Malte Simon Kurz

Dependence Modeling with Applications in Financial Econometrics

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

Eingereicht am 04.05.2018

Summary

The amount of data available in banking, finance and economics steadily increases due to the ongoing technological progress and the continuing digitalization. A key element of many econometric models for analyzing this data are methods for assessing dependencies, cross-sectionally as well as intertemporally. For this reason, the thesis is centered around statistical and econometric methods for dependence modeling with applications in financial econometrics.

The first part of this cumulative dissertation consists of three contributions. The first contribution provides a thorough explanation of the partial copula. It is a natural generalization of the partial correlation coefficient and several of its properties are investigated. In the second contribution, a different multivariate generalization of the partial correlation, the partial vine copula (PVC), is introduced. The PVC is a specific simplified vine copula (SVC) consisting of bivariate higher-order partial copulas, which are copula-based generalizations of sequentially computed partial correlations. Several properties of the PVC are presented and it is shown that, if SVCs are considered as approximations of multivariate distributions, the PVC has a special role as it is the limit of stepwise estimators. The third contribution introduces statistical tests for the simplifying assumption with a special focus on high-dimensional vine copulas. We propose a computationally feasible test for the simplifying assumption in high-dimensions, which is successfully applied to data sets with up to 49 dimensions. The novel test procedure is based on a decision tree which is used to identify the possibly strongest violation of the simplifying assumption. The asymptotic distribution of the test statistic is derived under consideration of estimation uncertainty in the copula parameters. The finite sample performance is analyzed in an extensive simulation study and the results show that the power of the test only slightly decreases in the dimensionality of the test problem.

In the second part of the dissertation, the assessment of risk measures is studied with a special focus on the financial return data used for estimation. It is shown that the choice of the sampling scheme can greatly affect the results of risk assessment procedures if the assessment frequency and forecasting horizon are fixed. Specifically, we study sequences of variance estimates and show that they exhibit spurious seasonality, if the assessment frequency is higher than the sampling frequency of non-overlapping return data. The root cause of spurious seasonality is identified by deriving the theoretical autocorrelation function of sequences of variance estimates under general assumptions. To overcome spurious seasonality, alternative variance estimators based on overlapping return data are suggested.

The third part of the dissertation is about state space methods for systems with lagged

states in the measurement equation. Recently, a low-dimensional modified Kalman filter and smoother for such systems was proposed in the literature. Special attention is paid to the modified Kalman smoother, for which it is shown that the suggested smoother in general does not minimize the mean squared error (MSE). The correct MSE-minimizing modified Kalman smoother is derived and computationally more efficient smoothing algorithms are discussed. Finally, a comparison of the competing smoothers with regards to the MSE is performed.

Zusammenfassung

Im Zuge des technologischen Fortschritts und der immer weiter voranschreitenden Digitalisierung werden immer größere Datenmengen im Bankwesen, der Finanzwirtschaft und den Wirtschaftswissenschaften erzeugt und gespeichert. Ein entscheidender Bestandteil vieler ökonometrischer Modelle zur Analyse dieser Daten sind Methoden zur Modellierung von Abhängigkeiten im Quer- und Längsschnitt. Aus diesem Grund werden in dieser Arbeit statistische und ökonometrische Methoden zur Abhängigkeitsmodellierung mit Anwendungen aus dem Bereich Finanzökonometrie entwickelt und analysiert.

