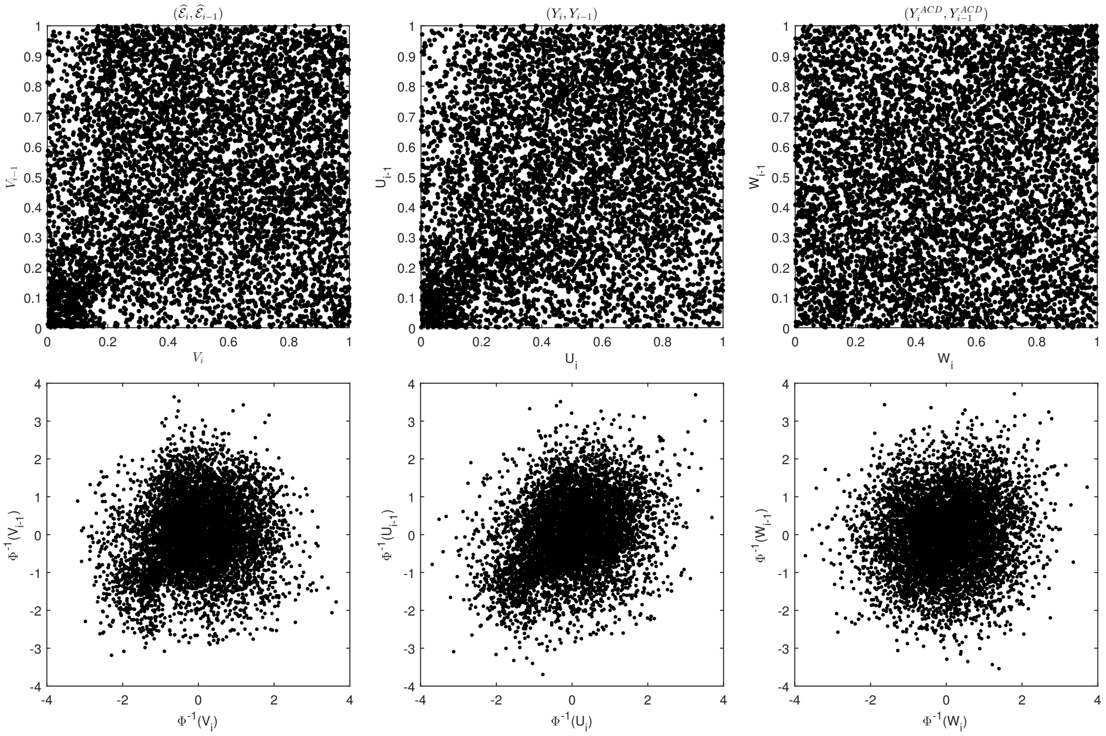
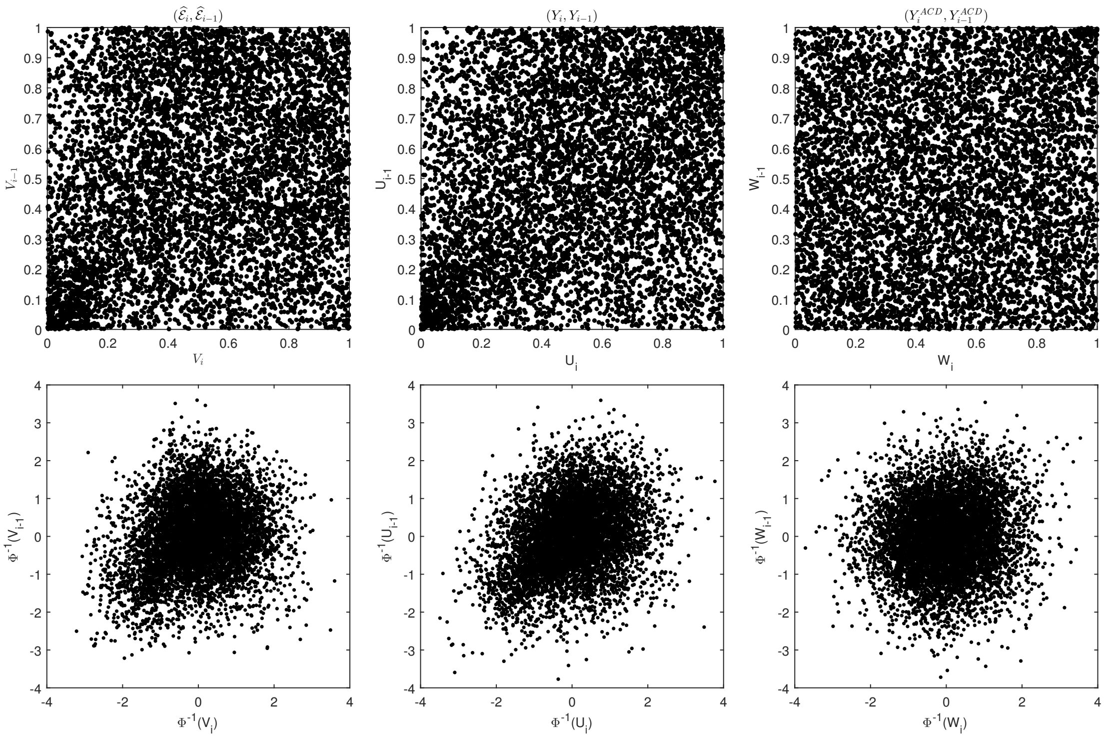


Figure 4.18: Intel



The first row shows copula scatter plots of consecutive residuals $\widehat{\mathcal{E}}$ of the BCACD(2,2)-GB2-LOGN model, consecutive price durations Y, and consecutive observations Y^{ACD} that are simulated from the BCACD(2,2)-GB2-LOGN model. The second row shows the scatter plots of the corresponding pseudo-$N(0,1)$ observations, i.e., the scatter plots that arise if the quantile function of the $N(0,1)$ distribution is applied to the normalized ranks. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data if the number of scatter plot observations exceeds 7000 observations.

A.4 Estimation results for the CMP models

Table 4.19: Apple: Estimation results for the truncated CMP(18) model with a lag function. Kendall's τ of $C_{3:18}$ is modeled by an exponential lag decay function, i.e., $\tau_i = \tau_3 \gamma^{i-3}$ for $i = 3, \ldots, 18$.

	Family	(w, θ) or θ	Kendall's τ	$\log \mathcal{L}$	AIC
C_1	CL-FR	(0.658, 0.342, 0.130, 4.061)	-	101.565	-197.130
C_2	rGU-GA	(0.052, 0.948, 2.424, 0.100)	-	28.118	-50.237
C_3	rCL	0.100	0.048	24.000	-52.405
C_4	rCL	0.091	0.044	5.240	-9.957
C_5	rCL	0.083	0.040	9.847	-17.426
C_6	rCL	0.076	0.037	-2.680	1.090
C_7	GU	1.035	0.034	7.880	-15.475
C_8	rCL	0.064	0.031	4.190	-5.991
C_9	rCL	0.058	0.028	2.631	-4.717
C_{10}	rCL	0.053	0.026	6.016	-9.983
C_{11}	GA	0.037	0.024	2.654	-4.117
C_{12}	GU	1.022	0.022	3.892	-5.781
C_{13}	CL	0.041	0.020	2.411	-2.951
C_{14}	rCL	0.038	0.018	2.952	-4.506
C_{15}	rCL	0.034	0.017	1.846	-1.101
C_{16}	rCL	0.032	0.016	0.507	0.357
C_{17}	GU	1.014	0.014	3.161	-4.080
C_{18}	GU	1.013	0.013	1.975	-2.119
lag-decay	Exponential	$(0.048, 0.917)$	-	-	-

Table 4.20: Cisco: Estimation results for the truncated CMP(18) model with a lag function. Kendall's τ of $C_{3:18}$ is modeled by an exponential lag decay function, i.e., $\tau_i = \tau_3 \gamma^{i-3}$ for $i = 3, \ldots, 18$.

	Family	(w, θ) or θ	Kendall's τ	$\log \mathcal{L}$	AIC
C_1	CL-rGU	(0.828, 0.172, 0.287, 2.034)	-	530.766	-1055.532
C_2	rGU-FR	(0.829, 0.171, 1.003, 5.883)	-	89.267	-172.534
C_3	rCL	0.050	0.024	4.223	-8.387
C_4	GU	1.024	0.023	20.441	-41.716
C_5	GU	1.023	0.022	3.726	-7.710
C_6	rCL	0.044	0.021	10.106	-17.316
C_7	GU	1.021	0.020	4.534	-7.692
C_8	GA	0.031	0.019	6.943	-12.189
C_9	rCL	0.038	0.019	5.917	-10.364
C_{10}	GU	1.018	0.018	9.185	-16.444
C_{11}	GA	0.027	0.017	7.351	-15.625
C_{12}	GU	1.016	0.016	5.341	-8.855
C_{13}	GA	0.024	0.015	4.981	-8.986
C_{14}	GA	0.023	0.015	3.226	-4.983
C_{15}	GU	1.014	0.014	-0.666	-0.108
C_{16}	GU	1.014	0.013	1.745	-2.517
C_{17}	GU	1.013	0.013	-1.741	1.616
C_{18}	GU	1.012	0.012	3.259	-4.507
lag-decay	Exponential	$(0.024, 0.955)$	-	-	-

The second column displays the copula family that is best wrt the AIC. GU, FR, GA are the Gumbel, Frank and Gaussian copula. r refers to the survival version of a copula and - indicates a mixture, i.e., rGU-FR is a mixture of the survival Gumbel and the Frank copula. The first 18 rows of the third column contain the weights $w = (w_1, w_2)$ and parameters $\theta = (\theta_1, \theta_2)$ of the mixture copula or the scalar parameter θ of the copula. The fourth column shows the values of Kendall's τ of the copulas $C_{3:18}$ which are implied by the fitted exponential lag function. $\log \mathcal{L}$ and AIC represent the log-likelihood or AIC value of the copula C_i. The third column in the second last row contains the parameters (τ_3, γ) of the fitted lag function.

Table 4.21: Hewlett-Packard: Estimation results for the truncated CMP(18) model with a lag function. Kendall's τ of $C_{3:18}$ is modeled by an exponential lag decay function, i.e., $\tau_i = \tau_3 \gamma^{i-3}$ for $i = 3, \ldots, 18$.

	Family	(w, θ) or θ	Kendall's τ	$\log \mathcal{L}$	AIC
C_1	CL-rGU	(0.793, 0.207, 0.244, 2.032)	-	482.592	-959.183
C_2	rGU-FR	(0.809, 0.191, 1.011, 5.769)	-	122.190	-238.380
C_3	rCL	0.078	0.038	26.285	-50.412
C_4	GU	1.037	0.035	25.040	-48.690
C_5	GU	1.034	0.033	24.969	-48.235
C_6	rCL	0.064	0.031	20.897	-40.034
C_7	rCL	0.060	0.029	15.782	-28.935
C_8	GA	0.043	0.027	10.265	-19.057
C_9	GU	1.026	0.026	9.996	-18.014
C_{10}	GU	1.025	0.024	8.901	-15.982
C_{11}	GA	0.036	0.023	6.799	-12.023
C_{12}	GU	1.022	0.021	4.317	-7.540
C_{13}	rCL	0.041	0.020	6.491	-11.078
C_{14}	rGU	1.019	0.019	-1.534	-1.194
C_{15}	GA	0.028	0.018	5.844	-10.004
C_{16}	GU	1.017	0.017	8.170	-16.053
C_{17}	rCL	0.032	0.016	1.547	-1.683
C_{18}	rGU	1.015	0.015	0.601	-1.623
lag-decay	Exponential	$(0.038, 0.940)$	-	-	-

Table 4.22: Intel: Estimation results for the truncated CMP(18) model with a lag function. Kendall's τ of $C_{3:18}$ is modeled by an exponential lag decay function, i.e., $\tau_i = \tau_3 \gamma^{i-3}$ for $i = 3, \ldots, 18$.

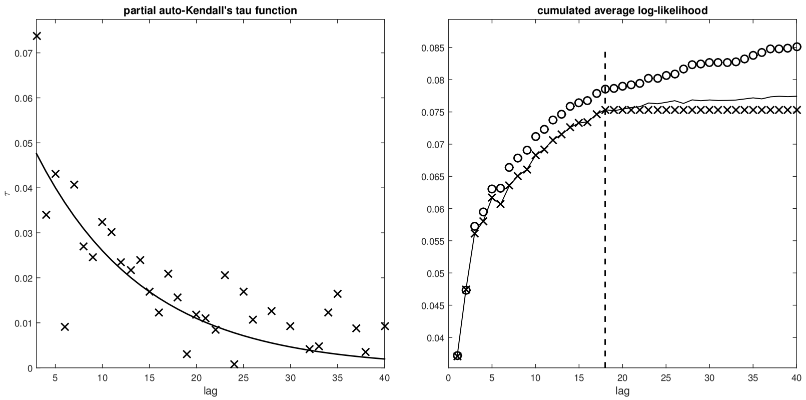
	Family	(w, θ) or θ	Kendall's τ	$\log \mathcal{L}$	AIC
C_1	CL-rGU	(0.757, 0.243, 0.263, 1.808)	-	693.510	-1381.020
C_2	CL-FR	(0.875, 0.125, 0.033, 7.883)	-	116.653	-227.305
C_3	rCL	0.054	0.026	12.705	-23.387
C_4	GU	1.026	0.025	32.764	-72.847
C_5	GU	1.025	0.024	2.706	-8.607
C_6	GU	1.024	0.023	16.357	-29.904
C_7	GU	1.023	0.022	6.180	-11.990
C_8	rCL	0.043	0.021	8.014	-13.290
C_9	rCL	0.041	0.020	5.082	-9.820
C_{10}	GU	1.020	0.019	11.165	-20.440
C_{11}	GU	1.019	0.019	3.838	-7.074
C_{12}	GU	1.018	0.018	8.543	-14.528
C_{13}	GU	1.017	0.017	4.395	-8.285
C_{14}	GU	1.017	0.016	7.599	-12.428
C_{15}	GA	0.025	0.016	8.144	-17.478
C_{16}	GU	1.015	0.015	3.241	-4.633
C_{17}	rGU	1.015	0.014	-1.948	0.490
C_{18}	rGU	1.014	0.014	5.847	-8.710
lag-decay	Exponential	$(0.026, 0.959)$	-	-	-

The second column displays the copula family that is best wrt the AIC. GU, FR, GA are the Gumbel, Frank and Gaussian copula. r refers to the survival version of a copula and - indicates a mixture, i.e., rGU-FR is a mixture of the survival Gumbel and the Frank copula. The first 18 rows of the third column contain the weights $w = (w_1, w_2)$ and parameters $\theta = (\theta_1, \theta_2)$ of the mixture copula or the scalar parameter θ of the copula. The fourth column shows the values of Kendall's τ of the copulas $C_{3:18}$ which are implied by the fitted exponential lag function. $\log \mathcal{L}$ and AIC represent the log-likelihood or AIC value of the copula C_i. The third column in the second last row contains the parameters (τ_3, γ) of the fitted lag function.

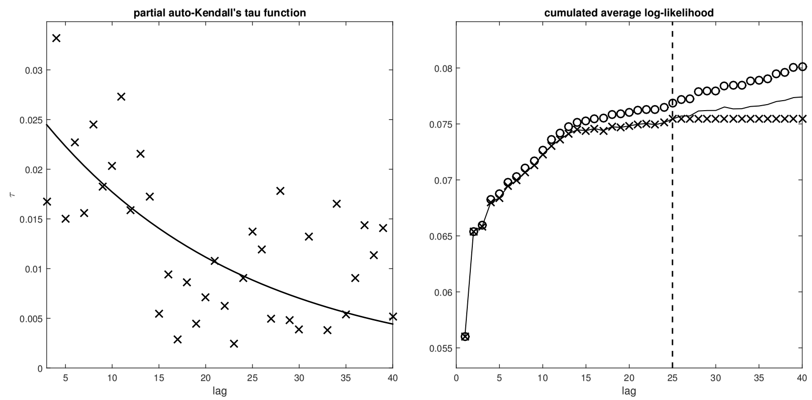
A.5 Visualization of the implied lag decays of the CMP models

Figure 4.19: Implied lag decays.

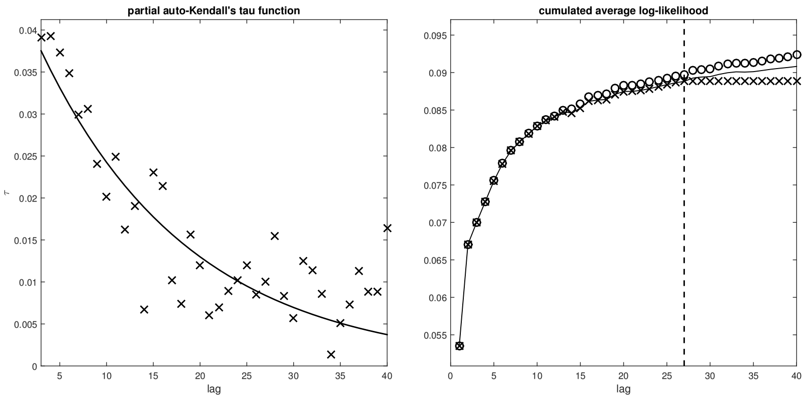
(a) Apple



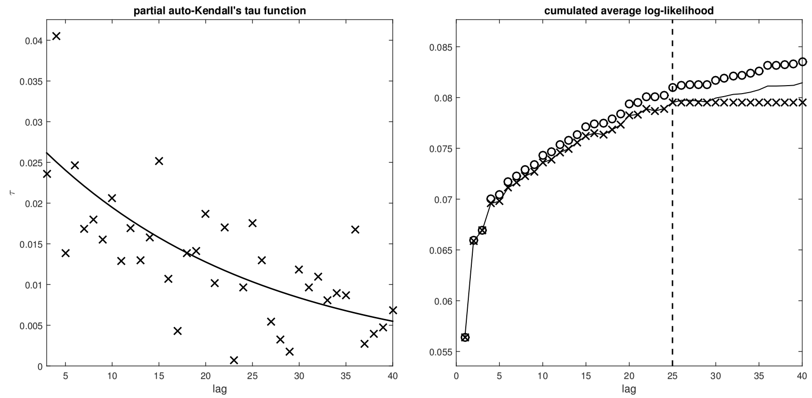
(b) Cisco



(c) Hewlett-Packard



(d) Intel



Left panel in each subfigure: The marks show the sequence $(\tau_j)_{j=3,\dots,40}$, where τ_j is the estimated Kendall's tau of C_j using the first four steps in Algorithm 3.4. The line refers to the corresponding $(\tau_j^{g_3})_{j=3,\dots,40}$ sequence which is implied by an exponential lag function g_3 using step 5 and 6 of Algorithm 3.4.

Right panel in each subfigure: The circles display the cumulated average log-likelihoods $\frac{1}{T}\sum_{i=1}^{j}\mathcal{L}_i(\tau_i)$ that correspond to the marks in the left panel. The line displays the cumulated average log-likelihoods $\frac{1}{T}\sum_{i=1}^{j}\mathcal{L}_i(\tau_i^{g_3})$ that are associated with the lines in the left panel. The marks indicate the cumulated average log-likelihoods that result from the truncated copula sequence that is obtained via Definition 3.8 if $\Delta = 0.0021$. The truncation point $p^\star$ is indicated by a vertical line.

A.6 Goodness-of-fit of the CMP models

Figure 4.20: Apple

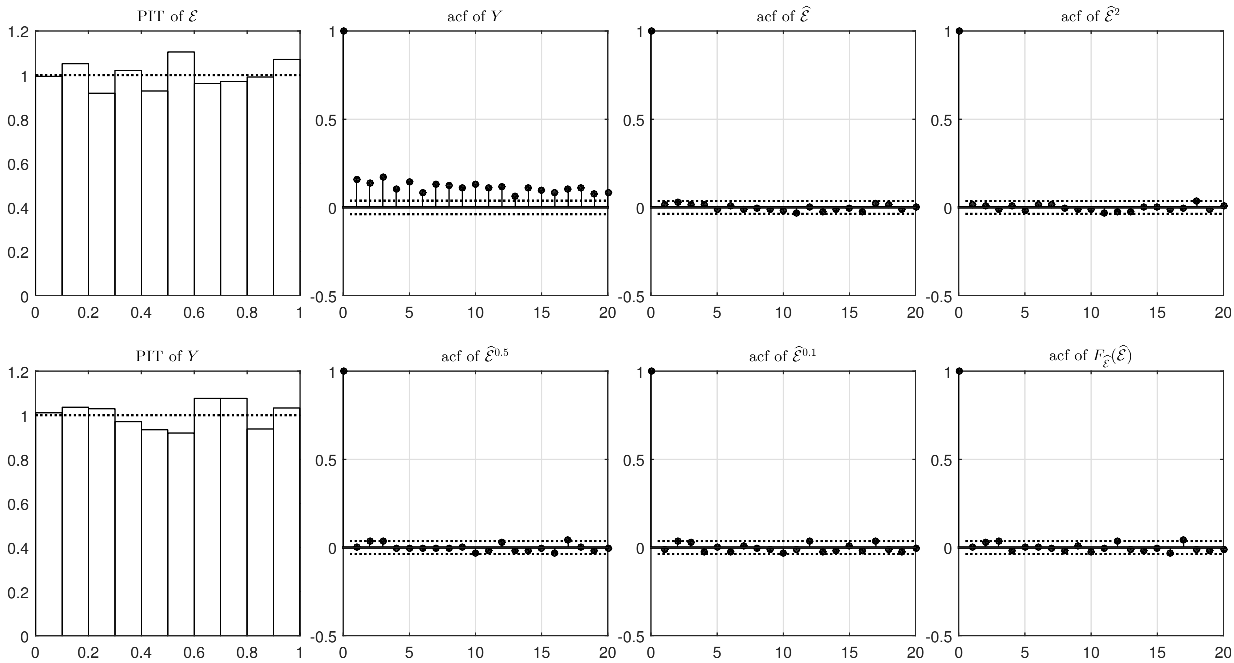
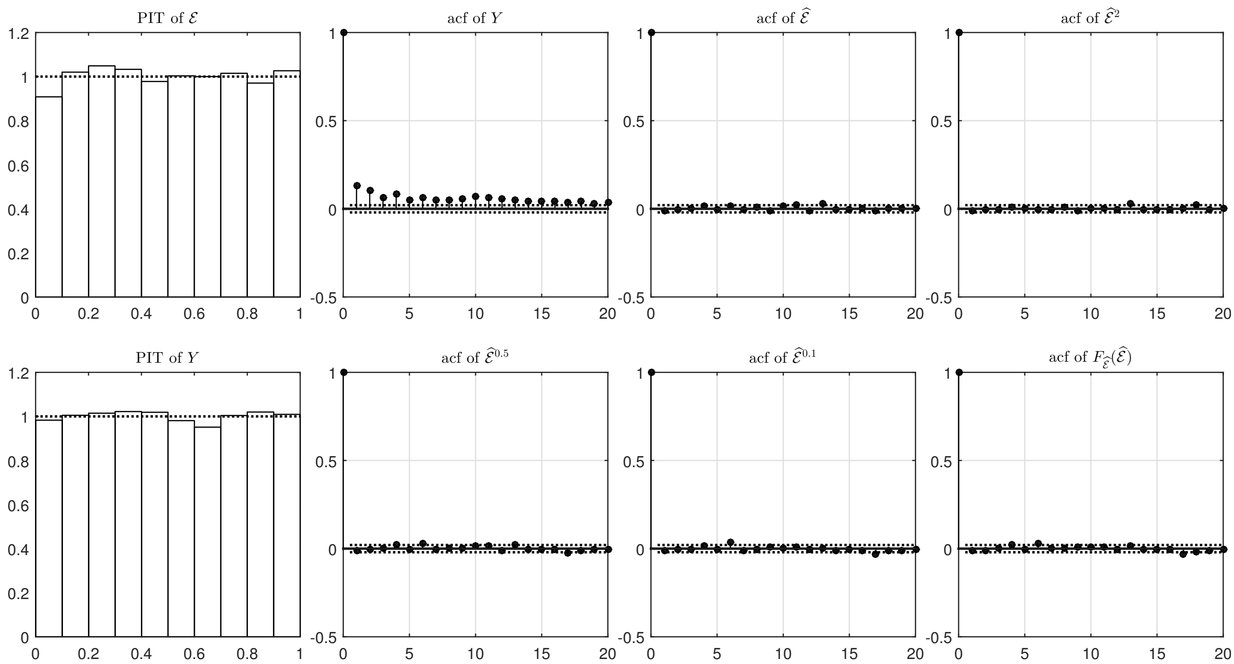


Figure 4.21: Cisco



The first column in each figure shows the histogram of the estimated PIT of the error $\mathcal{E}$ and the price duration Y. The other plots in each figure show the empirical autocorrelation function of Y and various transformations of $\widehat{\mathcal{E}}$ with approximate 95% confidence intervals. For $p = 2, 0.5, 0.1$, $\widehat{\mathcal{E}}^q$ is the sequence given by $\widehat{\mathcal{E}}^q = (\widehat{\mathcal{E}}_i^q)_{i=1,\ldots,N}$ and $F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}) = (F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}_i))_{i=1,\ldots,N}$.

Figure 4.22: Hewlett-Packard

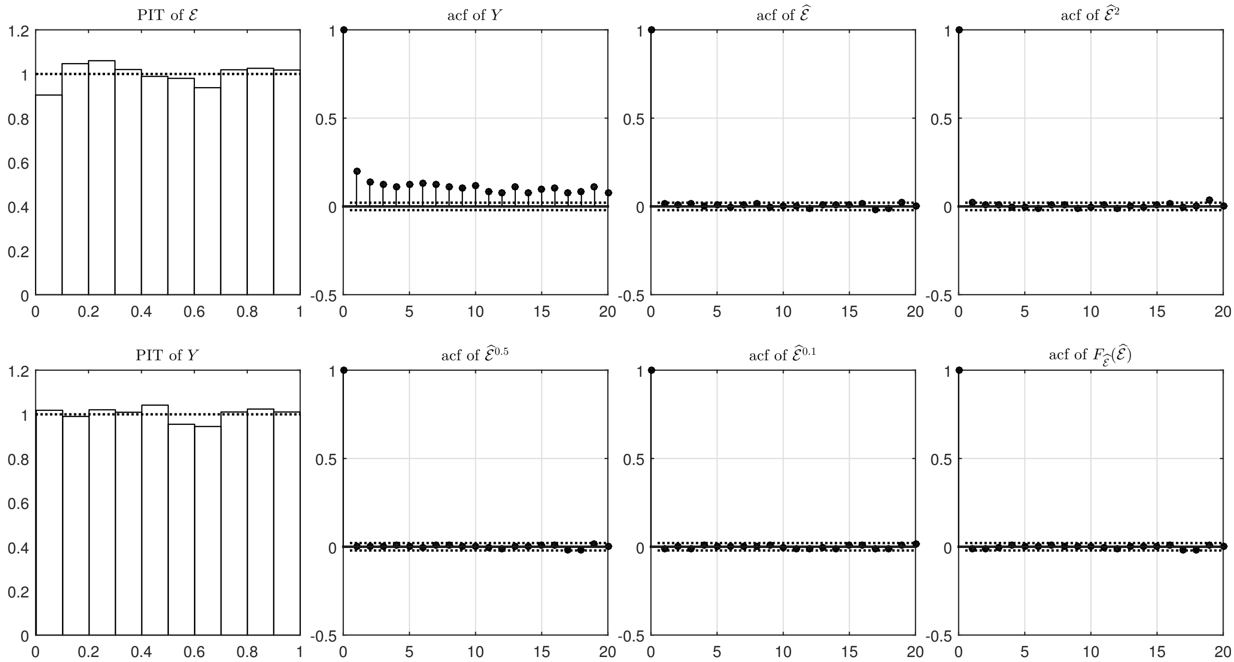
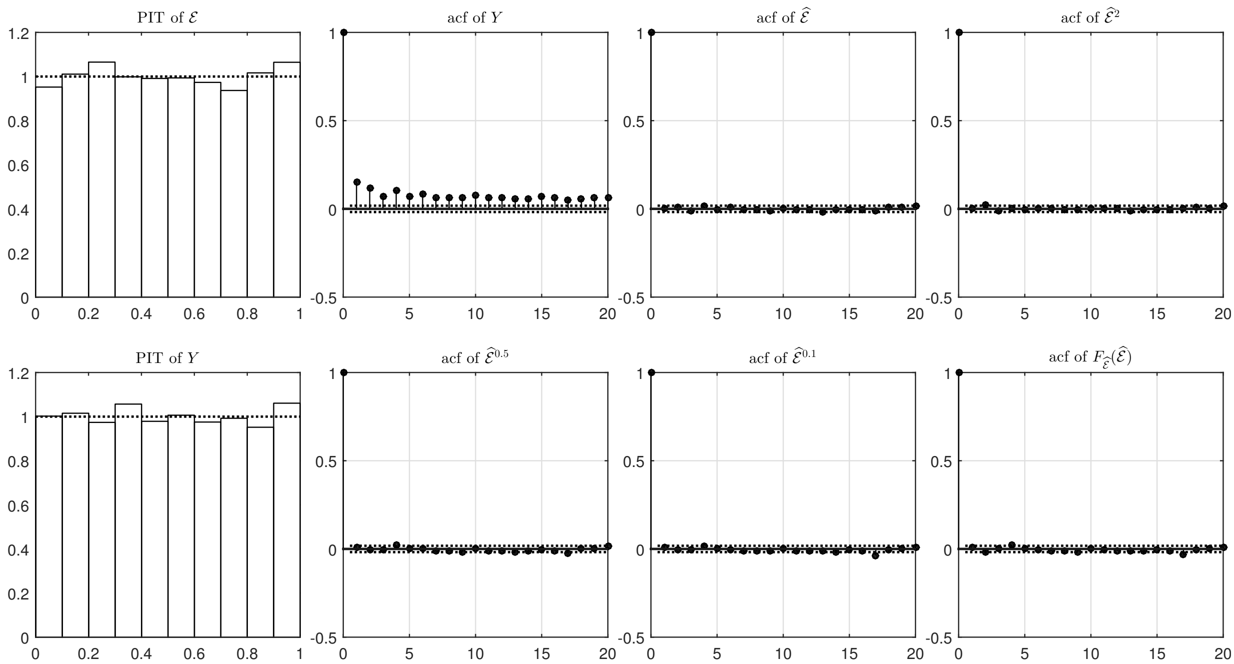


Figure 4.23: Intel



The first column in each figure shows the histogram of the estimated PIT of the error $\mathcal{E}$ and the price duration Y. The other plots in each figure show the empirical autocorrelation function of Y and various transformations of $\widehat{\mathcal{E}}$ with approximate 95% confidence intervals. For $p = 2, 0.5, 0.1$, $\widehat{\mathcal{E}}^q$ is the sequence given by $\widehat{\mathcal{E}}^q = (\widehat{\mathcal{E}}_i^q)_{i=1,\ldots,N}$ and $F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}) = (F_{\widehat{\mathcal{E}}}(\widehat{\mathcal{E}}_i))_{i=1,\ldots,N}$.

A.7 CMP-implied and true dependence of price durations

Figure 4.24: Apple

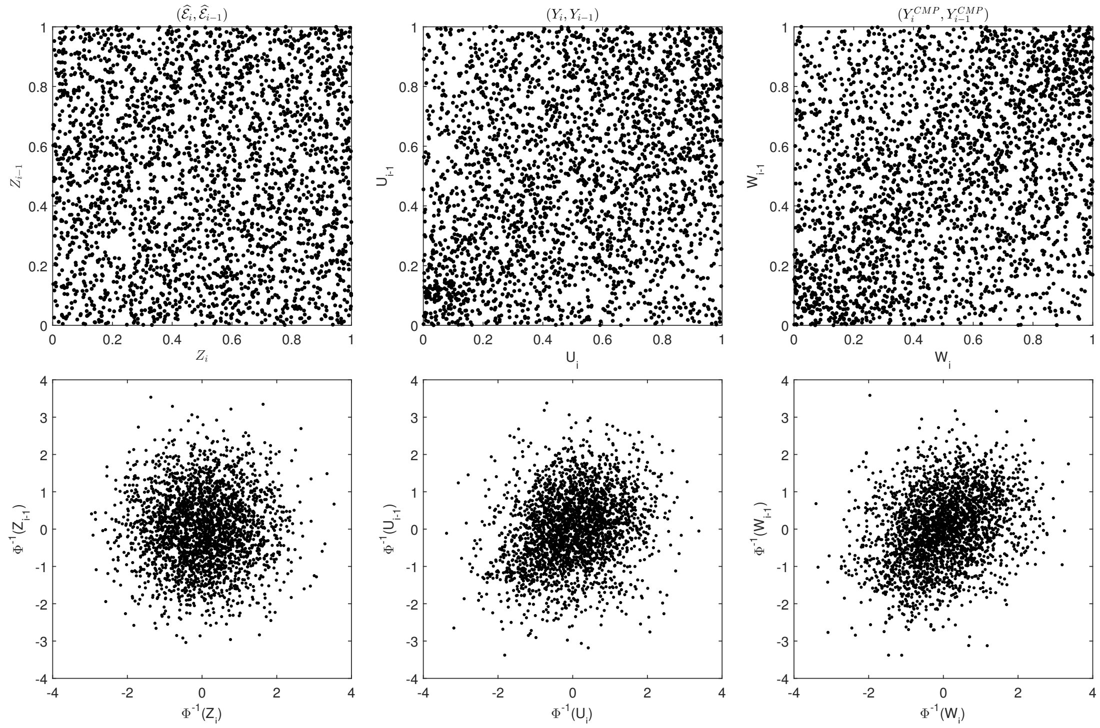
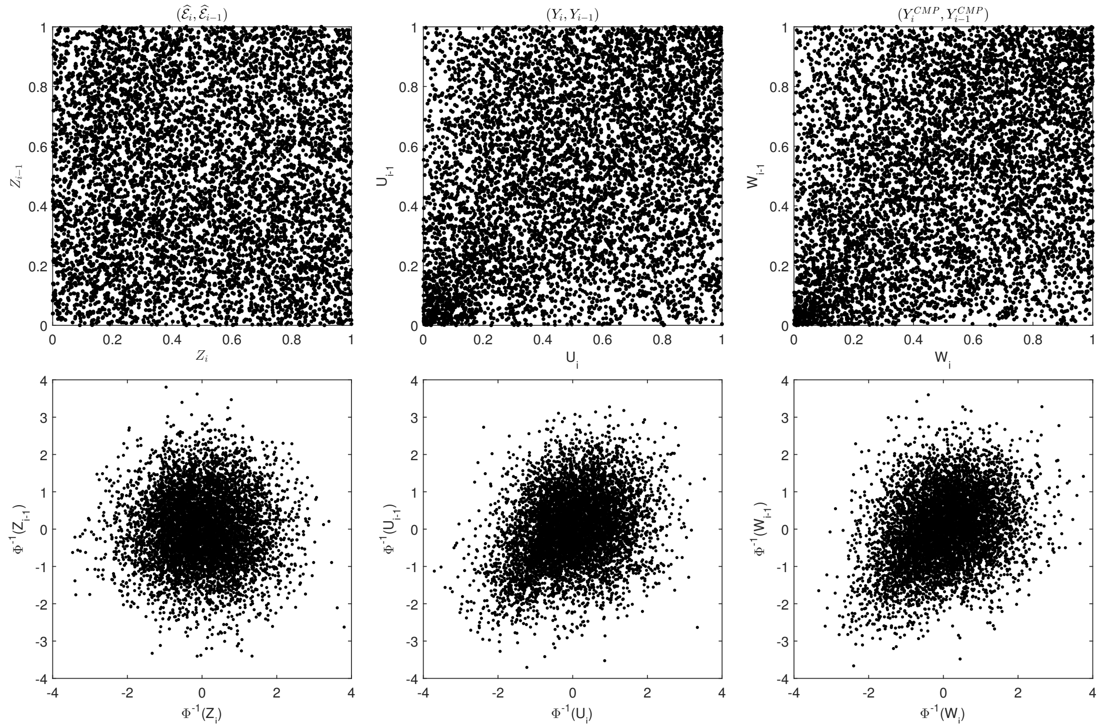


Figure 4.25: Cisco



The first row in each figure shows copula scatter plots of consecutive residuals $\widehat{\mathcal{E}}$ of the CMP model, consecutive price durations Y, and consecutive observations Y^{CMP} that are simulated from the CMP model. The second row shows the scatter plots of the corresponding pseudo-$N(0,1)$ observations, i.e., the scatter plots that arise if the quantile function of the $N(0,1)$ distribution is applied to the data. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data if the number of scatter plot observations exceeds 7000 observations.

Figure 4.26: Hewlett-Packard

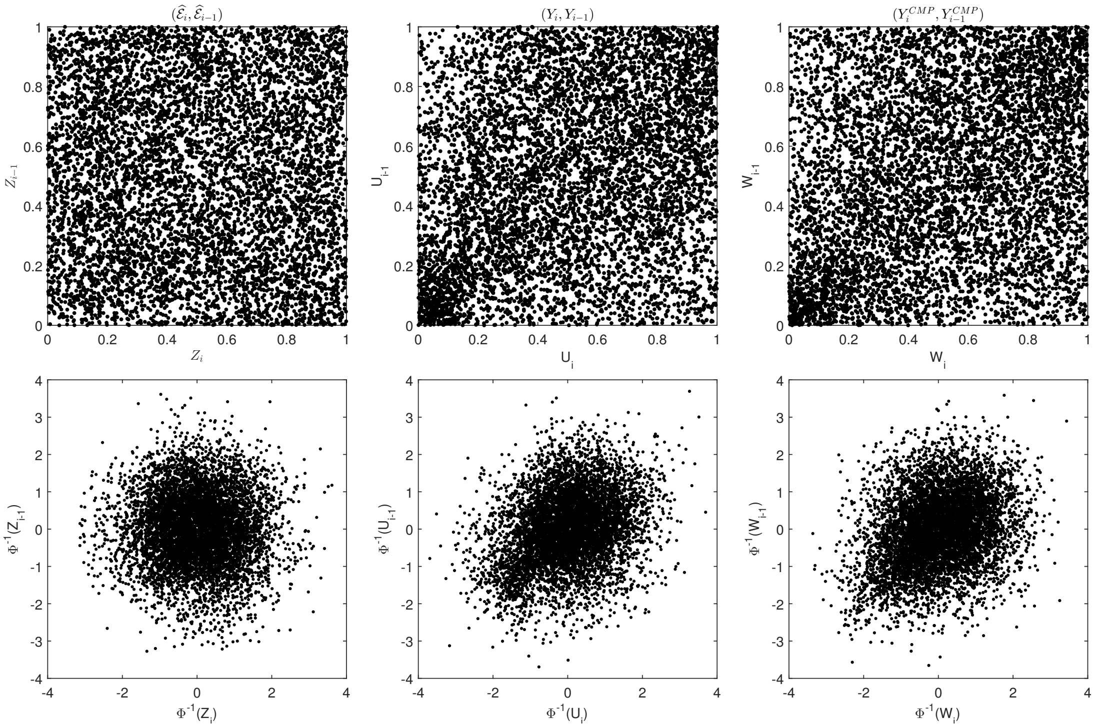
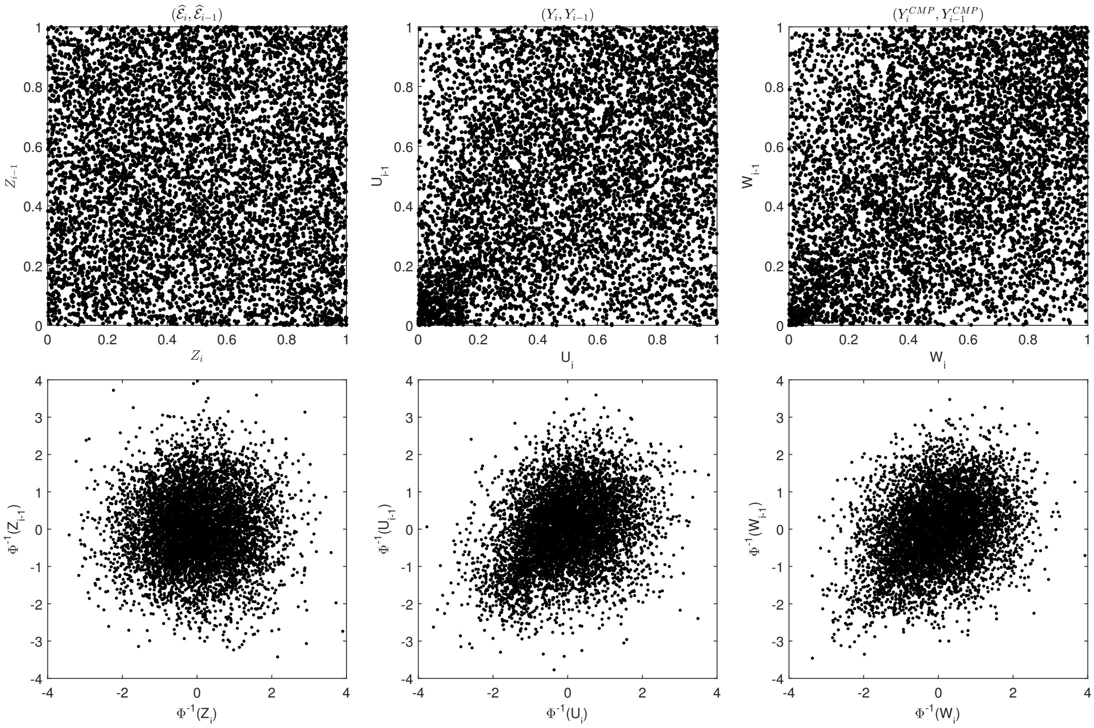


Figure 4.27: Intel

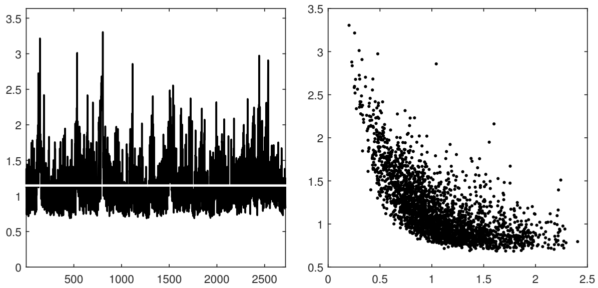


The first row in each figure shows copula scatter plots of consecutive residuals $\widehat{\mathcal{E}}$ of the CMP model, consecutive price durations Y, and consecutive observations Y^{CMP} that are simulated from the CMP model. The second row shows the scatter plots of the corresponding pseudo-$N(0,1)$ observations, i.e., the scatter plots that arise if the quantile function of the $N(0,1)$ distribution is applied to the data. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data if the number of scatter plot observations exceeds 7000 observations.

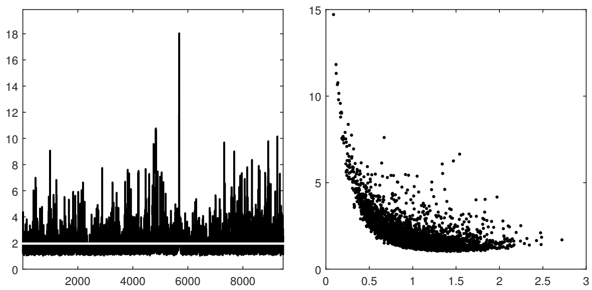
A.8 Time-varying dispersion of the CMP models

Figure 4.28: Time-varying dispersion of the CMP models.

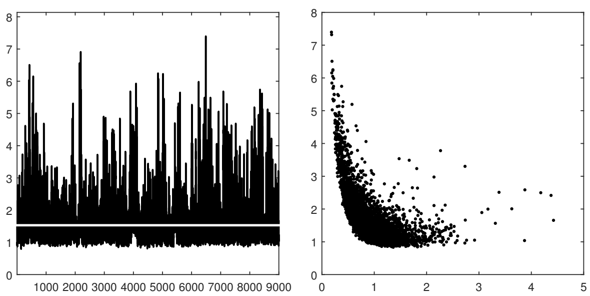
(a) Apple



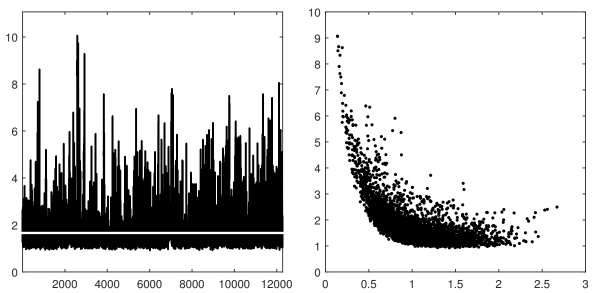
(b) Cisco



(c) Hewlett Packard



(d) Intel



The left plot in each subfigure shows the CMP-implied time-varying dispersion σ_i^2/μ_i^2 for the in-sample period. The horizontal white line is the constant dispersion of the BCACD(2,2)-GB2-LOGN model. The middle plot in each subfigure shows the relation between the mean (x-axis) and the dispersion (y-axis) of the transition distribution. The right plot in each subfigure shows the relation between the mean and the dispersion of the transition distribution on the copula scale. To avoid too much clutter, we randomly draw 7000 observations of the scatter plot data if the number of scatter plot observations exceeds 7000 observations.

5 Copulas for modeling financial returns

5.1 Motivation

A great number of studies have revealed a striking similarity in the statistical properties of time series of daily financial returns.[1] These common statistical properties of asset returns have become known as stylized facts, see Cont (2001) for an excellent review. Ever since the seminal work of Mandelbrot (1963) it has been known that the marginal distribution of a time series of daily returns can not adequately be described by a member of the normal distribution. First and foremost, daily returns exhibit so called fat tails, i.e., the probability of extreme returns is by far larger than the implied probability of a fitted normal distribution. On top of that, daily returns also cluster more around the median of the distribution. This implies that the density of the return distribution is more sharply peaked in center than the corresponding density of a fitted normal distribution. As a result, distributions of daily returns typically exhibit high excess kurtosis. The marginal distribution of daily returns is often approximately symmetric but, depending on the asset class, it can also be slightly skewed.

The dependence structure within a time series of daily financial returns is generally highly nonlinear. A vast number of studies have documented that, in liquid markets, there is no substantial linear correlation between financial returns at different points in time, which is in line with the efficient market hypothesis (cf. Fama, 1970). There is also empirical evidence that the mean of the transition distribution is often close to zero, and the assumption that time series of daily financial returns are indeed (approximately) generated by martingale difference sequences is widely accepted in theory and practice. But martingale difference sequences do not necessarily constitute sequences of independent random variables and financial returns are certainly not independent. Some simple non-linear transformation of daily returns, such as squaring or taking the absolute value, produces a time series with a moderate but a significant positive autocorrelation function which is also highly persistence. This implies that "large [price] changes tend to be followed by large changes – of either sign – and small changes tend to be followed by small changes" (Mandelbrot (1963)). This characteristic of daily financial returns is also described as volatility clustering because squared returns are proxies for the variance of the transition distribution if the underlying stochastic process is a martingale difference sequence. Consequently, the persistence in the autocorrelation function of squared returns implies that the transition variance of daily returns can be predicted by lagged squared returns. Another important stylized fact that is prevalent among many markets and financial instruments is the so

[1] Stylized facts of financial returns are dependent on the sample frequency of the data, with returns becoming more independent and their distribution closer to the normal if the frequency decreases.

called leverage effect which corresponds to a negative correlation between squared returns and lagged returns.