Der erste Teil dieser kumulativen Dissertation umfasst drei Beiträge. Im ersten Beitrag wird das Konzept der partiellen Copula untersucht und erläutert. Es wird gezeigt, dass die partielle Copula eine natürliche Verallgemeinerung der partiellen Korrelation darstellt und verschiedene Eigenschaften der partiellen Copula werden hergeleitet. Im zweiten Beitrag wird eine weitere multivariate Verallgemeinerung der partiellen Korrelation eingeführt, die sogenannte partielle *vine copula* (PVC). Die PVC ist eine spezielle *simplified vine copula* (SVC) und besteht aus bivariaten partiellen Copulas höherer Ordnung. Diese partiellen Copulas höherer Ordnung sind copulabasierte Verallgemeinerungen der sequentiell berechneten partiellen Korrelation. Verschiedene Eigenschaften der PVC werden analysiert und es wird gezeigt, dass die PVC eine besondere Rolle einnimmt, falls SVCs zur Approximation multivariater Verteilungen verwendet werden, da schrittweise Schätzer zur PVC konvergieren. Der dritte Beitrag präsentiert statistische Tests für die *simplifying assumption*, wobei der Schwerpunkt auf hochdimensionalen *vine copulas* liegt. In diesem Sinne wird ein numerisch effizienten Test für die *simplifying assumption* vorgeschlagen, welcher erfolgreich für bis zu 49-dimensionale Datensätze angewendet wird. Die neuartige Testprozedur nutzt Entscheidungsbäume um zu identifizieren in welcher Art und Weise die stärksten Abweichungen von der *simplifying assumption* vorliegen könnten. Die asymptotische Verteilung der Teststatistik wird unter Beachtung der Schätzunsicherheit von Copulaparametern hergeleitet. Die Analyse der Testgüte für endliche Stichproben erfolgt in einer umfassenden Simulationsstudie. Die Resultate zeigen, dass die *power* des Tests nur geringfügig sinkt, wenn die Dimensionalität des Testproblems erhöht wird.

Der zweite Teil der Dissertation befasst sich mit der Schätzung von Riskomaßen, wobei ein Schwerpunkt auf die verwendeten Finanzrenditen gelegt wird. Es wird gezeigt, dass die Wahl der Beobachtungsfrequenz und der Länge des Renditeintervalls einen großen Einfluss auf die Schätzung von Risiken haben kann, falls sowohl die Häufigkeit der Risikoerhebung als auch der Zeitraum für die Vorhersage festgelegt sind. Insbesondere Zeitreihen von Varianzschätzern können Scheinsaisonalitäten (*spurious seasonality*) aufweisen, falls

sie häufiger geschätzt werden als die Beobachtungsfrequenz und Intervalllänge der nicht überlappenden Finanzrenditen, die zur Schätzung verwendet werden. Um die Ursache für die Scheinsaisonalität zu identifizieren, wird die theoretische Autokorrelationsfunktion der Zeitreihen von Varianzschätzern unter gewissen Annahmen hergeleitet. Als Lösungsansatz für das Problem der Scheinsaisonalität werden alternative Varianzschätzer vorgeschlagen, welche auf überlappenden Finanzrenditen basieren.

Im dritten Teil der Dissertation werden spezielle Zustandsraummodelle diskutiert bei denen die Beobachtungsgleichung auch von vergangenen Zuständen abhängt. Vor Kurzem wurde in der Literatur ein adaptierter niedrigdimensionaler Kalman-Filter und Kalman-Glätter vorgeschlagen. Es wird gezeigt, dass der in der Literatur vorgeschlagene Kalman-Glätter im Allgemeinen nicht den mittleren quadratischen Abstand (MSE) minimiert. Der korrekte Glätter, welcher den mittleren quadratischen Abstand minimiert, wird hergeleitet, außerdem werden numerisch effizientere Algorithmen zur Berechnung des Glätters diskutiert. Abschließend wird der in der Literatur vorgeschlagene Glätter mit dem neuen Glätter auf Basis des mittleren quadratischen Abstands verglichen.

Acknowledgements

This thesis would not have been possible without the help, support and guidance of many people.

First, I would like to thank my supervisor Stefan Mittnik for giving me the opportunity to write my thesis under his supervision. I would like to thank you for introducing me to financial econometrics during my Master's studies and for the freedom to work on all the different topics during my PhD studies. Thanks for sharing your knowledge and experiences with me – I especially enjoyed the numerous anecdotes and stories which often helped as a guidance through live as a researcher.

I also would like to thank Michael Rockinger and Dorota Kurowicka for joining my PhD committee as external reviewers and Volker Schmid and Christian Heumann for agreeing to serve as internal members in the committee.

I thank my co-author Fabian Spanhel for all the joint research we did together. The close collaboration with Fabian over several years was characterized by interesting discussions and resulted in many great results. I enjoyed breaking new ground together.

Special thanks go to Markus Glaser, the supervisor of my Bachelor's thesis. As a student research assistant at his chairs at University of Konstanz and LMU Munich, I learned all the basics of academic work, the publishing process and so much more. Without this experience as student research assistant, I would have never thought about a career in academia. I also would like to thank my former colleagues at both chairs, particularly, Jan, Matthias, Peter, Steffen and Zwetelina from whom I learned a lot.

Everything has to start somehow – for me the Bachelor studies in Mathematical Finance at University of Konstanz was the irreplaceable basis for my research. I would like to thank all professors and assistants for teaching me the basics which made this dissertation possible.

I would like to thank all former and current colleagues at the Chair of Financial Econometrics for the great atmosphere and the joint lunch and coffee breaks. Special thanks go to my colleagues at Scalable Capital – I really enjoyed working in a team of brilliant financial econometricians. I also would like to thank my friends Jan and Sven, who were always available for a coffee and helped me a lot to have a smooth start into my PhD studies.

I further want to thank all developers contributing to the version control software *git*. During my PhD studies it was the most important tool and saved my life more than once. A big thanks goes to Chris for introducing me to *git*.

For the great support and encouragement over all the years, I would like to thank my family. Last and most importantly, I thank Clara for the wonderful support, which was unbounded and diverged in the recent years.

x

Contents

Chapter I

Introduction and methods

Introduction and methods: Contributing material, research questions, related literature and remarks

The ongoing technological progress and the continuing digitalization results in steadily increasing amounts of data available in banking, finance and economics. As a consequence, there is growing demand for econometric models for adequately assessing dependencies in high-dimensional settings. This cumulative dissertation provides contributions to econometric methods for dependence modeling. The five contributing articles can be grouped into three parts and we now give a short overview of the structure of the introduction.

A key element of many econometric models are methods for assessing dependencies cross-sectionally as well as intertemporally. The first three articles (Chapters II, III & IV: Spanhel and Kurz (2016), Spanhel and Kurz (2017) & Kurz and Spanhel (2018)) are related to copulas and vine copulas which in financial econometrics are mainly applied for cross-sectional dependence modeling. In Section 1 dependence measures, copulas and vine copulas are introduced and discussed. A special focus is on partial correlations and their copula-based generalizations, the partial copula and the partial vine copula. Furthermore, we will define the simplifying assumption for vine copulas and introduce ideas for testing its validity.

The fourth article (Chapter V: Kurz and Mittnik (2018)) discusses the phenomenon spurious seasonality which can be observed when assessing risk measures and considering intertemporal dependencies. In Section 2 methods for intertemporal dependence modeling in financial econometrics are discussed. A recapitulation of weak white noise processes, like the Gaussian white noise process and the GARCH process is provided. Furthermore, temporal aggregation of stochastic processes as well as implications of the usage of overlapping (return) data for statistical methods are reviewed. Finally, the phenomenon spurious seasonality in risk assessments is introduced.

The fifth article (Chapter VI: Kurz (2018a)) is about Kalman smoothers for state space systems with a lagged state in the measurement equation. In economics, state space models are often used to model systems with cross-sectional as well as intertemporal dependencies. Section 3 recapitulates the Kalman filter and smoother. Methods for state space systems with lagged states in the measurement equation are provided and Kalman smoothers for such systems are introduced and discussed.