(Generalized) autoregressive conditional heteroscedasticity (GARCH) models of Engle (1982) and Bollerslev (1986) are the most popular model class which effectively reproduce these stylized facts of daily financial returns. Obviously, a minimal requirement for the successful application of a statistical model of asset returns is its capability of reproducing these stylized facts. Therefore, we investigate in this chapter the construction of parametric copulas that give rise to a CMP which incorporates these general accepted statistical properties of daily financial returns. In particular, a suitable stochastic process for daily financial returns should exhibit the following properties:

- The marginal distribution exhibits excess kurtosis and fat tails.
- No dependence in the mean: The process is a martingale difference sequence.
- Volatility clustering: Squaring or taking the absolute value of each random variable at any point in time produces a sequence which exhibits a positive autocorrelation function.
- Leverage effect: The correlation between squared random variables and lagged random variables is negative.

Evidently, the possible separate modeling of the marginal distribution in the copula-based approach to time series analysis is highly beneficial to reproduce the features of the marginal distribution of daily returns. Here we can draw upon the vast literature that is concerned with the fitting of the marginal return distribution (see Behr and Pötter, 2009, for an overview) in order to obtain a time series model which reproduces the marginal features of financial returns in an excellent way.[2]

The other stylized facts of daily asset returns are neither a property of the copula nor a property of the marginal distribution. Indeed, the remaining stylized facts are a property of both the copula and the marginal distribution together and that is the reason why the copula-based approach to time series analysis encounters some difficulties when it comes to reproducing stylized facts that are related to this kind of dependence structure. For instance, the martingale difference sequence property of a time series is a property of the mean of its transition distribution. The mean of the transition distribution itself is determined by the copula and the marginal distribution and thus not invariant under transformations of the marginal distribution. Thus, except for the case of the independence copula, we can not find conditions for a copula such that for any marginal distribution

[2] Classical time series models might have some problems in reproducing the marginal features of the time series. Malmsten and Teräsvirta (2010) point out that GARCH models can not exhibit any combination of the kurtosis of the marginal distribution and the functional form of the autocorrelation function. The authors show that the GARCH(1,1) model and the EGARCH(1,1) model can not reproduce the stylized facts of high kurtosis and low-starting autocorrelation of financial returns.

the resulting time series has a zero autocorrelation function or is a martingale difference sequence. In other words, it is in general not possible to choose the copula independently of the marginal distribution if the resulting process should reproduce stylized facts that are related to the marginal distribution. However, if the transition distribution is symmetric, then its mean equals its median which is completely specified by the copula. Hence, one can find sufficient conditions such that the resulting process is a martingale difference sequence for all symmetric marginal distributions. On these grounds, we consider in this chapter the construction of conditionally symmetric martingale difference sequences that exhibit volatility clusters.

Besides the afore-mentioned stylized facts, many parametric models of financial returns additionally assume that the transition variance is increasing in the absolute value of each conditioning variable. Such a functional shape of the transition variance seems plausible since then the transition variance is never decreased if the absolute value of any lagged return is increased. In order to capture the leverage effect, parametric models of daily financial returns typically specify the transition variance as an asymmetric function of the past so that for a given absolute value of a lagged return the transition variance is larger if the lagged return is negative. Therefore, we also investigate whether we can find sufficient conditions for a copula such that the CMP exhibits the following additional properties:

- The transition variance is increasing in the absolute value of each conditioning variable.

- For a given absolute value of a lagged realization the transition variance is larger for the negative than for the positive value.

This chapter is outlined as follows. Section 5.2 establishes necessary and sufficient conditions for a conditionally symmetric martingale difference sequence and shows that popular copula families can not be used to obtain an adequate model of financial returns. Section 5.3 considers the construction of copula-based first-order Markov processes that exhibit volatility clustering. For that purpose, we define two important dependence properties of a bivariate copula, one of which is sufficient for a positive autocorrelation function of squared or absolute symmetric random variables and an increasing conditional variance in the absolute value of the conditioning variable. Moreover, several construction methods of copulas with the desired dependence properties are presented and compared. We also explore means to implement the leverage effect into the copula-based approach to time series analysis. Copulas for modeling higher-order Markov processes with volatility clustering are discussed in Section 5.4. In particular, we examine whether the results for first-order Markov processes can be generalized to higher-order Markov processes. Section 5.5 provides a discussion of the established results.

5.2 Copulas and martingale difference sequences

In this section we examine the relation between martingale difference sequences (MDS) and dependence properties of copulas. We show that a necessary condition for a martingale difference sequence is that all bivariate copulas of the conditional autocopula sequence are not strictly quadrant dependent. Since almost all commonly used copulas are strictly quadrant dependent, it follows that there is a need for developing new families of parametric copulas that are not strictly quadrant dependent. Moreover, we establish conditions that characterize conditionally symmetric martingale difference sequences (CSMDS). For that purpose we introduce the class of vertically symmetric copulas.

5.2.1 Necessary conditions for a MDS in terms of copulas

In the following, let Y be a stationary[3] stochastic process with parameter space $\mathbb{N}$ and state space $\mathcal{S}$ which is a connected subset of $\mathbb{R}$. We assume that the autocovariance function of Y and the variance of its transition distribution exists and denote $C_{1,1+i|2:i} = C_{12}$ if $i = 1$.

Definition 5.1 (Martingale difference sequence (MDS))

Y is a martingale (wrt to the natural filtration) if $\mathbb{E}[Y_t|Y_{t-1:1}] = 0$.

A martingale difference sequence is unpredictable in the mean and therefore consistent with the empirical evidence that conditional on past and current returns the expectation of future returns is approximately zero. Since $\mathbb{E}[Y_t|Y_{t-1:1}] = 0$ is equivalent to the condition that Y_t is uncorrelated with all measurable functions of $Y_{t-1:1}$, a martingale difference sequence is also a white noise process. That is, if Y is a martingale difference sequence then $\mathbb{E}[Y_t] = 0$ and $\mathbb{C}\mathrm{ov}[Y_t, Y_{t+j}] = 0$ for all $(t, j) \in \mathbb{N}^2$. Since quantile functions are increasing, Lemma 2.4 implies that $C_{i,1}$ can not be strictly quadrant dependent if $\mathbb{C}\mathrm{ov}[Y_i, Y_1] = 0$. It follows that a necessary condition for a martingale difference sequence is that all bivariate copulas $C_{i,1}, i \in \mathbb{N}$, of a time series are not strictly quadrant dependent. The following proposition also shows that the bivariate conditional copulas of the conditional autocopula sequence are not strictly quadrant dependent if they describe the dependence of a martingale difference sequence.

Proposition 5.1 (Necessary condition for a CMP to generate a MDS)

Let Y be a martingale difference sequence. Then $C_{1,1+i|2:i}$ is not strictly quadrant dependent for all $i \in \mathbb{N}$.

Proof. Let $i \in \mathbb{N}$. Wlog assume that $C_{1,1+i|2:i}$ is strictly PQD. From Lemma 2.4 it follows that $\mathbb{C}\mathrm{ov}[Y_t, Y_{t-i}|Y_{t-1:t-i+1}] > 0$. Consequently, $\mathbb{E}\big[\mathbb{C}\mathrm{ov}[Y_t, Y_{t-i}|Y_{t-1:t-i+1}]\big] > 0$ since the

[3] It is obvious how the following results can be generalized to non-stationary processes. However, we are interested in modeling stationary processes and the notation for stationary processes is much more readable.

integrand is almost surely non-negative and strictly positive on a set with positive Lebesgue measure. This implies that $\mathbb{C}\text{ov}[Y_t, Y_{t-i}] > 0$, since by the law of total covariance we have that

$$\begin{aligned}\mathbb{C}\text{ov}[Y_t, Y_{t-i}] &= \mathbb{C}\text{ov}\big[\mathbb{E}[Y_t|Y_{t-1:t-i+1}], \mathbb{E}[Y_{t-i}|Y_{t-1:t-i+1}]\big] + \mathbb{E}\big[\mathbb{C}\text{ov}[Y_t, Y_{t-i}|Y_{t-1:t-i+1}]\big] \\ &= \mathbb{C}\text{ov}\big[0, \mathbb{E}[Y_{t-i}|Y_{t-1:t-i+1}]\big] + \mathbb{E}\big[\mathbb{C}\text{ov}[Y_t, Y_{t-i}|Y_{t-1:t-i+1}]\big] \\ &= \mathbb{E}\big[\mathbb{C}\text{ov}[Y_t, Y_{t-i}|Y_{t-1:t-i+1}]\big]. \qquad \blacksquare\end{aligned}$$

Almost all popular families of copulas have the strict quadrant dependence property if they do not collapse to the product copula. This is true for the Gaussian, FGM, Plackett copula, all extreme-value copulas, and, under weak regularity conditions, Archimedean copulas such as the Clayton, Gumbel, Frank, or AMH copula (Joe, 1997, Theorem 4.2). The only exception is the Student-t copula which does not have the QD property for all finite degrees of freedom, see Appendix A.1 for details. Consequently, the only martingale difference sequence that one can construct with commonly used copula families is a sequence of independent random variables. Since the strict quadrant dependence property is implied by all popular dependence properties, the question arises, what properties should a copula exhibit to reproduce the characteristics of financial return series? Moreover, there is a need for developing new families of parametric copulas that have these desired properties. We address these questions in the following sections, starting with sufficient conditions for a copula to generate a martingale difference sequence.

5.2.2 Sufficient conditions for a MDS in terms of copulas

Contrary to the Markov property, which is a property of the copulas, a martingale difference sequence is characterized by the copulas and the marginal distribution of the process. In order to establish sufficient conditions for the MDS property in terms of the copulas, we have to impose the restriction that the marginal distribution is symmetric around zero. In this way, we can characterize the class of conditionally symmetric martingale difference sequences and obtain sufficient conditions for martingale difference sequences.

Definition 5.2 (Symmetric distribution)

A random variable Y_t has a symmetric distribution around $\mu \in \mathbb{R}$ if one of the following equivalent statements is true.

1) $Y_t - \mu \overset{d}{=} \mu - Y_t$.

2) $F_t(\mu + y_t) = 1 - F_t(\mu - y_t)$ for all $y_t \in \mathbb{R}$.

3) $f_t(\mu + y_t) = f_t(\mu - y_t)$ for almost all $y_t \in \mathbb{R}$.

Setting $y_t = 0$ shows that μ is the median and if $\mathbb{E}[Y_t]$ exists then $\mu = \mathbb{E}[Y_t]$. Moreover, we have the relation $F_t^{-1}(z) = 2\mu - F_t^{-1}(1-z)$ for all $z \in [0,1]$. We say that the stochastic process Y has a symmetric margin around μ if Y_t has a symmetric margin around μ for all $t \in \mathbb{N}$.

Definition 5.3 (Symmetric transition distribution around the median)

Y has a symmetric transition distribution around its median μ if for all $t \in \mathbb{N}$:

$$F_{t|t-1:1}(\mu + Y_t|Y_{t-1:1}) = 1 - F_{t|t-1:1}(\mu - Y_t|Y_{t-1:1}) \quad (F_{t-1:1}\text{-a.s.}).$$

It is easy to see that if the transition distribution of Y is symmetric around μ then Y has a symmetric marginal distribution around μ. If Y has a symmetric transition distribution then $F_{t|t-1:1}^{-1}(0.5|Y_{t-1:1}) = \mu$, (a.s.), for all $t \in \mathbb{N}$, i.e., the conditional median does not depend on the past. Note that, in general, a constant conditional median does not imply that the transition distribution is symmetric.

Definition 5.4 (Vertically, horizontally, and jointly symmetric copula)

Let $C^{\mathrm{V}}_{1K|2:K-1}\colon [0,1]^K \to [0,1]$, $C^{\mathrm{V}}_{1K|2:K-1}(u_1, u_K|u_{2:K-1}) = u_K - C_{1K|2:K-1}(1-u_1, u_K|u_{2:K-1})$, and $C^{\mathrm{H}}_{1K|2:K-1}\colon [0,1]^K \to [0,1]$, $C^{\mathrm{H}}_{1K|2:K-1}(u_1, u_K|u_{2:K-1}) = u_1 - C_{1K|2:K-1}(u_1, 1-u_K|u_{2:K-1})$. We call $C^{\mathrm{V}}_{1K|2:K-1}$ and $C^{\mathrm{H}}_{1K|2:K-1}$ the vertically and horizontally reflected version of $C_{1K|2:K-1}$, respectively.

$C_{1K|2:K-1}$ is vertically symmetric if it is equal to its vertically reflected version, i.e., one of the following equivalent statements is true:

1) $\Big((U_1, U_K)|U_{2:K-1} = u_{2:K-1}\Big) \overset{d}{=} \Big((1-U_1, U_K)|U_{2:K-1} = u_{2:K-1}\Big)$ for almost all $u_{2:K-1}$.

2) $C_{1K|2:K-1}(u_1, u_K|U_{2:K-1}) = C^{\mathrm{V}}_{1K|2:K-1}(u_1, u_K|U_{2:K-1})$ with probability one, for all $(u_1, u_K) \in [0,1]^2$.

3) For $C_{1:K}$-almost all $(u_{1:K}) \in [0, 0.5] \times [0,1]^{K-1}$:

$$c_{1K|2:K-1}(u_1, u_K|u_{2:K-1}) = c_{1K|2:K-1}(1-u_1, u_K|u_{2:K-1})$$

or

$$c_{1K|2:K-1}(0.5-u_1, u_K|u_{2:K-1}) = c_{1K|2:K-1}(0.5+u_1, u_K|u_{2:K-1}).$$

Accordingly, we call $C_{1K|2:K-1}$ horizontally symmetric if it is equal to its horizontally reflected version. A copula is jointly symmetric if it is vertically and horizontally symmetric.

We discuss the construction of vertically symmetric copulas later on in Section 5.3. An easy graphical interpretation of a vertically symmetric copula can be given if its density exists. For almost all $u_{2:K-1}$ its density is reflection symmetric across the vertical line $v := \{(u_1, u_K) \mid u_1 = 0.5, u_K \in [0,1]\}$. Moreover, a scatter plot of the copula data should display a symmetric scattering of the realizations around the line v. Also note that $C_{1K|2:K-1}(0.5, u_K|Y_{2:K-1}) = 0.5u_K$ for all $u_K \in (0,1)$ and if $F_{1|2:K-1}$ is symmetric around μ then $C_{1K|2:K-1}$ is vertically symmetric if and only if $\Big((Y_1 - \mu, Y_K)|Y_{2:K-1} = y_{2:K-1}\Big) \overset{d}{=} \Big((\mu - Y_1, Y_K)|Y_{2:K-1} = y_{2:K-1}\Big)$ for almost all $y_{2:K-1}$. The significance of vertically symmetric copulas results from the fact that they characterize time series processes with a symmetric transition distribution around μ.

Lemma 5.1 (Symmetric transition distribution around μ)

Y has a symmetric transition distribution around μ if and only if all bivariate conditional copulas of the conditional autocopula sequence $(C_{1,i+1|2:i})_{i\in\mathbb{N}}$ are vertically symmetric and Y has a symmetric margin around μ.

Proof. See Appendix A.2 ■

The transition distribution in Lemma 5.1 is always symmetric around the same value so that the median of the transition distribution is always μ. Consequently, we can derive the following necessary and sufficient conditions for Y to be a conditionally symmetric martingale difference sequence (CSMDS).

Proposition 5.2 (Characterization of CSMDS in terms of copulas)

Let the marginal distribution of Y be symmetric around zero. Then Y is a conditionally symmetric martingale difference sequence if and only if all bivariate conditional copulas of the conditional autocopula sequence are vertically symmetric.

Proof. From Lemma 5.1 it follows that Y has a symmetric transition distribution around zero. Since the conditional median is equal to the conditional mean if the conditional distribution is symmetric, the conclusion follows. ■

Proposition 5.2 generalizes Theorem 4 in Ibragimov (2009) in which first-order Markov processes are considered. Proposition 5.2 is not a necessary condition for a martingale difference sequence since only martingale difference sequences with a symmetric transition distribution are considered. This rules out martingale difference sequences with an asymmetric transition distribution which can have an asymmetric or symmetric marginal distribution.[4] However, the class of conditionally symmetric martingale difference sequences is very rich and plausible for financial returns. Thus, the sufficient condition in Proposition 5.2 is also almost necessary for a martingale difference sequence that generates financial returns.

[4] The process $Y_t = \mathcal{E}_t - \mathcal{E}_{t-1}$, where $\mathcal{E}_t$ is iid, has always a symmetric marginal distribution because $Y_t = -Y_t$. The transition distribution is given by $\mathcal{E}_{t-1}$ and asymmetric if $\mathcal{E}_{t-1}$ has an asymmetric distribution.

It follows from the definition of the partial autocorrelation function (see Brockwell and Davis, 2009, Definition 3.4.2) that the partial autocorrelation function is also zero if the autocorrelation function is zero. If the marginal distribution of Y is not symmetric around zero then the autocovariance function may not be equal to zero if all copulas of the conditional autocopula sequence are vertically symmetric. However, we can show that for measures of concordance to be zero, it is sufficient that all copulas of the conditional autocopula sequence are vertically or horizontally symmetric.

Proposition 5.3 (Zero (conditional and partial) concordance measures)

Let all copulas of the conditional autocopula sequence be either vertically or horizontally symmetric. Then the auto Spearman's rho function $(\rho_{C_{1,1+j}})_{j\in\mathbb{N}}$, the partial auto Spearman's rho function $(\rho_{C^{\partial j-1}_{1,1+j;\,2:j}})_{j\in\mathbb{N}}$, the auto Kendall's tau function $(\tau_{C_{1,1+j}})_{j\in\mathbb{N}}$, and the partial auto Kendall's tau function $(\tau_{C^{\partial j-1}_{1,1+j;\,2:j}})_{j\in\mathbb{N}}$, are zero. Moreover, the conditional Spearman's rho function $\left(\rho_{C_{1,1+j|2:j}}(Y_{2:j})\right)_{j\in\mathbb{N}}$ and the conditional Kendall's tau function $\left(\tau_{C_{1,1+j|2:j}}(Y_{2:j})\right)_{j\in\mathbb{N}}$ are almost-surely zero.

Proof. See Appendix A.3. ■

Once again, vertical or horizontal symmetry of the copulas of the conditional autocopula sequence is just a sufficient condition for zero concordance measures.[5] Moreover, CMPs with horizontal symmetric bivariate conditional copulas are not martingale difference sequences in general and provide means to construct processes with zero autocorrelation function but where the conditional expectation depends on the past.[6]

5.3 Copulas for modeling volatility clusters: The first-order Markov case

Although we have derived sufficient conditions for a CMP to be a martingale difference sequence, this does not imply that the process also exhibits volatility clustering. Therefore, we now examine properties of a copula that are useful for generating a CMP that can be used to model financial returns. For ease of exposition, and also because this case is more tractable, we begin with the first-order Markov case in this section. At first, we investigate conditions that imply a positive association between squared or absolute symmetric random variables. Subsequently, we explore the relation between the copula structure and the functional shape of the conditional variance function.

[5] For instance, consider the bivariate mixture copula $C = 0.5C^{\text{Cl}}(\theta_1) + 0.25C^{\text{V-Cl}}(\theta_2) + 0.25C^{\text{H-Cl}}(\theta_3)$. Spearman's rho is then given by $\rho_C = 0.5\rho_{C^{\text{Cl}}(\theta_1)} + 0.25\rho_{C^{\text{V-Cl}}(\theta_2)} + 0.25\rho_{C^{\text{H-Cl}}(\theta_3)}$. Thus, for given θ_1 we can find $\theta_i, i = 2, 3$, such that Spearman's rho is zero but the resulting copula is not vertically symmetric.

[6] For example, a first-order Markov process with $C_{t,t-1} = 0.5C^{\text{Cl}}(\theta) + 0.5C^{\text{H-Cl}}(\theta)$ has zero autocorrelation function but $\mathbb{E}[Y_t|Y_{t-1} = y_{t-1}]$ is rather negative if y_{t-1} is a large negative or positive value.

5.3.1 The PQD$_{(1,3)}$-NQD$_{(2,4)}$ property

So far we have not excluded martingale difference sequences that allow for a negative autocorrelation between squared or absolute random variables of the sequence. For instance, consider the random vector (X_1, X_2), with distribution given by

$$X_1 \sim N(0,1), \quad X_2 \stackrel{d}{=} 1/X_1.$$

Then, $(X_1, X_2) \stackrel{d}{=} (-X_1, X_2)$, which is equivalent to vertical symmetry of the copula C_{X_1,X_2} since X_1 has a symmetric distribution. Now, $X_2^2 = 1/X_1^2 =: g(X_1^2)$. Since g is an (a.s.) decreasing function of X_1^2, we conclude, by Theorem 1 (iii) in Lehmann (1966), that (X_1^2, X_2^2) is NQD. In fact, the copula of (X_1^2, X_2^2) is the lower Fréchet bound since X_1^2 and X_2^2 are continuous. Thus, if Y is a first-order Markov process with the copula of (X_1, X_2), then Y is a MDS with $\mathbb{C}\mathrm{ov}[Y_1^2, Y_2^2] = -1$, which follows from Theorem 5.25 in McNeil et al. (2005).[7] But it is well documented in the literature that squared and absolute returns of financial time series exhibit a positive autocorrelation function which reflects volatility clustering.

In order to exclude a negative correlation between squared or absolute random variables, we have to impose some structure on the copula. For that purpose, assume that (Y_1, Y_2) has symmetric margins around zero and copula C. Moreover, let Q_i refer to the i-th quadrant of the unit cube, i.e.,

$$Q_1 = (0.5,1)^2, \quad Q_2 = (0.5,1)\times(0,0.5), \quad Q_3 = (0,0.5)^2, \quad Q_4 = (0,0.5)\times(0.5,1),$$

and $G(y_i) \in \{|y_i|, y_i^2\}, i = 1,2$, with derivative $g(y_i)$. Using the generalization of Hoeffding's lemma (Lemma 2.3) and a change of variables, we can express the covariance between squared or absolute symmetric random variables in terms of the underlying copula as

$$\begin{aligned}\mathbb{C}\mathrm{ov}[G(Y_1), G(Y_2)] &= \sum_{i=1,2,3,4} \int_{Q_i} \left(\prod_{k=1,2} \frac{g(F_k^{-1}(u_k))}{f_k(F_k^{-1}(u_k))} \right) (C(u_1,u_2) - u_1 u_2)\, \mathrm{d}u_1 \mathrm{d}u_2 \qquad (5.3.1)\\ &= \sum_{i=1,2,3,4} \int_{Q_i} h(u_1,u_2)\,(C(u_1,u_2) - u_1u_2)\, \mathrm{d}u_1\mathrm{d}u_2.\end{aligned}$$

We note that

$$\begin{aligned}&\forall (u_1,u_2) \in Q_1 \cup Q_3 \colon h(u_1,u_2) \geq 0,\\ &\forall (u_1,u_2) \in Q_2 \cup Q_4 \colon h(u_1,u_2) \leq 0,\end{aligned}$$

[7] This is not possible in the classical GARCH framework without risking a negative conditional variance. To the best of our knowledge, Vries (1991) is the only one who constructed a MDS with negative correlation between consecutive random variables.

so that the integrand in (5.3.1), and so the covariance, is non-negative if

$$\forall (u_1, u_2) \in Q_1 \cup Q_3 \colon C(u_1, u_2) \geq u_1 u_2,$$
$$\forall (u_1, u_2) \in Q_2 \cup Q_4 \colon C(u_1, u_2) \leq u_1 u_2.$$

Therefore, the following property of a copula is sufficient for a non-negative correlation between pairs of squared or absolute symmetric random variables.

Definition 5.5 ($\mathbf{PQD_{(1,3)}}$-$\mathbf{NQD_{(2,4)}}$ property)

A bivariate copula C has the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property if

$$\forall (u_1, u_2) \in Q_1 \cup Q_3 \colon C(u_1, u_2) \geq u_1 u_2, \tag{5.3.2}$$

and

$$\forall (u_1, u_2) \in Q_2 \cup Q_4 \colon C(u_1, u_2) \leq u_1 u_2. \tag{5.3.3}$$

We call the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property strict if (5.3.2) and (5.3.3) hold and the statement

$$\exists i = 1, 2, 3, 4; \forall (u_1, u_K) \in Q_i \colon C(u_1, u_2) = u_1 u_2$$

is false.

A few remarks regarding the definition of the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property are in order. We have not explicitly defined the values of the copula for $(u_1, u_K) \in (0,1)^2 \backslash (\cup_{i=1}^4 Q_i)$. Since a copula is uniformly continuous (Nelsen, 2006, Theorem 2.2.4 and Corollary 2.2.6), it follows easily that $\forall u_1 \in (0,1) \colon C(u_1, 0.5) = 0.5u_1$ and $\forall u_2 \in (0,1) \colon C(0.5, u_2) = 0.5u_2$, if C has the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property.[8] We can, at the same time, replace the left- and the right-hand side of the inequality in (5.3.2) and (5.3.3) with $\mathbb{P}(U_1 \geq u_1, U_2 \geq u_2)$ and $\prod_{i=1,2} \mathbb{P}(U_i \geq u_i)$, respectively. If C is vertically symmetric then (5.3.2) and (5.3.3) are equivalent. Condition (5.3.2) implies (5.3.3) because if $(u_1, u_2) \in Q_2 \cup Q_4$ then $(1 - u_1, u_2) \in Q_1 \cup Q_3$, so that $C(u_1, u_2) = u_2 - C(1 - u_1, u_2) \leq u_2 - (1 - u_1)u_2 = u_1 u_2$, where the equality follows because C is vertically symmetric and the inequality if (5.3.2) is true. If $(u_1, u_2) \in Q_1 \cup Q_3$ then the previous inequality is reversed if (5.3.3) holds, so (5.3.2) and (5.3.3) are indeed equivalent if C is vertically symmetric.

Similar to the PQD property, which is the weakest positive dependence property of a copula, the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property is the weakest property that a copula can exhibit so that squared or absolute symmetric random variables are positively associated. If a copula has the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property it is locally PQD in the first and third unit quadrant and locally NQD in the second and fourth unit quadrant. Thus, if Y_1 and Y_2

[8] For instance, if $u_1 \in (0, 0.5]$, then $C(u_1, 0.5) = \lim_{\epsilon \to 0} C(u_1, 0.5 - \epsilon) \geq 0.5u_1$ and $C(u_1, 0.5) = \lim_{\epsilon \to 0} C(u_1, 0.5 + \epsilon) \leq 0.5u_1$, so that $C(u_1, 0.5) = 0.5u_1$.

have such a copula, then Y_1 and Y_2 are more likely to be large (small) together or to be close to their medians in the first (third) unit quadrant compared with $\tilde{Y}_1$ and $\tilde{Y}_2$, where $\tilde{Y}_i \overset{d}{=} Y_i, i = 1, 2$, and $\tilde{Y}_1 \perp \tilde{Y}_2$. In addition, Y_1 and Y_2 are more likely to be simultaneously large and small (small and large) or to be close to their medians in the second (fourth) unit quadrant than $\tilde{Y}_1$ and $\tilde{Y}_2$. Consequently, if a distribution has such a copula it is more likely that extreme realizations occur together and that realizations cluster around the center of the distribution. Indeed, the following proposition shows that then, for symmetric random variables, the conditional covariances in the first and third quadrant are positive while the conditional covariances in the second and fourth quadrant are negative. Moreover, the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property does not only imply a non-negative correlation for squared or absolute symmetric random variables but also that the corresponding copulas have the PQD property.

Lemma 5.2 (Implications of the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property)

1. If (Y_1, Y_2) has symmetric marginal distributions around zero and its copula the (strict) $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property then the copula of $(|Y_1|^i, |Y_2|^i), i = 1, 2$, has the (strict) PQD property.

2. Let (Y_1, Y_2) have symmetric marginal distributions around μ and a copula with the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property, and set

$$R_1 = (-\infty, \mu)^2, \quad R_3 = (\mu, \infty)^2,$$
$$R_2 = (\mu, \infty) \times (-\infty, \mu), \quad R_4 = (-\infty, \mu) \times (\mu, \infty).$$

Then,

$$\forall i = 1, 3\colon \mathbb{C}\text{ov}[Y_1, Y_2 | (Y_1, Y_2) \in R_i] \geq 0,$$
$$\forall i = 2, 4\colon \mathbb{C}\text{ov}[Y_1, Y_2 | (Y_1, Y_2) \in R_i] \geq 0,$$

with equality only if the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property is not strict.

Proof. See Appendix A.4 ■

5.3.2 The $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property

As the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property is a sufficient property for a copula to display positive association of squared or absolute symmetric random variables, it is natural to explore conditions that imply the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property. Stronger conditions are also of interest since a positive correlation between squared or absolute random variables does not pin down the functional form of the conditional variance. For instance, it is not necessarily true that the conditional variance increases with the absolute value of the conditioning variable. But such a behavior may be desired for certain applications. On these grounds, we now investigate the relation between the form of a copula and the implied conditional variance. In particular, we examine whether we can impose conditions on a vertically symmetric bivariate copula such that the conditional variance of one variable is a (strictly) increasing function of the absolute value of the other variable, i.e.,

$$\mathbb{V}\text{ar}[Y_1|Y_2 = y_2] =: \sigma^2_{Y_1}(y_2) \text{ is (strictly) increasing in } |y_2|.$$

In the following, a (strictly) increasing conditional variance means that the conditional variance is (strictly) increasing in the absolute value of the conditioning variable. Provided that $\mathbb{E}[Y_1|Y_2] = 0$ and $(Y_1, Y_2) \sim F$, we can express the conditional variance generally as

$$\sigma^2_{Y_1}(y_2) = 2\left(\int_{-\infty}^{0} y_1 F(y_1|y_2)\mathrm{d}y_1 + \int_{0}^{\infty} y_1(1 - F(y_1|y_2))\mathrm{d}y_1\right). \tag{5.3.4}$$

From (5.3.4) we see that the conditional variance does not depend on the sign of the conditioning value, i.e., $\sigma^2_{Y_1}(y_2) = \sigma^2_{Y_1}(-y_2)$ for all $y_2 \in \mathbb{R}$, if $F(y_1|y_2) = F(y_1|-y_2)$ for all $(y_1, y_2) \in \mathbb{R}^2$. This is equivalent to $(Y_1, Y_2) \stackrel{d}{=} (-Y_1, Y_2)$ and implies horizontal symmetry of the underlying copula, i.e., $C(u_1, u_2) = u_1 - C(u_1, 1 - u_2)$. In this case, (5.3.4) simplifies to

$$\sigma^2_{Y_1}(y_2) = 4\int_{0}^{\infty} y_1(1 - F(y_1|y_2))\mathrm{d}y_1 = -4\int_{-\infty}^{0} y_1 F(y_1|y_2)\mathrm{d}y_1,$$

Thus, if a bivariate distribution has a jointly symmetric copula and symmetric marginal distributions, then it can not produce leverage effects. To obtain a sufficient condition for a (strictly) increasing conditional variance we define the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property.

Definition 5.6 ($\mathbf{SI_{(1,3)}}$-$\mathbf{SD_{(2,4)}}$property)

A copula C has the (strict) $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property if C has the (strict) $\mathrm{SI}(U_1|U_2)$ property in the first and third quadrant and the (strict) $\mathrm{SD}(U_1|U_2)$ property in the second and fourth quadrant, i.e.,

$$\forall (u_1, u_2) \in Q_1 \cup Q_3 : C(u_1|u_2) \text{ is (strictly) decreasing in } u_2,$$
$$\forall (u_1, u_2) \in Q_2 \cup Q_4 : C(u_1|u_2) \text{ is (strictly) increasing in } u_2,$$

where $C(u_1|u_2) = F_{U_1|U_2}(u_1|u_2)$.

The definition of the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property does not make an explicit statement about $C(u_1|u_2)$ if $u_1 = 0.5$ or $u_2 = 0.5$. If a copula has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property then $C(0.5|u_2) = 0.5$ (a.s.), see Lemma 5.3. Moreover, since the conditional cdf $C(u_1|u_2)$ as a function of u_2 is only almost surely unique, the actual value of $C(u_1|0.5)$ does not matter. If C is vertically symmetric, then C has the (strict) $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property if C has the (strict) $\mathrm{SI}(U_1|U_2)$

property in either the first or third quadrant and the (strict) SD($U_1|U_2$) property in either the second or fourth quadrant. Obviously, if C is jointly symmetric, then it suffices to show that $C(u_1|u_2)$ has the desired functional form in one quadrant in order to verify the SI$_{(1,3)}$-SD$_{(2,4)}$ property. The following proposition easily follows from (5.3.4).

Proposition 5.4 (Strictly increasing conditional variance)

Let (Y_1, Y_2) have a vertically symmetric copula C and each marginal distribution be symmetric around μ. Then $\sigma^2_{Y_1}(y_2)$ is (strictly) increasing in the absolute value of y_2 if C has the (strict) SI$_{(1,3)}$-SD$_{(2,4)}$ property.

The following properties follow from the SI$_{(1,3)}$-SD$_{(2,4)}$ property and are important for establishing some dependence concepts later on.

Lemma 5.3 (Properties of copulas with the SI$_{(1,3)}$-SD$_{(2,4)}$ property)

Let C have the (strict) SI$_{(1,3)}$-SD$_{(2,4)}$ property. It holds:

1. C has the (strict) PQD$_{(1,3)}$-NQD$_{(2,4)}$ property.
2. The PQD$_{(1,3)}$-NQD$_{(2,4)}$ property is not a sufficient condition.
3. Let (Y_1, Y_2) have symmetric margins around zero and copula C. Then the copula of $(|Y_1|^i, |Y_2|^i), i = 1, 2$, has the (strict) SI($U_1|U_2$) property.
4. A necessary condition is that $\forall u_1 \in (0,1)\colon C(u_1, 0.5) = 0.5u_1$ and $\forall u_2 \in (0,1)\colon C(0.5, u_2) = 0.5u_2$.

Proof. See Appendix A.5. ■

Lemma 5.3 shows that the SI$_{(1,3)}$-SD$_{(2,4)}$ property is indeed a stronger property than the PQD$_{(1,3)}$-NQD$_{(2,4)}$ property. As a result, if two symmetric random variables have a copula with the SI$_{(1,3)}$-SD$_{(2,4)}$ property then the copula of squared or absolute symmetric random variables does not only have the PQD property but also the SI($U_1|U_2$) property. The necessary condition in 4) is actually a necessary condition for the PQD$_{(1,3)}$-NQD$_{(2,4)}$ property. We mention it here to emphasize that the SI$_{(1,3)}$-SD$_{(2,4)}$ property imposes a strong symmetry condition on the copula. This condition is equivalent to $C(0.5|u_i) = 0.5, i = 1, 2$, which means that the conditional median of $U_i|U_j = u_j, j \neq i$, must equal the unconditional median for all u_j.

Typically, we would expect that the sequence of squared or absolute random variables displays a positive but decreasing autocovariance function if the underlying sequence exhibits volatility clustering. If the process is first-order Markov, the SI$_{(1,3)}$-SD$_{(2,4)}$ property of the underlying copula is indeed a sufficient condition for this feature.

Proposition 5.5

Let Y be a stationary first-order Markov process with symmetric margin around zero and $j = 1, 2$. If C_{12} has the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property then, provided the integrals exist,

$$\forall i \in \mathbb{N}\colon \mathbb{C}\text{ov}[|Y_1|^j, |Y_i|^j] \geq \mathbb{C}\text{ov}[|Y_1|^j, |Y_{i+1}|^j] \geq 0.$$

Proof. Follows from Theorem 8.3 in Joe (1997) and Lemma 5.3, 3. ■

According to Joe (Joe, 1997, p. 271), the PQD property of the copula of (Y_t^2, Y_{t-1}^2) is not sufficient for a decreasing positive autocovariance function. Thus, Proposition 5.5 does not hold if C_{12} has only the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$property. Proposition 5.4 and Proposition 5.5 demonstrate that vertically symmetric copulas with the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property can be used to obtain symmetric first-order Markov processes that reproduce volatility clusters and exhibit a positive decreasing autocovariance function for the sequence of squared or absolute random variables. To the best of our knowledge, copulas with the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ have not been considered in the literature. Nevertheless, a subclass of a very popular copula family has indeed the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property.

Proposition 5.6

Let C be the Student-t copula with zero correlation parameter and dof $\nu < \infty$. Then C is jointly symmetric and has the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property.

Proof. That C is jointly symmetric is obvious. Let $T_\nu(\cdot)$ be the cdf of the univariate Student-t distribution with dof ν and denote $x_i = T_\nu^{-1}(u_i), t_\nu(u_i) = \partial T_\nu(u_i)$. If $(u_1, u_2) \in Q_3$ then $x_i < 0$, so that

$$\partial_2 C(u_1|u_2) = -\frac{1}{(\nu+1)t_\nu(x_2)} \left(\frac{\nu + x_2^2}{\nu+1}\right)^{-3/2} t_\nu\left(x_1 \middle/ \sqrt{\frac{\nu+x_2^2}{\nu+1}}\right) \prod_{i=1,2} x_i < 0,$$

which implies that $C(u_1|u_2)$ is strictly decreasing in u_2 for $(u_1, u_2) \in Q_3$. The conclusion follows since the Student-t copula with zero correlation parameter is jointly symmetric. ■

In fact, with an appropriate marginal distribution the Student-t copula can be used to obtain a CMP(1) process that can be represented as an AR(1)-ARCH(1) process. If X is a first-order Markov process such that (X_t, X_{t-1}) has a bivariate Student-t distribution, we obtain from the well known stochastic representation of the bivariate Student-t distribution (see Kotz and Nadarajah, 2004) that

$$X_t = \rho X_{t-1} + \sqrt{(1-\rho^2)\frac{\nu + X_{t-1}^2}{\nu+1}} T_{\nu+1}^{-1}(Z), \quad Z_t \overset{iid}{\sim} U(0,1)$$

$$= \rho X_{t-1} + \sqrt{(1-\rho^2)\frac{\nu + X_{t-1}^2}{\nu - 1}}\mathcal{E}_t, \quad \mathcal{E}_t := \sqrt{\frac{\nu-1}{\nu+1}}T_{\nu+1}^{-1}(Z_t).$$

Note that $\mathcal{E}_t$ has a Student-t distribution with $\nu+1$ dof and unit variance. This yields the following proposition.

Proposition 5.7 (AR(1)-ARCH(1) representation of CMP(1) with Student-t copula)

Let X be a stationary first-order Markov process. If the copula of (X_t, X_{t-1}) is the Student-t copula with dof ν and the marginal distribution of X is the Student-t distribution with dof ν, then X has the stochastic representation

$$X_t = \rho X_{t-1} + \sigma_t \mathcal{E}_t,$$
$$\sigma_t^2 = \mathbb{V}\text{ar}[X_t | X_{t-1:1}] = \nu a + a X_{t-1}^2, \quad a := \frac{1-\rho^2}{\nu - 1},$$

where $(\mathcal{E}_t)_{t\in\mathbb{N}}$ is an iid sequence such that $\mathcal{E}_t$ has a Student-t distribution with dof $\nu+1>3$ and unit variance.

To the best of our knowledge, the CMP(1) processes given in Proposition 5.7 are the only AR(1)-ARCH(1) processes with innovations from a Student-t distribution such that the marginal distribution is explicitly known. We observe that the constant in the variance equation must be proportional to the coefficient a of the lagged squared random variable so that the marginal distribution is a member of the Student-t family. The conditional variance of X_t is decreasing in ν for fixed x_{t-1}^2. If $\nu \to \infty$ the process collapses to a Gaussian AR(1) process with standard normal margins.

In general, the stochastic representation of a first-order Markov process with marginal distribution F_Y and such that two consecutive random variables have a Student-t copula with zero correlation parameter is given by

$$Y_t = F_Y^{-1} \circ T_\nu \left(\sqrt{\frac{\nu + X_{t-1}^2}{\nu+1}} T_{\nu+1}^{-1}(Z_t) \right), \quad Z_t \overset{iid}{\sim} U(0,1).$$

The conditional variance is in general a non-linear function of the lagged squared random variable and not tractable. However, because the Student-t copula with zero correlation parameter has the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property we know that it is a strictly increasing function if F_Y is a symmetric distribution. In order to gain some insights in the resulting non-linear dynamics, we set the dof of the Student-t copula to two and the marginal distribution to a Student-t distribution with one dof. Using the closed-form expressions for the cdf and quantile function of the Student-t distribution with one or two dof, we obtain that

$$Y_t = \tan\left(\pi\left(0.5 + \frac{g(Y_{t-1}, Z_t)}{2\sqrt{g(Y_{t-1}, Z_t)^2 + 2}}\right) - 0.5\right),$$

where

$$g(Y_{t-1}, Z_t) = \sqrt{\frac{2 + h(Y_{t-1})^2}{3}} T_3^{-1}(Z_t),$$

and

$$h(Y_{t-1}) = T_2^{-1} \circ T_1(Y_{t-1}) = \frac{\frac{2}{\pi}\tan^{-1}(Y_{t-1})}{\sqrt{2(0.25 - (\frac{1}{\pi}\tan^{-1}(Y_{t-1})^2)}}.$$

This illustrates that a small change in the marginal distribution already results in hardly tractable non-linear dynamics although the effective degree of non-linearity is rather mild in this case. On the other side, for general dof ν of the Student-t copula and arbitrary margin, the quantile function of $F_{t|t-1}$ is a strictly increasing transformation of the conditional quantile of the ARCH(1) representation of X and given by

$$F^{-1}_{Y_t|Y_{t-1}}(z|y_{t-1}) = F_Y^{-1} \circ T_\nu \circ F^{-1}_{X_t|X_{t-1}}\big(z\big|T_\nu^{-1}(F_Y(y_{t-1}))\big).$$

This representation can be used to investigate how the conditional variance is affected by a change in the marginal distribution. Assume that Y_t has finite variance and that its symmetric distribution F_Y has fatter tails than the Student-t distribution in the sense that $\forall w \leq 0$: $F_Y(w) \geq T_\nu(w)$, which implies that $|F_Y^{-1} \circ T_\nu(w)| \geq |w|$. Then $\forall z \in [0,1]$: $|F^{-1}_{Y_t|Y_{t-1}}(z|w)| \geq |F^{-1}_{X_t|X_{t-1}}(z|w)|$. Since $|F^{-1}_{X_t|X_{t-1}}(\cdot|w)|$ is increasing in the absolute value of w it also follows that $\forall z \in [0,1]$: $|F^{-1}_{X_t|X_{t-1}}(z|F_Y^{-1}(T_\nu(w))| \geq |F^{-1}_{X_t|X_{t-1}}(z|w)|$. We conclude that

$$\mathbb{V}\mathrm{ar}[Y_t|Y_{t-1} = w] = \int_{\mathbb{R}} F^{-1}_{Y_t|Y_{t-1}}(z|w)^2 \mathrm{d}z \geq \int_{\mathbb{R}} F^{-1}_{X_t|X_{t-1}}(z|w)^2 \mathrm{d}z = \mathbb{V}\mathrm{ar}[X_t|X_{t-1} = w],$$

i.e., the conditional variance is increased if the marginal distribution becomes more heavy-tailed.

5.3.3 Jointly symmetric copulas with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property

Among the popular copulas in the literature, the Student-t copula with zero correlation parameter is the only copula that is vertically symmetric and not the product copula.[9] The evaluation of the Student-t copula is very expensive due to the computation of the incomplete beta function and the inverse thereof. Moreover, the Student-t copula is jointly symmetric so the conditional variance only depends on the absolute value of the condi-

[9] Copulas of other bivariate elliptical distributions with zero correlation parameter, excluding the normal distribution, might also exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property.

tioning value but not on the sign. Therefore, we examine in the following the construction of copulas which are vertically symmetric and have the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. We first focus on the case of jointly symmetric copulas, since these copulas automatically satisfy the necessary symmetry property stated in Lemma 5.3. This rules out a negative association between the conditional variance and the conditioning variable. After that, we turn to copulas that reproduce asymmetric volatility responses.

Merged X-shaped version of a copula

Except for the Student-t copula with zero correlation or the product copula, commonly used copulas are not jointly symmetric. Therefore, we present in the following lemma a simple device that can be used to construct a jointly symmetric copula by merging a radially symmetric copula with its vertically reflected counterpart.

Definition and Lemma 5.1 (Merged X-shaped version of a copula)

Define the radially symmetric version of a copula D by

$$D^{\mathrm{RS}}(u_1, u_2) = 0.5(D(u_1, u_2) + D^{\mathrm{S}}(u_1, u_2)),$$

where

$$D^{\mathrm{S}}(u_1, u_2) = u_1 + u_2 - 1 + D(1 - u_1, 1 - u_2),$$

is the survival copula of D. Then $C^{\text{MX-D}}\colon [0,1]^2 \to [0,1]$, with

$$\begin{aligned} C^{\text{MX-D}}(u_1, u_2) &= 0.5(D^{\mathrm{RS}}(u_1, u_2) + u_2 - D^{\mathrm{RS}}(1 - u_1, u_2)) \\ &= 0.25(D(u_1, u_2) + D(1 - u_1, 1 - u_2) + 2(u_1 + u_2) \\ &\quad - (D(1 - u_1, u_2) + D(u_1, 1 - u_2) + 1)), \end{aligned}$$

is a radially and jointly symmetric copula. If D is exchangeable so is $C^{\text{MX-D}}$. We call $C^{\text{MX-D}}$ the merged X-shaped version of D.