1 Models for cross-sectional dependence in financial econometrics: Copulas and vine copulas

In the following, we will recapitulate copulas, dependence measures and vine copulas and will provide a historic as well as a recent literature review. The first three contributions of this cumulative dissertation are related to copulas: Spanhel and Kurz (2016) is about partial copulas, Spanhel and Kurz (2017) introduces and discusses the partial vine copula and Kurz and Spanhel (2018) is about tests of the simplifying assumption. Therefore, a special focus will be on partial copulas, the partial vine copula and the simplifying assumption to provide a detailed context and discussion of the research questions being studied.

In Section 1, we use the following notation. The cdf of a d-dimensional vector of random variables, $Y_{1:d} := (Y_1, \ldots, Y_d)^T$, is denoted by $F_{Y_{1:d}}(y_{1:d}) = \mathbb{P}(Y_1 \leq y_1, \ldots, Y_d \leq y_d)$. The cdf or copula of a random vector $U_{1:d}$ with uniformly distributed margins is denoted by $F_{1:d} = C_{1:d}$. If random variables X and Y are stochastically independent, we write $X \perp Y$. For a d-dimensional function, $f(x_{1:d})$, the partial derivative with respect to the i-th element is denoted by $\partial_i f(x_{1:d}) = \frac{\partial f}{\partial x_i}(x_{1:d})$.

1.1 Copulas

First findings which resulted in the introduction of the concept which is today know as a copula were derived by Hoeffding (1940) in his dissertation on scale-invariant correlation theory.[1] The introduction of the term *copula* goes back to Sklar (1959) who developed the famous Sklar's theorem which today forms the basis in every work on copulas. Copula theory in general and their application in the area of quantitative risk management are presented in Joe (1997), Joe (2014), Nelsen (2006) and McNeil et al. (2005). Applications in financial econometrics, like, for example, the copula-GARCH model (Jondeau and Rockinger (2006) and Patton (2006)), are presented in the review article by Patton (2012) and in Cherubini et al. (2005), Cherubini et al. (2012) and Remillard (2013).

Generally speaking, every d-dimensional distribution function of d random variables which are marginally uniformly distributed constitutes a d-dimensional copula.

Definition 1 (Copula). *A d-dimensional copula is a distribution function $C_{1:d}(u_1, \ldots, u_d)$ on $[0,1]^d$ with standard uniform marginal distributions.*

Equivalently, a copula can also be defined as a map from the d-dimensional hypercube into the unit interval satisfying specific properties, see, for example, Nelsen (2006).

---

[1] The original German title is "Maszstabinvariante Korrelationstheorie".

Sklar's theorem (Sklar 1959) which states that copulas as functions join, or *couple*, the joint distribution function $F_{Y_{1:d}}$ of a d-dimensional random variable with their marginal distribution functions $F_{Y_1}, \ldots, F_{Y_d}$ is a central result in copula theory.

Theorem 1 (Sklar's theorem – Sklar (1959)). *Let $F_{Y_{1:d}}$ be a d-dimensional continuous distribution function with marginals $F_{Y_1}, \ldots, F_{Y_d}$. Then there exists a unique copula function $C_{1:d} : [0,1]^d \to [0,1]$ with*

$$F_{Y_{1:d}}(y_1, \ldots, y_d) = C_{1:d}(F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d)),$$

for all $(y_1, \ldots, y_d) \in \mathbb{R}^d$.

Sklar's theorem provides two insights: First, every combination of marginal distributions $F_{Y_1}, \ldots, F_{Y_d}$ with an arbitrary copula defines a valid d-dimensional distribution $F_{Y_{1:d}}$ and second, any joint distribution $F_{Y_{1:d}}$ can be represented via a copula and the univariate marginal distributions. This is sometimes also referred as *the joy of copulas* (Genest and MacKay (1986) and Genest and Nešlehová (2014)). It can be shown that the copula $C_{1:d}$ in Sklar's theorem is unique for continuous distributions.

It is well known that probability integral transforms (PITs), $U_i = F_{Y_i}(Y_i)$, are standard uniformly distributed for continuous distributions $F_{Y_1}, \ldots, F_{Y_d}$. If the vector of random variables $(Y_1, \ldots, Y_d)^T$ has the joint distribution function $F_{Y_{1:d}}$ with marginals $F_{Y_1}, \ldots, F_{Y_d}$, Sklar's theorem directly implies that the corresponding copula $C_{1:d}$ is the joint distribution function of the PITs $(U_1, \ldots, U_d)^T = (F_{Y_1}(Y_1), \ldots, F_{Y_d}(Y_d))^T$.

1.2 Dependence measures

In financial econometrics many basic models for cross-sectional as well as intertemporal dependence are based on simple Pearson correlations. For two random variables with existing variances, the Pearson correlation, also called linear or product moment correlation, is defined as

$$\rho(X, Y) = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

While Pearson correlations are a natural choice for elliptical distributions, they are not necessarily useful and interpretable when leaving the elliptical world. As pointed out in McNeil et al. (2005), elliptical distributions are fully determined by a mean vector, a covariance matrix and a characteristic generator function. As a result, if one considers copulas which allow to separate the modeling of the margins from the dependence modeling, the mean

vector and the variances solely determine the margins. This means that copulas of ellipti-cal distributions are basically determined by the correlation matrix and the characteristic generator function. From this point of view it is obvious that the correlation matrix alone forms a rich knowledge about the overall multivariate distribution in case it is elliptical, but this is not the case in general. Constructing examples that clearly show that the joint distribution of random variables is by far not fully specified if the correlation matrix and the marginal distributions are specified is straightforward using copulas, see, for example, McNeil et al. (2005). Another important fact is that, given the univariate marginals, it is not always possible to construct a joint distribution function for every correlation in $[-1, 1]$. That means, that the marginals determine an interval of attainable correlations, see Theorem 5.25 in McNeil et al. (2005).

These drawbacks of correlations can be overcome by using copulas instead. Sklar's theorem implies that every combination of marginals and a copula determine a valid joint distribution and the joint distribution is fully specified by the combination of the marginals together with the copula. Moreover, it is well known that correlations are invariant with respect to strictly increasing linear transformations but in general not invariant with respect to non-linear strictly increasing transformations. In contrast, copulas are invariant with respect to every strictly increasing transformation (Schweizer and Wolff 1981, Theorem 3) and therefore perfectly suited as a scale-invariant dependence measure.

Copula functions greatly serve as dependence measure as they provide the full scale-invariant information about the dependence. However, often it is of practical interest to break the information down to a single number as a dependence measure. Examples for such copula-based measures of dependence in the bivariate case are Kendall's τ (Kendall 1938) and Spearman's ρ_S (Spearman 1904). For continuous random variables X and Y with unique copula C, Kendall's τ and Spearman's ρ_S are given by (Nelsen 2006)

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) \, \mathrm{d}C(u, v) - 1$$

and

$$\rho_S(X, Y) = 12 \int_0^1 \int_0^1 C(u, v) \, \mathrm{d}u \, \mathrm{d}v - 3.$$

By Sklar's theorem, the copula C is the joint distribution function of the two PITs $U = F_X(X)$ and $V = F_Y(Y)$ and it turns out that the scale-invariant Spearman's ρ_S is equal to the Pearson correlation of U and V, i.e.,

$$\rho_S(X, Y) = \mathrm{Corr}(F_X(X), F_Y(Y)).$$

Interestingly, as pointed out by Genest and Nešlehová (2014), scale invariant dependence measures like the rank correlations Spearman's ρ_S (Spearman 1904) and Kendall's τ (Kendall 1938), and also the work by Hoeffding (1940) have been developed long before the general concept of copulas was established. Today copula theory provides an easy and straightforward access to these scale-invariant dependence measures. However, Spearman's ρ and Kendall's τ can also be defined without copulas as measures of concordance and can be estimated using the ranks of associated data.