Proof. This is obvious. ■

We now elucidate why a merged X-shaped version of a copula D might be a reasonable copula for modeling volatility, provided that the underlying copula D exhibits the PQD or a stronger positive dependence property. If D is PQD then so is D^{RS}, i.e., large (small) values of one random variable tend to be associated with large (small) values of the other random variable. The converse holds for the vertically reflected version of D^{RS}. Since $C^{\text{MX-D}}$ in Definition and Lemma 5.1 is a mixture of two copulas with equal weights, we expect that small or large values of one random variable occur with either small or large values of the other random variable with 50% probability, respectively. In other words, extreme values of one random variable should be associated with extreme values of the

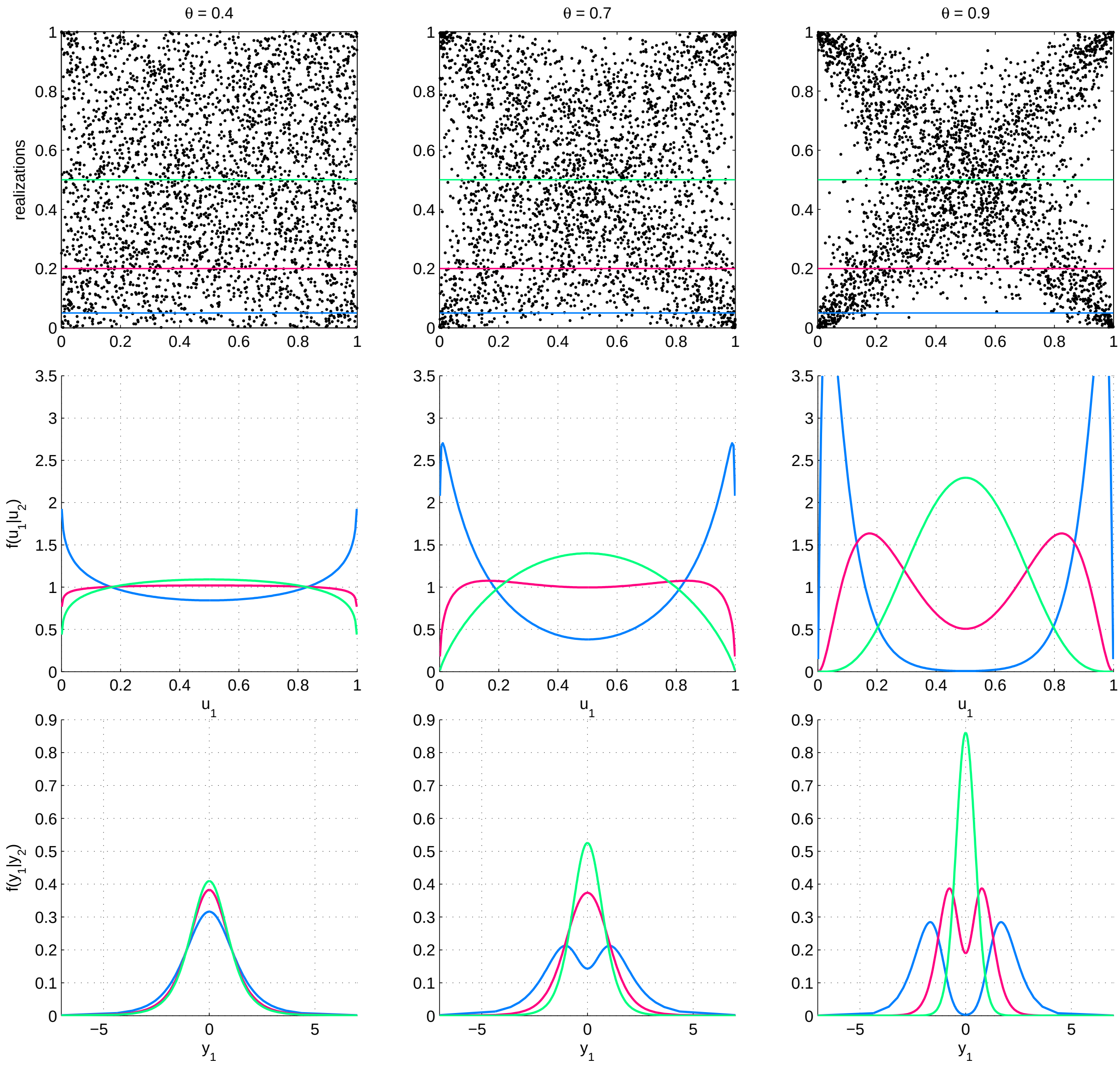


Figure 5.1: Illustration of the merged X-shaped version of the Gaussian copula with correlation parameter θ. The first row displays realizations from the copula whereas the second row depicts slices through the copula density along $u_2 \in \{0.05, 0.2, 0.5\}$. The third row shows the conditional density of $Y_1|Y_2 = y_2$ when y_2 attains its 5%, 20% or 50% quantile and (Y_1, Y_2) has this copula and Student-t margins with dof 4.

other random variable and the probability that both random variables take values near their medians should be large. As a result, the conditional variance of one variable might be (strictly) increasing in the absolute value of the other variable.

The first row of Figure 5.1 shows realizations from the merged X-shaped version of the Gaussian copula for different values of the dependence parameter θ. For the simulation we use the fact that if $(U_1, U_2) \sim D$ and $\forall i = 1, 2\colon S_i = 2\mathbb{1}_{\{Z_i \leq 0.5\}} - 1, Z_i \sim U(0, 1)$, $Z_1 \perp Z_2, (Z_1, Z_2) \perp (U_1, U_2)$, then $\big(0.5 + S_1(U_1 - 0.5), 0.5 + S_2(U_2 - 0.5)\big) \sim C^{\text{MX-D}}$. Due to the joint symmetry of the copula, the realizations are symmetrically scattered around the lines $v := \{(u_1, u_2)\colon u_1 = 0.5, u_2 \in [0, 1]\}$ and $h := \{(u_1, u_2)\colon u_2 = 0.5, u_1 \in [0, 1]\}$. We see that, for moderate dependence, the realizations are scattered around the lines of the letter

X, which is consistent with the fact that the underlying copula D has the PQD property. Indeed, if D is the upper (or lower) Fréchet bound, the realizations are located on the area $\{(u_1, u_2) \colon u_2 = u_1 \text{ or } u_2 = 1 - u_1\}$ which resembles the letter X.[10] The second row of Figure 5.1 shows slices through the copula density along $u_2 \in \{0.05, 0.2, 0.5\}$. Since we have uniform margins, this is equal to the conditional density of U_1 given $U_2 = u_2$. We observe that the probability of U_1 taking small and large values increases with θ and with U_2 also taking extreme values. This is also true if we transform the margins into Student-t distributions with dof 4. The third row in Figure 5.1 shows this case. Actually, for large values of θ the probability of Y_1 taking extreme values if Y_2 takes extremes values can be so large that the conditional density becomes bimodal. Note that the occurrence of a bimodal conditional density depends on the marginal distribution of Y_1, its likelihood decreases if the marginal distribution becomes more fat-tailed. For instance, if the margin of Y_1 is a Student-t distribution with dof 0.1, all conditional densities for $u_2 \in \{0.05, 0.2, 0.5\}$ are unimodal. The graphical illustrations in Figure 5.1 suggest that a merged X-shaped version of a copula with the PQD property can be used to obtain a bivariate distribution with the property that the conditional variance is (strictly) increasing. This is indeed true if $C^{\text{MX-D}}$ has the (strict) $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property, i.e., if

$$\begin{aligned} \forall (u_1, u_2) \in Q_3 \colon & D(u_1|u_2) + D(u_1|1-u_2) - (D(1-u_1|u_2) + D(1-u_1|1-u_2)) \\ & \text{is (strictly) decreasing in } u_2. \end{aligned} \tag{5.3.5}$$

This raises the question of whether one can impose conditions on D such that (5.3.5) is automatically satisfied so that there is no need to check (5.3.5) directly. Since many copulas exhibit the (strict) $\text{SI}(U_1|U_2)$ property it would be most helpful if this also implies that the merged X-shaped versions have the (strict) $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property. Unfortunately, there is in general no relation between the dependence properties of a copula and the dependence properties of its merged X-shaped version.

Lemma 5.4

Let $C^{\text{MX-D}}$ be the X-shaped version of a copula D. Then the following statements are true.

1. D having the $\text{SI}(U_1|U_2)$ property is not necessary for $C^{\text{MX-D}}$ to exhibit the strict $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property.
2. D having the strict $\text{SI}(U_1|U_2)$ property is not sufficient for $C^{\text{MX-D}}$ to have the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property.
3. D having a TP_2 density it not sufficient for $C^{\text{MX-D}}$ to exhibit the strict $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property.

Proof. See Appendix A.6 ■

[10] If D does not have the PQD property its merged X-shaped version may not resemble the letter X.

Thus, even the strongest positive dependence property, a TP_2 density of D, does neither ensure a strictly increasing conditional variance nor a positive correlation between squared or absolute symmetric random variables.[11] Moreover, the strict $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property of D is not sufficient for a non-negative correlation between squared or absolute symmetric random variables. The proof of Lemma 5.4 also provides a concrete example where the conditional variance is not increasing in this case, i.e., $\exists \overline{y_2} > \underline{y_2} > 0\colon \sigma^2_{Y_1}(|\overline{y_2}|) - \sigma^2_{Y_1}(|\underline{y_2}|) < 0$. On the other hand, a X-shaped version of a copula might have the strict $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$property even if the underlying copula does not exhibit the strict $\mathrm{SI}(U_1|U_2)$ property.[12] The general conclusion from Lemma 5.4 is that we can not choose an appropriate copula D and use Definition and Lemma 5.1 as a general construction principle to obtain a jointly symmetric copula with the strict $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. Instead, one has to check (5.3.5) directly for each individual copula D to obtain a statement about the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. Unfortunately, the direct verification of this condition is by no means straight forward for popular copula families. An important exception is the Gaussian copula.

Proposition 5.8

The merged X-shaped version of the Gaussian copula has the strict $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property if $\rho \neq 0$.

Proof. See Appendix A.7. ■

A different motivation for the merged X-shaped version of the Gaussian copula as a suitable copula to construct a CMP(1) that produces volatility clusters can be given if we assume (standard) normal margins. Let $Z_t \overset{iid}{\sim} U(0,1), I_t = \mathbb{1}_{\{Z_t \leq 0.5\}}$ and $\mathcal{E}_t \overset{iid}{\sim} N(0,1)$ such that $(\mathcal{E}_t, Z_t) \perp Y_{t-1:1}, \mathcal{E}_{t:1} \perp Z_{t:1}$. A stochastic representation is then given by

$$Y_t = (I_t\rho - (1 - I_t)\rho)Y_{t-1} + \sqrt{1-\rho^2}\mathcal{E}_t,$$

which can be interpreted as an AR(1) process with random autoregressive coefficient. It is easy to see that $\mathbb{E}[Y_t|Y_{t-1:1}] = 0$, and

$$\mathbb{V}\mathrm{ar}[Y_t|Y_{t-1:1}] = Y^2_{t-1}\mathbb{V}\mathrm{ar}[2I_t\rho - \rho] + \sqrt{1-\rho^2} = Y^2_{t-1}(2\rho)^2\mathbb{V}\mathrm{ar}[I_t] + \sqrt{1-\rho^2}.$$

Since $\mathbb{V}\mathrm{ar}[I_t] = 0.25$, the next proposition follows.

Proposition 5.9 (Merged X-shaped version of the Gaussian copula as semi-strong ARCH(1) process)

Let Y be a CMP(1) process with standard normal margins and its copula be the merged X-shaped version of the Gaussian copula. There exists a dependent sequence V with

[11] It is an open question whether this property is sufficient for the conditional variance to be increasing or for a non-negative correlation between squared and absolute symmetric random variables.
[12] However, we conjecture that the PQD property is a necessary condition.

$\mathbb{E}[V_t|V_{t-1:1}] = 0$ and $\mathbb{V}\text{ar}[V_t|V_{t-1:1}] = 1$ so that

$$Y_t = \sigma_t V_t,$$
$$\sigma_t^2 = \mathbb{V}\text{ar}[Y_t|Y_{t-1:1}] = \sqrt{1-\rho^2} + \rho^2 Y_{t-1}^2.$$

Note that the ARCH(1) representation is only semi-strong, i.e., V is not an independent sequence. The conditional variance increases quadratically in ρ and we also expect such a super-linear effect of the correlation parameter on the conditional variance if the marginal distribution is not standard normal but a different symmetric distribution. From a different point of view this suggests that a merged X-shaped version of the Gaussian copula with correlation parameter ρ is closer to the product copula than the Gaussian copula with the same correlation parameter. Since the merged X-shaped version of the Gaussian copula is frequently used in the empirical application in Chapter 6 we investigate this matter in more detail now.

By Lemma 5.3 and Lemma 5.2 it follows that any distribution with symmetric margins around μ and the merged X-shaped version of the Gaussian copula has positive correlation in R_1 and R_3 and negative correlation in R_2 and R_4. For each of these sections of the support, the local correlation can be computed by averaging the local correlations of two distributions which have a Gaussian copula with correlation parameter ρ and $-\rho$, respectively. On the other hand, the local correlation in R_i is the correlation of these random variables if the support is truncated to this section of the plane. If the margins are normal distributions one can use the formulas in Rosenbaum (1961) and Gajjar and Subrahmaniam (1978), which derive the correlation of a bivariate normal distribution when both variables are truncated, to obtain a closed-form expression for the local correlations of a distribution with normal margins and a merged X-shaped version of the Gaussian copula. Although there is a simple relation between Spearman's rho and the correlation of the Gaussian copula, there seems to be no closed-form expression for the relation between the correlation parameter of a truncated bivariate normal distribution and Spearman's rho of the corresponding copula.[13] Thus, we have to use numerical integration to approximate the local correlation of the merged X-shaped version of the Gaussian copula in the i-th

[13] Following the lines of the proof of Proposition 5.35 in McNeil et al. (2005) we obtain that Spearman's rho of a copula C, conditional on the event $\{U_1 \le a, U_2 \le a\}$, can be expressed as

$$\begin{aligned}\rho_C(Q_3) &= 3\int_{\mathbb{R}^2} \mathbb{P}\big(X_1 \le \min(x_1, \Phi^{-1}(a)), X_2 \le \min(x_2, \Phi^{-1}(a))\big)\phi(x_1)\phi(x_2)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \mathbb{E}\big[\mathbb{P}(X_1 \le \min(Z_1, \Phi^{-1}(a), X_2 \le \min(Z_2, \Phi^{-1}(a)|Z_1, Z_2))\big] = \mathbb{P}(Y_1(a) \le 0, Y_2(a) \le 0),\end{aligned}$$

where $Y_i(a) = X_i - \min(Z_i, \Phi^{-1}(a)), X_i = \Phi^{-1}(U_i), Z_i \overset{iid}{\sim} N(0,1), i = 1,2, Z_{1:2} \perp X_{1:2}$. To the best of our knowledge, this expression can only be evaluated by numerical methods in general. An exception is the case $a = 1$. Then $(Y_1(1), Y_2(1))$ is normally distributed and we obtain the well known relation between (unconditional) Spearman's rho and ρ.

unit quadrant. Since the local correlations in the second and fourth quadrant are equal to the negative local correlations in the first and third quadrant, we just calculate the local correlation in the third quadrant which is given by

$$\rho^{(3)}_{C^{\text{MX-D}}} = \frac{\mathbb{C}\text{ov}[U_1, U_2 | (U_1, U_2) \in Q_3]}{\mathbb{V}\text{ar}[U_1 | (U_1, U_2) \in Q_3]} = \frac{4 \int_{Q_3} u_1 u_2 c(u_1, u_2) \mathrm{d}u_{1:2} - 1/16}{1/12 - 1/16},$$

see Appendix A.8 for the derivation.

Figure 5.2 displays the relation between the correlation parameter of the merged X-shaped version of the Gaussian copula and the corresponding absolute local correlation in any of the four unit quadrants. We observe an exponential increase in absolute local

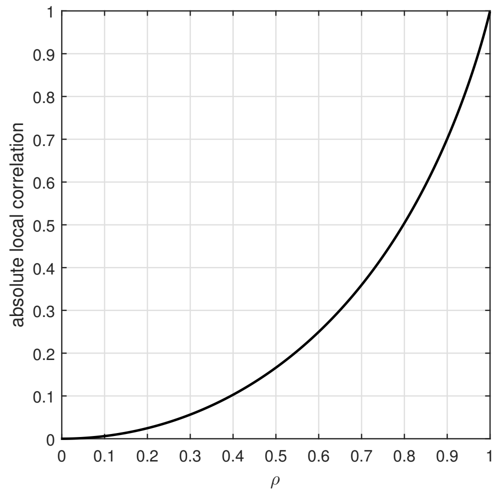


Figure 5.2: Absolute local correlation of the merged X-shaped version of the Gaussian copula in one unit quadrant as a function of ρ.

correlation if ρ is increased. If $\rho = 0.5$, the absolute local correlation is still smaller than 0.167. On the other hand, the absolute local correlation doubles when ρ is increased from 0.8 to 1. When $\rho = 1$, the truncated distribution in the first or third quadrant is the upper Fréchet bound whereas the truncated distribution in the second or fourth quadrant attains the lower Fréchet bound. Finally, we note that the absolute value of local Kendall's tau in any of the four quadrants is non-negative but smaller than the absolute value of local Spearman's rho. This follows from Theorem 5.2.8 in Nelsen (2006), which can be applied since the truncated distribution in any of the four quadrants is a copula which follows from the joint symmetry of the merged X-shaped version of the Gaussian copula.

If we consider merged X-shaped versions of other copulas, e.g., Archimedean copulas or the Student-t copula, it turns out that the analytical verification of the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ or $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property is not straight forward since it involves the verification of non-trivial inequalities. The verification of the SI property of a single copula can sometimes be pretty difficult (see Joe, 1997, p. 49) and it even gets worse for the merged X-shaped version of a copula, see (5.3.5), which is a mixture of four copulas. We investigated the merged X-shaped versions of the Student-t and the Clayton copula in detail but did not succeed in verifying or falsifying the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. However, numerical computations strongly suggest that these merged X-shaped versions have the strict $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. By numerical computations we mean that the conditions are evaluated over a (tight) grid of values of the relevant variables. But even if these copulas would not exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property, their functional form is pretty close to copulas with this property. Thus, they should still induce a strictly increasing conditional variance for many symmetric marginal distributions. For example, Figure 5.3 shows the implied conditional variance $\sigma^2_{Y_1}$ if the copula is the X-shaped version of the Clayton copula and the marginals are Student-t distributions with dof 4. In Appendix A.10 additional density plots of and realizations from various merged X-shaped versions of copulas are depicted and compared.

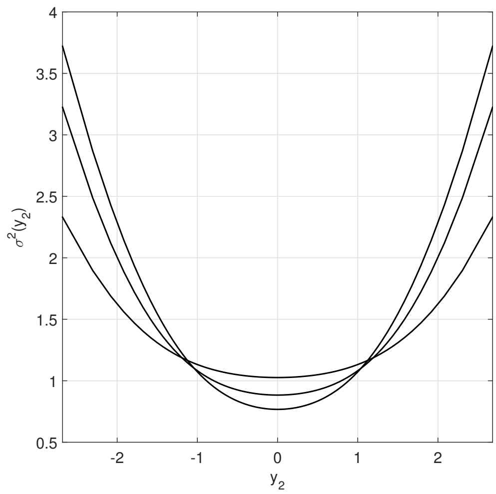


Figure 5.3: Conditional variance $\sigma^2(y_2) = \mathbb{V}\mathrm{ar}[Y_1|Y_2 = y_2]$ if the marginals are Student-t distributions with dof 4 and the copula is the merged X-shaped version of the Clayton copula with $\theta = 0.5, 1, 1.5$.

Patched X-shaped versions of copulas

A more direct approach to generate copulas with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property is introduced in the following. Recall that the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property is a property of the family of cdfs of $U_1|U_2 = u_2$ when they are restricted to the quadrants of the unit cube. Inspired by the work of Durante et al. (2009), we can obtain a copula with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property by placing a re-normalized copula with the appropriate $\mathrm{SI}(U_1|U_2)$ or $\mathrm{SD}(U_1|U_2)$ property on each quadrant of the unit cube. If D_i is a copula with the (strict) $\mathrm{SI}(U_1|U_2)$ property for $i = 1, 3$, and the (strict) $\mathrm{SD}(U_1|U_2)$ property for $i = 2, 4$, then

$$C(u_1, u_2) = \begin{cases} 0.25D_1(2(u_1 - 0.5), 2(u_2 - 0.5)) + 0.5(u_2 + u_1) - 0.25 & \text{if } (u_1, u_2) \in Q_1 \\ 0.25D_2(2(u_1 - 0.5), 2u_2) + 0.5u_2 & \text{if } (u_1, u_2) \in Q_2 \\ 0.25D_3(2u_1, 2u_1) & \text{if } (u_1, u_2) \in Q_3 \\ 0.25D_4(2u_1, 2(u_2 - 0.5)) + 0.5u_1 & \text{if } (u_1, u_2) \in Q_4 \end{cases} \tag{5.3.6}$$

is a copula (see Durante et al., 2009) with the (strict) $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property.[14] Note that C is not necessarily vertically symmetric. However, a jointly symmetric copula can easily be obtained.

Definition and Lemma 5.2 (Patched X-shaped version of a copula)

Let D be a copula with density d and let sgn denote the sign function.
Then $C^{\text{PX-D}}\colon [0,1]^2 \to [0,1]$, with

$$C^{\text{PX-D}}(u_1,u_2) = 0.25\Big(\operatorname{sgn}\Big(\prod_{i=1,2}(2u_i-1)\Big)D(|2u_1-1|,|2u_2-1|) + 2(u_1+u_2)-1\Big),$$

is a radially and jointly symmetric copula. If D is exchangeable so is $C^{\text{PX-D}}$. We call $C^{\text{PX-D}}$ the patched X-shaped version of D and its density is given by

$$d^{\text{PX-D}}(u_1,u_2) = d(|2u_1-1|,|2u_2-1|).$$

Proof. Directly follows by setting $D_1 = D, D_2 = D^{\text{H}}, D_3 = D^{\text{S}}, D_4 = D^{\text{V}}$ in (5.3.6). ■

The following corollary immediately follows since D^{S} has the (strict) $\mathrm{SI}(U_1|U_2)$ property and D^{H} and D^{V} have the (strict) $\mathrm{SD}(U_1|U_2)$ property if D has the (strict) $\mathrm{SI}(U_1|U_2)$ property.

Corollary 5.1 ($\mathbf{SI_{(1,3)}}$-$\mathbf{SD_{(2,4)}}$ property of a patched X-shaped version of a copula)

Let D have the (strict) $\mathrm{SI}(U_1|U_2)$ property. Then its patched X-shaped version has the (strict) $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property.

Patched X-shaped versions of copulas share some properties of the univariate Laplace distribution. The Laplace distribution arises if one patches and accordingly re-normalizes the distributions of X and $-X$, where X is exponentially distributed. Due to this construction, the density of the Laplace distribution is not differentiable at its mean. The same is true for the densities of patched X-shaped versions of copulas whose partial derivatives do not exist at the stitches, i.e., $c(u_1,u_2)$ is not differentiable in one variable at the point 0.5. In addition to that, we do not expect that copula D has lower tail dependence or much probability mass in the lower tail in practical applications. If D has a lower tail dependence coefficient of λ_l, then

$$\lim_{u\to 0.5}\frac{C^{\text{PX-D}}(u,u)}{u} = \lim_{u\searrow 0}\frac{C(u,u)^{\text{PX-}(D^{\text{S}})}}{u} = \lim_{u\searrow 0}\frac{1}{8}\frac{D(2u,2u)}{2u} = \frac{\lambda_l}{8},$$

i.e., there is a very strong clustering of values around the point $(0.5, 0.5)$. On the other

[14] The other way round, if D_i is a copula with the (strict) $\mathrm{SD}(U_1|U_2)$ property for $i = 1,3$, and the (strict) $\mathrm{SI}(U_1|U_2)$ property for $i = 2,4$, we obtain a copula which implies a non-positive correlation between squared or absolute symmetric random variables. Moreover, the conditional variance is maximized at the median of the conditioning variable and decreases in the absolute value of the conditioning variable.

hand, if D has an upper tail dependence coefficient of λ_u, the tail dependence coefficient of C in all corners is $\lambda_u/8$. To exclude a density of $C^{\text{PX-D}}$ that explodes on some points of the area $\{(u_1, u_2) \in [0,1]^2 \colon u_1 = 0.5 \text{ or } u_2 = 0.5\}$, we have to specify a copula for D which does not explode on the area $\{(u_1, u_2) \in [0,1]^2 \colon u_1 = 0 \text{ or } u_2 = 0\}$. Moreover, the copula D should be asymmetric in the sense that the probability mass is by far larger in the upper right corner than in the lower left corner. A suitable candidate is the survival version of the Clayton copula with dependence parameter θ because $c(u_1, 0) = (1+\theta)u_1^{\theta} \leq (1+\theta)$, so that the density of its patched X-shaped version is not larger than $(1+\theta)$ on the points of attachment. Figure 5.4 illustrates the patched X-shaped version of the survival version of

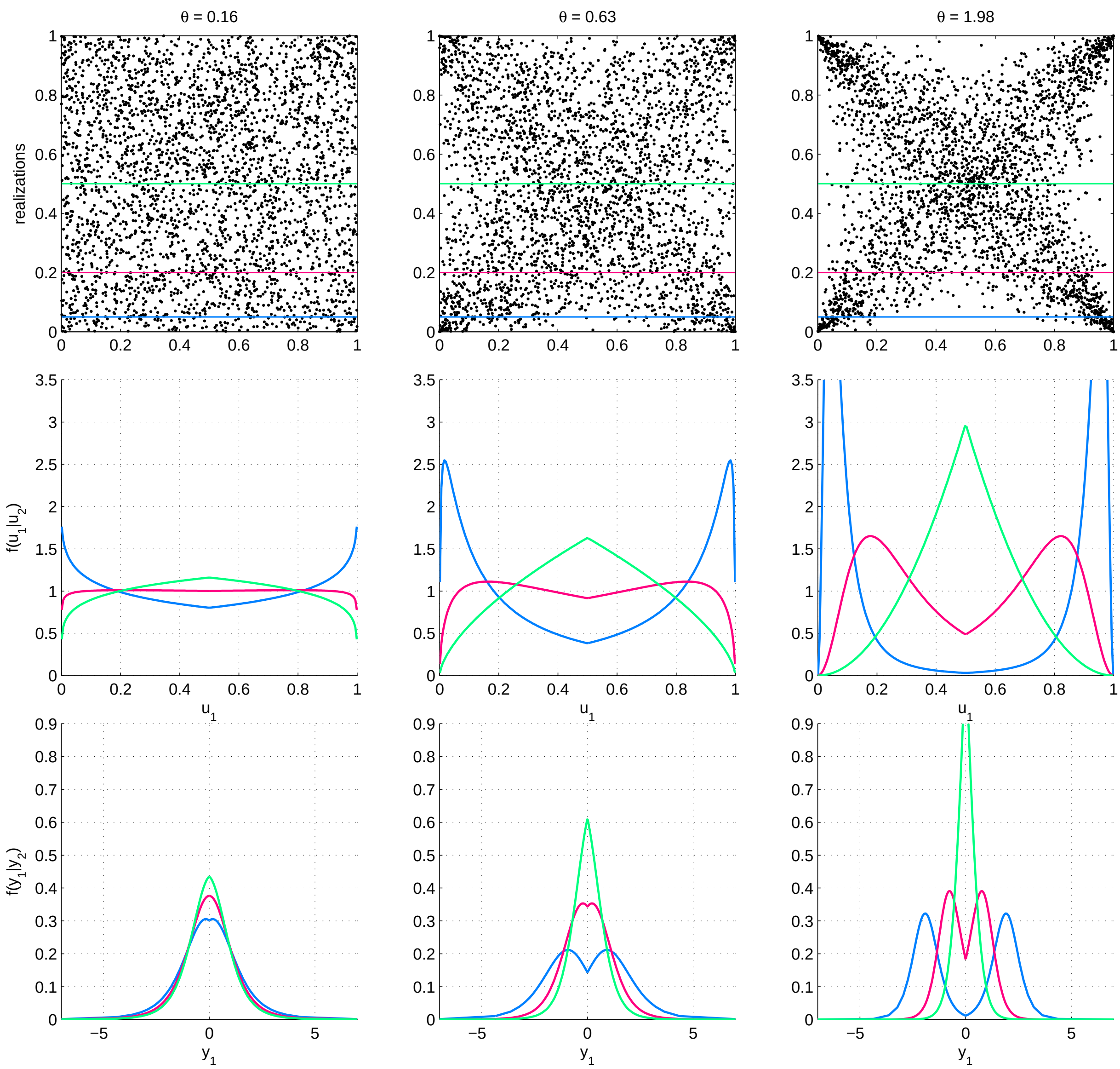


Figure 5.4: Illustration of of the patched X-shaped version of the survival version of the Clayton copula with dependence parameter θ. The value of θ in each column was chosen such that the mutual information of the copula is identical to the mutual information of the merged X-shaped version of the Gaussian copula in the corresponding column of Figure 5.1. The first row displays realizations from the copula whereas the second row depicts slices through the copula density along $u_2 \in \{0.05, 0.2, 0.5\}$. The third row shows the conditional density of $Y_1|Y_2 = y_2$ when y_2 attains its 5%, 20% or 50% quantile and (Y_1, Y_2) has this copula and Student-t margins with dof 4.

the Clayton copula. In the first row of Figure 5.4 realizations from the patched X-shaped versions are shown. For the simulation we use the following stochastic representation. Let $(U_1, U_2) \sim D$ and $\forall i = 1, 2$: $S_i = 2\mathbb{1}_{\{Z_i \leq 0.5\}} - 1, Z_1 \perp Z_2, (Z_1, Z_2) \perp (U_1, U_2)$, i.e. S_i is independently positive or negative with equal probability. Then $(0.5 + 0.5S_1U_1, 0.5 + 0.5S_2U_2) \sim C^{\text{PX-D}}$. The graphical appearance of the scatter is similar to the scatter of the merged X-shaped version of the Gaussian copula in Figure 5.1. Once again, if D approaches the upper (or lower) Fréchet bound, the realizations become more concentrated on the X-like area $\{(u_1, u_2) : u_2 = u_1 \text{ or } u_2 = 1 - u_1\}$. The dependence parameter θ in each column of Figure 5.4 is chosen such that the mutual information of the copula is identical to the mutual information of the merged X-shaped version of the Gaussian copula in the corresponding column of Figure 5.1. Thus, the second and the third row of Figure 5.4 reveal possible differences between the patched X-shaped version of the survival version of the Clayton copula and the merged X-shaped version of the Gaussian copula. Contrary to the merged X-shaped version of the Gaussian copula with identical mutual information, the conditional density of a distribution with a patched X-shaped version of the survival version of the Clayton copula is more peaked if the conditioning variable takes values close to its median. The non-differentiability of the density function in one variable at the median is also clearly visible. Moreover, the occurrence of a bimodal conditional density seems to be more likely. In Appendix A.10 density plots of various patched X-shaped versions of copulas and implied conditional variances are depicted and analyzed in detail.

Besides their simple sufficient condition for the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property, patched X-shaped versions of copulas give rise to a closed form expression of their conditional quantile function, provided the conditional quantile function of D has a closed form. Noting that $C^{\text{PX-D}}(u_1|u_2) > 0.5 \Leftrightarrow 2u_1 - 1 > 0$ and

$$C^{\text{PX-D}}(u_1|u_2) = 0.5\big(1 + \text{sgn}(2u_1 - 1)D(|2u_1 - 1|\big|(2u_2 - 1))\big),$$

it is easy to show that the conditional quantile function is given as

$$(C^{\text{PX-D}})^{-1}(z|u_2) = 0.5\big(1 + \text{sgn}(2z - 1)D^{-1}(|2z - 1| \;\big|\; |2u_2 - 1|)\big).$$

This is different from merged X-shaped versions of copulas where the conditional quantile function has to be computed by a numerical inversion of the conditional cdf.

5.3.4 Copulas that allow for an asymmetric shape of the conditional variance

The copulas we have considered so far exhibit very strong symmetry properties. They are jointly symmetric, which implies that they are horizontally symmetric, so that the conditional variance only depends on the absolute value of the conditioning variable but not on the sign. However, it has been found that, especially for equity markets, the conditional variance often increases more due to a negative lagged return than due to a positive lagged return. Because of that, we now examine whether it is possible to construct copulas that reproduce the leverage effect in the conditional variance. For this to be the case, the copula has to exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property, and, as a consequence, the conditional median of $U_2|U_1 = u_1$ must equal the unconditional median for all $u_1 \in (0,1)$, which is a strong restriction on the shape of the copula. Thus, one might wonder whether there even exists a copula with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property that is not horizontally symmetric. In other words, can a conditional variance that is strictly increasing in the absolute value of the conditioning variable, but not symmetric in the sign of the variable, be a property of a copula? The following section gives an affirmative answer but also shows that the construction of such copulas that are usable for applied work is difficult.

Extraction of a GJR-ARCH(1)-like copula

In the previous sections we have used certain operations on established copula families to construct vertically symmetric copulas with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. Alternatively, one can employ Sklar's theorem to extract vertically symmetric copulas with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property directly from bivariate distributions with the desired properties. That is, given a bivariate distribution function F_{12} we obtain a copula via

$$C(u_1, u_2) = F_{12}(F_1^{-1}(u_1), F_2^{-1}(u_2)). \tag{5.3.7}$$

C is vertically symmetric if $(\mu - Y_1, Y_2) \overset{d}{=} (\mu + Y_1, Y_2)$ and has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property if the margins of (Y_1, Y_2) are symmetric around μ and

$$\forall (y_1, y_2) \in R_1 \cup R_3\colon\ F_{1|2}(y_1|y_2) \text{ is strictly decreasing in } y_2,$$
$$\forall (y_1, y_2) \in R_2 \cup R_4\colon\ F_{1|2}(y_1|y_2) \text{ is strictly increasing in } y_2,$$

where R_i is defined as in Lemma 5.2. These conditions are satisfied by the distribution of two consecutive random variables of the stationary strong GJR-ARCH(1) process with symmetric innovations. The popular GJR-ARCH model is an ARCH variant that includes leverage terms for modeling asymmetric volatility clustering and was introduced by Glosten et al. (1993). The strong GJR-ARCH(1) model with independent innovations $\mathcal{E}_t$ is given by

$$Y_t = \sqrt{\omega + aY_{t-1}^2 + \mathbb{1}_{\{Y_{t-1}<0\}}\gamma Y_{t-1}^2}\,\mathcal{E}_t, \quad \omega > 0,\ a \geq 0,\ a + \gamma \geq 0,\ \mathbb{E}[\mathcal{E}_t] = 0,\ \mathbb{V}\mathrm{ar}[\mathcal{E}_t] = 1,$$

where γ is the leverage effect. If $\gamma > 0$, then the increase in the conditional variance is larger for negative than for positive lagged values. Under further constraints on the parameters and the innovation distribution, the GJR-ARCH(1) process is stationary (Lindner, 2009). The marginal distribution of a stationary GJR-ARCH(1) process is the distribution of

$$Y_t = \omega \sum_{i=0}^{\infty} \left(\prod_{j=0}^{i-1} A_{t-j} \right), \quad A_t := (a + \mathbb{1}_{\{\mathcal{E}_{t-1}<0\}}\gamma)\mathcal{E}_{t-1}^2.$$

As with almost any stationary ARCH(1) process, the actual computation of the marginal distribution is infeasible and can only be approximated by simulation. Therefore, the extraction of the GJR-ARCH(1) copula is not possible. However, we can extract a GJR-ARCH(1)-like copula by leaving the time series context and considering the following data generating process. Let X and $\mathcal{E}$ be independent random variables, each with zero mean and unit variance. Define

$$Y = \sqrt{\omega + aX^2 + \mathbb{1}_{\{X<0\}}\gamma X^2}\,\mathcal{E}, \quad \omega > 0,\ a \geq 0,\ a + \gamma \geq 0.$$

The marginal distribution of Y can be computed using

$$F_Y(y) = \int_{-\infty}^{\infty} F_{\mathcal{E}} \left(\frac{y}{\sqrt{\omega + ax^2 + \mathbb{1}_{\{x<0\}}\gamma x^2}} \right) f_X(x)\mathrm{d}x. \tag{5.3.8}$$

Let $\theta := (\omega, \alpha, \gamma)$. A GJR-ARCH(1)-like copula can then be extracted via

$$\begin{aligned} C(u_1, u_2; \mathcal{E}, X, \theta) &= F_{YX}(F_Y^{-1}(u_1), F_X^{-1}(u_2)) \\ &= \int_0^{F_X^{-1}(u_2)} F_{\mathcal{E}} \left(\frac{F_Y^{-1}(u_1)}{\sqrt{\omega + ax^2 + \mathbb{1}_{\{x<0\}}\gamma x^2}} \right) f_X(x)\mathrm{d}x. \end{aligned}$$

It is easy to check that if $\mathcal{E}$ has a symmetric distribution, the GJR-ARCH(1)-like copula is vertically symmetric.[15] Obviously,

$$\begin{aligned} &\forall (y, x) \in R_1 \cup R_3\colon F_{Y|X}(y|x) \text{ is strictly decreasing in } x, \\ &\forall (y, x) \in R_2 \cup R_4\colon F_{Y|X}(y|x) \text{ is strictly increasing in } x, \end{aligned}$$

[15] If $\mathcal{E}$ is symmetric around zero then Y is symmetric around zero. If Y is symmetric around zero then $F_Y^{-1}(1 - u_1) = -F_Y^{-1}(u_1)$ and

$$\begin{aligned} u_2 - C(1 - u_1, u_2) &= F_X^{-1}(u_2) - \int_0^{F_X^{-1}(u_2)} F_{\mathcal{E}} \left(\frac{F_Y^{-1}(1 - u_1)}{\sqrt{\omega + ax^2 + \mathbb{1}_{\{x<0\}}\gamma x^2}} \right) f_X(x)\mathrm{d}x \\ &= F_X^{-1}(u_2) - \int_0^{F_X^{-1}(u_2)} F_{\mathcal{E}} \left(\frac{-F_Y^{-1}(u_1)}{\sqrt{\omega + ax^2 + \mathbb{1}_{\{x<0\}}\gamma x^2}} \right) f_X(x)\mathrm{d}x \\ &= F_X^{-1}(u_2) - \int_0^{F_X^{-1}(u_2)} \left(1 - F_{\mathcal{E}} \left(\frac{F_Y^{-1}(u_1)}{\sqrt{\omega + ax^2 + \mathbb{1}_{\{x<0\}}\gamma x^2}} \right) \right) f_X(x)\mathrm{d}x \\ &= C(u_1, u_2). \end{aligned}$$

which implies that the GJR-ARCH(1)-like copula has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property if $\mathcal{E}$ has a symmetric distribution and the median of X is zero. Note that if $\mathcal{E}$ has an asymmetric distribution then C is not vertically symmetric. Moreover, if either the median of $\mathcal{E}$ or X is not zero then C does not have the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. In the following, let $\mathcal{E}$ and X have symmetric margins around zero. If (Z_1, Z_2) has symmetric margins around zero and a GJR-ARCH(1)-like copula $C(\mathcal{E}, X, \theta)$ with $\gamma > 0$, the implied conditional variance exhibits a leverage effect in the sense that the conditional variance is larger for negative than for positive values, i.e.,

$$\forall z_2 > 0\colon \sigma^2_{Z_1}(-z_2) > \sigma^2_{Z_1}(z_2) > \sigma^2_{Z_1}(0).$$

This follows since

$$\forall (y, x) \in R_3\colon F_{Y|X}(y|x) > F_{Y|X}(y|-x)$$

and

$$F_X^{-1}(F_{Z_2}(z_2)) = F_X^{-1}(1 - F_{Z_2}(-z_2)) = -F_X^{-1}(F_{Z_2}(-z_2)),$$

so that for all $(z_1, z_2) \in R_3$:

$$\begin{aligned} C(F_{Z_1}(z_1)|F_{Z_2}(z_2)) &= F_{Y|X}\big(F_Y^{-1}(F_{Z_1}(z_1))\big|F_X^{-1}(F_{Z_2}(z_2))\big) \\ &= F_{Y|X}\big(F_Y^{-1}(F_{Z_1}(z_1))\big|-F_X^{-1}(F_{Z_2}(-z_2))\big) \\ &> F_{Y|X}(F_Y^{-1}(F_{Z_1}(z_1))\big|F_X^{-1}(F_{Z_2}(-z_2))\big) = C\big(F_{Z_1}(z_1)\big|F_{Z_2}(-z_2)\big), \end{aligned}$$

where the inequality follows because $F_{Z_2}(-z_2) > 0.5$, so that $\big(F_Y^{-1}(F_{Z_1}(z_1)), -F_X^{-1}(F_{Z_2}(-z_2))\big)$ $\in R_3$. Consequently,

$$\forall z_2 < 0\colon \sigma^2_{Z_1}(z_2) = -2\int_{-\infty}^{0} z_1 C(F_{Z_1}(z_1)|F_{Z_2}(z_2))\mathrm{d}z_1 > \sigma^2_{Z_1}(-z_2).$$

Note that C is always a vertically symmetric copula but C is jointly symmetric if and only if there is no leverage effect. We summarize these results in the following proposition.

Proposition 5.10 (GJR-ARCH(1)-like copula with the $\mathbf{SI}_{(1,3)}$-$\mathbf{SD}_{(2,4)}$ property and leverage)

Let X and $\mathcal{E}$ be independent random variables with symmetric distributions around zero and unit variances, and

$$Y = \sqrt{\omega + aX^2 + \mathbb{1}_{\{X<0\}}\gamma X^2}\mathcal{E}, \quad \omega > 0,\ a \geq 0,\ a + \gamma \geq 0.$$

Then the copula C of (Y, X) is vertically symmetric and has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. C is jointly symmetric if and only if $\gamma = 0$. If (Z_1, Z_2) has symmetric margins around zero and copula C and $\gamma > 0$ then

$$\forall z_2 > 0\colon \sigma^2_{Z_1}(-z_2) > \sigma^2_{Z_1}(z_2) > \sigma^2_{Z_1}(0).$$

Although a GJR-ARCH(1)-like copula exhibits nice and tractable properties, its use for applied work is limited. To extract the copula via (5.3.7) we have to compute the quantile function of Y which is unknown. Instead, the quantile function of Y must be computed by inverting the marginal distribution of Y given in (5.3.8) which is a one-dimensional integral. This integral has in general no closed form solution and has to be evaluated by numerical methods. Thus, the evaluation of the quantile function of Y and consequently the computation of the copula is computationally very expensive. The evaluation of the copula density, which is required for ML estimation, is computationally even more complex since also the density of Y must be computed by numerically approximating a one-dimension integral. To render ML estimation of the GJR-ARCH(1)-like copula feasible one should specify $F_{\mathcal{E}}$ and f_X such that the marginal distribution and density of Y can be rapidly evaluated. For instance, one can specify the logistic distribution for $\mathcal{E}$ and X.[16] But even if the density of $\mathcal{E}$ and the distribution of X have a simple form, the evaluation time of the copula density is still excessively long for moderate sample sizes.[17] Note that this is not a particular problem of the GJR-ARCH(1)-like copula. In general, it is very difficult, or even impossible, to construct a vertically but not horizontally symmetric bivariate distribution that exhibits the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property and a closed form expression for both marginal distributions.

For this reason and also because the fit of other copulas might be superior, we investigate in the following alternatives to the GJR-ARCH(1)-like copula which are vertically but not jointly symmetric so that the implied conditional variance depends not only on the absolute value but also on the sign of the conditioning variable. It turns out that the construction of a copula C with the following properties is a quite intricate problem:

1) C is vertically but not jointly symmetric.

2) C has the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ or $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property.

3) C has a continuous density.

4) C is not extracted from a bivariate distribution with the desired properties as it is the case for the GJR-ARCH(1)-like copula.

[16] If $\mathcal{E}$ and X have logistic margins then simulations of the GJR-ARCH(1)-like copula show a strong similarity to the copula of the GJR-ARCH(1) process if the innovation distribution is Student-t with dof 4.

[17] If the sample size is 5000 and $\mathcal{E}$ and X have logistic distributions, then the fitting of a GJR-ARCH(1)-like copula takes about 400 times longer than the estimation of a merged X-shaped version of the Gaussian copula which is also rather computationally expensive.

In fact, the copulas we have proposed so far and introduce in the following, violate exactly one of these four conditions. Except for one case, all presented copula constructions do not exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property in general but satisfy the other conditions. The main difficulty is that for the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ or $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property it is necessary that, for all $i, j = 1, 2,\ j \neq i$, the median of $U_i|U_j = u_j$ must be equal to 0.5 for all $u_j \in (0, 1)$ (see Lemma 5.3). These conditions are automatically satisfied for jointly symmetric copulas. But if the copula is only vertically but not horizontally symmetric, then the necessary condition that the median of $U_2|U_1 = u_1$ equals 0.5 for all $u_1 \in (0, 1)$ is not automatically satisfied. This necessary condition is satisfied for the GJR-ARCH(1)-like copula with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property, but it appears to be very difficult to fulfill this condition if the copula has to satisfy condition 4). For that reason, most vertically symmetric copulas that are not jointly symmetric fail to satisfy the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. If a distribution has symmetric marginal distributions around zero and such a copula, it can happen that the conditional variance is not minimized if the value of the conditioning variable is zero. Thus, the conditional variance can not be strictly increasing (decreasing) for all positive (negative) values of the conditioning variable. It might even happen that the conditional variance is not strictly increasing (decreasing) for values that are located to the right (left) of the value where the minimal variance is attained. However, the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property is only a sufficient but not necessary condition for a "well-behaved" conditional variance. We demonstrate in the following that for reasonable symmetric marginal distributions, copula families and amount of leverage, the minimum of the implied conditional variance is approximately zero and that the conditional variance is strictly increasing (decreasing) for values that are larger (smaller) than the value at which the minimal variance is attained. Thus, the violation of the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property for the proposed copulas seems not to be relevant for modeling financial returns, especially if the copula is close to a jointly symmetric copula with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property.[18]

Merging a non-radially symmetric copula and its vertically reflected counterpart

In Definition and Lemma 5.1 we have constructed a jointly symmetric copula by merging a radially symmetric copula with its vertically reflected counterpart. To obtain a vertically symmetric copula which is not horizontally symmetric, we can simply merge a copula, which is not radially symmetric, with its vertically reflected counterpart. That is, if $(U_1, U_2) \sim D$ and $(1 - U_1, 1 - U_2) \overset{d}{\neq} (1 - U_1, 1 - U_2)$, then the copula C, which is defined by

$$C(u_1, u_2) = 0.5D(u_1, u_2) + 0.5(u_2 - D(1 - u_1, u_2)), \tag{5.3.9}$$

[18] In this regard, we point to the EGARCH(1,1) model of Nelson (1991). If the leverage effect is very strong, the conditional variance is not decreasing for negative values of the lagged innovation. In particular, the conditional variance first increases but eventually decreases if the lagged innovation becomes more negative. However, such strong leverage and extreme negative values of the innovation are typically not observed in practice so that this is not an issue in applications.