1.3 Partial correlations and the partial copula

In many applications, e.g., in financial econometrics, the interest lies in the dependence of two random variables conditional on a third random variable or vector. For illustration, we consider a four-dimensional random vector, i.e., we are looking for a dependence model for the random variables Y_1 and Y_4 given $Y_{23} := (Y_2, Y_3)^T$. A straightforward measure for the conditional dependence is the conditional correlation

$$\text{Corr}(Y_1, Y_4 | Y_{23}) = \frac{\mathbb{E}(Y_1 Y_4 | Y_{23}) - \mathbb{E}(Y_1 | Y_{23}) \mathbb{E}(Y_4 | Y_{23})}{\sqrt{\text{Var}(Y_1 | Y_{23}) \text{Var}(Y_4 | Y_{23})}}$$

between Y_1 and Y_4 given Y_{23}. The conditional correlation has the disadvantage that it is a function in Y_{23} and often one is interested in summarizing the measure in a single number. One attempt to do so, is the so-called partial correlation coefficient. Assume w.l.o.g. that all marginals of $Y_{1:4} := (Y_1, Y_2, Y_3, Y_4)^T$ have zero mean and that their variance is finite. Further denote the best linear predictor of Y given X by $\mathcal{P}(Y|X)$, i.e., $\mathcal{P}(Y|X)$ is a linear function in X and within this class of functions minimizes the mean-squared error $\mathbb{E}((Y - \mathcal{P}(Y|X))^2)$.

For the two variables of interest Y_1 and Y_4, we remove all linear effects of the conditioning vector Y_{23} by considering the prediction errors

$$\tilde{\mathcal{E}}_{1|23} = Y_1 - \mathcal{P}(Y_1 | Y_{23}), \qquad\qquad \tilde{\mathcal{E}}_{4|23} = Y_4 - \mathcal{P}(Y_4 | Y_{23}).$$

The partial correlation is then defined as the Pearson correlation of these prediction errors

$$\rho_{14;23} = \text{Corr}(\tilde{\mathcal{E}}_{1|23}, \tilde{\mathcal{E}}_{4|23}),$$

i.e., after decorrelating the two variables of interest with respect to the conditioning vector, we measure the dependence using the Pearson correlation.

As in the unconditional case, the partial correlation is a natural choice as measure
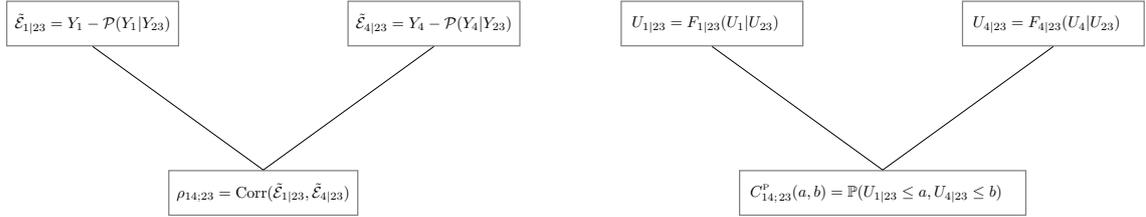
Figure 1: On the left, the partial correlation $\rho_{14;23}$ of Y_1 and Y_4 given Y_{23} is graphically illustrated. The right plot shows the corresponding partial copula for the PITs U_1 and U_4 given U_{23}.

of conditional dependence in the case of a joint elliptical distribution for $Y_{1:4}$, but not necessarily outside the world of elliptical distributions. To generalize the partial correlation using copulas, we can consider conditional probability integral transforms (CPITs). Assume w.l.o.g. that all marginals $Y_{1:4}$ have already been transferred to the copula level, i.e., we consider the standard uniformly distributed random variables $U_i = F_{Y_i}(Y_i)$, for $i = 1, \ldots, 4$. Then, we can define CPITs, the conditional copula and the partial copula.