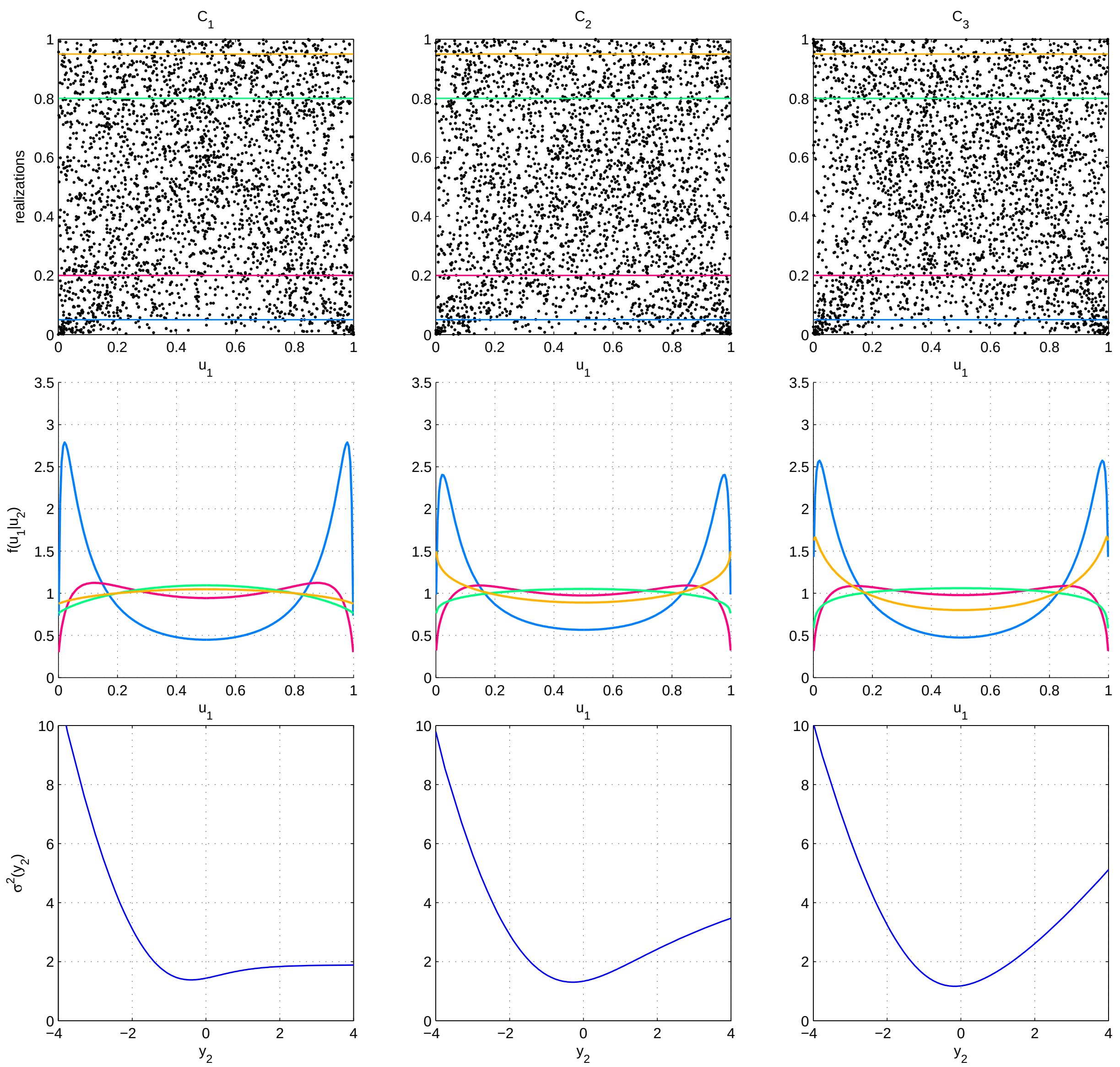


Figure 5.5: Illustration of copulas that are constructed via (5.3.9). C_1 is based on the Clayton copula with $\theta = 0.82$. C_2 originates from the survival version of the Gumbel copula with $\theta = 1.4$. C_3 is based on an equally weighted mixture of the Gaussian copula with $\theta = 0.61$ and the Clayton copula with $\theta = 0.93$. The mutual information of all copulas is 0.05 to render the plots comparable. The first row displays realizations from the copula whereas the second row depicts slices through the copula density along $u_2 \in \{0.05, 0.2, 0.8, 0.95\}$. The third row shows the variance of $Y_1|Y_2 = y_2$ when (Y_1, Y_2) has this copula and Student-t margins with dof 4.

is vertically but not jointly symmetric. Similar to the case when D is a radially symmetric copula, we expect that extreme values of one random variable are associated with extreme values of the other random variable. Moreover, the probability that both random variables take values near their medians should be rather large if a distribution has the copula in (5.3.9) and D has the PQD property. However, the implied conditional variance will not only depend on the absolute value but also on the sign of the conditioning variable since the copula is not horizontally symmetric. Figure 5.5 illustrates some resulting copulas if D is a non-radially symmetric copula. For identical mutual information, we can generate a variety of asymmetric functional shapes of the conditional variance if a distribution has such a copula and Student-t margins with dof 4. The conditional variance is not minimized at zero but at a negative value $y_2^\star$, which implies that none of these copulas has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property.[19] However, $y_2^\star$ is close to zero and the conditional variance is strictly increasing (decreasing) for values that are larger (smaller) than $y_2^\star$. Similar results hold for other symmetric marginal distributions and copulas that are based on (5.3.9) with D having the PQD property. Thus, if D has the PQD property and the amount of radial non-symmetry is not too strong, the construction in (5.3.9) often yields a vertically symmetric copula that exhibits a mild degree of horizontal non-symmetry and a "well-behaved" conditional variance.

Patching different copulas

Let E_1 be a copula with the strict $\mathrm{SI}(U_1|U_2)$ property and E_2 a copula with the strict $\mathrm{SD}(U_1|U_2)$ property. Further assume that $E_1(u_1, u_2) \neq u_1 - E_2(u_1, 1 - u_2)$. If we set $D_1 = E_1, D_4 = E_1^{\mathrm{H}}$, $D_2 = E_2$, and $D_3 = E_2^{\mathrm{H}}$ in (5.3.6), we obtain a vertically symmetric copula C that is not horizontally symmetric and has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. However, the density of C is in general not continuous (a.s.). For instance, the function $g(u_1) = c(u_1, u_2)$ is not continuous at $u_1 = 0.5$ for fixed $u_2 \in (0, 1)$. A necessary condition so that the density of C is continuous is that $\forall u_1 \in [0, 1], u_2 = 0\colon e_1(u_1, u_2) = e_2(1 - u_1, u_2)$ and $\forall u_2 \in [0, 1], u_1 = 0\colon e_1(u_1, u_2) = e_2(u_1, 1 - u_2)$, which is hard to satisfy if $E_1(u_1, u_2) \neq u_1 - E_2(u_1, 1 - u_2)$. Empirical investigations confirm the expectation that a copula with such a non-continuous density is not useful for practical applications, so that we do not consider this approach in the following.

Convex combinations with location dependent coefficients

The next construction can be used to obtain a vertically but not horizontally symmetric copula from a jointly symmetric copula. A general way to obtain an asymmetric density is to take the convex combination of two densities and let the coefficients (of the convex combination) vary with the values at which the densities of the components are evaluated. If the two densities of the convex combinations are copula densities then the resulting density is in general not a copula density. However, the following proposition shows that a convex combination of two vertically symmetric copula densities is a copula density even if the coefficients depend on one variable.

---

[19] Indeed, if D belongs to the family of Clayton copulas with dependence parameter θ, then $C(0.2, 0.5) \neq 0.1$, which is a necessary condition for the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property. On the other side, for many values of θ it appears that either we have $\mathrm{SI}(U_1|U_2)$ in $Q_1 \cup Q_3$ or $\mathrm{SD}(U_1|U_2)$ in $Q_2 \cup Q_4$, but never both conditions simultaneously. In terms of D, a necessary condition is that for all $u_1 \in (0, 1)$ we have that $D(u_1, 0.5) = u_1 - 0.5 + D(1 - u_1, 0.5)$ or, equivalently, $D(u_1, 0.5) = P(U_1 \geq u_1, U_2 \geq 0.5)$.

Proposition 5.11 (Convex combination of vertically symmetric copulas with varying coefficients)

For $i = 1, 2$, let D_i be a vertically symmetric copula with density d_i and $g\colon [0, 1] \to [0, 1]$ be a continuous function with the property $\exists a \geq 0, \forall u_2 \in [0, 0.5]\colon g(0.5+u_2) = a - g(0.5-u_2)$. Then

$$C(u_1, u_2) = \int_0^{u_2} \big[g(t_2)D_1(u_1|t_2) + (1 - g(t_2))D_2(u_1|t_2)\big]\mathrm{d}t_2,$$

with density

$$c(u_1, u_2) = g(u_2)d_1(u_1, u_2) + (1 - g(u_2))d_2(u_1, u_2),$$

is a vertically symmetric copula. If D_i is jointly symmetric for $i = 1, 2$, then C is horizontally symmetric only if g is a constant function.

Proof. See Appendix A.9. ■

Setting $g(u_2) = w \in [0, 1]$ yields an ordinary mixture copula of two vertically symmetric copulas. However, due to the fact that the conditional mean of $U_1|U_2 = u_2$ is constant if the underlying copula is vertically symmetric, the weights can also vary with the location of u_2 so that C is still a copula. As a result, if D_1 and D_2 are jointly symmetric copulas, the resulting copula C in Proposition 5.11 is only jointly symmetric if g is a constant function. If we set $a = 1$, then any cdf of a symmetric random variable with support on the unit interval can be specified as function g in order to obtain a vertically symmetric copula which is not jointly symmetric. The motivation for the construction given in Proposition 5.11 can be explained as follows. Consider the ordinary mixture copula with density given by

$$\tilde{c}(u_1, u_2) = \mathbb{E}[g(U_2)]d_1(u_1, u_2) + (1 - \mathbb{E}[g(U_2)])d_2(u_1, u_2),$$

where D_1 and D_2 are jointly symmetric. If g is an increasing function of u_2, then the density of C in Proposition 5.11 approaches the density of D_2 if u_2 goes to zero. If the conditional density $d_2(u_1|u_2)$ of D_2 has more probability mass in the tails than the corresponding conditional density $d_1(u_1|u_2)$ of D_1 when u_2 is near zero, we expect that the conditional density $c(u_1|u_2)$ has more probability mass in the tails than $\tilde{c}(u_1|u_2)$, which is the density of a jointly symmetric copula. If u_2 is close to one, the conditional density $d_2(u_1|u_2)$ makes the major contribution to the conditional density $c(u_1|u_2)$ which then should have less probability mass in the tails than $\tilde{c}(u_1|u_2)$. Consequently, we expect that $c(u_1|u_2)$ has more probability mass in the tails than $c(u_1|1 - u_2)$ if u_2 is near zero. Thus, for a given absolute value of the conditioning variable, the conditional variance of a distribution with this copula should be larger if the sign of the conditioning variable was negative. A more rigorous statement about the functional form of the conditional variance does not seem to be possible. In general, the functional form of the conditional variance depends on the

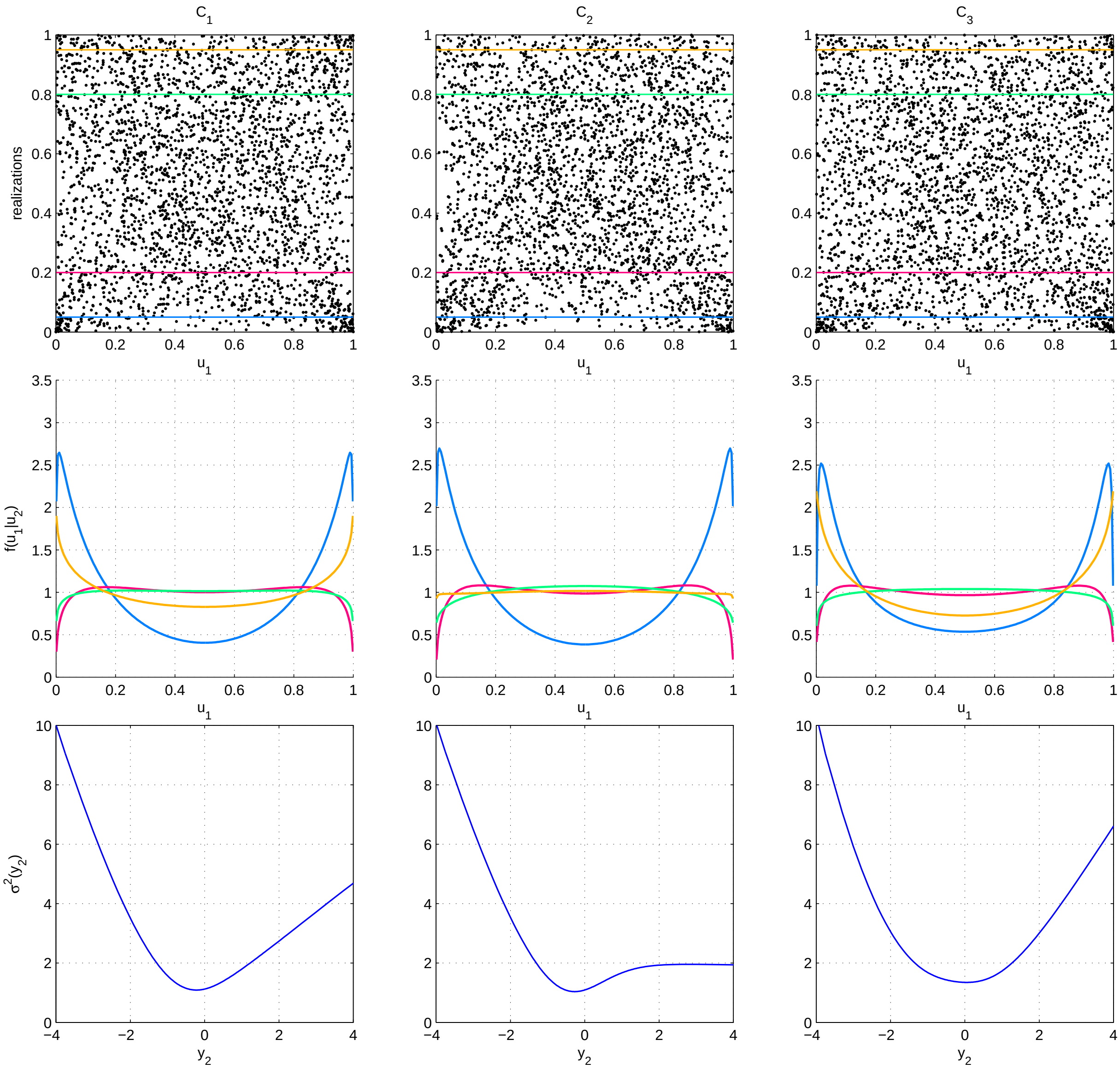


Figure 5.6: Illustration of copulas that are constructed via Proposition 5.11. In all cases $g(u_2) = u_2$. C_1 is based on $D_i, i = 1, 2$, which is the merged X-shaped version of the Gaussian copula with $\theta_1 = 0.39$ and $\theta_2 = 0.7$. C_2 originates from D_1 which is the merged X-shaped version of the survival version of the Clayton copula with $\theta = 0.8$ and from D_2 which is the merged X-shaped version of the Gaussian copula with $\theta = 0.7$. C_3 is based on D_1 which is the merged X-shaped version of the Gaussian copula with $\theta = 0.52$ and D_2 which is the merged X-shaped version of the survival version of the Clayton copula with $\theta = 0.7$. The mutual information of all copulas is 0.05 to render the plots comparable. The first row displays realizations from the copula whereas the second row depicts slices through the copula density along $u_2 \in \{0.05, 0.2, 0.8, 0.95\}$. The third row shows the variance of $Y_1|Y_2 = y_2$ when (Y_1, Y_2) has this copula and Student-t margins with dof 4.

particular chosen copulas D_1 and D_2 and the marginal distribution. Figure 5.6 illustrates some copulas (C_1, C_2, C_3) that originate from the construction in Proposition 5.11. We observe that for C_1 and C_2 the conditional variance is minimized for negative values of y_2, whereas for C_3 the minimum of the variance is attained for a positive value of y_2. Thus, all copulas do not have the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property.[20] Nevertheless, in all cases, the minimum of the variance is attained at a value close to zero which we denote by $y_2^\star$. Moreover,

[20] It is very unlikely that the necessary condition for the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property mentioned in Lemma 5.3 can be satisfied for copulas defined in Proposition 5.11 if g is not a constant function.

the conditional variance seems to be strictly increasing (decreasing) for values that are located to the right (left) of $y_2^\star$. However, this is not strictly true for C_2. Although it is hardly visible in the plot, the implied conditional variance for C_2 slightly decreases when y_2 becomes larger than its 98% quantile.[21] Altogether, the construction in Proposition 5.11 seems to be adequate for modeling the leverage effect in practical applications.

5.4 Copulas for modeling volatility clusters: The higher-order Markov case

In empirical work, the estimated autocorrelation function of squared or absolute financial returns is slowly decaying which is often taken as evidence that the underlying process is not first-order Markov but has a longer memory. On these grounds, we now investigate to what extent we can construct a copula-based higher-order Markov process that exhibits volatility clustering. We are especially interested in the generalization of Proposition 5.4 that provides, for a conditionally symmetric first-order Markov process, sufficient conditions for a non-negative autocorrelation function of the squared or absolute random variables and a strictly increasing conditional variance.

5.4.1 General remarks

For higher-order Markov processes, the autocovariance function of squared or absolute symmetric random variables is non-negative if all bivariate copulas $C_{t,t+j}, j \in \mathbb{N}$, have the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property. In general, the transition variance is given by

$$\sigma^2(y_{t-1:1}) := \mathbb{V}\mathrm{ar}[Y_t|Y_{t-1:1} = y_{t-1:1}] = 2\int_{\mathbb{R}} y_t\big(\mathbb{1}_{\{y_t\geq 0\}} - F_{t|t-1:1}(y_t|y_{t-1:1})\big)\mathrm{d}y_t.$$

Volatility clustering occurs when $\sigma^2(y_{t-1:1})$ is large (small) when the absolute values of lagged realizations are large (small). This is obviously true if $\sigma^2(y_{t-1:1})$ is (strictly) increasing in the absolute value of each lagged variable. Provided Y has a symmetric margin, this is the case if, for all $t \geq 2$,

$$\begin{aligned}&\forall u_{t:1} \in (0.5,1)^t \cup (0,0.5)^t:\\ &F_{U_t|U_{t-1:1}}(u_t|u_{t-1:1}) \text{ is (strictly) decreasing in each element of } u_{t-1:1},\end{aligned}$$

[21] We have investigated this issue in more detail. It seems to be rarely the case that the conditional variance is not strictly increasing for $y_2 > y_2^\star$. When this is the case, the conditional variance only decreases for extreme values of y_2 and the decrease is very mild. If the marginal distribution is too platykurtic, i.e., it has negative excess kurtosis, the likelihood that the conditional variance is not strictly increasing for $y_2 > y_2^\star$ increases. For instance, if $D_i, i = 1, 2$, is the X-shaped version of the Gaussian copula with correlation parameter $\theta_i = \rho_i$ and the margins are uniformly distributed, then the conditional variance may have a second (local) minima near $u_2 = 1$ or it may be decreasing in u_2 for values of u_2 that are close to 1 if $\rho_2 - \rho_1$ is very large and ρ_1 is rather small.

and

$$\forall u_{t:1} \in (0.5,1)\times(0,0.5)^{t-1} \cup (0,0.5)\times(0.5,1)^{t-1}:$$
$$F_{U_t|U_{t-1:1}}(u_t|u_{t-1:1}) \text{ is (strictly) increasing in each element of } u_{t-1:1},$$

where $U_t = F_t(Y_t)$.[22] A sufficient condition for a non-negative autocorrelation function of squared or absolute symmetric random variables and an increasing transition variance in the absolute value of each lagged variable is given next.

Definition 5.7 (CIS$_{(1,3)}$-CDS$_{(2,4)}$ property)

Let $U_{1:K} \sim C_{1:K}$. A copula $C_{1:K}$ has the (strict) CIS$_{(1,3)}$-CDS$_{(2,4)}(U_{K:1})$ property if for all $i = 2,\dots,K, j = 1,\dots,i-1$,

$$\forall(u_i,u_j) \in Q_1 \cup Q_3\colon F_{U_i|U_{1:i-1}}(u_i|u_{1:i-1}) \text{ is (strictly) decreasing in } u_j,$$

and

$$\forall(u_i,u_j) \in Q_2 \cup Q_4\colon F_{U_i|U_{1:i-1}}(u_i|u_{1:i-1}) \text{ is (strictly) increasing in } u_j.$$

Since the CIS$_{(1,3)}$-CDS$_{(2,4)}$ property is a natural adaption of the CIS property to the case of martingale difference sequences the following proposition follows.

Proposition 5.12 (Non-negative autocorrelation and increasing transition variance)

Let Y be a p-th order Markov process with a symmetric margin and $j = 1,2$. If the copula $C_{1:p+1}$ has the CIS$_{(1,3)}$-CDS$_{(2,4)}(U_{p+1:1})$ property then, for all $i \in \mathbb{N}$, $\mathbb{C}\text{ov}[|Y_i|^j, |Y_1|^j] \geq 0$, and $\sigma^2(y_{t-1:1})$ is (strictly) increasing in the absolute value of each lagged variable.

Proof. Follows from similar arguments as in Corollary 3.2 and Lemma 3.1. ■

The following generalization of a GJR-ARCH-like copula to the (p+1)-dimensional case has the CIS$_{(1,3)}$-CDS$_{(2,4)}$ property. For all $i = 1,\dots,p$, let $\mathcal{E}_i$ and Y_1 have symmetric marginal distributions. If we define, for $i = 1,\dots,p+1$,

$$Y_i = \sqrt{\omega_i + \sum_{k=1}^{i-1}(a_{ki} + \mathbb{1}_{\{Y_k \leq 0\}}\gamma_{ki})Y_k^2}\,\mathcal{E}_i,$$

where

$$\mathcal{E}_i \perp Y_{1:i-1},\ \mathbb{E}[\mathcal{E}_i] = 0,\ \mathbb{V}\text{ar}[\mathcal{E}_i] = 1,\ \omega_i > 0,\ a_{ki} \geq 0,\ \gamma_{ki} \geq 0,$$

then the copula of $Y_{1:p+1}$ obviously exhibits the CIS$_{(1,3)}$-CDS$_{(2,4)}(U_{p+1:1})$ property. However, this GJR-ARCH(p)-like copula can only be applied for very small p since the required

[22] It can be easily checked that this strong condition is satisfied by every ARCH(p) process with iid innovations that are symmetric around zero, i.e., $Y_t = \sqrt{\omega + \sum_{i=1}^{p} a_i Y_{t-1}^2}\,\mathcal{E}_t$, such that $a_i \geq 0$, for all $i = 1,\dots,p$.

marginal distribution of Y_i is a $(i-1)$-dimensional integral for all $i = 2, \ldots, p$. Another copula that exhibits the $\text{CIS}_{(1,3)}$-$\text{CDS}_{(2,4)}$ property is the multivariate Student-t copula with unit correlation matrix. For instance, if $X_{1:3}$ is a trivariate Student-t distribution with unit covariance matrix and dof ν, then

$$F_{3|12}(x_3|x_{1:2}) = T_{\nu+2}\left(\frac{x_3}{\sqrt{\frac{\nu}{\nu+2}}\sqrt{1+\frac{1}{\nu}(x_2^2+x_1^2)}}\right).$$

It is easy to check that $F_{3|12}(x_3|x_{1:2})$ is strictly decreasing (or increasing) in $x_i, i = 1, 2$, if $(x_3, x_i) \in Q_1 \cup Q_3$ (or $(x_3, x_i) \in Q_2 \cup Q_4$). Since (X_2, X_1) also has the strict $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property it follows that the copula of $X_{1:3}$ has the strict $\text{CIS}_{(1,3)}$-$\text{CDS}_{(2,4)}(X_{3:1})$ property. However, the usefulness of the copula of the $(p+1)$-variate Student-t distribution for practical applications is seriously limited since $(U_1, U_2) \stackrel{d}{=} (U_1, U_i)$ for all $i = 2, \ldots, p$, so that the dependence may only decay after the p-th lag if the process is Markov of order p. In general, it appears that copulas with the $\text{CIS}_{(1,3)}$-$\text{CDS}_{(2,4)}$ property are inflexible or computationally not feasible for practical applications.

5.4.2 Modeling volatility clusters with vine copulas

The construction methods of bivariate jointly symmetric copulas in Section 5.3.3 can be generalized to obtain multivariate jointly symmetric copulas. That is, one can apply operations on a $(p+1)$-dimensional copula D to obtain a copula C with the property that, for all $i = 1, \ldots, p+1$, and $u_{1:p+1} \in (0,1)^{p+1}$,

$$\begin{aligned}&C(u_1, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots, u_{p+1})\\&= C(u_1, \ldots, u_{i-1}, 1, u_{i+1}, \ldots, u_{p+1}) - C(u_1, \ldots, u_{i-1}, 1-u_i, u_{i+1}, \ldots, u_{p+1}).\end{aligned}$$

A distribution with a multivariate jointly symmetric copula has a unit correlation matrix but is not necessarily compatible with the conditions for a stationary process if $p \geq 2$. Thus, care has to be taken with the choice of the base copula D to obtain a stationary p-th order Markov process. This is also true if we use a $(p+1)$-dimensional copula D to construct a copula C with the property that

$$C(u_{1:p+1}) = C(u_{1:p}, 1) - C(u_{1:p}, 1-u_{p+1}), \tag{5.4.1}$$

which is the generalization of a bivariate vertically symmetric copula that allows for an asymmetric shape of the conditional variance.[23] In order to satisfy the conditions of stationarity and to obtain a flexible copula one can specify a vine copula for D. However,

[23] Note that only if $(U_1, U_2) \stackrel{d}{=} (U_1, U_i), i = 3, \ldots, p+1$, which follows from stationarity, we can conclude that a distribution with this copula and symmetric margins has a unit correlation matrix.

in this thesis, we do not pursue this approach of obtaining a multivariate copula C that can be used to model time series with volatility clustering. Instead, we use bivariate vertically symmetric copulas as building blocks for the SD-vine copula in order to construct a higher-order Markov process that exhibits volatility clustering.[24] That is, we model the density of C by means of

$$c(u_{1:p+1}) = \prod_{j=1}^{p} \prod_{i=1}^{p+1-j} c_j\big(F^u_{i|s(i,j)}(u_i|u_{s(i,j)}), F^u_{i+j|s(i,j)}(u_{i+j}|u_{s(i,j)})\big|u_{s(i,j)}\big),$$

where $c_j, j = 1,\dots,p+1$, is the density of a vertically symmetric copula.

In order to obtain a higher-order Markov process that produces volatility clustering, it seems reasonable to assume that all copulas $C_i \in \mathbb{C}_p$ of the SD-vine copula exhibit the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property or are close to copulas with the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property if the copulas are not horizontally symmetric. Unfortunately, by the same arguments as in Section 3.7.1, the resulting $(p+1)$-dimensional copula has in general not the $\text{CIS}_{(1,3)}$-$\text{CDS}_{(2,4)}$ property if all copulas $C_i \in \mathbb{C}_p$ exhibit the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property. On the other side, the $\text{CIS}_{(1,3)}$-$\text{CDS}_{(2,4)}$ property is a very strong sufficient condition for the implications of Proposition 5.12. Simulation studies indicate that the autocorrelation function of the squared or absolute random variables is non-negative if copulas with the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property are used as building blocks for the SD-vine copula. Moreover, if every copula $C_i \in \mathbb{C}_p$ of the SD-vine copula exhibits the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property, this is a sufficient condition for a strongly related dependence property.

Proposition 5.13 (Positive conditional autocorrelation of absolute and squared random variables)

Let $j = 1, 2$, and define $G_j(x) = |x|^j$. Assume that Y has a symmetric marginal distribution around zero and that $\mathbb{E}\big[G_j(Y_1Y_{1+i})\big]$ and $\mathbb{E}\big[G_j(Y_i)\big]$ exist for all $i \in \mathbb{N}$. If, for all $i \in \mathbb{N}$, $C_{1,1+i|2:i}$ has the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property, i.e.,

$$\forall (u_1, u_{1+i}) \in Q_1 \cup Q_3 \colon C_{1,1+i|2:i}(u_1, u_{1+i}|U_{2:i}) \geq u_1 u_{1+i} \quad \text{(a.s.)},$$
$$\forall (u_1, u_{1+i}) \in Q_2 \cup Q_4 \colon C_{1,1+i|2:i}(u_1, u_{1+i}|U_{2:i}) \leq u_1 u_{1+i} \quad \text{(a.s.)},$$

[24] These two approaches are not equivalent if the order of the process is larger than one and simplified vine copulas are used. For instance, if $p = 2$, then the copula C with density

$$c(u_{1:3}) = 0.5(d(u_{1:3}) + d(1 - u_1, u_{2:3}))$$

satisfies (5.4.1). If we specify a simplified D-vine copula for D, the density of C is given by

$$c(u_{1:3}) = 0.5\big(\sum_{i=0,1} d_{13|2}(D_{1|2}(\tilde{u}_i|u_2), D_{3|2}(u_3|u_2))d_{12}(\tilde{u}_i, u_2)d_{23}(u_2, u_3)\big),$$

where $\tilde{u}_0 = u_1$, $\tilde{u}_1 = 1 - u_1$. Obviously, the conditional copula of $(U_1, U_3)|U_2 = u_2$ is in general a function of u_2.

then $\mathbb{C}\mathrm{ov}\big[G_j(Y_1), G_j(Y_{1+i})\big|Y_{2:i}\big] \geq 0$ for all $i \in \mathbb{N}$. If, in addition, $C_{1,2}$ has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property and is exchangeable then $\mathbb{C}\mathrm{ov}[G_j(Y_1), G_j(Y_3)] \geq 0$.

Proof. By Lemma 5.2 (1) it follows that the conditional autocorrelation function is non-negative because $F_{G_j(Y_1),G_j(Y_{1+i})|Y_{2:i}}$ has the PQD property for all $i \in \mathbb{N}$. An application of Lemma 3.3 proves that $\mathbb{C}\mathrm{ov}[G_j(Y_1), G_j(Y_3)] \geq 0$, since $(G_j(Y_1), G_j(Y_2))$ has the $\mathrm{SI}(U_1|U_2)$ property, see Lemma 5.2 (3), and also the $\mathrm{SI}(U_2|U_1)$ property because $C_{1,2}$ is exchangeable. ■

Proposition 5.13 establishes the conditions such that, conditional on the random variables in between, the correlation between squared or absolute symmetric random variables is non-negative, which is a plausible property of a time series with volatility clustering. Moreover, under some stronger assumptions on C_1 we can exclude a negative correlation between squared or absolute symmetric random variables which are two periods apart. Altogether, the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property represents an important property that vertically symmetric bivariate (conditional) copulas of an SD-vine copula should exhibit if the resulting CMP should be used as a model for financial returns. We strongly expect that time series with this property generate a positive autocorrelation function of squared or absolute symmetric random variables in the majority of cases.

5.4.3 The case of (non-)increasing transition variances

It is quite intuitive that the transition variance of a process with volatility clustering is increasing in the absolute value of each conditioning variable. If we have a p-th order Markov process with symmetric margin and such that all $C_i \in \mathbb{C}_p$ are vertically symmetric, then it is easy to show that the transition variance is increasing in the absolute value of the p-th lagged random variable if and only if the partial copula of C_p has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property (cf. Lemma 3.3). However, in general, nothing can be said about whether the transition variance is also increasing in the absolute value of the first $(p-1)$ lagged random variables. Therefore, we investigate the functional shape of the transition variance as a function of the first lagged variable for a second-order Markov process in more detail.

First, we examine the transition variance if both copulas are jointly symmetric and have the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. Figure 5.7 shows the transition variance as a function of the first lagged variable for different values of the second lagged variable if the process has a Student-t margin with dof 4. In this case, the transition variance appears to be strictly increasing in the absolute value of the first lag for all values of the second lag. This also seems to be true if we use other copulas for $C_i, i = 1, 2$, that exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. Moreover, we see that the transition variance increases with the absolute value of the second lag, which is implied by the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property of C_2. Thus, joint symmetry and the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property of C_1 and C_2 seem to imply a reasonable functional shape for the transition variance in many cases.

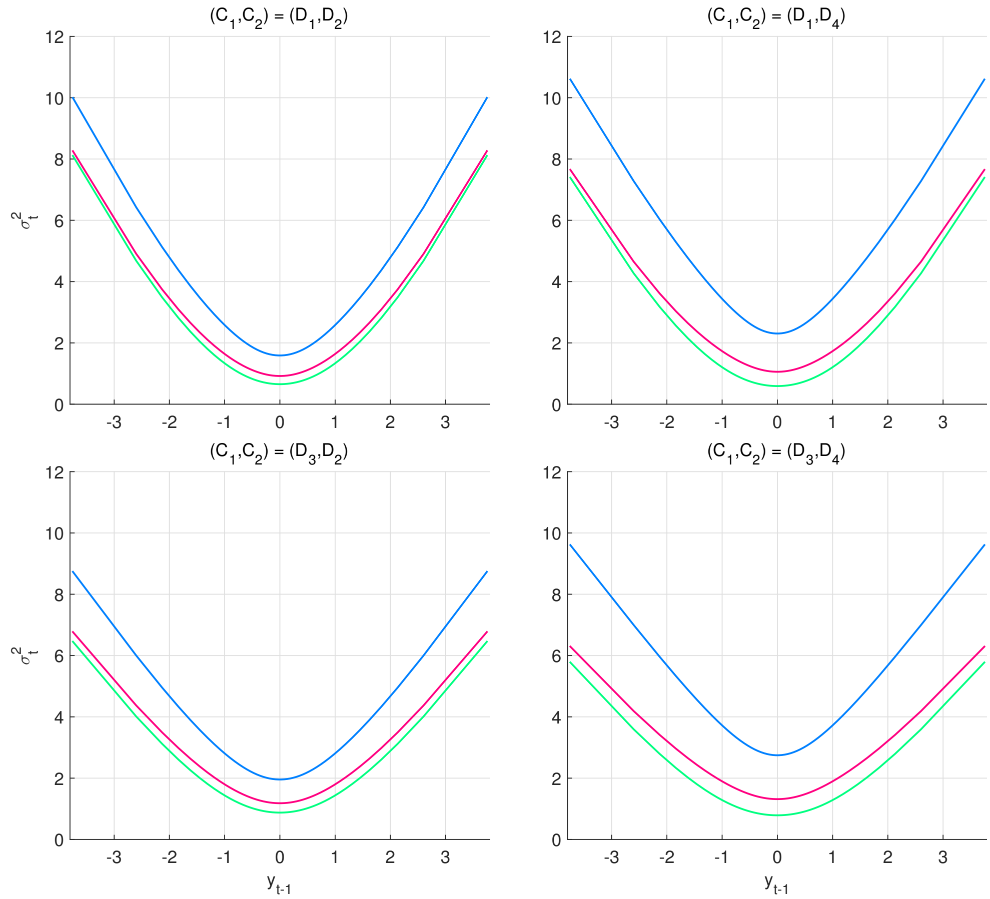


Figure 5.7: Illustration of $\sigma_t^2 = \mathbb{V}\mathrm{ar}[Y_t|Y_{t-1:t-2} = y_{t-1:t-2}]$ if Y is a stationary CMP(2) process having a Student-t margin with dof 4 and dependence structure $(C_1, C_2) \subset \mathbb{C}_2$ and y_{t-2} attains its 1%, 10% or 50% quantile. For $i = 1, 2, 3, 4$, D_i is the merged X-shaped version of the Gaussian copula with parameter θ_i, where $\theta_{1:4} = (0.7, 0.3, 0.6, 0.4)$.

This picture changes when at least one of the copulas does not exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. Except for the plot in the top left corner, the plots of Figure 5.8 show the implied transition variance if at least one of the copulas is only vertically symmetric but does not have the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. In these cases the transition variance is not increasing in the first lag for all values of the second lag. Indeed, if the value of the second lag is a rather large negative number, the transition variance has a local maximum near zero which seems to be an unlikely feature of a process with volatility clustering.[25] However, the copulas that do not have the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property exhibit a rather strong amount of horizontal asymmetry, so that the implied transition variances might not be representative for a CMP that has been fitted to data. We investigate this issue in more detail in Chapter 6 when we apply a CMP to model financial returns.

In order to explore if the non-increasingness of the transition variance is due to the fact that one copula does not have the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property or an inherent problem of the vine copula framework, we investigate the case when both vertically symmetric copulas have the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property but are not necessarily jointly symmetric. Except for the plot in

[25] Note that the variance conditional on the first lagged random variable is always increasing if C_1 has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property, irrespective of the order of the process.

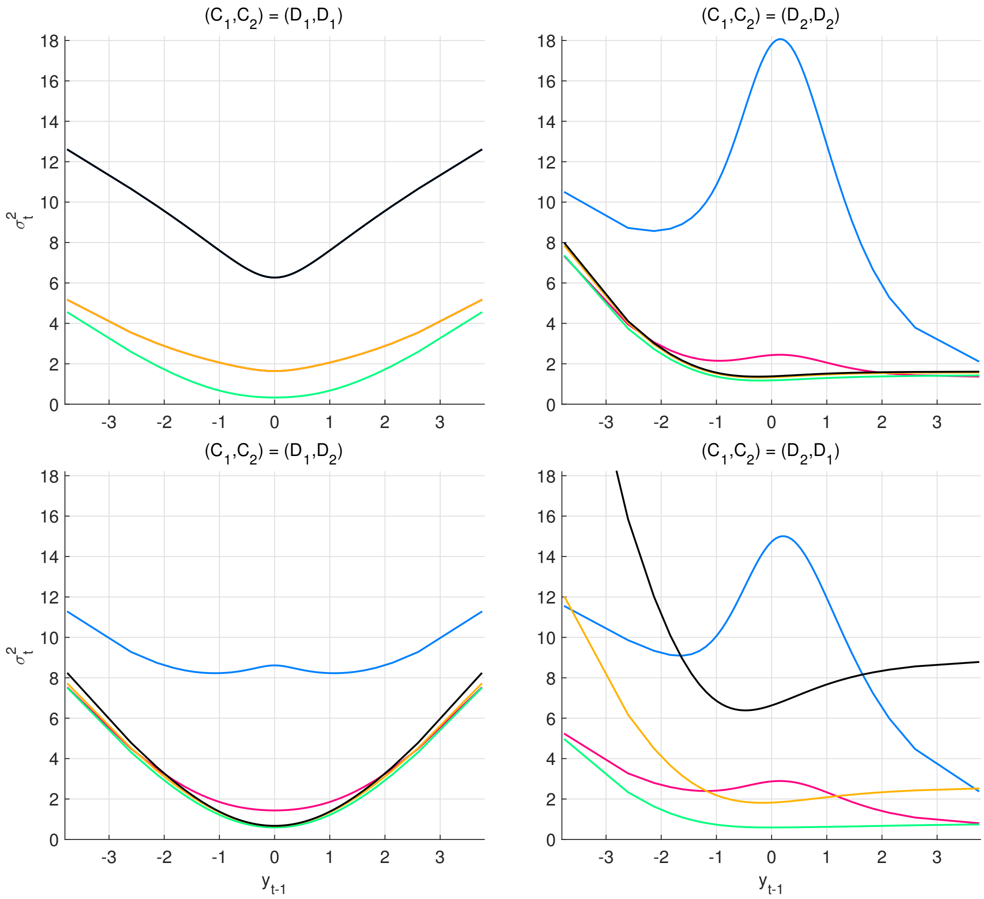


Figure 5.8: Illustration of $\sigma_t^2 = \mathbb{V}\mathrm{ar}[Y_t | Y_{t-1:t-2} = y_{t-1:t-2}]$ if Y is a stationary CMP(2) process having a Student-t margin with dof 4 and dependence structure $(C_1, C_2) \subset \mathbb{C}_2$ and y_{t-2} attains its 1%, 10%, 50%, 90% or 99% quantile. D_1 is the merged X-shaped version of the Gaussian copula with parameter $\theta = 0.63$, D_2 is an equal mixture of the Clayton copula with parameter $\theta = 0.7$ and its vertically reflected version. C_1 and C_2 have the same mutual information of 0.05.

the top left corner, Figure 5.9 shows the implied transition variance of a CMP(2) if at least one vertically symmetric copula with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property is not jointly symmetric. In all cases, a local maximum of the transition variance is not present. Altogether, the implied functional shape of the transition variance as a function of the first lag appears reasonable. Figure 5.8 and Figure 5.9 also illustrate that the vine copula framework gives rise to a transition variance that is not additive in the squares of the conditioning variables. This might be useful to obtain a better model for financial returns.

5.5 Conclusion

We have shown that, except for the Student-t copula with zero correlation parameter, commonly used copula families can not be used to model financial returns. That is mainly because most copula families have been developed for modeling monotonically related random variables. As a consequence, the only possible martingale difference sequence that can be generated by many popular copula families is a sequence of independent random variables. By the introduction of vertically symmetric copulas we characterized condi-

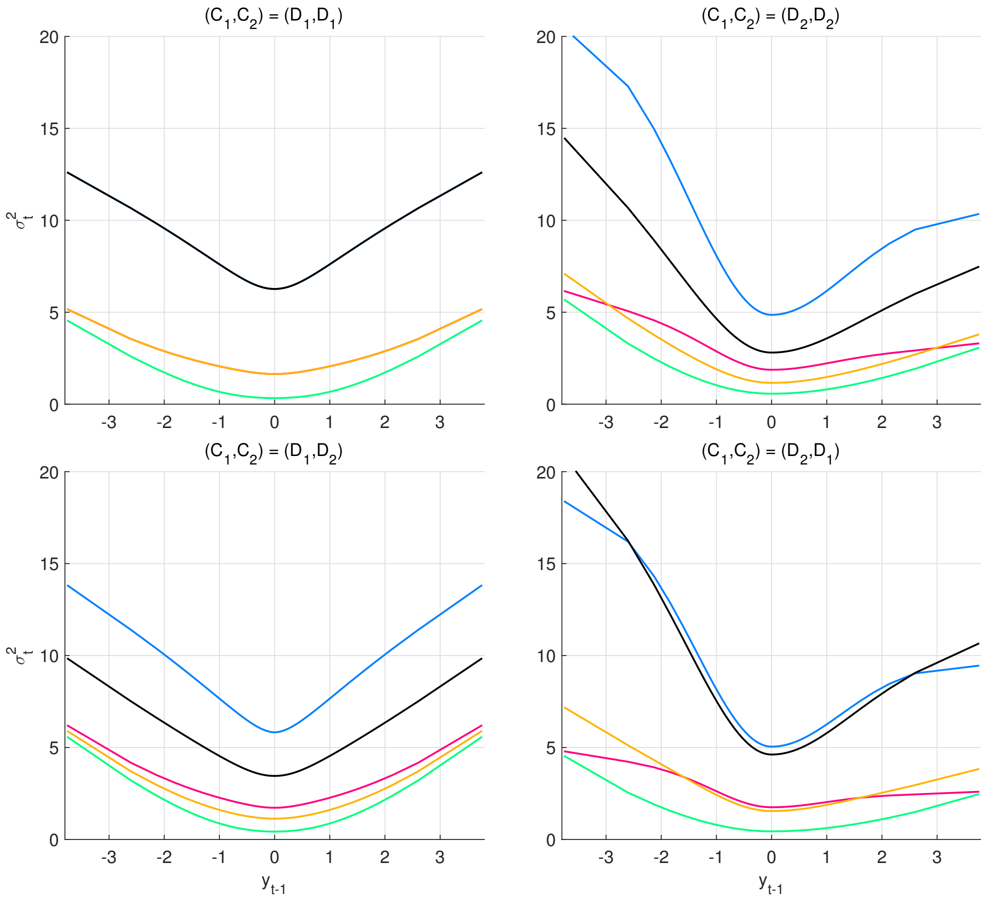


Figure 5.9: Illustration of $\sigma_t^2 = \mathbb{V}\mathrm{ar}[Y_t|Y_{t-1:t-2} = y_{t-1:t-2}]$ if Y is a stationary CMP(2) process having a Student-t margin with dof 4 and dependence structure $(C_1, C_2) \subset \mathbb{C}_2$ and y_{t-2} attains its 1%, 10%, 50%, 90% or 99% quantile. D_1 is the merged X-shaped version of the Gaussian copula with parameter $\theta = 0.63$, D_2 is a GJR-ARCH(1)-like copula with parameter $(\omega, a, \gamma) = (1, 0.62, 0.6)$ and where X and $\mathcal{E}$ have logistic distributions. D_1 and D_2 have the same mutual information of 0.05.

tionally symmetric martingale difference sequences in terms of the conditional autocopula sequence. In order to reproduce the volatility clustering of time series of financial returns, we defined the PQD$_{(1,3)}$-NQD$_{(2,4)}$ and the SI$_{(1,3)}$-SD$_{(2,4)}$ property of a bivariate copula. We demonstrated that the PQD$_{(1,3)}$-NQD$_{(2,4)}$ property implies a non-negative correlation between squared or absolute symmetric random variables and that the SI$_{(1,3)}$-SD$_{(2,4)}$ property is sufficient for a first-order Markov process to display a non-negative autocorrelation function of squared and absolute symmetric random variables and an increasing transition variance. We pointed out that the Student-t copula with zero correlation parameter has the SI$_{(1,3)}$-SD$_{(2,4)}$ property and that a first-order Markov process with this copula can be represented as a transformed ARCH(1) process.

The main part of this chapter was concerned with the construction of parametric copulas that exhibit the SI$_{(1,3)}$-SD$_{(2,4)}$ property. We proposed three different methods to construct jointly symmetric copulas with the SI$_{(1,3)}$-SD$_{(2,4)}$ property. The first two methods, the merged and patched X-shaped version of a copula, are based on transformations of established copulas families. While the SI$_{(1,3)}$-SD$_{(2,4)}$ property of the patched X-shaped version of a copula is inherited from the underlying copula, we demonstrated that this is not the case for the merged X-shaped version of a copula. We established the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property for the merged X-shaped version of the Gaussian copula which we also investigated in more detail. Moreover, we extracted a GJR-ARCH(1)-like copula with the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property that can be jointly or only vertically symmetric. The resulting GJR-ARCH(1)-like copula, that is only vertically symmetric, gives rise to an asymmetric shape of the transition variance and thus can be used to reproduce the leverage effect in financial returns.

We also investigated different ways to obtain only vertically symmetric copulas that exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. In this regard, we proved that a convex combination of jointly symmetric copulas with location dependent weights is a vertically symmetric copula. But this construction, as well as the natural generalization of the merged X-shaped version of a copula, does not exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property in general. We pointed out that this is because the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property imposes a strong symmetry constraint on the copula which is hard to satisfy if the copula is not jointly symmetric. In fact, it is an open question whether it is possible to construct a (non-degenerate) copula that is only vertically symmetric, exhibits the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property, and is not a GJR-ARCH(1)-like copula which is computationally very expensive.

Finally, we explored the construction of higher-order Markov processes that exhibit volatility clustering. We defined the $\mathrm{CIS}_{(1,3)}$-$\mathrm{CDS}_{(2,4)}$ property of a multivariate copula that is sufficient for a non-negative autocorrelation of squared or absolute symmetric random variables and a transition variance that is increasing in the absolute value of each conditioning variable. However, the usefulness of copulas with the $\mathrm{CIS}_{(1,3)}$-$\mathrm{CDS}_{(2,4)}$ property for practical applications is limited since these copulas are either inflexible or require high-dimensional integration.

In order to obtain a flexible copula model that can be used in practical applications, we advocated the use of vine copulas which typically do not exhibit the $\mathrm{CIS}_{(1,3)}$-$\mathrm{CDS}_{(2,4)}$ property. We derived sufficient conditions for the conditional autocopula sequence such that, conditional on the random variables in between, the correlation between squared or absolute random variables is non-negative. This property should be adequate to reproduce a positive autocorrelation of squared or absolute returns in many applications. However, this property does not imply that the transition variance is increasing in the absolute value of each conditioning variable. In general, if the dependence structure of a p-th order Markov process is modeled by a $(p+1)$-dimensional SD-vine copula, the functional shape of the transition variance is only tractable in the p-th lag. For this reason, we investigated the functional shape of the transition variance as a function of the first lagged variable in more detail for the case of a second-order Markov process. If both vertically symmetric copulas of the simplified SD-vine copula exhibit the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property, the transition variance seems to be increasing in the absolute value of the first lag. On the other side, if at least one vertically symmetric copula does not have the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property, the transition variance may not be increasing in the absolute value of the first lag.

5.6 Appendix

A.1 Proof that the Student-t copula is not QD

To the best of our knowledge, conditions under which elliptical copulas are QD are not known, except for the Gaussian copula, see Abdous et al. (2005) which also discuss why PQD is often a minimum dependence property that is required in many applications. Therefore, we provide a simple proof that the Student-t copula is not QD. Lemma 5.5 shows that a copula can not be QD if tail dependence is present in the corners of the first and the second or fourth quadrant or in the corners of the third and the second or fourth quadrant. Thus, if an elliptical copula has tail dependence in all corners of the unit cube, then the copula can not be PQD, see Abdous et al. (2005) and Schmidt (2002) for sufficient conditions that an elliptical copula has tail dependence in all corners of the unit cube.

Lemma 5.5 (The Student-t copula is not QD)

The Student-t copula with $\nu < \infty$ degrees of freedom is neither PQD nor NQD for all $\rho \in (-1, 1)$.

Proof. We first show that the Student-t copula has tail dependence in all corners of the unit cube. The lower (and upper) tail dependence coefficient of the Student-t copula (McNeil et al., 2005, Example 5.33), is given by

$$\lambda(C^t_{U_1,U_2}(\rho,\nu)) = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right) > 0, \quad \rho \in (-1,1] \tag{5.6.1}$$

The tail dependence coefficient in the corner of the second (and fourth) quadrant is given by

$$\begin{aligned}\lim_{q\searrow 0} P(U_2 \geq 1-q|U_1 < q) &= \lim_{q\searrow 0} \frac{q - C^t_{U_1,U_2}(1-q,q;\rho,\nu)}{q} = \lim_{q\searrow 0} \frac{C^t_{1-U_1,U_1}(q,q;-\rho,\nu)}{q} \\ &= \lambda(C^t_{1-U_1,U_2}(-\rho,\nu)) > 0, \end{aligned} \tag{5.6.2}$$

where the second equality follows because $C^t_{U_1,U_2}(u_1,u_2;\rho,\nu) = u_2 - C^t_{1-U_1,U_2}(1-u_1,u_2;-\rho,\nu)$.

Wlog assume that in the following $\rho \leq 0$. If $C^t_{U_1,U_2}(\rho,\nu)$ is NQD, then

$$\lambda(C^t_{U_1,U_2}(\rho,\nu)) = \lim_{q\searrow 0} \frac{C^t_{U_1,U_2}(q,q;\rho,\nu)}{q} \leq \lim_{q\searrow 0} \frac{qq}{q} = 0,$$

which contradicts (5.6.1). If $C^t_{U_1,U_2}(\rho,\nu)$ is PQD, then the tail dependence coefficient in the corner of the second quadrant yields

$$\lim_{q\searrow 0} \mathbb{P}(U_2 \geq 1-q|U_1 < q) = \lim_{q\searrow 0} \frac{q - C^t_{U_1,U_2}(1-q,q;\rho,\nu)}{q} = \lim_{q\searrow 0} \frac{C^t_{1-U_1,U_1}(q,q;-\rho,\nu)}{q}$$

$$\leq \lim_{q \searrow 0} \frac{qq}{q} = 0,$$

which contradicts (5.6.2).

■

A.2 Proof of Lemma 5.1

According to Definition 5.3, Y has a symmetric transition distribution around it median μ if and only if for all $t \in \mathbb{N}$

$$F_{t|t-1:1}(\mu + Y_t | Y_{t-1:1}) = 1 - F_{t|t-1:1}(\mu - Y_t | Y_{t-1:1}) \quad (F_{t-1:1}\text{-a.s.}). \tag{5.6.3}$$

In terms of copulas, this is equivalent to

$$\begin{aligned} &\partial_2 C_{t1|t-1:2}\big(F_{t|t-1:2}(\mu + Y_t|Y_{t-1:2}), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}\big) \\ &= 1 - \partial_2 C_{t1|t-1:2}\big(1 - F_{t|t-1:2}(\mu - Y_t|Y_{t-1:2}), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}\big). \end{aligned} \tag{5.6.4}$$

Thus if (5.6.3) holds for all $t \in \mathbb{N}$ then (5.6.4) becomes for all $t \in \mathbb{N}$

$$\begin{aligned} &\partial_2 C_{t1|t-1:2}\big(F_{t|t-1:2}(\mu + Y_t|Y_{t-1:2}), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}\big) \\ &= 1 - \partial_2 C_{t1|t-1:2}\big(1 - F_{t|t-1:2}(\mu + Y_t|Y_{t-1:2}), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}\big). \end{aligned} \tag{5.6.5}$$

This can only be true if for all $t \in \mathbb{N}$

$$\begin{aligned} &C_{t1|t-1:2}(F_{t|t-1:2}(\mu + Y_t|Y_{t-1:2}), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}) \\ &= U_1 - C_{t1|t-1:2}(1 - F_{t|t-1:2}(\mu + Y_t|Y_{t-1:2}), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}), \end{aligned}$$

and we conclude that all copulas must be vertically symmetric if Y has a symmetric transition distribution.

To show the converse, we use induction. Assume that all copulas are vertically symmetric and Y has a symmetric marginal distribution. Then for all $t \in \mathbb{N}$

$$\begin{aligned} \partial_2 C_{t,t-1}\big(F_t(\mu + Y_t), F_{t-1}(Y_{t-1})\big) &= 1 - \partial_2 C_{t,t-1}\big(1 - F_t(\mu - Y_t), F_{t-1}(Y_{t-1})\big) \\ &= 1 - \partial_2 C_{t,t-1}\big(1 - (1 - F_t(\mu + Y_t)), F_{t-1}(Y_{t-1})\big), \end{aligned}$$

which is equivalent to

$$F_{t|t-1}(\mu + Y_t|Y_{t-1}) = 1 - F_{t|t-1}(\mu - Y_t|Y_{t-1}).$$

Now suppose that $F_{t|t-1:2}(\mu + Y_t|Y_{t-1:2}) = 1 - F_{t|t-1:2}(\mu - Y_t|Y_{t-1:2})$. Then

$$\begin{aligned}
&\partial_2 C_{t1|t-1:2}\big(F_{t|t-1:2}(\mu + Y_t|Y_{t-1:2}), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}\big)\\
&= 1 - \partial_2 C_{t1|t-1:2}\big(1 - F_{t|t-1:2}(\mu + Y_t|Y_{t-1:2}), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}\big)\\
&= 1 - \partial_2 C_{t1|t-1:2}\big(1 - (1 - F_{t|t-1:2}(\mu - Y_t|Y_{t-1:2})), F_{1|t-1:2}(Y_1|Y_{t-1:2})|Y_{t-1:2}\big),
\end{aligned}$$

which is equivalent to

$$F_{t|t-1:1}(\mu + Y_t|Y_{t-1:1}) = 1 - F_{t|t-1:1}(\mu - Y_t|Y_{t-1:1}).$$

Consequently, $F_{t|t-1:1}(\mu + Y_t|Y_{t-1:1}) = 1 - F_{t|t-1:1}(\mu - Y_t|Y_{t-1:1})$ for all $t \in \mathbb{N}$.

A.3 Proof of Proposition 5.3

We first show that conditional Kendall's tau or Spearman's rho for a vertically symmetric (or horizontally) symmetric copula is zero. Let $(U_1, U_K)|U_{2:K-1} \sim C_{1K|2:K-1}$ and $(1 - U_1, U_K)|U_{2:K-1} \sim D_{1K|2:K-1}$. Using Definition 5.4 we obtain

$$\begin{aligned}
&\tau_{D_{1K|2:K-1}}(U_{2:K-1})\\
&= 4\int_{[0,1]^2} D_{1K|2:K-1}(t_1, t_K|U_{2:K-1})\mathrm{d}C_{1K|2:K-1}(t_1, t_K|U_{2:K-1}) - 1\\
&= 4\int_{[0,1]^2} \big(t_K - C_{1K|2:K-1}(1 - t_1, t_K|U_{2:K-1})\big)\mathrm{d}C_{1K|2:K-1}(t_1, t_K|U_{2:K-1}) - 1\\
&= 4\int_{[0,1]} t_K dt_K - 4\int_{[0,1]^2} C_{1K|2:K-1}(t_1, t_K|U_{2:K-1})c_{1K|2:K-1}(t_1, t_K|U_{2:K-1})\mathrm{d}t_1\mathrm{d}t_K - 1\\
&\overset{t_1=1-z_1}{=} 2 + 4\int_0^1\int_1^0 C_{1K|2:K-1}(z_1, t_K|U_{2:K-1})c_{1K|2:K-1}(1 - z_1, t_K|U_{2:K-1})\mathrm{d}z_1\mathrm{d}t_K - 1\\
&\overset{\text{Definition 5.4 (iii)}}{=} -\Big(4\int_{[0,1]^2} C_{1K|2:K-1}(t_1, t_K|U_{2:K-1})\mathrm{d}C_{1K|2:K-1}(t_1, t_K|U_{2:K-1}) - 1\Big)\\
&= -\tau_{C_{1K|2:K-1}}(U_{2:K-1}),
\end{aligned}$$

and

$$\begin{aligned}
&\rho_{D_{1K|2:K-1}}(U_{2:K-1})\\
&= 12\int_{[0,1]^2} D_{1K|2:K-1}(t_1, t_K)\mathrm{d}t_1\mathrm{d}t_K - 3\\
&= 12\int_{[0,1]^2} \big(t_k - C_{1K|2:K-1}(1 - t_1, t_K)\big)\mathrm{d}t_1\mathrm{d}t_K - 3\\
&\overset{t_1=1-z_1}{=} 12\int_{[0,1]^2} t_k\mathrm{d}t_1\mathrm{d}t_k - 12\int_{[0,1]^2} C_{1K|2:K-1}(z_1, t_K)\mathrm{d}z_1\mathrm{d}t_K - 3\\
&= -\Big(12\int_{[0,1]^2} C_{1K|2:K-1}(t_1, t_K)\mathrm{d}t_1\mathrm{d}t_K - 3\Big)\\
&= -\rho_{C_{1K|2:K-1}}(U_{2:K-1}).
\end{aligned}$$

Thus, if $(U_1, U_K)|U_{2:K-1} \overset{d}{=} (1 - U_1, U_K)|U_{2:K-1}$ we conclude that

$$\tau_{D_{1K|2:K-1}}(U_{2:K-1}) = \tau_{C_{1K|2:K-1}}(U_{2:K-1}) = \rho_{D_{1K|2:K-1}}(U_{2:K-1}) = \rho_{C_{1K|2:K-1}}(U_{2:K-1}) = 0 \text{ (a.s.)}.$$

It follows that the conditional auto Spearman's rho and the conditional auto Kendall's tau function are zero.

In order to verify that auto Spearman's rho and auto Kendall's tau function are zero, we show that if all bivariate conditional copulas are vertically symmetric then $(U_j, U_1) \overset{d}{=} (1 - U_j, U_1)$ for all $j \in \mathbb{N}$.

$$\begin{aligned}
&C_{1K}(u_1, u_K)\\
&= \int_{[0,1]^{K-2}} C_{1K|2:K-1}\big(F_{1|2:K-1}(u_1|u_{2:K-1}), F_{K|2:K-1}(u_K|u_{2:K-1})\big|u_{2:K-1}\big)\mathrm{d}C_{2:K-1}(u_{2:K-1})\\
&= \int_{[0,1]^{K-2}} C_{1K|2:K-1}\big(1 - F_{1|2:K-1}(1 - u_1|u_{2:K-1}), F_{K|2:K-1}(u_K|u_{2:K-1})|u_{2:K-1}\big)\mathrm{d}C_{2:K-1}(u_{2:K-1})\\
&= \int_{[0,1]^{K-2}} \Big(F_{K|2:K-1}(u_K|u_{2:K-1})\\
&\quad - C_{1K|2:K-1}\big(F_{1|2:K-1}(1 - u_1|u_{2:K-1}), F_{K|2:K-1}(u_K|u_{2:K-1})|u_{2:K-1}\big)\Big)\mathrm{d}C_{2:K-1}(u_{2:K-1})\\
&= u_K - C_{1K}(1 - u_1, u_K),
\end{aligned}$$

where the third and fourth equality follow from the assumption that $C_{1K-1|2:K-2}$ and $C_{1K|2:K-1}$ are vertically symmetric.

The partial auto Spearman's rho and partial auto Kendall's tau function are also zero since a bivariate (higher-order) partial copula is vertically symmetric if the corresponding bivariate conditional copula is vertically symmetric as we show now. For $i = 1, K$, define the function $K_i(t_i|t_{2:K_1}) = F_{i|2:K-1}\big((F^{\partial^{K-3}}_{i|2:K-1})^{-1}(t_i|t_{2:K-1})\big|t_{2:K-1}\big)$, where $(F^{\partial^{K-3}}_{i|2:K-1})^{-1}$ is the inverse of $F^{\partial^{K-3}}_{i|2:K-1}$ wrt to the first argument. The (K–2)-th order partial copula of $C_{1K|2:K-1}$ (Definition 2.6) is given by

$$\begin{aligned}
&D^{\partial^{K-2}}_{1K;\,2:K-1}(a, b)\\
&= \int_{[0,1]^{K-2}} D_{1K|2:K-1}\big(K_1(a|t_{2:K-1}), K_K(b|t_{2:K-1})|t_{2:K-1}\big)c_{2:K-1}(t_{2:K-1})\mathrm{d}t_{2:K-1}\\
&= \int_{[0,1]^{K-2}} \Big(K_K(b|t_{2:K-1}) - C_{1K|2:K-1}\big(1 - K_1(a|t_{2:K-1}), K_K(b|t_{2:K-1})\big|t_{2:K-1}\big)\Big)\\
&\quad \times c_{2:K-1}(t_{2:K-1})\mathrm{d}t_{2:K-1}\\
&= b - C^{\partial^{K-2}}_{1K;\,2:K-1}(1 - a, b),
\end{aligned}$$

where the last equality follows because $U^{\partial^{K-3}}_{K|2:K-1} \sim U(0,1)$ and

$$\begin{aligned}
&\int_{[0,1]^{K-2}} K_K(b|t_{2:K-1})c_{2:K-1}(t_{2:K-1})\mathrm{d}t_{2:K-1} \\
&= \int_{[0,1]^{K-2}} \mathbb{P}(U_K \leq (F^{\partial^{K-3}}_{K|2:K-1})^{-1}(b|t_{2:K-1})|U_{2:K-1} = t_{2:K-1})c_{2:K-1}(t_{2:K-1})\mathrm{d}t_{2:K-1} \\
&= \mathbb{P}(F^{\partial^{K-3}}_{K|2:K-1}(U_K|U_{2:K-1}) \leq b) = \mathbb{P}(U^{\partial^{K-3}}_{K|2:K-1} \leq b) = b.
\end{aligned}$$

A.4 Proof of Lemma 5.2

1. If and only if (Y_1, Y_2) has symmetric marginal distributions around zero, then $\forall (y_1, y_2) \in [0, \infty)^2\colon F_i(y_i) \geq 0.5$ and $F_i(-y_i) \leq 0.5$. Thus,

$$\begin{aligned}
&\exists (u_1, u_2) \in Q_1\colon F_{12}(y_1, y_2) = C(u_1, u_2), \\
&\exists (u_1, u_2) \in Q_3\colon F_{12}(-y_1, -y_2) = C(u_1, u_2), \\
&\exists (u_1, u_2) \in Q_4\colon F_{12}(y_1, -y_2) = C(u_1, u_2), \\
&\exists (u_1, u_2) \in Q_2\colon F_{12}(-y_1, y_2) = C(u_1, u_2).
\end{aligned}$$

   By the PQD$_{(1,3)}$-NQD$_{(2,4)}$ property we conclude that for all $(y_1, y_2) \in [0, \infty)^2$,

$$\begin{aligned}
F_{12}(y_1, y_2) &\geq F_1(y_1)F_2(y_2), \\
F_{12}(-y_1, -y_2) &\geq F_1(-y_1)F_2(-y_2), \\
F_{12}(y_1, -y_2) &\leq F_1(y_1)F_2(-y_2), \\
F_{12}(-y_1, y_2) &\leq F_1(-y_1)F_2(y_2).
\end{aligned}$$

   Let $(Y_1, Y_2) \sim F_{12}$ and $G(Y_i) = |Y_i|, i = 1, 2$, then

$$\begin{aligned}
F_{G(Y_1),G(Y_2)}(a, b) &= \mathbb{P}(-a \leq Y_1 \leq a, -b \leq Y_2 \leq b) \\
&= F_{12}(a, b) + F_{12}(-a, -b) - F_{12}(a, -b) - F_{12}(-a, b). \\
&\geq F_1(a)F_2(b) + F_1(-a)F_2(-b) - F_1(a)F_2(-b) - F_1(-a)F_2(b) \\
&= \mathbb{P}(-a \leq Y_1 \leq a)\mathbb{P}(-b \leq Y_2 \leq b) \\
&= F_{G(Y_1)}(a)F_{G(Y_2)}(b),
\end{aligned}$$

   where the inequality follows from the previous considerations and because $(a, b) \in [0, \infty)^2$. Moreover, the inequality can not be an equality if C has the strict PQD$_{(1,3)}$-NQD$_{(2,4)}$ property. Since the PQD property is property of the copula if the margins are symmetric, the conclusion follows. The proof for $G(Y_i) = Y_i^2, i = 1, 2$, is similar.

2. Wlog assume that $\mu = 0$. If (Y_1, Y_2) has symmetric margins and its copula exhibits

the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property then

$$\mathbb{P}((Y_1,Y_2)\in R_3)=F(0,0)=C(F_1(0),F_2(0))=C(0.5,0.5)=0.5^2.$$

Hoeffding's lemma yields

$$\begin{aligned}
&\mathbb{C}\text{ov}[Y_1,Y_2|(Y_1,Y_2)\in R_3]\\
&=\int_{\mathbb{R}^2}\big(\mathbb{P}(Y_i\le y_i,\forall i=1,2|(Y_1,Y_2)\in R_3)-\prod_{i=1,2}\mathbb{P}(Y_i\le y_i|(Y_1,Y_2)\in R_3)\big)\mathrm{d}y_1\mathrm{d}y_2\\
&=\frac{1}{F(0,0)}\int_{(-\infty,0]^2}\left(F(y_1,y_2)-\frac{F(y_1,0)F(0,y_2)}{F(0,0)}\right)\mathrm{d}y_1\mathrm{d}y_2, \qquad (5.6.6)
\end{aligned}$$

where we have used for the second equality that

$$\begin{aligned}
&\mathbb{P}(Y_i\le y_i,\forall i=1,2|(Y_1,Y_2)\in R_3)-\prod_{i=1,2}\mathbb{P}(Y_i\le y_i|(Y_1,Y_2)\in R_3)\\
&=\frac{\mathbb{P}(Y_i\le \min(y_i,0),\forall i=1,2)}{\mathbb{P}(Y_i\le 0,\forall i=1,2)}-\prod_{i,j=1,2;\ j\ne i}\frac{P(Y_i\le\min(y_i,0),Y_j\le 0)}{\mathbb{P}(Y_i\le 0,\forall i=1,2)}\\
&=\left(\frac{F(y_1,y_2)}{F(0,0)}-\frac{F(y_1,0)F(0,y_2)}{F(0,0)^2}\right)\mathbb{1}_{\{y_i\le 0,\forall i=1,2\}}.
\end{aligned}$$

If the copula C of (Y_1,Y_2) has the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property then for all $(u_1,u_2)\in Q_3$,

$$C(0.5,0.5)C(u_1,u_2)=0.5^2C(u_1,u_2)\ge 0.5^2u_1u_2=C(u_1,0.5)C(0.5,u_2),$$

so that for all $(y_1,y_2)\in(-\infty,0]^2$,

$$F(0,0)F(y_1,y_2)\ge F(y_1,0)F(0,y_2),$$

if the margins of (Y_1,Y_2) are symmetric. Consequently, the integrand in (5.6.6) is non-negative with equality only if C has not the strict $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property. The remaining statements for $\mathbb{C}\text{ov}[Y_1,Y_2|(Y_1,Y_2)\in R_i], i=1,2,4$, can be shown in a similar manner.

A.5 Proof of Lemma 5.3

1. We only need to show that $\forall(u_1,u_2)\in Q_3\colon C(u_1,u_2)\ge u_1u_2$. Let $(U_1,U_2)\sim C$ and consider the truncated distribution $(\tilde U_1,\tilde U_2)=(U_1,U_2)|(U_1,U_2)\in Q_3$, with distribution function

$$\mathbb{P}(\tilde U_1\le u_1,\tilde U_2\le u_2)=\mathbb{P}(U_1\le u_1,U_2\le u_2|(U_1,U_2)\in Q_3)$$

$$= \frac{C(\min(u_1, 0.5), \min(u_2, 0.5))}{C(0.5, 0.5)},$$

and conditional cdf

$$\mathbb{P}(\tilde{U}_1 \leq u_1 | \tilde{U}_2 = u_2) = \frac{C(\min(u_1, 0.5)|u_2)}{C(0.5, 0.5)} \mathbb{1}_{\{0 \leq u_2 \leq 0.5\}},$$

which is (strictly) decreasing in u_2 for all $u_1 \in (0, 0.5)$ if C has the (strict) $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. Thus, $(\tilde{U}_1, \tilde{U}_2)$ is (strictly) $\mathrm{SI}(\tilde{U}_1|\tilde{U}_2)$ and (strictly) PQD. Since for all $(u_1, u_2) \in Q_3$ it holds that

$$C(u_1, u_2) \geq u_1 u_2 \Leftrightarrow \mathbb{P}(\tilde{U}_1 \leq u_1, \tilde{U}_2 \leq u_2) \geq \prod_{i=1,2} \mathbb{P}(\tilde{U}_i \leq u_1),$$

it follows that $\forall (u_1, u_2) \in Q_3 \colon C(u_1, u_2) \geq u_1 u_2$.

For the remaining statements note that

$$\begin{aligned} C^{\mathrm{V}}(u_1|u_2) &= 1 - C(1 - u_1|u_2), \\ C^{\mathrm{H}}(u_1|u_2) &= C(u_1|1 - u_2), \\ C^{\mathrm{S}}(u_1|u_2) &= 1 - C(1 - u_1|1 - u_2), \end{aligned}$$

are (strictly) decreasing in u_2 for $(u_1, u_2) \in Q_3$ if C has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property. Consequently, using the same arguments as for the conclusion that $\forall (u_1, u_2) \in Q_3 :$ $C(u_1, u_2) \geq u_1 u_2$, we have that for all $(u_1, u_2) \in Q_3$

$$C^{\mathrm{V}}(u_1, u_2) \geq u_1 u_2, \quad C^{\mathrm{H}}(u_1, u_2) \geq u_1 u_2, \quad C^{\mathrm{S}}(u_1, u_2) \geq u_1 u_2.$$

Thus, the remaining statements follow because

$$\begin{aligned} \forall (u_1, u_2) \in Q_3 \colon C^{\mathrm{V}}(u_1, u_2) \geq u_1 u_2 &\Leftrightarrow \forall (u_1, u_2) \in Q_2 : C(u_1, u_2) \leq u_1 u_2, \\ \forall (u_1, u_2) \in Q_3 \colon C^{\mathrm{H}}(u_1, u_2) \geq u_1 u_2 &\Leftrightarrow \forall (u_1, u_2) \in Q_4 : C(u_1, u_2) \leq u_1 u_2, \\ \forall (u_1, u_2) \in Q_3 \colon C^{\mathrm{S}}(u_1, u_2) \geq u_1 u_2 &\Leftrightarrow \forall (u_1, u_2) \in Q_1 : C(u_1, u_2) \geq u_1 u_2. \end{aligned}$$

2. This easily follows by choosing in Definition and Lemma 5.2 for D a copula which has the PQD but not the $\mathrm{SI}(U_1|U_2)$ property.

3. Let $(Y_1, Y_2) \sim F_{12}$, then

$$\begin{aligned} F_{Y_1^2, Y_2^2}(a, b) &= P(-\sqrt{a} \leq Y_1 \leq \sqrt{a}, -\sqrt{b} \leq Y_2 \leq \sqrt{b}) \\ &= F_{12}(\sqrt{a}, \sqrt{b}) + F_{12}(-\sqrt{a}, -\sqrt{b}) - F_{12}(\sqrt{a}, -\sqrt{b}) - F_{12}(-\sqrt{a}, \sqrt{b}). \end{aligned}$$

The conditional cdf is given by

$$
\begin{aligned}
&F_{Y_1^2|Y_2^2}(a|b)\\
&= \frac{1}{2\sqrt{b}}\Big(F_{1|2}(\sqrt{a}|\sqrt{b}) - F_{1|2}(-\sqrt{a}|-\sqrt{b}) + F_{1|2}(\sqrt{a}|-\sqrt{b}) - F_{1|2}(-\sqrt{a}|\sqrt{b})\Big)\\
&= \frac{1}{2\sqrt{b}}\Big(C_{1|2}(F_1(\sqrt{a})|F_2(\sqrt{b})) - C_{1|2}(F_1(-\sqrt{a})|F_2(-\sqrt{b})) + C_{1|2}(F_1(\sqrt{a})|F_2(-\sqrt{b}))\\
&\quad - C_{1|2}(F_1(-\sqrt{a})|F_2(\sqrt{b}))\Big).
\end{aligned}
$$

If $F_i, i = 1,2,$ is symmetric around zero then $F_1(-\sqrt{a}) < 0.5,\ F_1(\sqrt{a}) > 0.5,\ F_2(-\sqrt{b}) < 0.5,\ F_2(\sqrt{b}) > 0.5$ for $(a,b) \in (0,\infty)^2$. It is straight forward to show that $F_{Y_1^2|Y_2^2}(a|b)$ is then (strictly) decreasing in b if C_{12} has the (strict) $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property.

Let $X_i := |Y_i|, i = 1,2$. It can also be readily verified that F_{X_1,X_2} has the $\mathrm{SI}(U_1|U_2)$ property if F has symmetric margins around zero and C_{12} has the $\mathrm{SI}_{(1,3)}$-$\mathrm{SD}_{(2,4)}$ property because

$$F_{X_1,X_2}(a,b) = F_{12}(a,b) + F_{12}(-a,-b) - F_{12}(a,-b) - F_{12}(-a,b),$$

so that

$$F_{X_1|X_2}(a,b) = F_{1|2}(a|b) - F_{1|2}(-a|-b) + F_{1|2}(a|-b) - F_{1|2}(-a|b)$$

is strictly decreasing under the stated conditions.

4. This follows from the definition of the $\mathrm{PQD}_{(1,3)}$-$\mathrm{NQD}_{(2,4)}$ property, see the remarks after Definition 5.5.

A.6 Proof of Lemma 5.4

In the following, the unit cube is partitioned in 9 equally sized rectangles $N_{i,j}$ which are defined by

$$
\begin{aligned}
K_i &:= \left(\frac{i-1}{3}, \frac{i}{3}\right), \quad i = 1,2,3,\\
N_{i,j} &:= K_i \times K_j, \qquad i,j = 1,2,3.
\end{aligned}
$$

1. We first show that the SI property of D is not a necessary condition for (5.3.5). Define the density of the exchangeable and radially symmetric checkerboard copula

D as

$$d(u_1,u_2) := \begin{cases} 2 & \text{if } (u_1,u_2) \in N_{1,1} \cup N_{3,3} \\ 1 & \text{if } (u_1,u_2) \in N_{1,3} \cup N_{3,1} \\ 3 & \text{if } (u_1,u_2) \in N_{2,2} \\ 0 & \text{else} \end{cases}.$$

Sketching the copula density, it can be easily verified that D is PQD but not SI since, e.g., $D(0.3|0.3) > D(0.3|0.4)$. The merged X-shaped version of D has the density

$$c(u_1,u_2) := \begin{cases} 1.5 & \text{if } (u_1,u_2) \in N_{1,1} \cup N_{3,3} \cup N_{1,3} \cup N_{3,1} \\ 3 & \text{if } (u_1,u_2) \in N_{2,2} \\ 0 & \text{else} \end{cases},$$

so that

$$C(u_1|u_2) = \begin{cases} \frac{3}{2}\int_0^{\min(u_1,1/3)} dt & \text{if } u_2 < 1/3, u_1 < 0.5 \\ 0 & \text{if } 1/3 < u_2 < 0.5, u_1 < 1/3 \\ 3\int_{1/3}^{u_1} dt & \text{if } 1/3 < u_i < 0.5, i = 1,2, \end{cases}$$

which shows that C has the strict $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property.

2. On the other side, the strict $\text{SI}(U_1|U_2)$ property of D is not sufficient for the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property as the following counterexample shows.

   Consider the exchangeable and radially symmetric checkerboard copula D with density given by

$$d(u_1,u_2) = \begin{cases} 0 & \text{if } (u_1,u_2) \in N_{31} \cup N_{22} \cup N_{13} \\ 1.5 & \text{if } (u_1,u_2) \in [0,1]^2 \backslash (N_{31} \cup N_{22} \cup N_{13}) \end{cases}.$$

   It is easy to check that D is strictly SI. The merged X-shaped version of D has the density

$$c(u_1,u_2) = \begin{cases} 1.5 & \text{if } (u_1,u_2) \in N_{12} \cup N_{21} \cup N_{23} \cup N_{32} \\ 0.75 & \text{if } (u_1,u_2) \in N_{11} \cup N_{13} \cup N_{31} \cup N_{33} \\ 0 & \text{if } (u_1,u_2) \in N_{22}, \end{cases}$$

   so that

$$C(1/3|1/3) = \int_0^{1/3} 0.75\mathrm{d}t = 0.25,$$
$$C(1/3|1/2) = \int_0^{1/3} 1.5\mathrm{d}t = 0.5,$$

which shows that $C(u_1|u_2)$ is not strictly decreasing in u_2 for $(u_1, u_2) \in Q_3$. In fact, elementary computations show that

$$\mathbb{V}\text{ar}[U_1|U_2 = 1/3] < \mathbb{V}\text{ar}[U_1|U_2 = 1/2],$$

so that the conditional variance decreases in the absolute value of the conditioning variable. This is also true if (Y_1, Y_2) has standard normal margins. Moreover, $C(1/3, 1/3) = (2/3)(1/3)^2 = 1/12 \leq 1/9 = C^{\perp}(1/3, 1/3)$, so that C does not have the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property.

However, the density of D is not TP_2 since $d(0,0)d(0.5,0.5) = 3/2\times 0 \leq= 3/2\times 3/2 = d(0,0.5)d(0.5,0)$.[26]

3. The FGM copula has a TP_2 density and simple computations show that its merged X-shaped version is the product copula so that $C(u_1|u_2)$ is not *strictly* decreasing in u_2 for $(u_1, u_2) \in Q_3$. Thus, D having a TP_2 density does not ensure that the conditional variance is *strictly* increasing.

A.7 Proof of Proposition 5.8

Let D be the Gaussian copula with $\rho > 0$. Since D is radially symmetric we have that

$$\partial_2 C(u_1|u_2;\rho) = 0.5(\partial_2 D(u_1|u_2;\rho) + \partial_2 D(u_1|u_2;-\rho)).$$

Moreover,

$$\begin{aligned} D(u_1|u_2;\rho) &= \Phi\left(\frac{\Phi^{-1}(u_1) - \rho\Phi^{-1}(u_2)}{\sqrt{1-\rho^2}}\right), \\ \partial_2 D(u_1|u_2;\rho) &= -\frac{\rho}{\sqrt{1-\rho^2}}\phi\left(\frac{\Phi^{-1}(u_1) - \rho\Phi^{-1}(u_2)}{\sqrt{1-\rho^2}}\right)/\phi(\Phi^{-1}(u_2)), \end{aligned}$$

where $\partial_2 D(u_1|u_2;\rho) := \frac{\partial D(u_1|u_2;\rho)}{\partial u_2}$. Thus,

$$\begin{aligned} 2\partial_2 C(u_1|u_2;\rho) &= -\frac{\rho(1-\rho^2)^{-0.5}}{\phi(\Phi^{-1}(u_2))}\left(\phi\left(\frac{\Phi^{-1}(u_1) - \rho\Phi^{-1}(u_2)}{\sqrt{1-\rho^2}}\right) - \phi\left(\frac{\Phi^{-1}(u_1) + \rho\Phi^{-1}(u_2)}{\sqrt{1-\rho^2}}\right)\right) \\ &=: -\frac{\rho(1-\rho^2)^{-0.5}}{\phi(\Phi^{-1}(u_2))}\big(g_1(u_1,u_2) - g_2(u_1,u_2)\big). \end{aligned}$$

Since $\partial_1\phi(x) = -x\phi(x)$ is positive for $x < 0$, negative for $x > 0$, and zero only at $x = 0$, we have that $\phi(x') > \phi(x)$ if $|x'| < |x|$. Let $(u_1, u_2) \in Q_1 \cup Q_3$, it is straightforward to show that $|\Phi^{-1}(u_1) - \rho\Phi^{-1}(u_2)| < |\Phi^{-1}(u_1) + \rho\Phi^{-1}(u_2)|$. Consequently, $g_1 > g_2$ and

[26] We did not find a copula D with TP_2 density such that $C(u_1|u_2)$ is not decreasing in u_2 for $(u_1, u_2) \in Q_3$.

$\partial_2 C(u_1|u_2;\rho)$ is negative. By the same reasoning we obtain that $\partial_2 C(u_1|u_2;\rho)$ is positive if $(u_1,u_2)\in Q_2\cup Q_4$.

A.8 Local correlation of the merged X-shaped version of the Gaussian copula in the third quadrant

Let C be the merged X-shaped version of the Gaussian copula. Since C is jointly symmetric, it holds that $\mathbb{P}((U_1,U_2)\in Q_3)=0.25$ and $\partial_1 C(u_1,0.5;\rho)=0.5$ so that

$$\begin{aligned}
g(u_1,u_2;Q_3) &:= f(u_1,u_2|(U_1,U_2)\in Q_3)\\
&= \frac{1}{P((U_1,U_2)\in Q_3)}c(u_1,u_2)\mathbb{1}_{\{(u_1,u_2)\in Q_3\}} = 4c(u_1,u_2)\mathbb{1}_{\{(u_1,u_2)\in Q_3\}},\\
h(u_1;Q_3) &:= f(u_1|(U_1,U_2)\in Q_3)\\
&= \int_0^{0.5} f(u_1,u_2|(U_1,U_2)\in Q_3)\mathrm{d}u_2 = 4\partial_1 C(u_1,0.5;\rho)\mathbb{1}_{\{u_1\le 0.5\}} = 2.
\end{aligned}$$

Moreover,

$$\int_0^{0.5} u_1 h(u_1;Q_3)\mathrm{d}u_1 = 2(0.5(0.5^2)) = 0.25,$$

and

$$\int_0^{0.5} u_1^2 h(u_1;Q_3)\mathrm{d}u_1 = 2(1/3(0.5^3)) = 1/12.$$

Since

$$\begin{aligned}
\mathbb{C}\mathrm{ov}[U_1,U_2|(U_1,U_2)\in Q_3] &= \int_{Q_3} u_1u_2 g(u_1,u_2;Q_3)\mathrm{d}u_1\mathrm{d}u_2 - \left(\int_0^{0.5} u_1 h(u_1;Q_3)\mathrm{d}u_1\right)^2,\\
\mathbb{V}\mathrm{ar}[U_1|(U_1,U_2)\in Q_3] &= \int_0^{0.5} u_1^2 h(u_1;Q_3)\mathrm{d}u_1 - \left(\int_0^{0.5} u_1 h(u_1;Q_3)\mathrm{d}u_1\right)^2,
\end{aligned}$$

we conclude that

$$\rho_C^{(3)} = \frac{\mathbb{C}\mathrm{ov}[U_1,U_2|(U_1,U_2)\in Q_3]}{\mathbb{V}\mathrm{ar}[U_1|(U_1,U_2)\in Q_3]} = \frac{4\int_{Q_3} u_1u_2 c(u_1,u_2)\mathrm{d}u_{1:2} - 1/16}{1/12-1/16}.$$

A.9 Proof of Proposition 5.11

Since $0\le g\le 1$, it follows that $g(u_2)$ and $1-g(u_2)$ are non-negative for all u_2 so that $c(u_1,u_2)$ is non-negative. The density of U_2 is given by

$$f(u_2) = \int_0^1 c(u_1,u_2)\mathrm{d}u_1 = g(u_2)+1-g(u_2) = 1.$$

The density of U_1 is given by

$$f(u_1) = \mathbb{E}_{D_1}[g(U_2)|U_1] + \mathbb{E}_{D_2}[1 - g(U_2)|U_1] = 1 + \mathbb{E}_{D_1}[g(U_2)|U_1] - \mathbb{E}_{D_2}[g(U_2)|U_1],$$

with

$$\begin{aligned}
&\mathbb{E}_{D_i}[g(U_2)|U_1)] \\
&= \int_0^1 g(u_2) d_i(u_1, u_2) \mathrm{d}u_2 = \int_0^{0.5} g(u_2) d_i(u_1, u_2) \mathrm{d}u_2 + \int_{0.5}^1 g(u_2) d_i(u_1, u_2) \mathrm{d}u_2 \\
&= \int_0^{0.5} g(0.5 - u_2) d_i(u_1, 0.5 - u_2) \mathrm{d}u_2 + \int_0^{0.5} g(0.5 + u_2) d_i(u_1, 0.5 + u_2) \mathrm{d}u_2 \\
&= \int_0^{0.5} g(0.5 - u_2) d_i(u_1, 0.5 - u_2) \mathrm{d}u_2 + \int_0^{0.5} (a - g(0.5 - u_2)) d_i(u_1, 0.5 - u_2) \mathrm{d}u_2 \\
&= a \int_0^{0.5} d_i(u_1, u_2) \mathrm{d}u_2 = a \partial_1 D_i(u_1, 0.5) = 0.5a,
\end{aligned}$$

and the last equality follows because D_i is vertically symmetric. Thus, $f(u_1) = 1$. It is also evident that $C(0,0) = 0$ and that $C(1,1) = \int_{[0,1]^2} c(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2 = 1$, so that c is indeed a proper copula density. C is vertically symmetric since it can be readily verified that $c(u_1, u_2) = c(1 - u_1, u_2)$ holds for all $(u_1, u_2) \in (0,1)^2$.

A.10 Graphical comparison of jointly symmetric copulas

In the following, we compare different jointly symmetric copulas. For that purpose, we choose three different parameter values for each copula such that the mutual information of a copula is equal to the mutual information of the Gaussian copula with correlation $0.1, 0.3$, or 0.5. Thus, the columns of the following figures are comparable across different copulas. Despite apparent differences in their densities and simulated realizations, the implied conditional variances of different copulas are pretty similar. A closer examination reveals that the conditional variances of the Student-t copula with zero correlation parameter and the merged X-shaped versions of the Gaussian and Clayton copula can be well approximated by a quadratic function if the margins are Student-t with dof 6. This is not true for the patched X-shaped versions of the survival version of the Clayton copula and the Gumbel copula. The patched X-shaped versions of copulas place more probability mass in the center than the merged X-shaped versions of copulas and, by construction, the conditional variance is not differentiable at zero.

Student-t copula with zero correlation parameter and dof θ

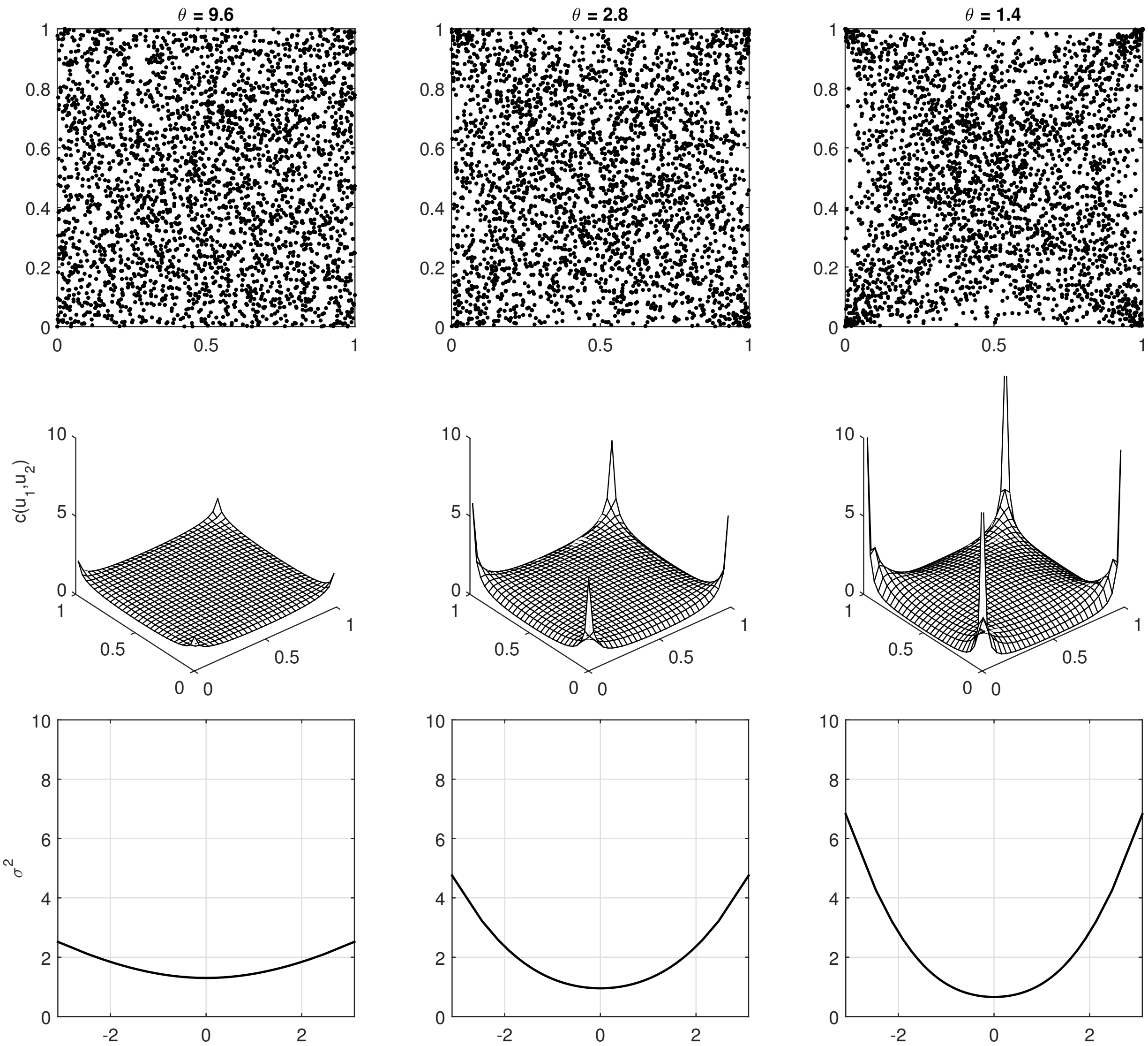


Figure 5.10: Illustration of the Student-t copula with zero correlation parameter and dof θ. The first row displays realizations from the copula whereas the second row depicts surface plots of its density. The third row shows the conditional variance of Y_1 given $Y_2 = y_2$ when (Y_1, Y_2) has this copula and Student-t margins with dof 6.

Merged X-shaped version of the Gaussian copula

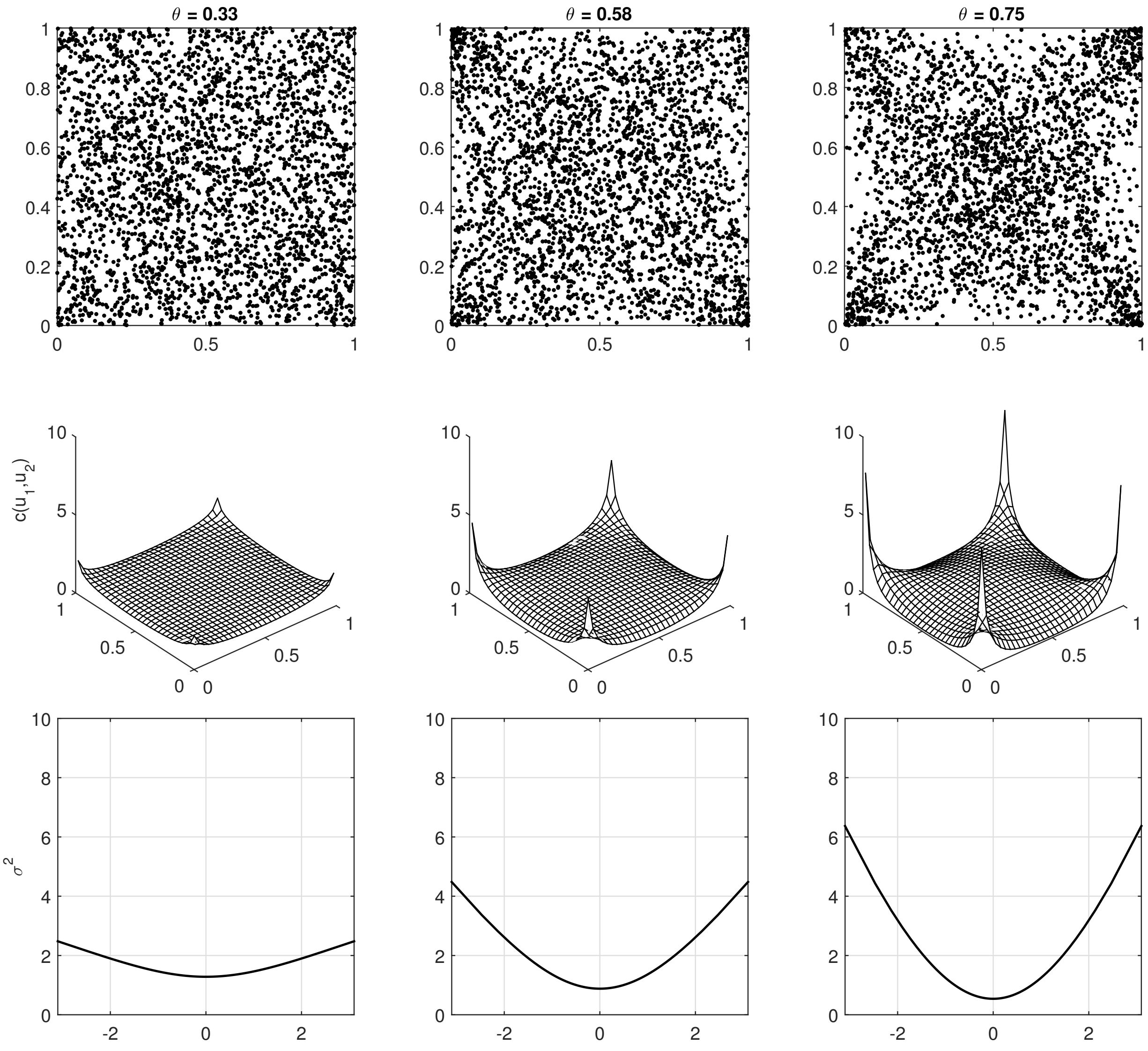


Figure 5.11: Illustration of the Merged X-shaped version of the Gaussian copula with correlation parameter θ. The first row displays realizations from the copula whereas the second row depicts surface plots of its density. The third row shows the conditional variance of Y_1 given $Y_2 = y_2$ when (Y_1, Y_2) has this copula and Student-t margins with dof 6.

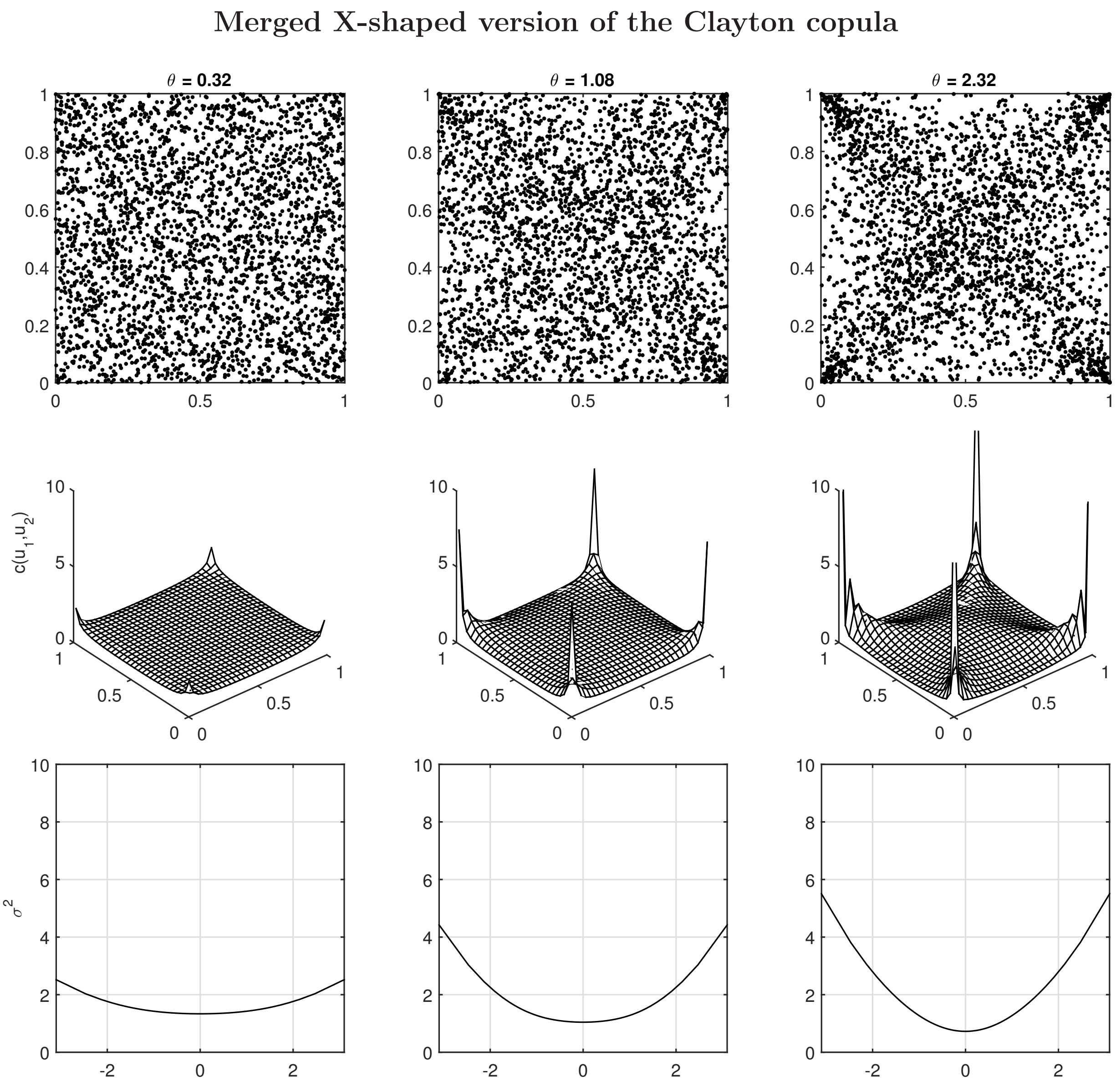


Figure 5.12: Illustration of the Merged X-shaped version of the Clayton copula with dependence parameter θ. The first row displays realizations from the copula whereas the second row depicts surface plots of its density. The third row shows the conditional variance of Y_1 given $Y_2 = y_2$ when (Y_1, Y_2) has this copula and Student-t margins with dof 6.

Patched X-shaped version of the survival version of the Clayton copula

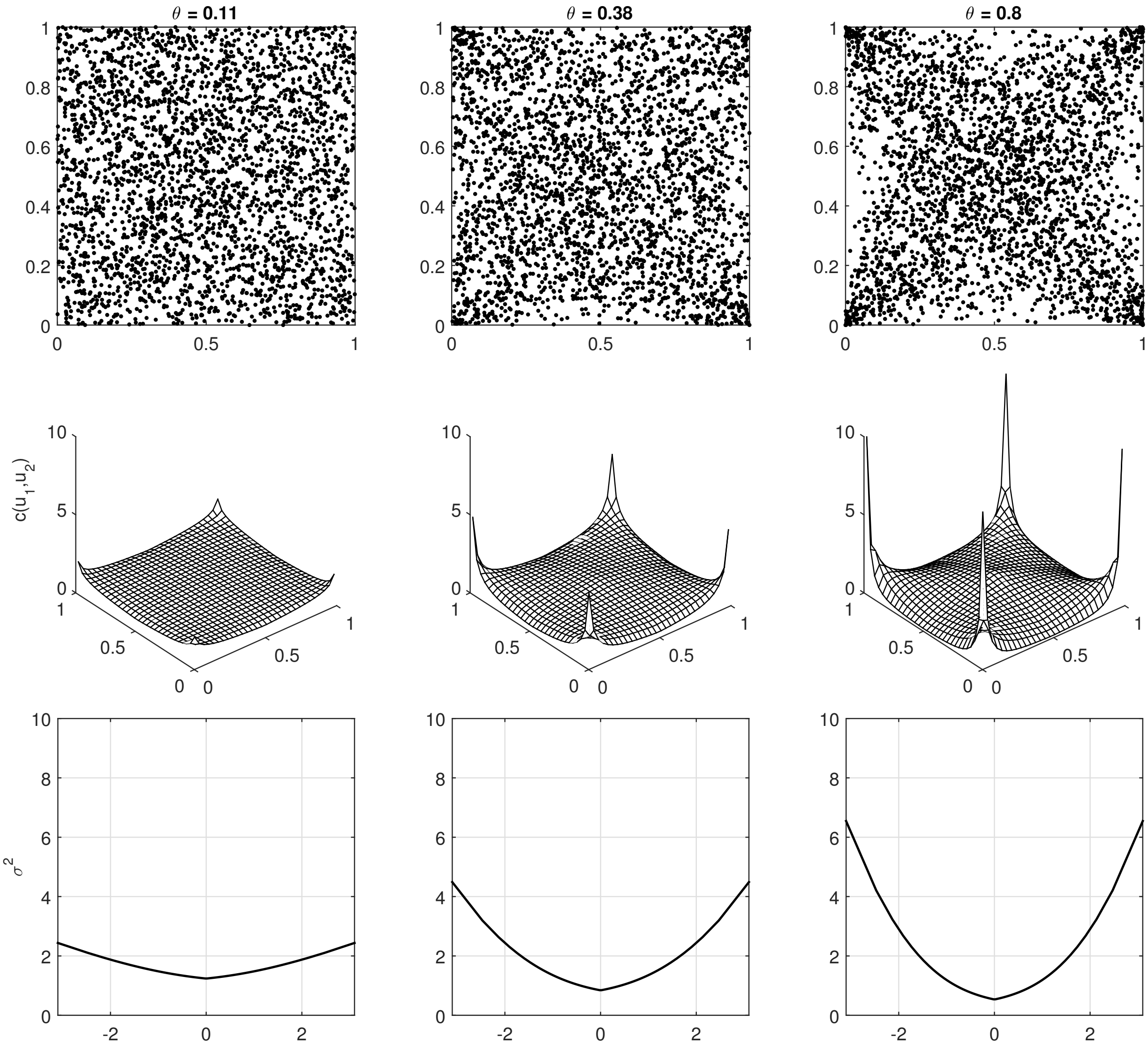


Figure 5.13: Illustration of the Patched X-shaped version of the survival version of the Clayton copula with dependence parameter θ. The first row displays realizations from the copula whereas the second row depicts surface plots of its density. The third row shows the conditional variance of Y_1 given $Y_2 = y_2$ when (Y_1, Y_2) has this copula and Student-t margins with dof 6.

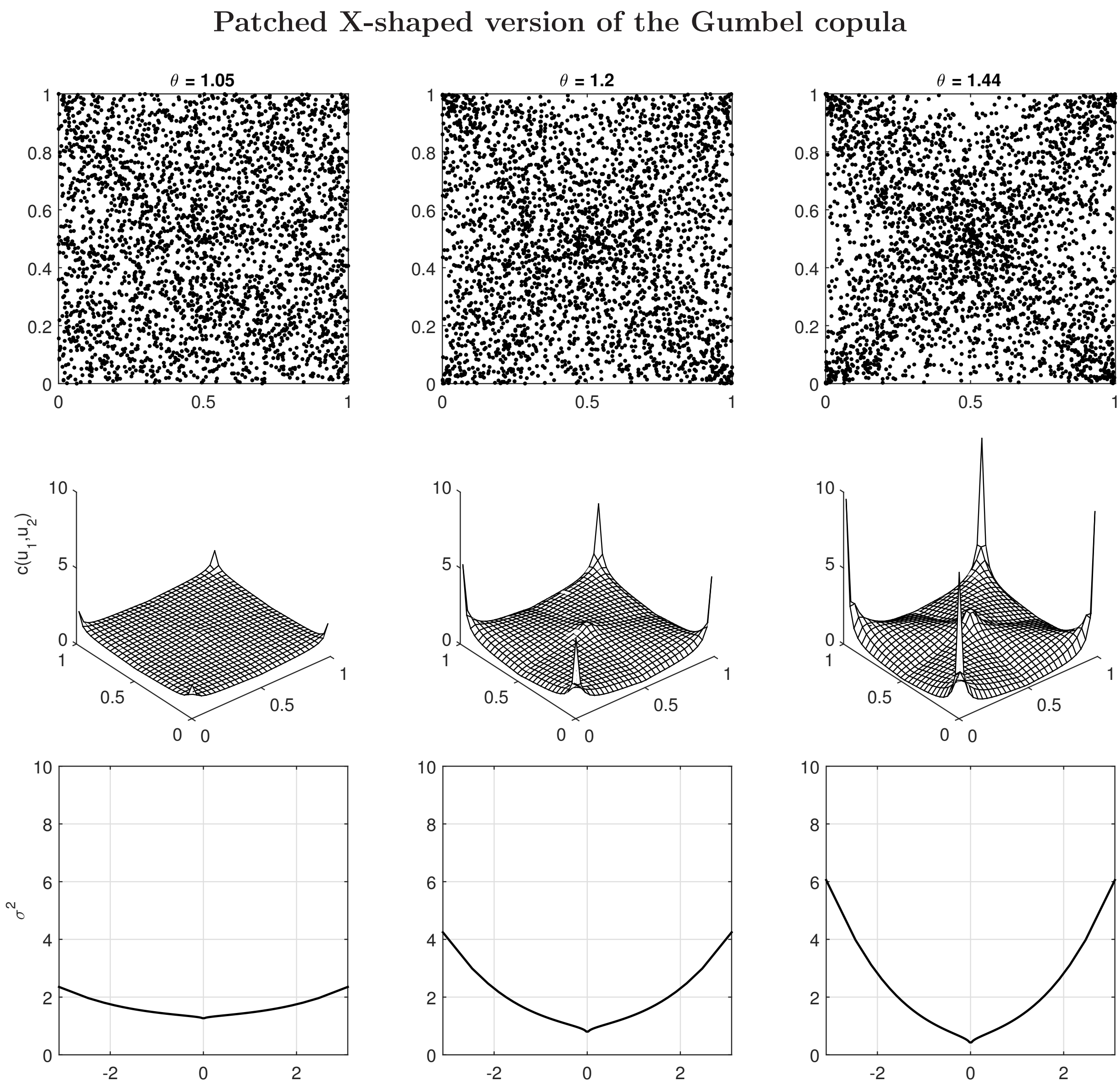


Figure 5.14: Illustration of the Patched X-shaped version of the Gumbel copula with dependence parameter θ. The first row displays realizations from the copula whereas the second row depicts surface plots of its density. The third row shows the conditional variance of Y_1 given $Y_2 = y_2$ when (Y_1, Y_2) has this copula and Student-t margins with dof 6.

6 Modeling financial returns with copula-based processes

6.1 Introduction

In this chapter, we investigate the fit of the copulas that are proposed in Chapter 5 for modeling clusters of extreme observations. Moreover, we model financial returns with simplified SD-vine copula-based Markov processes and evaluate their performance. For this purpose, we draw a comparison with GARCH models and illustrate the similarities and differences between both model classes.

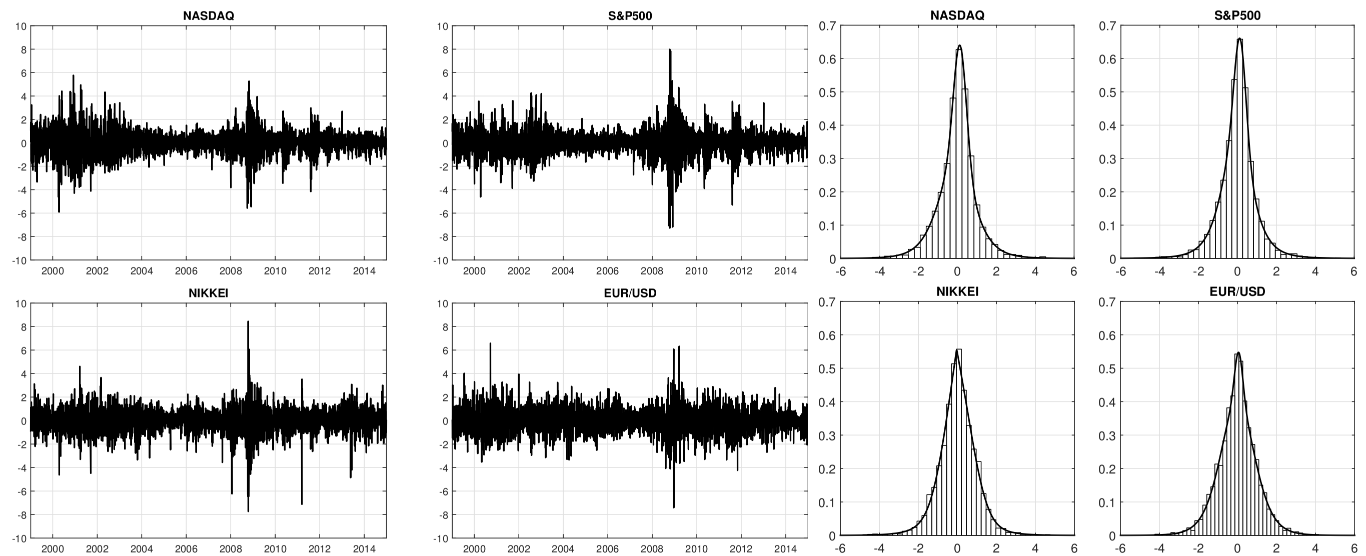
We use daily closing prices of the NASDAQ composite, the S&P 500 and the NIKKEI 225 stock index, which are obtained from Yahoo finance, and mid-market rates of the EUR/USD currency exchange rate, which are provided by the European Central Bank. The considered time period of the time series is January 4, 1999, to December 31, 2014. For the empirical analysis, we use AR(1)-filtered percentage log returns, i.e., if c is the constant and a the autoregressive coefficient of an estimated AR(1) model for the series of percentage log returns $(\tilde{Y}_t)_{t\in\mathbb{N}}$, then $Y_t = \tilde{Y}_t - (c + a\tilde{Y}_{t-1})$ is the AR(1)-filtered log return at period t.[1] In the following, we always refer to these AR(1)-filtered returns when we speak of financial returns.

Figure 6.1a shows the observed time series of financial returns. The returns on the three stock indices display very similar features, starting with a period of high volatility in the years 2000 to 2003, which is caused by the bursting of the dot-com bubble. The subsequent periods are characterized by a calm period with low volatility, until the financial crisis leads to a second period of large fluctuations which starts in the second half of the year 2008. The returns on the EUR/USD exchange rate does not exhibit such pronounced periods of low and large volatility. This is also confirmed by the sample autocorrelation functions of squared returns which are displayed in Figure 6.1c. All time series feature a significant autocorrelation for the first 20 lags which is slowly decaying. However, while the values of the autocorrelation functions of the squared returns cluster around 0.2 for the stock indices, the values for the EUR/USD exchange rate are much lower. The histograms of the financial returns in Figure 6.1b strongly indicate that there is a much larger probability for extreme observations and observations around zero than under the assumption of a normal distribution. Moreover, all marginal distributions are slightly skewed to the left. Figure 6.1d displays the copula scatter plots of consecutive returns. For the returns on the NASDAQ and S&P 500 index, the copula pseudo-observations seem to cluster more in the

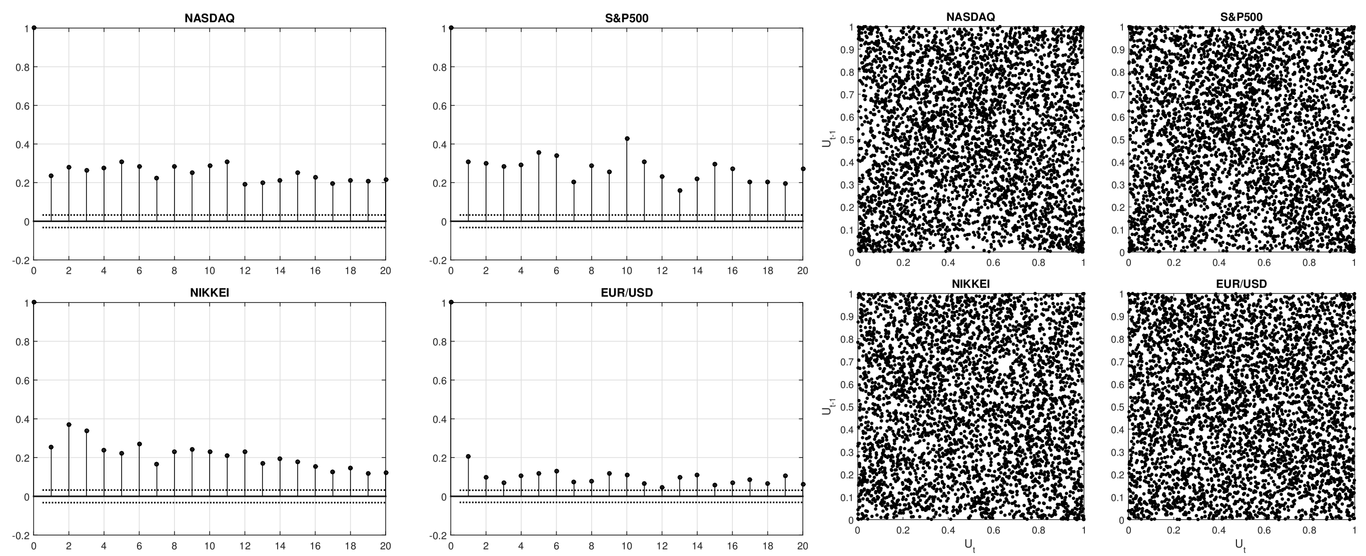
[1] Other models for the conditional mean such as MA(1) or ARMA(1,1) models do not result in a better fit.

Figure 6.1: Illustration of the time series of financial returns. For purposes of comparison, standardized financial returns with zero mean and unit variance are plotted in Figure 6.1b. The pseudo-observations in Figure 6.1d are estimated via the rescaled empirical distribution function.

(a) Time series of financial returns. **(b)** Histogram of returns & estimated marginal density.



(c) Sample autocorrelation function of squared financial returns. **(d)** Copula scatter plot of consecutive returns.



four corners of the unit cube than under the assumption of a product copula. Moreover, the clustering is more pronounced in the lower left and lower right corner of the unit cube, indicating that volatility increases more for a given absolute value of the lagged return if the return was negative. For the copula scatter plots of the returns on the NIKKEI index and the EUR/USD exchange rate no clear pattern is visible.

This chapter is structured as follows. Section 6.2 analyzes the performance of the copulas for modeling financial returns that are proposed in Chapter 5. We also investigate whether vertically symmetric copulas, which exhibit horizontal asymmetry, are preferred over jointly symmetric copulas. Section 6.3 addresses the specification and estimation of a CMP model of financial returns. The CMP model is compared with the GARCH model and its performance and properties are evaluated. Section 6.4 summarizes the main results.

6.2 Preliminary analysis

Specifying the marginal distribution

We now set up models for the marginal distribution of financial returns in order to specify a copula-based time series model. Since Hansen's skewed Student-t distribution (Hansen, 1994) does not provide an adequate fit, we consider mixture distributions with two components. In particular, we specify the Normal, Student-t, and the Laplace distribution as possible components of the mixture distributions. Table 6.1 reports the estimation results and shows that all marginal models pass Neyman's smooth test. Note that although the components of the mixture distributions are symmetric, the resulting mixture distributions are not symmetric, since the means of the components are not equal. The densities of the estimated mixture distributions are displayed in Figure 6.1b.

Table 6.1: Results for the estimated marginal mixture distributions. The last parameter of each distribution (Normal, Student-t, Laplace) is the corresponding mixing weight. $\log \mathcal{L}$ and AIC denote the log-likelihood and the AIC of the fitted mixture distribution. p-val is the p-value of Neyman's smooth test with four components.

	NASDAQ	S&P500	NIKKEI	EUR/USD
Normal (μ, σ^2, w_1)	$(0.267, 0.574, 0.303)$	$(0.191, 0.391, 0.232)$	$(0.115, 1.241, 0.519)$	$(0.037, 0.132, 0.076)$
Student-t $(\mu, \sigma^2, \nu, w_2)$	$(-0.092, 1.612, 5.372, 0.697)$	$(-0.031, 1.040, 3.917, 0.768)$	-	$(-0.002, 0.568, 7.643, 0.924)$
Laplace (μ, σ^2, w_3)	-	-	$(-0.054, 1.262, 0.481)$	-
$\log \mathcal{L}$	-7113.528	-5919.416	-6840.869	-3844.383
AIC	14239.057	11850.832	13691.739	7700.766
p-val	0.919	0.903	0.528	0.854

Jointly versus vertically symmetric copulas

Before we consider the full specification of a CMP model for financial returns, we investigate and compare the performance of the proposed copulas in Chapter 5 that are constructed for modeling volatility clustering. To this end, we take a closer look at the first two elements of the copula sequence. We first analyze the performance of jointly symmetric copulas and then investigate whether the use of vertically symmetric but horizontally asymmetric copulas improves the fit.

The first eights rows of Table 6.2 explain the jointly symmetric copulas that we consider. In addition to merged and patched X-shaped versions of copulas, we use the Student-t copula with zero correlation parameter and the GJR-ARCH(1)-like copula without leverage. Table 6.3 reports the estimation results of the jointly symmetric copulas for the first copula C_1 of the copula sequence. For the stock indices, the Student-t copula with zero correlation

Table 6.2: Explanation of jointly and vertically symmetric copula families. If D is a copula family then D^{VS} refers to its merged vertically symmetric version i.e., $C^{\mathrm{D}^{\mathrm{VS}}}(u_1,u_2) = 0.5C^{\mathrm{D}}(u_1,u_2) + 0.5C^{\text{V-D}}(u_1,u_2)$, where V-D is its vertically reflected version.

Family	Explanation		
t-zero	Student-t copula with zero correlation parameter and dof θ.		
$\mathrm{Ga}^{\mathrm{MX}}$	merged X-shaped version of the Gaussian copula.		
$\mathrm{Cl}^{\mathrm{MX}}$	merged X-shaped version of the Clayton copula.		
$\mathrm{Gu}^{\mathrm{MX}}$	merged X-shaped version of the Gumbel copula.		
$\mathrm{Fr}^{\mathrm{MX}}$	merged X-shaped version of the Frank copula.		
$(\text{S-Cl})^{\mathrm{PX}}$	patched X-shaped version of the survival Clayton copula.		
$\mathrm{Gu}^{\mathrm{PX}}$	patched X-shaped version of the Gumbel copula.		
GJR-Sym	GJR-ARCH(1)-like copula without leverage which is derived from $Y = \sqrt{\theta_1 + \theta_2 X^2}\mathcal{E}$, with a standard logistic distribution for $\mathcal{E}$ and X.		
		Family C_1	Family C_2
$(\text{t-zero})\text{-}\mathrm{Cl}^{\mathrm{VS}}$	A copula C which results from a convex combination of a jointly symmetric copula C_1 and a copula C_2 which is only vertically symmetric. The cdf of the copula C is given by $C(u_1,u_2) = \theta_3 C_1(u_1,u_2;\theta_1) + (1-\theta_3)C_2(u_1,u_2;\theta_2)$, see (5.3.9).	t-zero	$\mathrm{Cl}^{\mathrm{VS}}$
$\mathrm{Ga}^{\mathrm{MX}}\text{-}\mathrm{Cl}^{\mathrm{VS}}$		$\mathrm{Ga}^{\mathrm{MX}}$	$\mathrm{Cl}^{\mathrm{VS}}$
$\mathrm{Cl}^{\mathrm{MX}}\text{-}\mathrm{Cl}^{\mathrm{VS}}$		$\mathrm{Cl}^{\mathrm{MX}}$	$\mathrm{Cl}^{\mathrm{VS}}$
$\mathrm{Gu}^{\mathrm{MX}}\text{-}\mathrm{Cl}^{\mathrm{VS}}$		$\mathrm{Gu}^{\mathrm{MX}}$	$\mathrm{Cl}^{\mathrm{VS}}$
$(\text{S-Cl})^{\mathrm{PX}}\text{-}\mathrm{Cl}^{\mathrm{VS}}$		$(\text{S-Cl})^{\mathrm{PX}}$	$\mathrm{Cl}^{\mathrm{VS}}$
$\text{t-zero-}(\text{S-Gu})^{\mathrm{VS}}$		t-zero	$(\text{S-Gu})^{\mathrm{VS}}$
$\mathrm{Ga}^{\mathrm{MX}}\text{-}(\text{S-Gu})^{\mathrm{VS}}$		$\mathrm{Ga}^{\mathrm{MX}}$	$(\text{S-Gu})^{\mathrm{VS}}$
$\mathrm{Cl}^{\mathrm{MX}}\text{-}(\text{S-Gu})^{\mathrm{VS}}$		$\mathrm{Cl}^{\mathrm{MX}}$	$(\text{S-Gu})^{\mathrm{VS}}$
$\mathrm{Gu}^{\mathrm{MX}}\text{-}(\text{S-Gu})^{\mathrm{VS}}$		$\mathrm{Gu}^{\mathrm{MX}}$	$(\text{S-Gu})^{\mathrm{VS}}$
$(\text{S-Cl})^{\mathrm{PX}}\text{-}(\text{S-Gu})^{\mathrm{VS}}$		$(\text{S-Cl})^{\mathrm{PX}}$	$(\text{S-Gu})^{\mathrm{VS}}$
$(\text{Ga-Ga})^{\mathrm{VW}}$	A copula C which arises from a convex combination of two jointly symmetric copulas with varying mixing weights, i.e., the density is given by $c(u_1,u_2) = \theta_3 u_2 c_1(u_1,u_2;\theta_1) + (1-\theta_3 u_2)c_2(u_1,u_2;\theta_2)$, see Proposition 5.11.	$\mathrm{Ga}^{\mathrm{MX}}$	$\mathrm{Ga}^{\mathrm{MX}}$
$(\text{Cl-Ga})^{\mathrm{VW}}$		$\mathrm{Cl}^{\mathrm{MX}}$	$\mathrm{Ga}^{\mathrm{MX}}$
$(\text{Ga-Gl})^{\mathrm{VW}}$		$\mathrm{Ga}^{\mathrm{MX}}$	$\mathrm{Cl}^{\mathrm{MX}}$
GJR	The GJR-ARCH(1)-like copula C defined in Proposition 5.10 with a standard logistic distribution for $\mathcal{E}$ and X, i.e. C is derived from $Y = \sqrt{\theta_1 + \theta_2 X^2 + \mathbb{1}_{\{X<0\}}\theta_3 X^2}\mathcal{E}$.		

Table 6.3: Specification of C_1: Estimation results for jointly symmetric copulas. The best AIC value for each asset is given in bold. For an explanation of the copula families see Table 6.2.

	NASDAQ		S&P500		NIKKEI		EUR/USD	
family	θ	AIC	θ	AIC	θ	AIC	θ	AIC
t-zero	3.895	-198.376	4.760	**-150.401**	9.093	-61.802	9.571	-44.455
$\mathrm{Ga}^{\mathrm{MX}}$	0.500	-197.124	0.444	-137.022	0.330	-53.340	0.319	-39.728
$\mathrm{Cl}^{\mathrm{MX}}$	0.692	-184.308	0.568	-148.023	0.310	**-64.154**	0.314	**-45.325**
$\mathrm{Gu}^{\mathrm{MX}}$	1.360	-191.250	1.272	-146.451	1.123	-58.893	1.132	-44.982
$\mathrm{Fr}^{\mathrm{MX}}$	3.666	-130.796	2.923	-60.417	1.651	-5.934	2.059	-16.303
$(\text{S-Cl})^{\mathrm{PX}}$	0.269	**-198.662**	0.213	-141.146	0.107	-51.877	0.103	-40.184
$\mathrm{Gu}^{\mathrm{PX}}$	1.133	-185.159	1.103	-143.070	1.043	-51.813	1.049	-44.488
GJR-Sym	(0.504,0.221)	-189.202	(1.203,0.400)	-143.359	(0.484,0.067)	-58.682	(0.104,0.014)	-42.351

parameter is among the best two copulas wrt the AIC value, and also for the EUR/USD exchange rate, the AIC value of the Student-t copula with zero correlation parameter is almost indistinguishable from the smallest AIC value which is attained by the merged X-shaped version of the Clayton copula. Except for the merged X-shaped version of the Frank copula, the AIC values of the proposed copulas are comparable. The merged X-shaped version of the Frank copula performs poorly since a Frank copula with the PQD property places less probability mass in the lower left and upper right corner of the unit cube than the other copulas. As a result, the probability of joint extreme observations is lower for the merged X-shaped version of the Frank copula. This property is not supported by the data which exhibits a strong clustering of joint extreme observations.

For each asset we select for C_1 the jointly symmetric copula in Table 6.3 which yields the lowest AIC and compute the pseudo-observations from C_2. Table 6.4 displays the estimation results of the jointly symmetric copulas for the second copula C_2 of the copula sequence. The Student-t copula with zero correlation is still a strong competitor among the proposed copulas but outperformed by the patched X-shaped version of the survival Clayton copula for the NASDAQ and S&P 500 stock index. Altogether, the AIC values of the proposed copulas are similar, except for merged X-shaped versions of the Clayton and Frank copula which exhibit larger AIC values in general. It is striking that the AIC values

Table 6.4: Specification of C_2: Estimation results for jointly symmetric copulas. The best AIC value for each asset is given in bold. For an explanation of the copula families see Table 6.2.

	NASDAQ		S&P500		NIKKEI		EUR/USD	
family	θ	AIC	θ	AIC	θ	AIC	θ	AIC
t-zero	3.315	-258.423	3.000	-234.958	5.662	**-131.236**	13.487	-17.191
Ga^{MX}	0.545	-266.534	0.564	-241.640	0.424	-124.114	0.275	-17.702
Cl^{MX}	0.812	-232.998	0.877	-207.829	0.483	-122.591	0.254	-15.650
Gu^{MX}	1.439	-253.590	1.474	-227.640	1.239	-125.020	1.107	-16.801
Fr^{MX}	4.204	-204.888	4.242	-195.115	2.845	-58.839	2.143	-19.597
$(\text{S-Cl})^{\text{PX}}$	0.323	**-269.525**	0.353	**-248.080**	0.181	-125.612	0.075	-18.569
Gu^{PX}	1.163	-251.932	1.179	-237.883	1.082	-115.193	1.041	**-20.043**
GJR-Sym	(0.931,0.586)	-259.607	(0.452,0.400)	-228.814	(1.991,0.559)	-127.540	(0.103,0.013)	-14.316

of the fitted copulas for the stock indices are much lower for C_2 than for C_1. Thus, under the simplifying assumption, Y_t and Y_{t-2} conditional on Y_{t-1} are more dependent in terms of mutual information than Y_t and Y_{t-1}. In this regard, Y_{t-2} seems to be more important for explaining the transition distribution of financial returns than Y_{t-1}. For the NASDAQ and NIKKEI stock index this is also confirmed by the sample autocorrelation function of squared returns in Figure 6.1c. The correlation between Y_t^2 and Y_{t-2}^2 is larger than the correlation between Y_t^2 and Y_{t-1}^2.

We now turn to the estimation of vertically symmetric copulas which are not jointly symmetric and give rise to an asymmetric response of the conditional variance to positive and negative lagged returns. The rows 9-23 of Table 6.2 explain the vertically symmetric copulas that we take into consideration. The used copulas can be separated into three

Table 6.5: Specification of C_1: Estimation results for vertically symmetric copulas. The best AIC value for each asset is given in bold. For an explanation of the copula families see Table 6.2.

	NASDAQ		S&P500		NIKKEI		EUR/USD	
family	θ	AIC	θ	AIC	θ	AIC	θ	AIC
(t-zero)-Cl^{VS}	(0.611,2.380,0.389)	-222.833	(0.606,2.921,0.394)	**-164.460**	(0.718,5.098,0.282)	-70.227	(0.406,5.961,0.594)	-41.097
Ga^{MX}-Cl^{VS}	(0.639,-0.661,0.361)	**-227.658**	(0.628,-0.559,0.372)	-156.435	(0.684,0.405,0.316)	-66.286	(0.604,0.495,0.396)	-36.446
Cl^{MX}-Cl^{VS}	(0.610,0.934,0.390)	-211.388	(0.507,0.677,0.493)	-160.247	(0.586,0.375,0.414)	**-70.333**	(0.000,0.314,1.000)	-41.325
Gu^{MX}-Cl^{VS}	(0.635,1.608,0.365)	-219.208	(0.548,1.370,0.452)	-160.323	(0.631,1.159,0.369)	-68.212	(0.000,1.132,1.000)	-40.982
$(S\text{-}Cl)^{PX}$-Cl^{VS}	(0.610,0.478,0.390)	-226.806	(0.687,0.455,0.313)	-162.044	(0.829,0.293,0.171)	-67.040	(0.645,0.288,0.355)	-38.976
t-zero-$(S\text{-}Gu)^{VS}$	(0.607,2.877,0.393)	-210.725	(0.000,4.760,1.000)	-146.401	(0.583,6.544,0.417)	-62.769	(0.000,9.572,1.000)	-40.455
Ga^{MX}-$(S\text{-}Gu)^{VS}$	(0.636,0.627,0.364)	-214.636	(0.678,-0.549,0.322)	-151.890	(0.626,-0.388,0.374)	-59.481	(0.000,-0.319,1.000)	-35.728
Cl^{MX}-$(S\text{-}Gu)^{VS}$	(0.684,0.768,0.316)	-203.247	(0.523,0.591,0.477)	-154.788	(0.404,0.320,0.596)	-62.958	(0.000,0.314,1.000)	-41.325
Gu^{MX}-$(S\text{-}Gu)^{VS}$	(0.677,1.522,0.323)	-207.007	(0.582,1.329,0.418)	-154.275	(0.549,1.137,0.451)	-60.089	(0.000,1.132,1.000)	-40.982
$(S\text{-}Cl)^{PX}$-$(S\text{-}Gu)^{VS}$	(0.605,0.409,0.395)	-213.956	(0.700,0.406,0.300)	-156.037	(0.767,0.221,0.233)	-59.322	(0.632,0.284,0.368)	-38.884
$(Ga\text{-}Ga)^{VW}$	(0.281,0.598,0.768)	-214.682	(0.255,0.502,0.622)	-139.631	(0.261,0.381,1.000)	-53.032	(0.813,0.264,0.100)	-39.467
$(Cl\text{-}Ga)^{VW}$	(0.472,0.601,1.000)	-220.120	(0.452,0.504,1.000)	-148.202	(0.253,0.382,1.000)	-58.576	(0.446,0.275,0.631)	-42.227
$(Ga\text{-}Gl)^{VW}$	(0.377,0.818,0.648)	-185.742	(0.136,0.619,0.250)	-145.098	(0.100,0.333,0.240)	-60.521	(0.487,0.274,0.259)	**-42.681**
GJR	(0.822,0.200,0.394)	-218.827	(2.000,0.410,0.599)	-160.281	(1.905,0.144,0.240)	-67.828	(1.906,0.265,0.003)	-42.392

Table 6.6: Specification of C_2: Estimation results for vertically symmetric copulas. The best AIC value for each asset is given in bold. For an explanation of the copula families see Table 6.2.

	NASDAQ		S&P500		NIKKEI		EUR/USD	
family	θ	AIC	θ	AIC	θ	AIC	θ	AIC
(t-zero)-Cl^{VS}	(0.232,3.493,0.768)	-272.270	(0.274,3.247,0.726)	-242.727	(0.485,3.600,0.515)	-140.099	(0.880,1.236,0.120)	-15.256
Ga^{MX}-Cl^{VS}	(0.482,-0.662,0.518)	-284.649	(0.484,0.664,0.516)	**-258.357**	(0.518,-0.530,0.482)	-138.006	(0.000,0.277,1.000)	-13.089
Cl^{MX}-Cl^{VS}	(0.452,1.030,0.548)	-254.859	(0.398,0.864,0.602)	-222.619	(0.498,0.694,0.502)	-132.729	(0.032,0.194,0.968)	-14.088
Gu^{MX}-Cl^{VS}	(0.472,1.664,0.528)	-273.330	(0.444,1.599,0.556)	-239.971	(0.000,1.266,1.000)	-126.011	(0.872,2.207,0.128)	-18.002
$(S\text{-}Cl)^{PX}$-Cl^{VS}	(0.254,0.327,0.746)	-283.523	(0.286,0.357,0.714)	-256.733	(0.525,0.299,0.475)	-138.631	(0.861,0.791,0.139)	-16.883
t-zero-$(S\text{-}Gu)^{VS}$	(0.369,4.097,0.631)	-274.231	(0.311,3.960,0.689)	-242.698	(0.322,5.000,0.678)	-135.992	(0.047,24.876,0.953)	-18.544
Ga^{MX}-$(S\text{-}Gu)^{VS}$	(0.471,-0.603,0.529)	-281.951	(0.473,-0.639,0.527)	-253.244	(0.443,-0.485,0.557)	-134.421	(0.045,-0.209,0.955)	-18.815
Cl^{MX}-$(S\text{-}Gu)^{VS}$	(0.527,0.542,0.473)	-263.960	(0.485,0.542,0.515)	-230.926	(0.390,0.468,0.610)	-129.050	(0.819,1.476,0.181)	-15.256
Gu^{MX}-$(S\text{-}Gu)^{VS}$	(0.537,1.559,0.463)	-271.238	(0.475,1.487,0.525)	-236.692	(0.471,1.318,0.529)	-131.221	(0.872,2.207,0.128)	-18.002
$(S\text{-}Cl)^{PX}$-$(S\text{-}Gu)^{VS}$	(0.348,0.300,0.652)	-281.832	(0.286,0.321,0.714)	-253.269	(0.438,0.230,0.562)	-134.610	(0.861,0.791,0.139)	-16.884
$(Ga\text{-}Ga)^{VW}$	(0.478,0.610,1.000)	-277.043	(0.472,0.598,1.000)	-249.786	(0.169,0.490,0.483)	-130.473	(0.899,0.211,0.100)	**-19.434**
$(Cl\text{-}Ga)^{VW}$	(0.247,0.602,0.450)	-274.354	(0.292,0.596,0.496)	-249.765	(0.432,0.493,1.000)	-137.856	(2.080,0.241,0.110)	-16.031
$(Ga\text{-}Gl)^{VW}$	(0.502,1.055,1.000)	-261.476	(0.506,0.940,1.000)	-227.728	(0.357,0.566,0.378)	-122.864	(0.299,0.221,1.000)	-12.525
GJR	(0.500,0.200,0.270)	**-291.008**	(0.530,0.215,0.207)	-253.800	(1.999,0.362,0.367)	**-142.889**	(1.974,0.190,0.008)	-13.351

different classes. The largest class consists of convex combinations of jointly symmetric copulas and copulas which are only vertically symmetric, see rows 9-19 in Table 6.2. Apart from that, we consider convex combinations of two jointly symmetric copulas with varying mixing weights, see rows 19-22 in Table 6.2. The last class is given by the GJR-ARCH(1)-like copula with leverage effect, see row 23 in Table 6.2.

Table 6.5 contains the estimation results of the vertically symmetric copulas for the first copula C_1 of the copula sequence. Overall, the GJR(1)-ARCH-like copula and convex combinations of a merged or patched X-shaped version of a copula and a vertically symmetric version of the Clayton copula constitute the best performing copulas in terms of the AIC. For the stock indices, convex combinations of a merged X-shaped version of a copula and a vertically symmetric version of the Clayton copula provide the best fit. The performance of copulas which are based on a convex combination with varying weights is not convincing, except for the EUR/USD exchange rate where all copulas perform equally well.

The introduction of horizontal asymmetry considerably improves the fit of the copulas for the stock indices. For the NASDAQ index, the AIC value of the best jointly symmetric copula for C_1 is -198.662, while the AIC value of the best vertically symmetric copula is -227.658. On the other side, we do not find evidence for horizontal asymmetry of C_1 for the EUR/USD exchange rate. The AIC value for the best vertically symmetric copula is -42.681 which is larger than the AIC value of the best jointly symmetric copula which is given by -45.325. Moreover, the estimated leverage parameter of the GJR-ARCH(1)-like copula is close to zero which indicates that there is no significant leverage effect present for C_1.

For each asset we select for C_1 the vertically symmetric copula in Table 6.5 which yields the lowest AIC and compute the pseudo-observations from C_2. Table 6.6 displays the estimation results of the vertically symmetric copulas for the second copula C_2 of the copula sequence. For the stock indices, the GJR(1)-ARCH-like copula is in two out of three cases the best copula in terms of AIC. A strong competitor is the mixture of a merged X-shaped version of the Gaussian copula and the vertically symmetric version of the Clayton copula, which gives the lowest AIC value for the S&P 500 stock index. For the stock indices, similar to the case of jointly symmetric copulas, the estimated vertically symmetric copulas for C_2 have a larger mutual information than the estimated vertically symmetric copulas for C_1. Moreover, the AIC values of the estimated copulas for C_2 that allow for a non-symmetric conditional variance are much lower than the AIC values of the estimated jointly symmetric copulas. The exception is once again the EUR/USD exchange rate. According to the AIC, the best jointly symmetric copula is preferred over the best vertically symmetric copula, indicating that a significant leverage effect is not present in the first two elements of the copula sequence.

All in all, the proposed jointly and vertically symmetric copulas in Chapter 5 perform well in comparison with the GJR-ARCH(1)-like copula. For the stock indices, copulas that allow for a different response of the conditional variance to positive and negative returns lead to a substantial decrease in the AIC values of the fitted copulas for the first two elements of the copula sequence. This is not true for the EUR/USD exchange rate where, according to the AIC, jointly symmetric copulas are superior over vertically symmetric copulas. A further analysis of the copula sequences of the four time series shows that the following copulas $C_i, i \geq 3$, display no significant horizontal asymmetry and can be modeled with jointly symmetric copulas.

Analyzing the conditional variance of the CMP(2) models

Section 5.4.3 shows that $\sigma_t^2 = \mathbb{V}\text{ar}[Y_t | Y_{t-1:t-2} = y_{t-1:t-2}]$, which is the conditional variance of a CMP(2) model, may not be increasing in the absolute value of y_{t-1} if at least one copula of the SD-vine copula (C_1, C_2) does not exhibit the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property and the amount of horizontal asymmetry is strongly pronounced. We now investigate whether this is also the case for the three stock indices since the best CMP(2) models contain at least one copula which does not have the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property, see Table 6.5 and Table 6.6.

Figure 6.2: Conditional variances of the CMP(2) models as a function of y_{t-1} if y_{t-2} attains its 1%, 10%, 50%, 90%, or 99% quantile. The marginal distribution of a CMP(2) model is specified in Table 6.1 and the elements of its SD-vine copula (C_1, C_2) are given by the estimated copulas in Table 6.5 and Table 6.6 that yield the lowest AIC values.

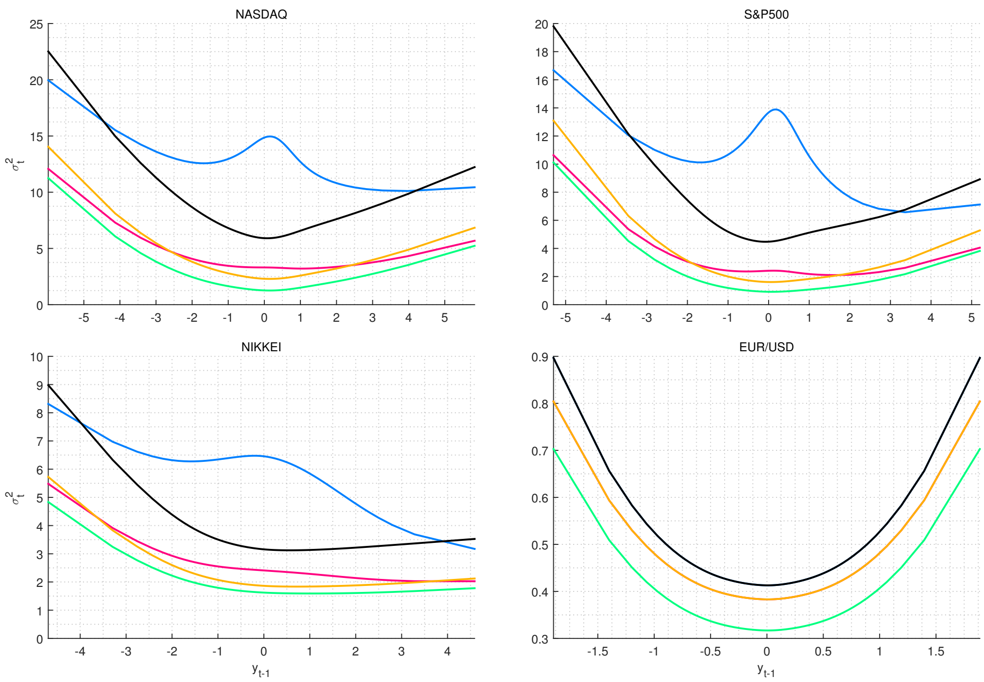


Figure 6.2 shows for all four time series the conditional variances of the CMP(2) models. For the EUR/USD exchange rate, the conditional variance seems to be increasing in the absolute value of y_{t-1}. In this case, both copulas of the SD-vine copula are jointly symmetric and exhibit the $\text{SI}_{(1,3)}$-$\text{SD}_{(2,4)}$ property. For the stock indices, the conditional variances are not always increasing in the absolute value of y_{t-1}. If y_{t-2} attains its 1% quantile, the

conditional variances have a local maximum near $y_{t-1} = 0$. Moreover, for the NIKKEI stock index, the conditional variance appears to be monotonically decreasing in $y_{t-1} > 0$ if y_{t-2} attains its 1% quantile, which is counterintuitive from an economic point of view.

There are several reasons that might explain why the conditional variances of the CMP(2) models of the stock indices are not increasing in the absolute value of y_{t-1}. First of all, at least one of the chosen copula families of the SD-vine copulas does not have the $SI_{(1,3)}$-$SD_{(2,4)}$ property. However, the choice of two GJR-ARCH(1)-like copulas that exhibit the $SI_{(1,3)}$-$SD_{(2,4)}$ property is not supported by the data, according to the AIC.[2] Although it may be possible that other horizontally asymmetric copulas with the $SI_{(1,3)}$-$SD_{(2,4)}$ property perform better, it seems unlikely that one can find other parametric copulas that clearly outperform the copulas that are considered in this study. The use of simplified vine copula approximations to model the SD-vine copulas could also explain the non-increasing behavior of the conditional variances. It may be possible to decrease the likelihood of non-increasing conditional variances if the SD-vine copulas are modeled by non-simplified vine copula models.

On the other side, we can not exclude the possibility that the implied conditional variances reflect the conditional variances of the data generating processes. It may be possible that the conditional variance of the NASDAQ index is reduced if y_{t-2} attains its 1% quantile and y_{t-1} becomes more positive. Finally, the conditional variance may not be a good measure to assess the likeliness of extreme observations. For instance, it may be true that the conditional variance decreases but that the difference between the 90% and the 10% quantile increases. For this reason, plots of the conditional quantiles of the CMP(2) models as a function of y_{t-1} for given y_{t-2} are illustrated in Appendix A.1. If y_{t-2} is larger than its 10% quantile, the conditional quantiles are more spread out if the absolute value of y_{t-1} increases. However, if y_{t-2} attains its 1% quantile, the differences between the conditional quantiles of the CMP(2) models also have a local maximum if y_{t-1} is close to zero. Additional plots of the conditional densities of the CMP(2) models are given in Appendix A.2.

[2] It may be possible that GJR-ARCH(1)-like copulas provide a better fit for the data if y_{t-2} is smaller than its 1% quantile, but that the fit is worse if y_{t-2} is larger than its 1% quantile, so that the overall fit is inferior.

6.3 Modeling financial returns with CMP models

6.3.1 Specification and estimation of the CMP models

We set $P = 100$ as maximal order and estimate truncated simplified SD-vine copula-based Markov models with lag functions. Due to the excessive computational demand, we exclude the GJR-ARCH(1)-like copula as a possible member of the copula sequences.[3]

For each time series, we choose for C_1 and C_2 the copula families in Table 6.2, excluding the GJR-ARCH(1)-like copula, that yield the smallest AIC values (see Tables 6.3-6.6). The next elements $(C_i)_{i=3,\dots,100}$ of the estimated partial autocopula sequences are almost always given by jointly symmetric copulas if the selection of the copula families is based on the AIC. Moreover, the mutual information is roughly decreasing in $3 \leq i \leq 100$. Thus, it is reasonable to parameterize the copulas $(C_i)_{i=3,\dots,100}$ of the CMP models by a lag function. To this end, it is necessary that the copulas in $(C_i)_{i=3,\dots,100}$ are of the same family, since the parameter values of the considered copula families are not equal for the same mutual information.[4] As a result, a smooth lag function can not be used to effectively capture the decrease in mutual information if different copula families are used. Therefore, we specify for each time series only one copula family for $(C_i)_{i=3,\dots,100}$. For all jointly symmetric copula families that are mentioned in Table 6.2, excluding the GJR-Sym copula, we perform a joint ML estimation of the resulting model of $(C_i)_{i=3,\dots,100}$ and select the copula family that yields the lowest AIC value. In all cases, the merged X-shaped version of the Gaussian copula is selected as the family of $(C_i)_{i=3,\dots,100}$.

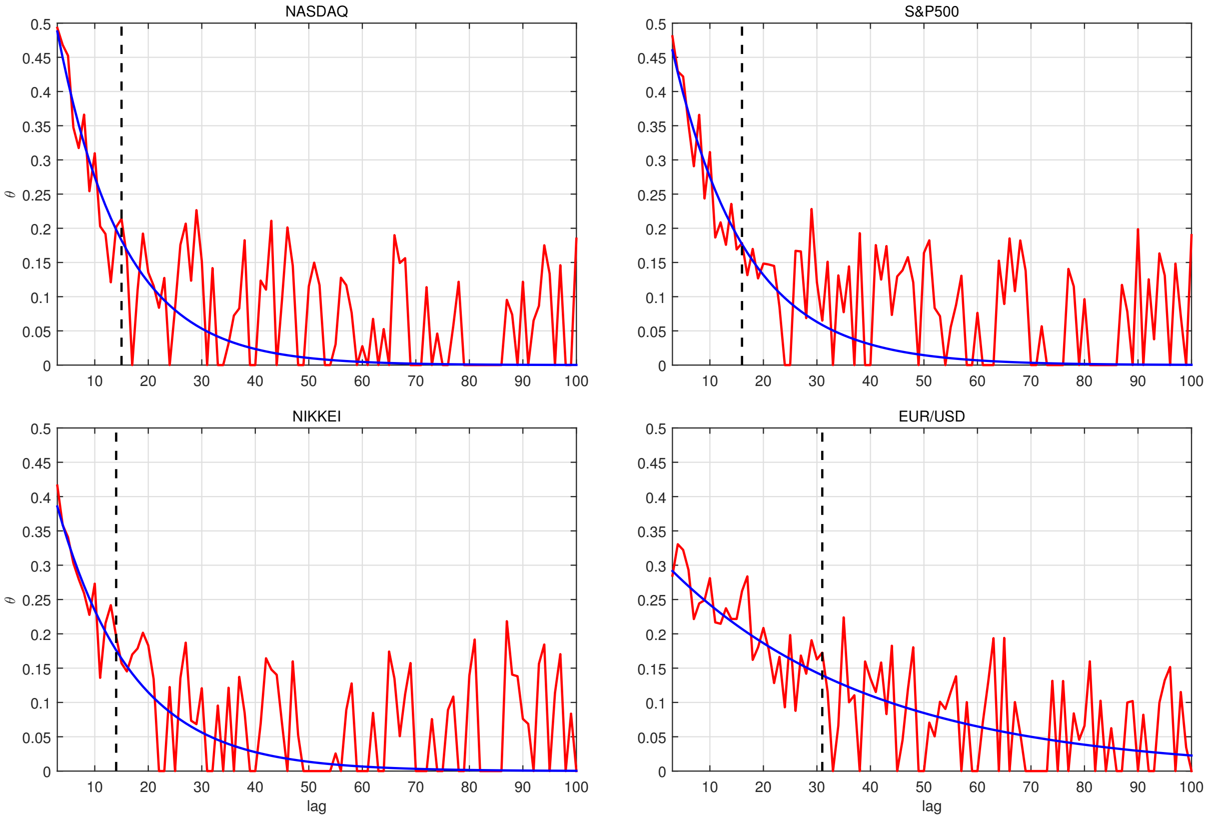
Figure 6.3a shows the estimated parameters θ_i of the sequence $(C_i)_{i=3,\dots,100}$, that are obtained if a joint ML estimation is performed and merged X-shaped versions of the Gaussian copula are used. For all time series, the estimated parameters θ_i show a strong fluctuation within the interval $[0, 0.2)$ for $i \geq 30$, which is persistent and does not decline with the lag. That is because the mutual information of the fitted merged X-shaped versions of the Gaussian copula is very close to zero for $i \geq 30$. A small change in the mutual information then leads to a large change in the dependence parameter of the merged X-shaped version of the Gaussian copula, resulting in strongly varying parameter estimates for $i \geq 30$. Figure 6.3b depicts the corresponding cumulated average log-likelihoods and shows that the increase in the cumulated average log-likelihoods is minor for $i \geq 30$.

[3] The effective calculation time for estimating a GJR-ARCH(1)-like copula using the MATLAB programming language takes quite an amount of time but is acceptable if one only estimates a couple of GJR-ARCH(1)-like copulas. However, numerical approximations or the implementation of its density and partial derivatives in a low-level language are required so that the estimation of CMP models with the GJR-ARCH(1)-like copula are feasible. In two out of eight cases, the GJR-ARCH(1)-like copula is chosen as the best copula, according to Table 6.5 and Table 6.6. However, the differences between the AIC values of the GJR-ARCH(1)-like copulas and the next best copulas are only minor. Moreover, there are no large differences in the implied conditional variances or quantiles of the CMP(2) models when the GJR-ARCH(1)-like copulas are replaced by the second best copulas.

[4] It may be possible to use different copula families by considering a family-invariant dependence measure. For copulas with the $\text{PQD}_{(1,3)}$-$\text{NQD}_{(2,4)}$ property, one could compute the absolute value of Spearman's rho in each copula quadrant and use the mean of these absolute values.

Figure 6.3: Lag function, truncation point, and cumulated average log-likelihoods. The vertical dashed lines indicate the last member of the truncated copula sequences.

(a) The red lines represent the jointly ML-estimated parameters $(\theta_i)_{i=3,\ldots,100}$ of the sequences $(C_i)_{i=3,\ldots,100}$ which consist of merged X-shaped versions of the Gaussian copula. The blue lines show the sequences $(\theta_i^{g_3})_{i=3,\ldots,100}$ that are given by fitted exponential lag functions g_3 which start with the third lag (Algorithm 3.4).



(b) The black circles and solid blue lines display the first 100 cumulated average log-likelihoods of $(C_i(\theta_i))_{i=1,\ldots,100}$ and $(C_1(\theta_1), C_2(\theta_2), (C_i(\theta_i^{g_3}))_{i=3,\ldots,100}$, respectively. The red marks visualize the first 100 cumulated average log-likelihoods that result from the truncated copula sequences.

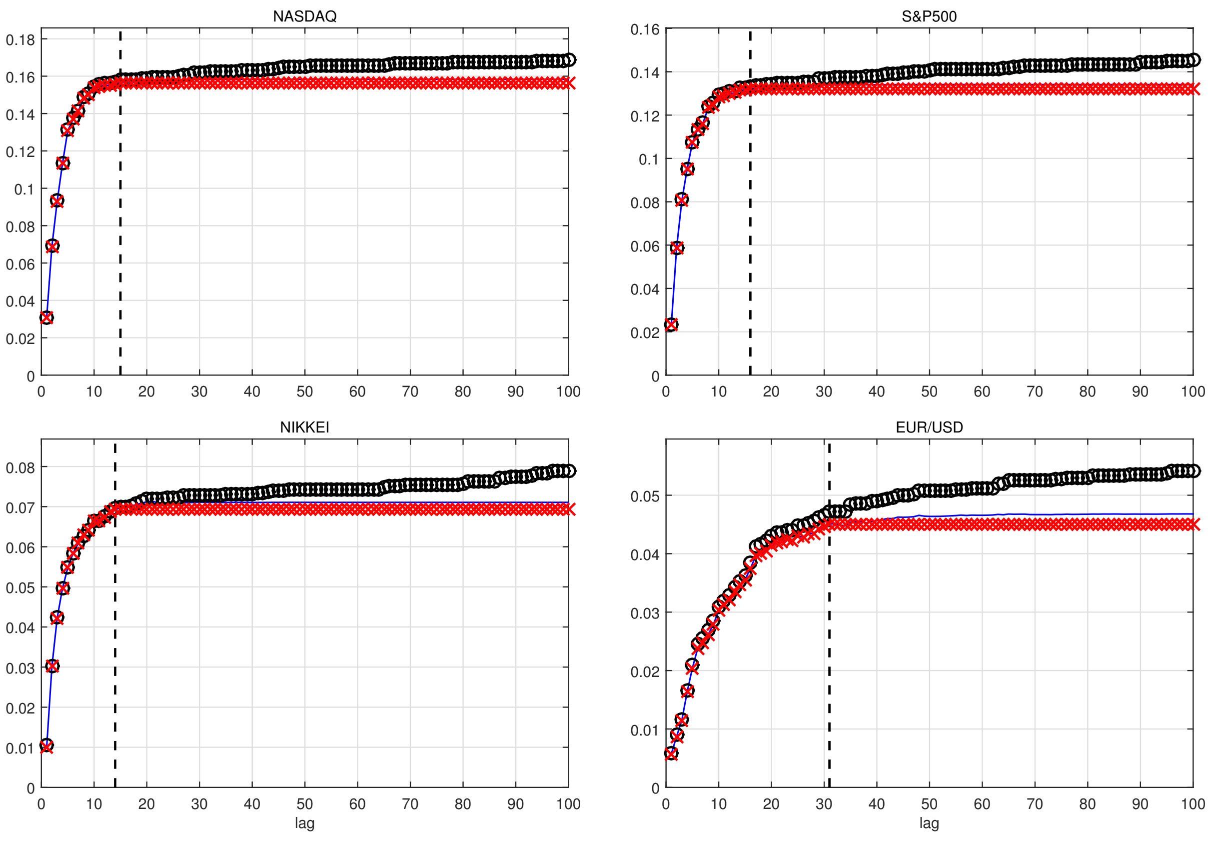


Table 6.7: Estimation results of the truncated CMP models with lag function for AR(1)-filtered returns using Algorithm 3.4. For an explanation of the copula families see Table 6.2. The threshold Δ for the truncation of the copula sequence in Definition 3.8 is 0.0021.

	NASDAQ		S&P 500		NIKKEI		EUR/USD	
Normal (μ, σ^2, w_1)	$(0.278, 0.536, 0.270)$		$(0.201, 0.401, 0.246)$		$(0.141, 1.230, 0.509)$		$(0.041, 0.117, 0.062)$	
Student-t (μ, σ^2, w_2)	$(-0.079, 1.499, 4.467, 0.730)$		$(-0.048, 1.1245, 3.478, 0.754)$		-		$(-0.002, 0.539, 6.703, 0.938)$	
Laplace (μ, σ^2, w_3)	-		-		$(-0.059, 1.189, 0.491)$		-	
	family	ψ_i	family	ψ_i	family	ψ_i	family	ψ_i
C_1	Ga^{MX}-Cl^{VS}	(0.67,0.41,0.36)	(t-zero)-Cl^{VS}	(2.20,0.47,0.30)	Cl^{MX}-Cl^{VS}	(0.36,0.25,0.09)	Cl^{MX}	0.30
C_2	Ga^{MX}-Cl^{VS}	(0.70,0.40,0.45)	Ga^{MX}-Cl^{VS}	(0.71,0.44,0.42)	(t-zero)-Cl^{VS}	(2.85,0.29,0.30)	Gu^{PX}	1.039
$C_{3:p}$	Ga^{MX}	(0.50,0.92,15)	Ga^{MX}	(0.48,0.92,16)	Ga^{MX}	(0.38,0.93,14)	Ga^{MX}	(0.30,0.97,31)
p-val	0.965		0.457		0.046		0.812	
$\log \mathcal{L}$(margin)	-7115.337		-5933.887		-6843.522		-3846.723	
$\log \mathcal{L}$(copulas)	599.701		511.877		261.540		185.767	
$\log \mathcal{L}$	-6515.637		-5422.010		-6581.982		-3660.956	
AIC	13059.273		10872.021		13189.965		7341.913	

The first three rows report the fitted marginal distributions of the CMP models. The next three rows show the results for the copula sequences $(C_i)_{i=1,\ldots,p}$. The families and parameters of the first two copulas, $C_1(\theta_1)$ and $C_2(\theta_2)$, are given in the respective columns with $\theta_i = \psi_i$ for $i = 1, 2$. $C_{3:p} = (C_i(\theta_i))_{i=3,\ldots,p}$, where θ_i is the dependence parameter of C_i, is modeled by an exponential lag function such that $\theta_i = \psi_{3,1}\psi_{3,2}{}^{i-3}$, where $\psi_{3,i}$ refers to the i-th element of ψ_3. The last element of ψ_3 is the order of the truncated CMP model, i.e., $\psi_{3,3} = p$. p-val is the p-value of Neyman's smooth test with four components for the marginal distribution of the innovation sequence of the CMP model. $\log \mathcal{L}$(margin) reports the log-likelihood of the marginal distribution and $\log \mathcal{L}$(copulas) the log-likelihood of the copula sequence. $\log \mathcal{L}$ and AIC are the log-likelihood and the AIC value of the CMP model.

In order to obtain a parsimonious representation, we model the parameters of the sequence $(C_i(\theta_i))_{i=3,\ldots,100}$ with an exponential lag function g_3 such that $\theta_i = \theta_i^{g_3}$.[5] Figure 6.3a and Figure 6.3b show the parameters $(\theta_i^{g_3})_{i=3,\ldots,100}$ of $(C_i)_{i=3,\ldots,100}$ and the resulting cumulated average log-likelihoods. For the stock indices, the decay of the sequences $(\theta_i^{g_3})_{i=3,\ldots,100}$ are pretty similar. The decay of the sequence $(\theta_i^{g_3})_{i=3,\ldots,100}$ is much slower for the EUR/USD exchange rate, suggesting a longer memory for this time series. Finally, we truncate the copula sequence of each CMP model such that the loss in average log-likelihood is no worse than $\Delta = 0.0021$. The truncation points are indicated by vertical dashed lines in Figure 6.3a and Figure 6.3b. For the NASDAQ, S&P 500 and NIKKEI stock index, the orders of the truncated CMP models are given by 15, 16 and 14, respectively. The order of the truncated CMP model of the EUR/USD exchange rate is 31. Table 6.7 summarizes the estimation results for the truncated CMP models with exponential lag functions. According to the log-likelihoods of the copula sequences, the temporal dependence is strongest for the NASDAQ and S&P500 stock index and weakest for the EUR/USD exchange rate. Nevertheless, the order of the truncated CMP model of the EUR/USD exchange rate is twice as large as the orders of the other truncated CMP models.

Since the estimated marginal distributions of the AR(1)-filtered returns are not symmetric, the resulting CMP models are not martingale difference sequences in a strict sense. However, the estimated marginal distributions are only slightly skewed, so that we expect that the conditional means of the AR(1)-filtered returns are only slightly time-varying. The upper plots in Figures 6.4a-6.4b show the conditional means of the raw returns that are implied by the estimated AR(1) models. The lower plots in Figures 6.4a-6.4b depict the conditional means of the AR(1)-filtered returns that are derived from the CMP models. Note that the conditional means of the raw returns are given by the sum of these two conditional means. For the NASDAQ index, the conditional mean of the CMP model is much more persistent than the conditional mean of the AR(1) model. While the conditional mean of the AR(1) model traces the trajectory of the raw returns, with a delay of one lag, the conditional mean of the CMP model is rather negative in periods of high volatility and rather positive in calm periods. This seems to be plausible and not to be an artefact that is caused by the combination of an asymmetric marginal distribution and vertically symmetric copulas. Regarding the other time series, the variations of the conditional means of the CMP models are negligible in relation to the AR(1) models.

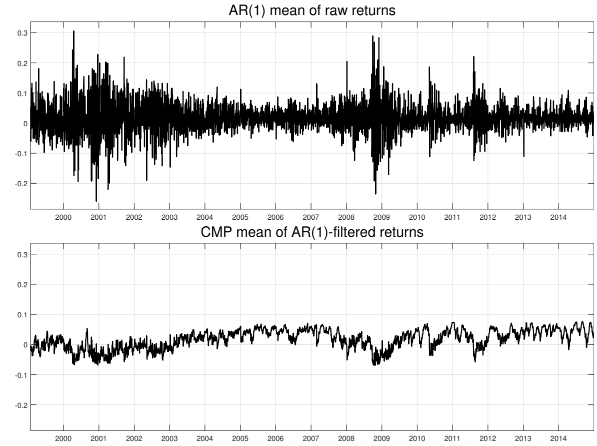
6.3.2 Comparison with GARCH models

In the following, we draw a comparison between CMP models and GARCH models in order to evaluate the performance and competitiveness of the copula-based approach to modeling financial returns. We consider GJR-GARCH(k, k, k) models with $k = 1, 2, 3$,

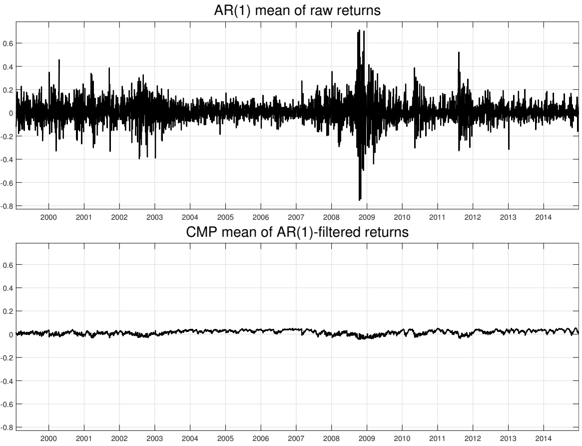
[5] Other lag function do not improve the fit.

Figure 6.4: Conditional means of the CMP models. The upper plots in Figures 6.4a-6.4d show the conditional means of the raw returns that are given by estimated AR(1) models. The lower plots in Figures 6.4a-6.4d display the conditional means of the AR(1)-filtered returns that are given by the estimated CMP models.

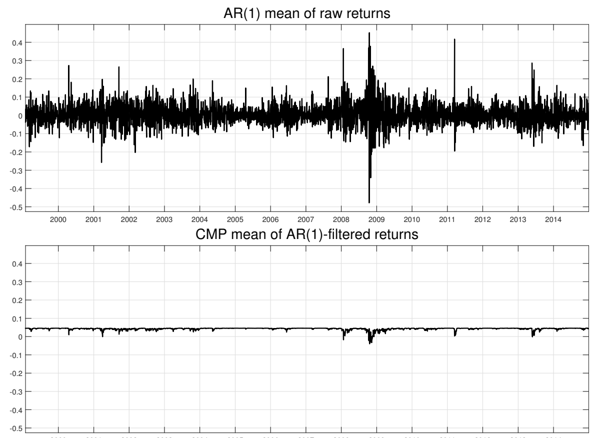
(a) NASDAQ



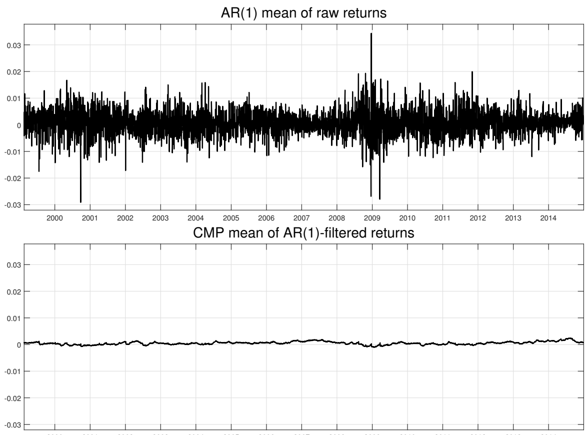
(b) S&P500



(c) NIKKEI



(d) EUR/USD



and for the error distribution Hansen's skewed t-distribution and mixture distributions with two components. The same components for the mixture distributions are taken into account as for the marginal distributions of the CMP models, i.e., the normal distribution, the Student-t distribution, and the Laplace distribution. Table 6.8 shows the estimation results of the fitted GARCH models that yield the lowest AIC values.

The best model for the error distribution is in two out of four cases Hansen's skewed t-distribution. Moreover, in the other two cases, the consideration of a mixture distribution decreases the AIC value by no more than two numbers. In this regard, the marginal distributions of the errors of the GARCH models are simpler than the marginal distributions of financial returns where mixture distributions are required. For the stock indices, the γ coefficients are much larger than the α coefficients, indicating pronounced leverage effects for these time series. The first coefficient of the leverage effect is negative for the

Table 6.8: GARCH estimation results. The estimated GJR-GARCH(2,2,2) models for the AR(1)-filtered returns are given by $Y_t = \sigma_t \mathcal{E}_t$, with $\sigma_t^2 = \omega + \sum_{i=1}^2 \alpha_i Y_{t-i}^2 + \sum_{i=1}^2 \gamma_i \mathbb{1}_{\{Y_{t-i}<0\}} Y_{t-i}^2 + \sum_{i=1}^2 \beta_i \sigma_{t-i}^2$. Conditions for stationarity are imposed. Hansen's skewed-t, or the mixture of the normal and the Laplace distribution refer to the marginal distribution of $\mathcal{E}$. p-val is the p-value of Neyman's smooth test with four components for the marginal distribution of $\mathcal{E}$. $\log \mathcal{L}$ is the value of the maximized log-likelihood function and AIC reports the value of the AIC.

	NASDAQ	S&P 500	NIKKEI	EUR/USD
α_1	0.000	0.000	0.000	0.016
α_2	0.027	0.000	0.050	0.015
γ_1	0.081	0.065	0.083	-0.016
γ_2	0.074	0.165	0.010	0.018
β_1	0.644	0.503	0.879	0.966
β_2	0.244	0.366	0.000	0.000
Hansen's skewed-t (ν, λ)	$(12.083, -0.155)$	$(8.057, -0.157)$	-	-
$N(\mu, \sigma^2, w_1)$	-	-	$(0.028, 0.977, 0.708)$	$(-0.025, 0.985, 0.667)$
Laplace(μ, σ^2, w_2)	-	-	$(-0.013, 0.740, 0.292)$	$(0.056, 0.725, 0.333)$
p-val	0.813	0.833	0.326	0.878
$\log \mathcal{L}$	-6539.018	-5422.739	-6582.388	-3659.314
AIC	13096.035	10863.478	13188.776	7342.628

EUR/USD exchange rate, implying that the conditional variance increases more if the lagged return was positive. The inclusion of σ_{t-2}^2 and $\mathbb{1}_{\{Y_{t-2}<0\}} Y_{t-2}^2$ is especially relevant for the NASDAQ and S&P500 index.

In-sample fit

We now compare the in-sample fit of the CMP and GARCH models. Figure 6.5 shows the sample autocorrelation functions of the squared residuals of the CMP and GARCH models. The p-values of the Ljung-Box test for the presence of no serial correlation in the first 30 lags are given in Table 6.9. The sample autocorrelation functions in Figures 6.5a, 6.5b, and 6.5d, show that the CMP models completely remove the autocorrelation of squared returns for the NASDAQ and S&P 500 stock index as well as for the EUR/USD exchange rate. The p-values of the Ljung-Box test also provide no evidence that the CMP-filtered squared returns still exhibit any significant autocorrelation for these time series. However, for the NIKKEI stock index, a significant correlation between consecutive squared residuals is visible in Figure 6.5c. Moreover, the null hypothesis of no serial correlation in the first 30 lags of the squared residuals is rejected at a 5% significance level.

The sample autocorrelation functions of the GARCH models show no significant autocorrelations for the S&P 500 and NIKKEI stock index and the EUR/USD exchange rate. The Ljung-Box test also confirms that, contrary to the CMP model, the squared residuals of the GARCH model display no significant autocorrelation for the NIKKEI stock index. On the other side, a negative correlation between consecutive squared residuals can be recognized in Figure 6.5a for the NASDAQ stock index. Moreover, the Ljung-Box test rejects the null hypothesis of no serial correlation at a 5% significance level. Overall, both model classes show comparable results when it comes to removing the serial correlation in squared returns.

Figure 6.5: Sample autocorrelation function of the squared residuals of the CMP and GARCH models. The upper plots in Figures 6.5a-6.5d show the autocorrelation functions of the CMP models. The lower plots in Figures 6.5a-6.5d show the autocorrelations of the GARCH models. The residual of a CMP model at period t is given by $F^{-1}(z_t)$, where $z_t = F_{t|t-1:1}(y_t|y_{t-1:1})$ is the estimated conditional probability integral transform and F the estimated marginal distribution.

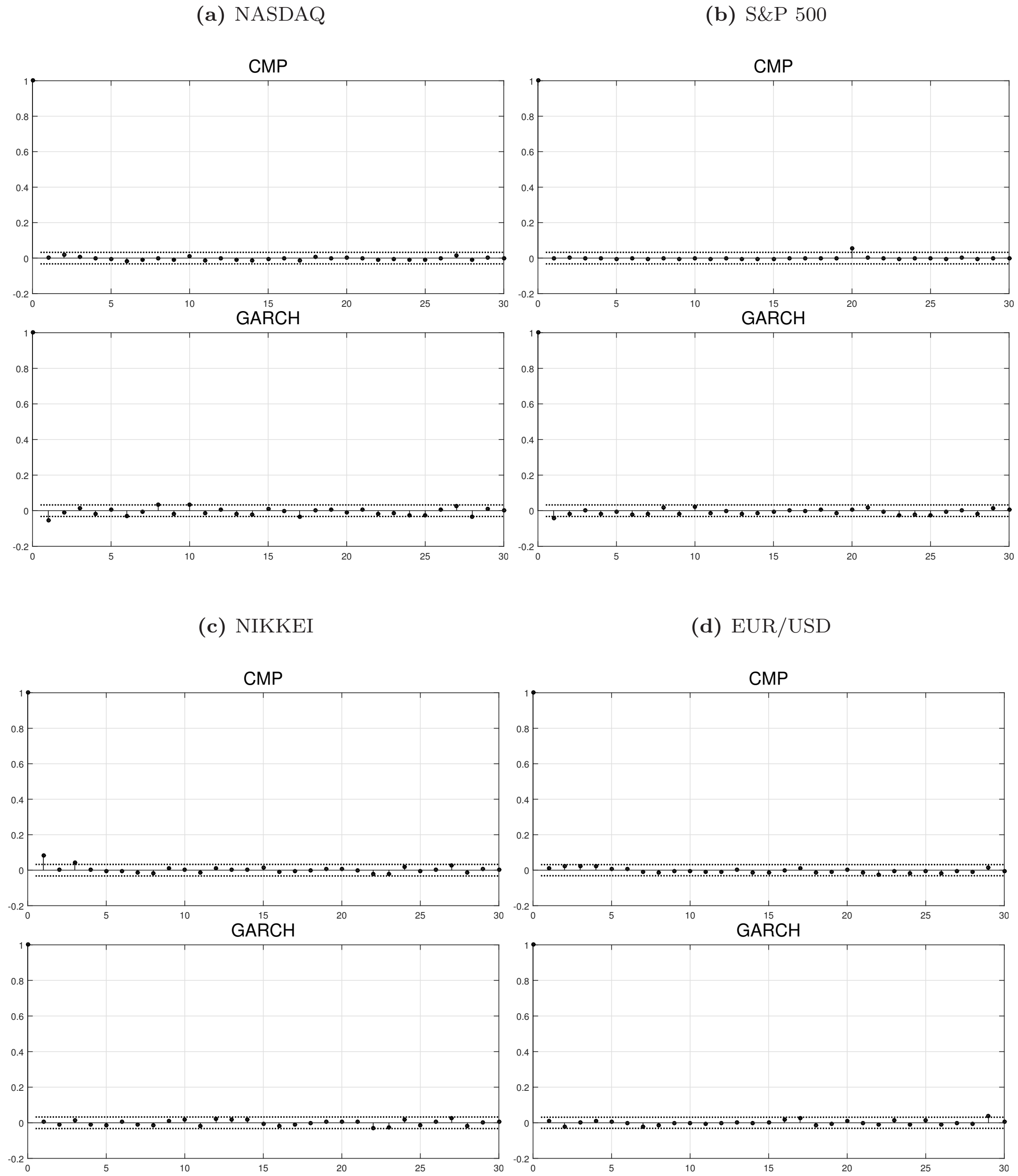


Table 6.9: p-values of the Ljung-Box test for the presence of no serial correlation in the first 30 lags of the squared residuals.

	NASDAQ	S&P 500	NIKKEI	EUR/USD
p-val(CMP)	0.999	0.998	0.017	0.933
p-val(GARCH)	0.014	0.449	0.724	0.904

Table 6.10 compares the in-sample fit of the CMP and GARCH models. Altogether, the CMP models are competitive with GARCH models. For the NASDAQ stock index, the

AIC value of the CMP model is considerably smaller. Regarding the S&P 500 stock index, the log-likelihood values of both model classes are very similar. However, the marginal distribution of the CMP model requires much more parameters than the innovation distribution of the GARCH model. As a result, the GARCH model performs better in terms of AIC for the S&P 500 stock index. For the NASDAQ index and the EUR/USD exchange rate, the differences in the AIC and log-likelihood values of both model classes are negligible.

Table 6.10: In-sample performances of CMP and GARCH models. $\log \mathcal{L}$(model) is the log-likelihood of the model and AIC(model) the AIC value of the model. The AIC values refer to the AR(1)-filtered returns, i.e., the estimated parameters of the AR(1) processes are not taken into account.

	NASDAQ	S&P 500	NIKKEI	EUR/USD
$\log \mathcal{L}$(CMP)	-6515.637	-5422.010	-6581.982	-3660.956
$\log \mathcal{L}$(GARCH)	-6539.018	-5422.739	-6582.388	-3659.314
AIC(CMP)	13059.273	10872.0211	13189.965	7341.913
AIC(GARCH)	13096.035	10863.478	13188.776	7342.628

Volatility and quantiles of the transition distribution

Although the CMP and GARCH models perform equally well in terms of in-sample fit, their transition distributions may be considerably different. In order to investigate this issue, we compare the trajectories of the volatilities and the 1% quantile of the transition distributions. Figure 6.6 shows the conditional volatilities of the CMP and GARCH models. The trajectories of the conditional volatilities of the CMP models are in general much more noisy, especially in periods of high volatility. It is possible that this is due to the fact that the CMP models only have a finite Markov order whereas the GARCH models can be recognized as Markov processes with an infinite order. Another explanation could be that the dependence of the transition distribution on the lagged conditional variance is weaker for the CMP models. In periods of low volatility, the trajectories of the conditional volatilities of both model classes are very similar. However, in periods of high volatility, we observe distinct patterns. The conditional volatilities of the CMP models are more strongly varying in these periods. As a result, the largest conditional volatility for each time series is observed for the CMP model. At first sight, it also appears that the peaks of the conditional volatilities are larger for the CMP models. A closer look reveals that this is not strictly true. For instance, for the NASDAQ index, the largest peak of the conditional volatility of the GARCH model, which occurs around the beginning of the second quarter of the year 2000, is larger than the corresponding value of the conditional volatility of the CMP model. Moreover, for the NIKKEI index, the second largest peak of the conditional volatility of the GARCH model, which appears at the end of the first quarter of the year 2011, is 1.25 times larger than the value of the conditional volatility of the CMP model at the same time period.

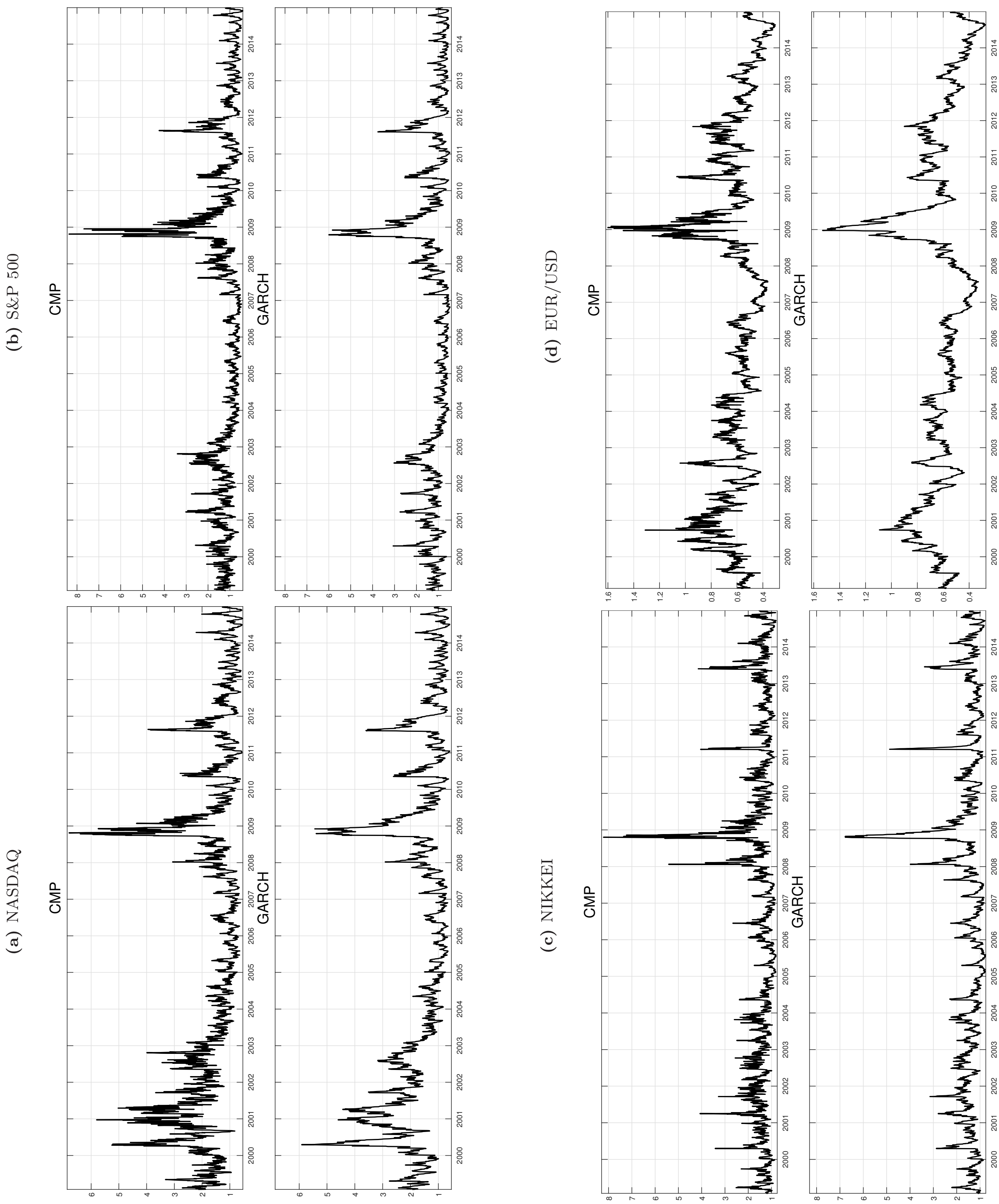


Figure 6.6: Volatilities of the transition distributions of the CMP and GARCH models.

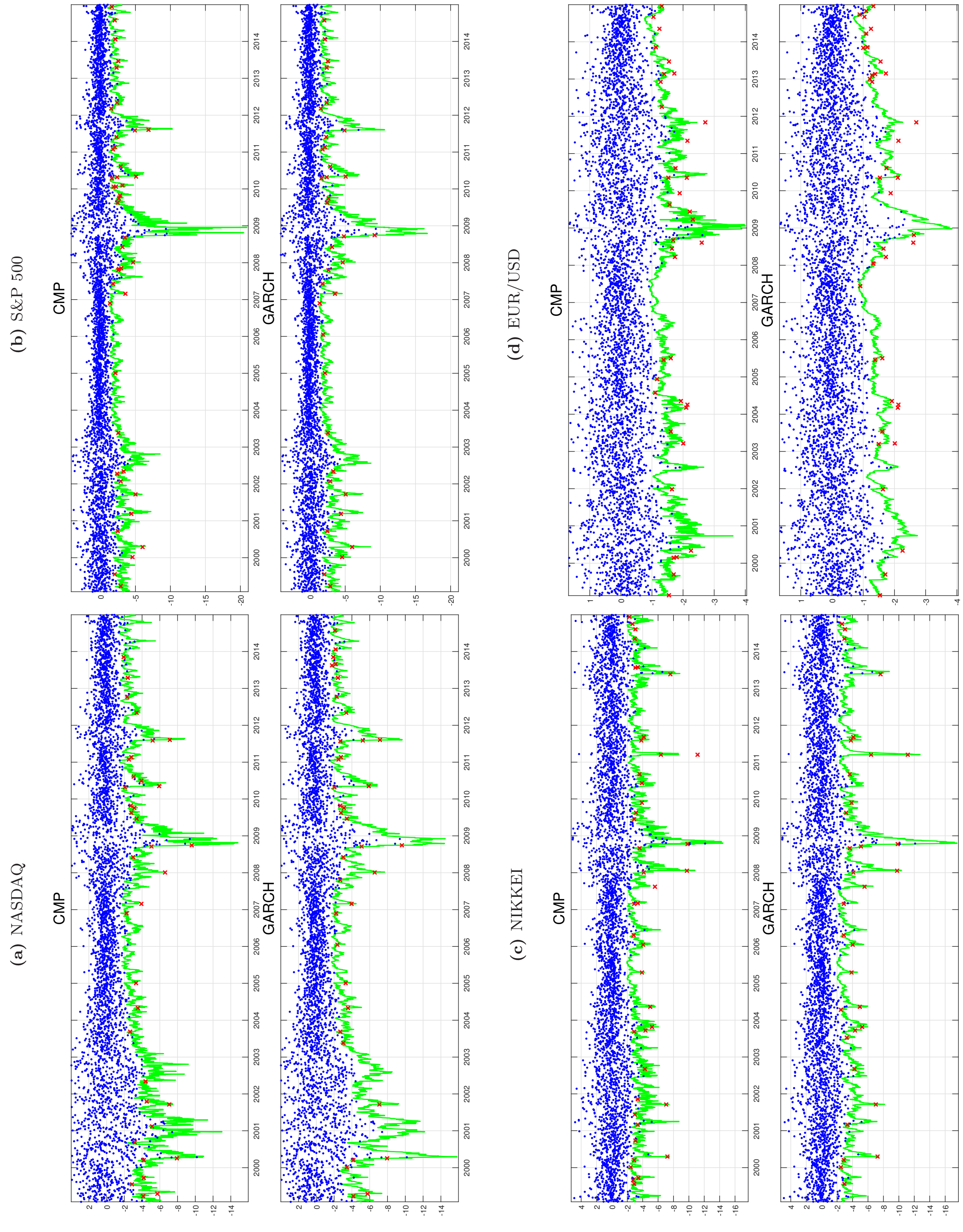


Figure 6.7: 1% quantiles of the transition distributions of the CMP and GARCH models. The blue dots represent the time series and the green lines show the 1% quantiles of the transition distributions. Red marks indicate returns that are smaller than the 1% quantile of the transition distribution.

Figure 6.7 shows the 1% quantiles of the transition distributions for both model classes. Since the conditional 1% quantile of a GARCH model is proportional to the volatility of its transition distribution, the conditional 1% quantiles of the GARCH models in Figure 6.7 reflect the trajectories of the conditional volatilities. This is also roughly true for the 1% quantiles of the CMP models, but there is some evidence that the 1% quantiles are not strictly proportional to the conditional volatilities of the CMP models. For the NASDAQ index, the maximum of the conditional volatility is larger for the CMP model, but the smallest 1% quantile of the transition distribution occurs for the GARCH model. Furthermore, at the beginning of the fourth quarter of the year 2008, the conditional conditional volatility of the CMP model of the NASDAQ index is 1.25 times larger than the corresponding conditional volatility of the GARCH model. However, the 1% conditional quantiles of both models are very similar, with the 1% conditional quantile of the CMP model being only 1.03 times as large. Consequently, it appears that the transition distribution of the CMP model of the NASDAQ index is not only driven by its conditional volatility.

Transition distribution of the standardized error

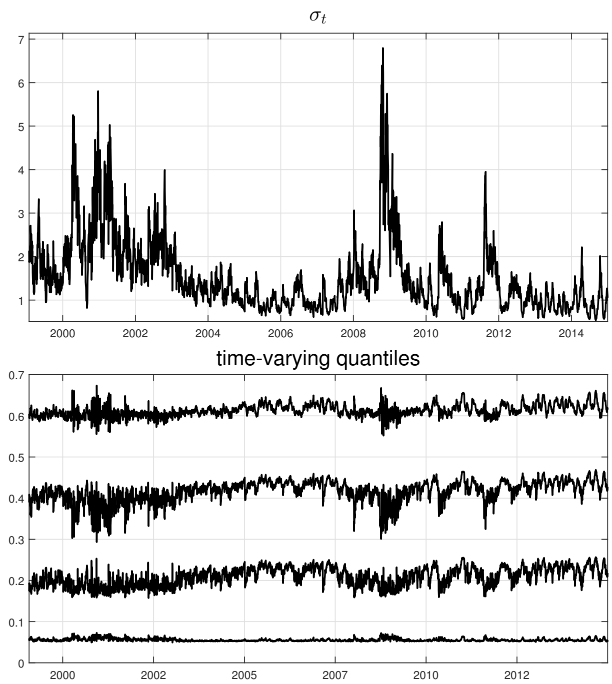
Obviously, the possible benefit of a copula-based approach is limited if the evolution of the transition distribution of financial returns is determined by its volatility. In this case, copula-based models might be competitive with GARCH models but will not result in a considerable improvement. We now investigate in more detail to what extent the transition distribution of the CMP model is time-varying if the effect of its conditional volatility has been accounted for. For that purpose, we define the estimated standardized error of the CMP model at period t by $\widehat{\mathcal{E}}_t = (Y_t - \widehat{\mu}_t)/\widehat{\sigma}_t$, where $\widehat{\mu}_t$ and $\widehat{\sigma}_t$ are the estimated mean and volatility of the transition distribution of the CMP model. The estimated transition distribution of the standardized error can be computed using $F_{\widehat{\mathcal{E}}_t|Y_{t-1:1}}(\hat{\epsilon}_t|y_{t-1:1}) = F_{Y_t|Y_{t-1:t-1}}(\widehat{\mu}_t + \widehat{\sigma}_t\hat{\epsilon}_t|y_{t-1:1})$, where $F_{Y_t|Y_{t-1:t-1}}$ denotes the transition distribution of the CMP model. If the transition distribution of the CMP model is characterized by its mean and volatility, the estimated quantiles of the transition distribution of the standardized error should not be time-varying.

Figure 6.8 illustrates the time variation of the estimated conditional quantiles of the standardized errors of the CMP models. Since conditional quantile values of the standardized error are hard to interpret, we plot the values of relative conditional quantiles which set the conditional quantiles in relation to the unconditional quantiles of the standardized error. The estimated relative conditional α-quantile at period t is given by $F_{\widehat{\mathcal{E}}} \circ F^{-1}_{\widehat{\mathcal{E}}_t|Y_{t-1:1}}(\alpha|y_{t-1:1})$, where $F_{\widehat{\mathcal{E}}}$ is the estimated marginal distribution of the standardized error sequence and $F^{-1}_{\widehat{\mathcal{E}}_t|Y_{t-1:1}}$ is the inverse function of $F_{\widehat{\mathcal{E}}_t|Y_{t-1:1}}$. Thus, if the standardized error sequence is a sequence of iid random variables, then the relative conditional α-quantile equals α. Moreover, the mean of the relative conditional α-quantile equals α.

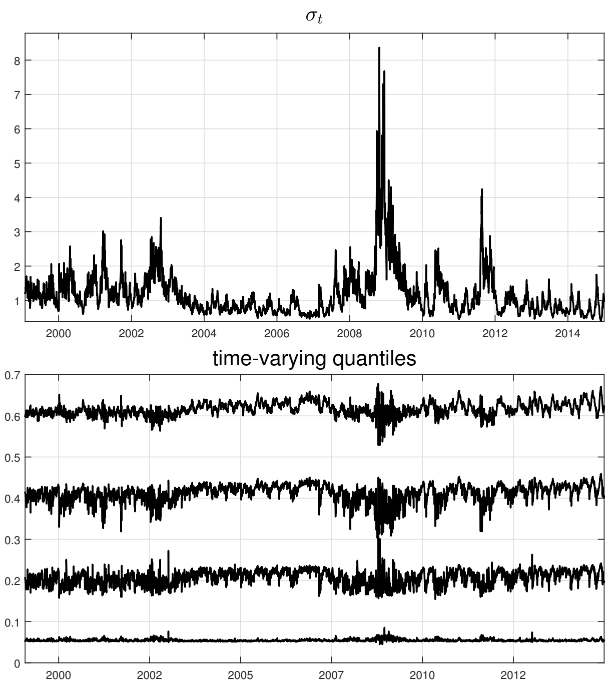
Especially for the NASDAQ and S&P 500 stock index, the trajectories of the relative

Figure 6.8: Time-varying conditional relative quantiles of the standardized errors of the CMP models. The upper plots in Figures 6.8a-6.8d show the conditional volatilities of the CMP models. The lower plots in Figures 6.8a-6.8d illustrate the estimated relative conditional α%-quantiles of the standardized error for $\alpha \in \{0.05, 0.2, 0.4, 0.6, 0.8, 0.95\}$.

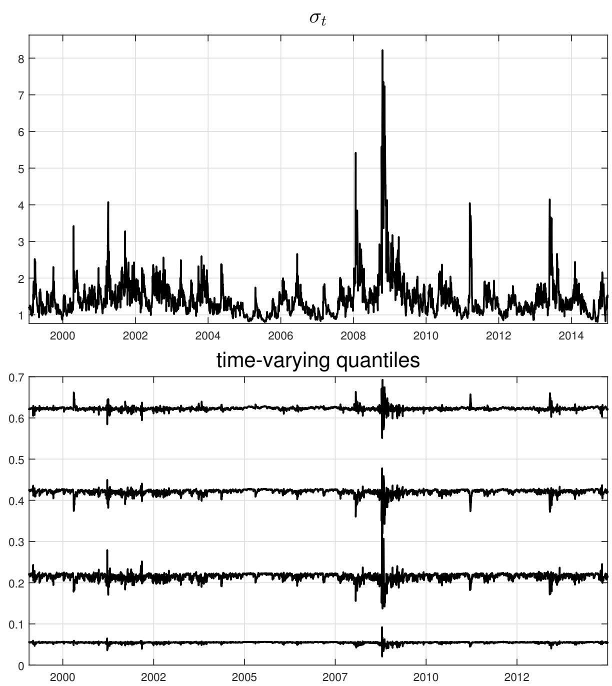
(a) NASDAQ



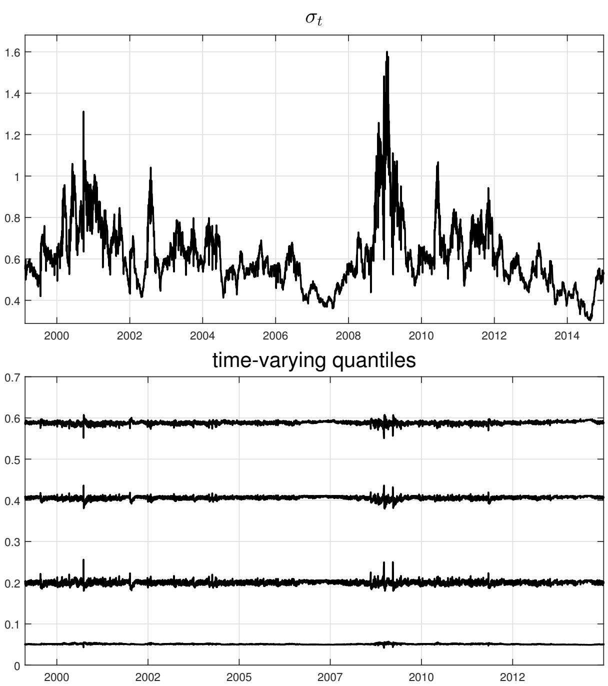
(b) S&P 500



(c) NIKKEI



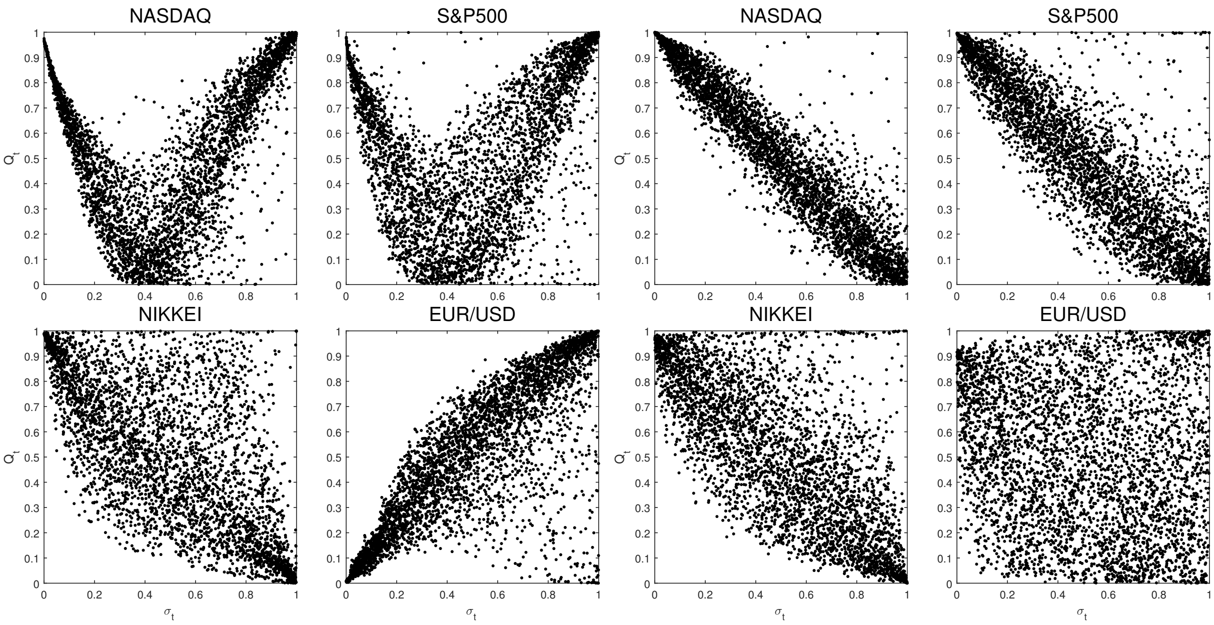
(d) EUR/USD



conditional quantiles show a strong variation. The variation is less pronounced for the NIKKEI stock index and rather negligible for the EUR/USD exchange rate. For all time series, the fluctuation of the relative conditional quantiles is increased during periods of high volatility. For the NASDAQ and S&P 500 stock index, the ranges of the relative 0.2, 0.4, 0.6, and 0.8 conditional quantiles are around 0.1 during the peak of the financial crisis, implying a great variation in the center of the transition distribution even if the effects of high volatility have been removed. The 0.2 and 0.4 conditional quantiles of the standardized error are also smaller than the corresponding unconditional quantiles in periods of high volatility. This indicates that the conditional probability of negative returns is larger in these cases than under the assumption of an iid sequence of standardized errors. On the contrary, the 0.05 conditional quantile of the standardized error seems to be slightly larger during periods of crisis for the NASDAQ and S&P500 stock index. The relation between

Figure 6.9: Copula scatter plots of conditional volatilities and conditional α-quantiles of the standardized errors.

(a) $\alpha = 0.05$ **(b)** $\alpha = 0.2$



lower quantiles and volatility for the NASDAQ and S&P 500 stock index is confirmed in Figure 6.9. For the stock indices, the copula scatter plots of conditional volatilities and conditional 0.2-quantiles indicate a pronounced negative relationship. However, the conditional 0.01-quantiles increase in periods of low and large volatility for the NASDAQ and S&P 500 stock index.

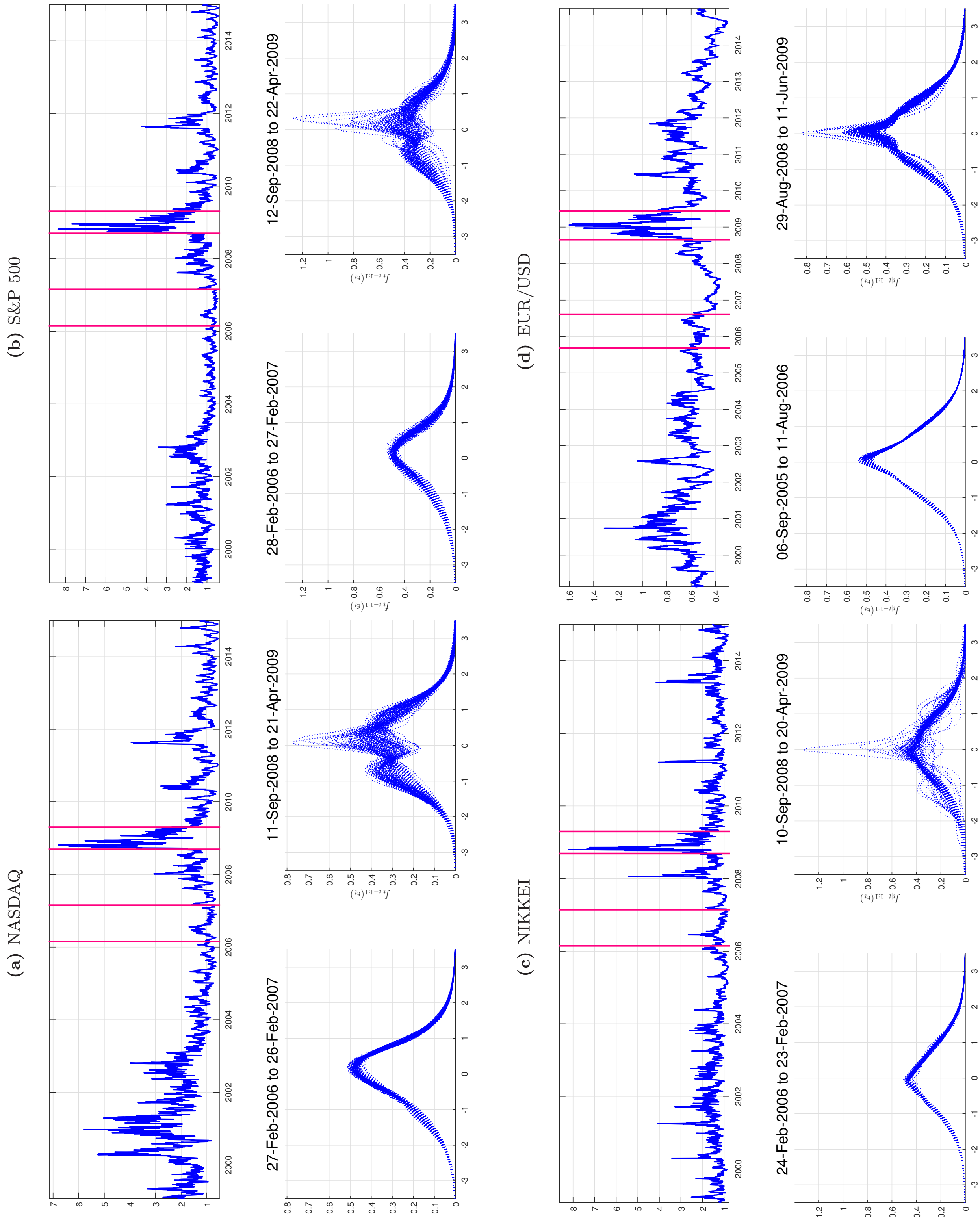


Figure 6.10: Transition densities of the standardized errors of the CMP models in periods of low and high volatility. The upper plots in Figures 6.10a-6.10d show the conditional volatilities of the CMP models. The starting and ending point of the periods of low and high volatility are indicated by vertical lines. The lower left and lower right plots in Figures 6.10a-6.10d depict the transition densities in periods of low and high volatility, respectively.

Figure 6.10 shows the transition densities of the standardized errors of the CMP models for periods of low and high volatility. The periods of low volatility last approximately one year whereas the periods of high volatility last around three quarters of a year. In the periods of low volatility the transition densities of the standardized errors show no great variation. Thus, during periods of low volatility, the assumption that the transition distribution of financial returns is driven by its volatility is supported by the CMP models. However, the shapes of the transition densities considerably change for the NASDAQ and the S&P 500 index during turbulent market periods. In periods of high volatility, the transition densities can be bimodal or exhibit a pronounced peak in the center which, to the best of our knowledge, is not considered by popular volatility models. It is plausible that, during financial crises, the probability of extreme returns can be far larger than the probability of small absolute returns, giving rise to a bimodal transition density. On the other side, the possible peak of the transition density of the standardized error indicates that returns scatter more strongly around the center of the distribution if the effect of volatility has been accounted for. The frequency of these peaked transition densities is very low and considerably smaller than the frequency of bimodal transition densities. During periods of very high volatility, the conditional 1% quantiles of the CMP models are on average too large under the assumption that the conditional volatility completely determines the transition distribution. Thus, it seems that these peaked transition densities occur in periods of very high volatility and compensate for the larger conditional volatility that is implied by the CMP models in these periods.

A variation of the transition density of the standardized error in periods of high volatility is also visible for the NIKKEI stock index, though the frequency of a bimodal or peaked density is considerably smaller. In the majority of cases, the transition density in periods of high volatility strongly resembles the transition density in periods of low volatility. However, in a small number of cases, the transition density has even three modes. Regarding the EUR/USD exchange rate, there is only a minor variation in the transition density during periods of crisis.

Although a possible bimodal shape of the transition density during a financial crisis appears plausible, the pronounced bimodal transition densities of the CMP model for the NASDAQ and S&P 500 stock index could also result from the model assumptions.[6] In order to investigate this matter in more detail, we have fitted CMP models to data which is generated from the estimated GARCH models of the NASDAQ and S&P 500 stock index. In the majority of cases, the transition densities of fitted CMP models are unimodal, also during periods of high volatility. However, in some few cases, the transition density of a CMP model can be bimodal, although this is not true for the data generating process. Consequently, the bimodal shape of the transition density could also result from the used

6 Because of the large sample size, we assume that the bimodality is not due to chance. However, a test for the statistical significance of the bimodality is an interesting research question.

parametric copula families. Another reason could be the specification of a simplified vine copula model for the SD-vine copula. On the other side, the more flexible shapes of the transition density might explain why the CMP model exhibits a better in-sample fit for the NASDAQ stock index.

Out-of-sample performance

Finally, we take a brief look at the out-of-sample performance of the CMP and GARCH model class. For this purpose, we conduct out-of-sample forecasts of the transition distribution, employing recursive window estimation schemes. The model specifications are based on the previous in-sample specifications which are given in Table 6.7 and Table 6.8.[7] We use an initial window size of 1000 observations, estimate the model parameters, and forecast the transition distribution $F_{1001|1000:1}(y_{1001}|y_{1000:1})$. After this and each subsequent forecast, the window size is increased by one observation, the model parameters are re-estimated, and the next transition distribution is predicted. We compare the out-of-sample performance of the CMP and GARCH model using the difference in logarithmic scores (see Section 4.5). Table 6.11 reports the average logarithmic scores of the two models and shows the results of the Diebold-Mariano type tests for superior out-of-sample specification of the transition distribution. Except for the NASDAQ stock index, the differences in the average logarithmic scores of both model classes are negligible and not significant at a 10% level. However, for the NASDAQ stock index, the average logarithmic score of the CMP model is significantly larger at a 10% level.

Table 6.11: Results of the log-score test for superior out-of-sample specification of the transition distribution. The p-value refers to the null hypothesis that the GARCH model produces larger log-scores.

	NASDAQ	S&P 500	NIKKEI	EUR/USD
average log-score(GARCH)	-1.594	-1.466	-1.788	-0.840
average log-score(CMP)	-1.587	-1.466	-1.791	-0.839
test statistic (4.5.1)	1.623	0.007	-0.620	0.305
p-value	0.052	0.497	0.732	0.380

6.4 Conclusion

We compared the performance of various jointly symmetric and vertically symmetric copulas for the two elements of a three-dimensional SD-vine copula. Overall, we found that the proposed copulas in Chapter 5 provide a good fit to financial data. Regarding jointly symmetric copulas, the Student-t copula with zero correlation parameter and merged or patched X-shaped versions of copulas perform comparably well. However, only for the

[7] The results do not change substantially if the model specifications are based on the first 1000 observations.

EUR/USD exchange rate, jointly vertically symmetric copulas are preferred. For the stock indices, the usage of vertically symmetric copulas with horizontal asymmetry results in a great improvement. The GJR(1)-ARCH-like copula and mixture copulas, consisting of a merged or patched X-shaped version of a copula and a vertically symmetric version of the Clayton copula, constitute the best vertically symmetric copulas in terms of AIC.

An analysis of the resulting CMP(2) processes revealed that the implied conditional variances of the stock indices are not increasing in the absolute value of the first lagged return. Instead, the conditional variances attain a local maximum if the first lagged return is close to zero and the second lagged return attains its 1% quantile. This issue is counter-intuitive from an economic point of view and needs further investigation. Nevertheless, the estimated truncated CMP models with lag functions provide adequate models of financial returns.

In order to evaluate the performance and properties of the CMP models, we drew a comparison with GARCH models and illustrated the similarities and differences between both approaches. In terms of log-likelihood and AIC, both model classes perform equally well, with the CMP model being slightly superior for the NASDAQ index. Despite their similar in-sample performance, we found substantial differences in the transition distributions of both model classes during periods of high volatility. While the trajectories of the conditional volatilities of both model classes are very similar in calm market periods, the conditional volatility of the CMP model is more strongly varying in turbulent market periods. Moreover, we found evidence that the quantiles of the transition distribution of the CMP model are not proportional to its volatility for the NASDAQ and S&P 500 index.

A closer look at the transition densities of the CMP models demonstrated that their shapes can be much more complex than under the assumption of a location scale model. For the NASDAQ and S&P 500 index, we discovered strongly time-varying conditional quantiles of the standardized errors of the CMP models during periods of high volatility. The transition densities of the standardized errors can be bimodal or exhibit a pronounced peak in the center, implying a great variation in center of the distribution even if the effects of high volatility have been accounted for. The more flexible shapes of the transition densities might explain why the CMP model yields a better in-sample fit for the NASDAQ stock index. For the NIKKEI index and the EUR/USD exchange rate, we only found a slight variation in the transition densities of the standardized errors during turbulent market periods. For these time series, the CMP models support the assumption that the transition distribution of financial returns is determined by its volatility.

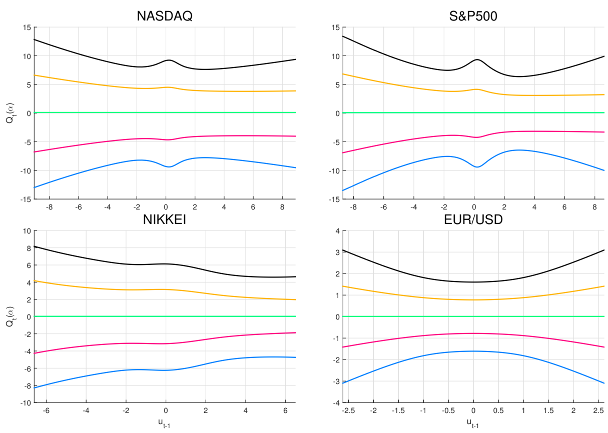
An out-of-sample evaluation of forecasted transition distributions further showed that CMP models of financial returns can be competitive with GARCH models. Moreover, for the NASDAQ stock index, we also found some evidence that a copula-based approach might improve the modeling of financial returns.

6.5 Appendix

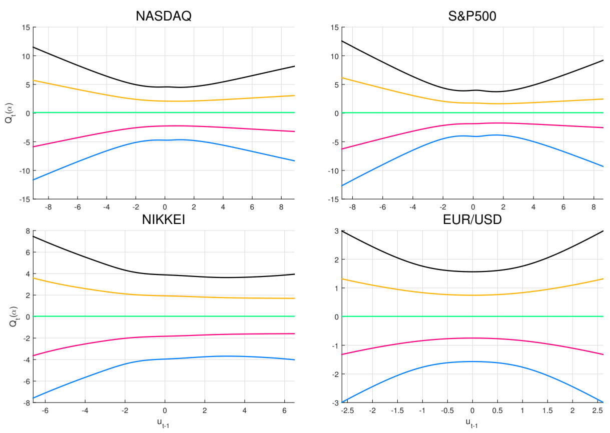
A.1 Quantiles of the CMP(2) models

Figure 6.11: Conditional quantiles of the CMP(2) models as a function of y_{t-1} for given y_{t-2}. The lines in each plot correspond to probability levels of 1%, 10%, 50%, 90%, and 99%. The marginal distribution of a CMP(2) model is specified in Table 6.1 and the elements of its SD-vine copula (C_1, C_2) are given by the estimated copulas in Table 6.5 and Table 6.6 that yield the lowest AIC values.

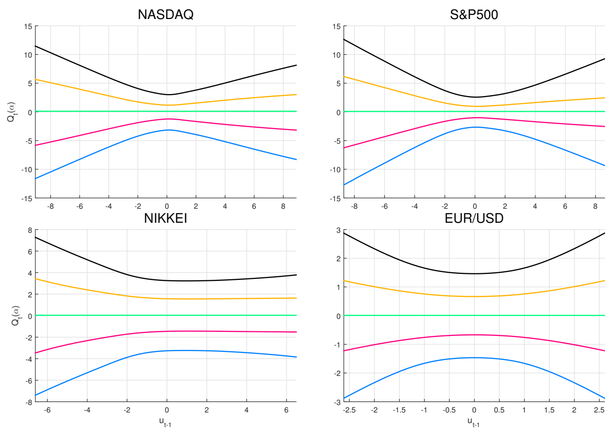
(a) Quantiles when y_{t-2} attains its 1% quantile.



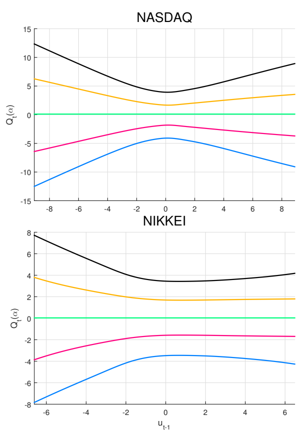
(b) Quantiles when y_{t-2} attains its 10% quantile.



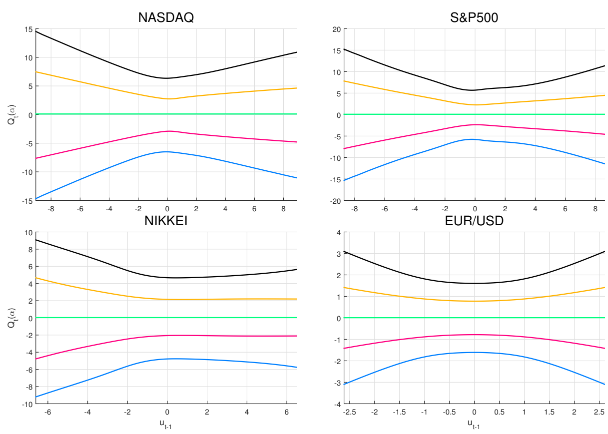
(c) Quantiles when y_{t-2} attains its 50% quantile.



(d) Quantiles when y_{t-2} attains its 90% quantile.



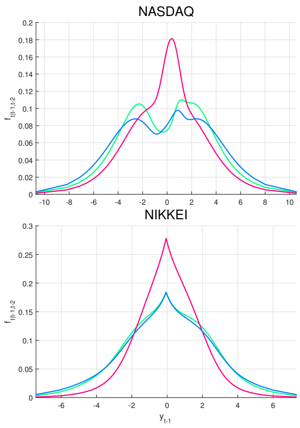
(e) Quantiles when y_{t-2} attains its 99% quantile.



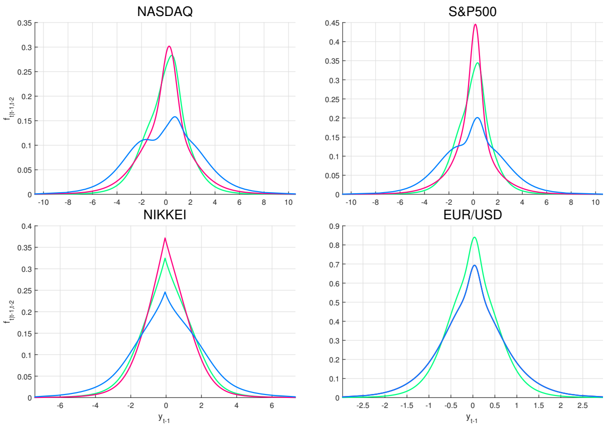
A.2 Conditional densities of the CMP(2) models

Figure 6.12: Conditional densities of the CMP(2) models when y_{t-1} attains its 1%, 50%, and 99% quantile and given y_{t-2}. The marginal distribution of a CMP(2) model is specified in Table 6.1 and the elements of its SD-vine copula (C_1, C_2) are given by the estimated copulas in Table 6.5 and Table 6.6 that yield the lowest AIC values.

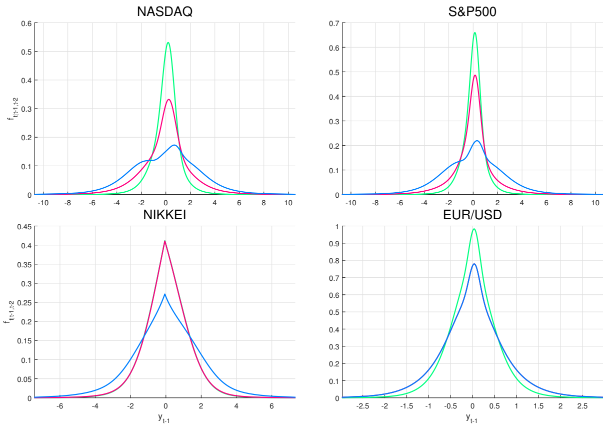
(a) $f_{t|t-1,t-2}$ when y_{t-2} attains its 1% quantile.



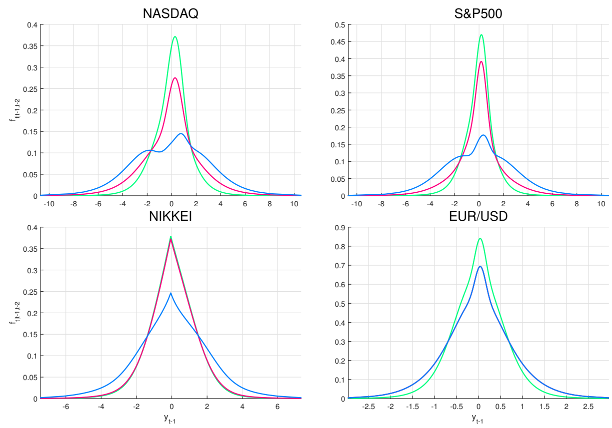
(b) $f_{t|t-1,t-2}$ when y_{t-2} attains its 10% quantile.



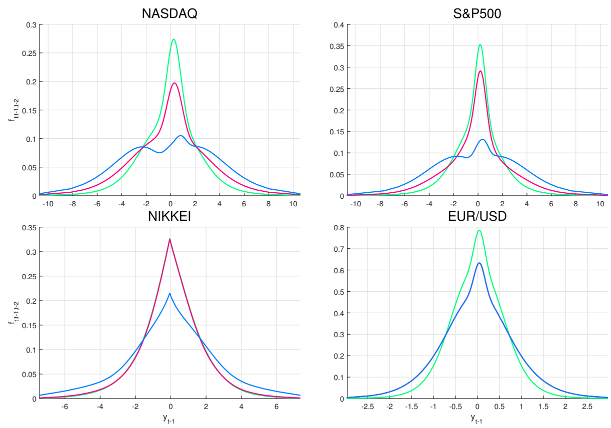
(c) $f_{t|t-1,t-2}$ when y_{t-2} attains its 50% quantile.



(d) $f_{t|t-1,t-2}$ when y_{t-2} attains its 90% quantile.



(e) $f_{t|t-1,t-2}$ when y_{t-2} attains its 99% quantile.



7 Conclusion

7.1 Summary and contributions of this thesis

In 2006, Thomas Mikosch provided compelling arguments that "copulas completely fail in describing complex space-time dependence structure" and "are not useful for modeling dependence through time" (Mikosch, 2006, p. 18-19). Indeed, the available framework at this time did not allow for the construction of copula-based models that are competitive with established models of time series analysis. However, this thesis showed that the recent introduction of vine copulas has opened the doors for a successful application of copula-based time series models that may even be superior to classical time series models.

We provided a thorough investigation of univariate stationary higher-order Markov processes in terms of copulas, with a focus on regular vine copulas, and developed a copula-based approach for modeling univariate time series. In order to model Markov processes with long memory, we introduced methods that allow for a parsimonious representation of an SD-vine copula-based stationary Markov processes. In addition, a detailed analysis of the dependence properties of SD-vine copula-based stationary Markov processes was conducted, which also contributes to the understanding of dependence in regular vine copula models in general. An application to time series of price durations demonstrated the advantages of our framework and revealed a strong superiority of the copula-based approach to the popular class of ACD models in terms of in-sample fit and different out-of-sample criteria.

In order to model financial returns, we derived sufficient and necessary conditions for a conditionally symmetric martingale difference sequence in terms of copulas. Moreover, we established a theory of bivariate copulas that can be used to model volatility clustering and constructed parametric copulas that have the desired dependence properties. An application of the copula-based time series model to the returns of three major stock indices and one currency exchange rate documents the competitiveness of our approach with established GARCH models.

In summary, we found the following strengths of our copula-based approach to time series analysis.

- It is possible to separate the modeling of the marginal distribution from the modeling of the dependence structure. This is helpful because, in general, the marginal distribution contributes the major part of the log-likelihood of a time series model.

- Exploratory data analysis can be utilized to set up a copula-based time series model which captures central features of the dependence structure.

- The transition distribution of the Markov process is derived from its SD-vine copula and its marginal distribution, resulting in flexible transition distributions that can not be expressed by location and scale models.

The use of exploratory data analysis to specify the transition distribution resulted in a major improvement in the modeling of the transition distribution of price durations. Our copula-based models can represent more flexible transition distributions, allowing for a time-varying dispersion, and this explains the superiority to ACD models with iid innovations. Obviously, a copula-based approach can only result in a major improvement if the transition distribution can not be adequately described by a location and scale model. This might explain why our copula-based approach did not result in a considerable improvement in the modeling of financial returns. Although we found some evidence that the transition distribution may be bimodal during turbulent market periods, it appears that the transition distribution of financial returns can be adequately described by its volatility.

7.2 Possible directions for future research

The developed univariate copula-based time series models in this thesis can be used as building blocks in the process of modeling a multivariate time series process. One can model each univariate time series by an SD-vine copula-based Markov model and then join the residuals of the univariate time series with a (conditional) copula, in order to obtain a multivariate time series model. This approach neglects non-instantaneous relations between the time series, but is popular if the univariate time series are modeled by GARCH models. The construction of copula-based multivariate time series models that allow for more complex cross-sectional dependencies has been considered by several authors recently (Smith, 2013; Brechmann and Czado, 2014; Beare and Seo, 2014). In the multivariate case, there is no unique copula that is compatible with the conditions for stationarity, but in all studies it is proposed to model the instantaneous cross-sectional relations as the first edge of a simplified vine copula after which the serial dependence is considered. Due to this construction, it is not possible to obtain a parsimonious representation of a Markov process with long memory.

In future research, we will investigate the modeling of multivariate time series by simplified vine copulas that first specify the univariate serial dependence of each time series and then turn to the cross-sectional relations. This would allow for a parsimonious representation of a higher-order Markov process by means of lag functions and is in line with the assumption that an adequate modeling of the serial dependence of each time series is more important. Moreover, we will apply the copula-based models to various financial data with positive support, e.g., realized volatilities or price ranges. Another important research topic is the construction of copula-based time series models which follow the spirit of ARMA or GARCH models. We will investigate copula-based processes that are based on latent variables, such as the conditional mean or the variance, to obtain infinite-order Markov processes that can be represented by low-dimensional vine copulas. Altogether, copula-based time series models constitute an alternative and useful approach to time series analysis, especially when the interest lies in the transition distribution, and are a promising field for future research.

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-Copula Constructions of Multiple Dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

Abdous, B., Genest, C., and Rémillard, B. (2005). Dependence Properties of Meta-Elliptical Distributions. In *Statistical Modeling and Analysis for Complex Data Problems*, pages 1–15. Springer US.

Abegaz, F. and Naik-Nimbalkar, U. (2008a). Modeling statistical dependence of Markov chains via copula models. *Journal of Statistical Planning and Inference*, 138(4):1131–1146.

Abegaz, F. and Naik-Nimbalkar, U. V. (2008b). Dynamic Copula-Based Markov Time Series. *Communications in Statistics - Theory and Methods*, 37(15):2447–2460.

Abegaz, F. and Naik-Nimbalkar, U. V. (2011). Semiparametric score test for varying copula parameter in Markov time series. *Statistics: A Journal of Theoretical and Applied Statistics*, 45(3):209–222.

Acar, E. F., Genest, C., and Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90.

Bai, J. and Ng, S. (2005). Tests for Skewness, Kurtosis, and Normality for Time Series Data. *Journal of Business & Economic Statistics*, 23(1):49–60.

Balakrishnan, N. and Lai, C.-D. (2009). *Continuous bivariate distributions*. Springer, Dordrecht, Heidelberg.

Bauwens, L. and Hautsch, N. (2006). Stochastic Conditional Intensity Processes. *Journal of Financial Econometrics*, 4(3):450–493.

Beare, B. K. (2009). A generalization of Hoeffding's lemma, and a new class of covariance inequalities. *Statistics & Probability Letters*, 79(5):637–642.

Beare, B. K. (2010). Copulas and Temporal Dependence. *Econometrica*, 78(1):395–410.

Beare, B. K. (2012). ARCHIMEDEAN COPULAS AND TEMPORAL DEPENDENCE. *Econometric Theory*, 28(6):1165–1185.

Beare, B. K. and Seo, J. (2014). TIME IRREVERSIBLE COPULA-BASED MARKOV MODELS. *Econometric Theory*, 30(05):923–960.

Bedford, T. and Cooke, R. M. (2002). Vines: A New Graphical Model for Dependent Random Variables. *The Annals of Statistics*, 30(4):1031–1068.

Behr, A. and Pötter, U. (2009). Alternatives to the normal model of stock returns: Gaussian mixture, generalised logF and generalised hyperbolic models. *Annals of Finance*, 5(1):49–68.

Bochkina, N. A. (2007). Wavelet goodness-of-fit test for dependent data. *Journal of Statistical Planning and Inference*, 137(8):2593–2612.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2):143–172.

Brechmann, E. C. and Czado, C. (2014). COPAR-multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry*, Forthcoming.

Brockwell, P. J. and Davis, R. A. (2009). *Time series*. Springer series in statistics. Springer, New York.

Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385.

Chen, X. and Fan, Y. (2004). Evaluating density forecasts via the copula approach. *Finance Research Letters*, 1(1):74–84.

Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125–154.

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

Chen, X., Koenker, R., and Xiao, Z. (2009a). Copula-based nonlinear quantile autoregression. *Econometrics Journal*, 12:50–67.

Chen, X., Wu, W. B., and Yi, Y. (2009b). Efficient estimation of copula-based semiparametric Markov models. *The Annals of Statistics*, 37(6B):4214–4253.

Cherubini, U. (2012). *Dynamic copula methods in finance*. Wiley, Chichester, West Sussex, U.K.

Cherubini, U., Mulinacci, S., and Romagnoli, S. (2011). A copula-based model of speculative price dynamics in discrete time. *Journal of Multivariate Analysis*, 102(6):1047–1063.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Corsi, F. (2008). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Darsow, W. F., Nguyen, B., and Olsen, E. T. (1992). Copulas and Markov processes. *Illinois J. Math.*, 36(4):600–642.

Davies, R. B. (1980). Algorithm AS 155: The Distribution of a Linear Combination of Chi-Square Random Variables. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 29(3):323–333.

De Luca, G. and Gallo, G. M. (2004). Mixture Processes for Financial Intradaily Durations. *Studies in Nonlinear Dynamics & Econometrics*, 8(2):1–20.

De Luca, G. and Gallo, G. M. (2008). Time-Varying Mixing Weights in Mixture Autoregressive Conditional Duration Models. *Econometric Reviews*, 28(1-3):102–120.

De Luca, G. and Zuccolotto, P. (2003). Finite and infinite mixtures for financial durations. *Metron - International Journal of Statistics*, 61(3):431–455.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–883.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(13):253–263.

Diks, C., Panchenko, V., and van Dijk, D. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, 163(2):215–230.

Dißmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

Domma, F., Giordano, S., and Perri, P. F. (2009). Statistical Modeling of Temporal Dependence in Financial Data via a Copula Function. *Communications in Statistics - Simulation and Computation*, 38(4):703–728.

Doukhan, P. (1994). *Mixing*, volume 85 of *Lecture notes in statistics*. Springer, New York.

Drost, F. C. and Werker, B. J. M. (2004). Semiparametric Duration Models. *Journal of Business & Economic Statistics*, 22(1):40–50.

Dufour, A. and Engle, R. F. (2000). The ACD model: predictibility of the time between consecutive trades. *Working Paper, ISMA Centre, University of Reading.*

Durante, F., Saminger-Platz, S., and Sarkoci, P. (2009). Rectangular Patchwork for Bivariate Copulas and Tail Dependence. *Communications in Statistics - Theory and Methods*, 38(15):2515–2527.

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

Engle, R. F. and González-Rivera, G. (1991). Semiparametric ARCH Models. *Journal of Business & Economic Statistics*, 9(4):345–359.

Engle, R. F. and Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, 66(5):1127–1162.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383–417.

Fan, Y. and Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *Journal of Nonparametric Statistics*, 11(1-3):337–360.

Fernandes, M. and Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, 130(1):1–23.

Francq, C. and Zakoïan, J.-M. (2013). Estimating the Marginal Law of a Time Series With Applications to Heavy-Tailed Distributions. *Journal of Business & Economic Statistics*, 31(4):412–425.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer series in statistics. Springer, New York.

Gajjar, A. and Subrahmaniam, K. (1978). On the sample correlation coefficient in the truncated bivariate normal population. *Communications in Statistics - Simulation and Computation*, 7(5):455–477.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368.

Ghysels, E., Gouriéroux, C., and Jasiak, J. (2004). Stochastic volatility duration models. *Journal of Econometrics*, 119(2):413–433.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26(1):53–90.

Giacomini, R. and White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, 74(6):1545–1578.

Giot, P. (1999). Time Transformations, Intraday Data And Volatility Models. *Journal of Computational Finance*, 4:31–62.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48(5):1779–1801.

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

Grammig, J. and Maurer, K.-O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *The Econometrics Journal*, 3(1):16–38.

Haas, M. (2004). A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Hansen, B. E. (1994). Autoregressive Conditional Density Estimation. *International Economic Review*, 35(3):705–730.

Hautsch, N. (2012). *Econometrics of financial high-frequency data*. Springer Berlin, Berlin.

Hobæk Haff, I. (2012). Comparison of estimators for pair-copula constructions. *Journal of Multivariate Analysis*, 110:91–105.

Hobæk Haff, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli*, 19(2):462–491.

Hobæk Haff, I., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction – Simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Ibragimov, R. (2009). COPULA-BASED CHARACTERIZATIONS FOR HIGHER ORDER MARKOV PROCESSES. *Econometric Theory*, 25(3):819–846.

Ibragimov, R. and Lentzas, G. (2008). Copulas and Long Memory. *Harvard Institute of Economic Research Discussion Paper No. 2160*.

Ignaccolo, R. (2004). Goodness-of-fit tests for dependent data. *Journal of Nonparametric Statistics*, 16(1-2):19–38.

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall, London.

Joe, H. (2014). *Dependence modelling with copulas*, volume 133 of *Monographs on statistics and applied probability*. Chapman and Hall/CRC.

Kallenberg, W. C. M. and Ledwina, T. (1997). Data-Driven Smooth Tests When the Hypothesis Is Composite. *Journal of the American Statistical Association*, 92(439):1094.

Koop, G., Pesaran, M., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119–147.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t distributions and their applications*. Cambridge Univ. Press, Cambridge.

Kruskal, W. H. (1958). Ordinal Measures of Association. *Journal of the American Statistical Association*, 53(284):814–861.

Kurowicka, D. and Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling*. Wiley, Chichester, West Sussex, U.K.

Kurowicka, D. and Joe, H., editors (2011). *Dependence modeling. Vine copula handbook.* World Scientific, Singapore.

Lagerås, A. N. (2010). Copulas for Markovian dependence. *Bernoulli*, 16(2):331–342.

Ledwina, T. (1994). Data-Driven Version of Neyman's Smooth Test of Fit. *Journal of the American Statistical Association*, 89(427):1000–1005.

Ledwina, T. and Inglot, T. (1996). Asymptotic optimality of data-driven Neyman's tests for uniformity. *The Annals of Statistics*, 24(5):1982–2019.

Lehmann, E. L. (1966). Some Concepts of Dependence. *The Annals of Mathematical Statistics*, 37(5):1137–1153.

Leucht, A. (2012). Characteristic function-based hypothesis tests under weak dependence. *Journal of Multivariate Analysis*, 108:67–89.

Lindner, A. M. (2009). Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes. In *Handbook of financial time series*, pages 43–69. Springer.

Longla, M. and Peligrad, M. (2012). Some aspects of modeling dependence in copula-based Markov chains. *Journal of Multivariate Analysis*, 111:234–240.

Malmsten, H. and Teräsvirta, T. (2010). Stylized facts of financial time series and three popular models of volatility. *European Journal of pure and applied mathematics*, 3(3):443–477.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4):394–419.

Mardia, K. V. and Thompson, J. W. (1972). Unified Treatment of Moment-Formulae. *Sankhya: The Indian Journal of Statistics, Series A*, 34(2):121–132.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1):77–91.

Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in Random variables: Theory and applications*, volume 126 of *Statistics*. Dekker, New York.

McCulloch, J. H. and Percy, E. R. J. (2013). Extended Neyman smooth goodness-of-fit tests, applied to competing heavy-tailed distributions. *Journal of Econometrics*, 172(2):275–282.

McDonald, J. B. (1984). Some Generalized Functions for the Size Distribution of Income. *Econometrica*, 52(3):647–663.

McLachlan, G. J. and Peel, D. (2000). *Finite mixture models*. Wiley, New York.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management*. Princeton series in finance. Princeton Univ. Press, Princeton NJ.

Mendes, B. V. M. d. and Aíube, C. (2011). Copula based models for serial dependence. *International Journal of Managerial Finance*, 7(1):68–82.

Meyn, S. P. and Tweedie, R. L. (2009). *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, New York.

Mikosch, T. (2006). Copulas: Tales and facts. *Extremes*, 9(1):3–20.

Müller, A. and Scarsini, M. (2005). Archimedean copulae and positive dependence. *Journal of Multivariate Analysis*, 93(2):434–445.

Munk, A., Giese, G., Stockis, J.-P., and Valeinis, J. (2011). Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data. *Annals of the Institute of Statistical Mathematics*, 63(5):939–959.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer series in statistics. Springer, New York.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2):347–370.

Neumann, M. H. and Paparoditis, E. (2000). On bootstrapping L2-type statistics in density testing. *Statistics & Probability Letters*, 50(2):137–147.

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708.

Newey, W. K. and West, K. D. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *The Review of Economic Studies*, 61(4):631–653.

Ning, C., Xu, D., and Wirjanto, T. S. (2010). Modeling Asymmetric Volatility Clusters Using Copulas and High Frequency Data. *Working Paper*.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013). Properties of hierarchical Archimedean copulas. *Statistics & Risk Modeling*, 30(1):21–54.

Pacurar, M. (2008). Autoregressive Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature. *Journal of Economic Surveys*, 22(4):711–751.

Patton, A. J. (2006). Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, 47(2):527–556.

Patton, A. J. (2009). Copula–based models for financial time series. In *Handbook of financial time series*, pages 767–785. Springer Berlin Heidelberg.

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18.

Rakonczai, P., Márkus, L., and Zempléni, A. (2012). Autocopulas: Investigating the Interdependence Structure of Stationary Time Series. *Methodology and Computing in Applied Probability*, 14(1):149–167.

Rayner, J. C. W., Best, D. J., and Thas, O. (2009). *Smooth tests of goodness of fit*. Wiley, Hoboken, NJ.

Rémillard, B., Papageorgiou, N., and Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis*, 110:30–42.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554.

Rosenbaum, S. (1961). Moments of a Truncated Bivariate Normal Distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(2):405–408.

Rüschendorf, L. and Valk, V. d. (1993). On regression representations of stochastic processes. *Stochastic Processes and their Applications*, 46(2):183–198.

Savu, C. and Ng, W. L. (2005). The SCoD Model: Analyzing Durations with a Semiparametric Copula Approach. *International Review of Finance*, 5(1-2):55–74.

Scarsini, M. (1984). On measures of concordance. *Stochastica: revista de matemática pura y aplicada*, 8(3):201–218.

Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research*, 55(2):301–327.

Smith, M., Min, A., Almeida, C., and Czado, C. (2010). Modeling Longitudinal Data Using a Pair-Copula Decomposition of Serial Dependence. *Journal of the American Statistical Association*, 105(492):1467–1479.

Smith, M. S. (2013). Copula Modelling of Dependence in Multivariate Time Series. *Working Paper.*

Sokolinskiy, O. and van Dijk, D. (2011). Forecasting Volatility with Copula-Based Time Series Models. *Tinbergen Institute Discussion Paper 125/4.*

Solow, R. M. (1960). On a Family of Lag Distributions. *Econometrica*, 28(2):393.

Spanhel, F. and Kurz, M. (2015). Simplified vine copula models: Approximations based on the simplifying assumption. *Working Paper, LMU Munich.*

Tinkl, F. and Reichert, K. (2011). Dynamic copula-based Markov chains at work: Theory, testing and performance in modeling daily stock returns. *IWQW Discussion Paper Series, FAU Erlangen-Nürnberg*, 9.

Tong, H. (1996). *Non-linear time series: A dynamical system approach*, volume 6 of *Oxford statistical science series.* Clarendon Press, Oxford.

Tse, Y.-k. and Yang, T. T. (2012). Estimation of High-Frequency Volatility: An Autoregressive Conditional Duration Approach. *Journal of Business & Economic Statistics*, 30(4):533–545.

Vaz Melo Mendes, B. d. and Accioly, V. B. (2013). Robust pair-copula based forecasts of realized volatility. *Applied Stochastic Models in Business and Industry*, 30(2):183–199.

Vries, C. G. d. (1991). On the relation between GARCH and stable processes. *Journal of Econometrics*, 48(3):313–324.

Yule, G. U. (1927). On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226:267–298.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

Spanhel, Fabian

....

Name, Vorname

....

Ort, Datum   Unterschrift Doktorand/in