Definition 2 (Bivariate conditional and partial copula). *Let $U_{1:4} \sim F_{1:4} = C_{1:4}$.*

(i) *$U_{k|23} := F_{k|23}(U_k|U_{23})$ is the conditional probability integral transform (CPIT) of U_k with respect to U_{23}, for $k = 1, 4$.*

(ii) *The bivariate conditional copula $C_{14;23}$ of $F_{14|23}$ (Patton (2006)) is defined as*

$$C_{14;23}(a, b|u_{23}) := \mathbb{P}(U_{1|23} \leq a, U_{4|23} \leq b|U_{23} = u_{23}).$$

(iii) *The bivariate partial copula $C_{14;23}^P$ of $F_{14|23}$ (Bergsma 2004, Gijbels et al. 2015b, Spanhel and Kurz 2016) is defined as*

$$C_{14;23}^P(a, b) := \mathbb{P}(U_{1|23} \leq a, U_{4|23} \leq b).$$

One can show that CPITs are standard uniformly distributed, i.e., $U_{1|23}, U_{4|23} \sim \mathcal{U}([0, 1])$, and that it holds $U_{1|23} \perp U_{23}$ and $U_{4|23} \perp U_{23}$. This means that by considering CPITs instead of the errors from a linear prediction, one obtains a random variable that it is not only uncorrelated, but even independent of the conditioning vector U_{23}. In this sense, the conditional copula can be seen as a generalization of the conditional correlation and the partial copula can be seen as a natural generalization of the partial correlation coefficient. This is also summarized in Figure 1, where the left plot illustrates the partial correlation coefficient as correlation of the errors from a linear prediction and the right plot shows the partial copula, which is the joint unconditional copula of the corresponding CPITs. The

first contribution (Chapter II: Spanhel and Kurz (2016)) of this cumulative dissertation is a thorough explanation of the partial copula and an investigation of its properties. While in

> ### Chapter II:
> ### The partial copula: Properties and associated dependence measures
>
> F. Spanhel and M. S. Kurz. 2016. "The partial copula: Properties and associated dependence measures". `https://doi.org/10.1016/j.spl.2016.07.014`, *Statistics & Probability Letters* 119: 76–83.

Spanhel and Kurz (2016) most results have been obtained using parametric copula families, independently, Gijbels et al. (2015b) provided an analysis of the partial copula in the non-parametric context. Gijbels et al. (2015b) also investigate the so-called average copula which coincides with the partial copula for continuous distributions, but not in general.

In Spanhel and Kurz (2016), examples for parametric partial copulas are derived. For example, the partial Frank copula, which corresponds to the partial copula that one obtains when conditioning on one variable of a three-dimensional Frank copula. Interestingly, while the corresponding conditional copula, the Ali-Mikhail-Haq copula, is an Archimedean copula (Mesfioui and Quessy 2008), the partial copula is not an Archimedean copula. A direct consequence is that the partial copula is in general not a member of the same copula family as the conditional copula. It is further shown, that the partial copula minimizes the Kullback-Leibler (KL) divergence to the conditional copula but not the L^2 distance. Moreover, the conditional distribution induced by the partial copula in general does not minimize the KL distance to the corresponding conditional distribution. Consequently, if one estimates the parameters of the CPITs and the partial copula, joint and stepwise ML estimates in general do not converge to the same probability limit. Relations between conditional correlations, partial correlations and the partial copula are presented. Furthermore, as the partial copula is an unconditional copula, one can also consider the corresponding dependence measures, like Spearman's ρ_S, Kendall's τ and the tail dependence coefficients (see Sibuya (1960) and Joe (1997) for the definition of lower and upper tail dependence). In Gijbels et al. (2015b) as well as in Spanhel and Kurz (2016), properties for those dependence measures induced by the partial copula are derived.

1.4 Vine copulas and the simplifying assumption

In the bivariate case there exist numerous copulas that can be used for dependence modeling, see Joe (2014), Nelsen (2006) or Balakrishnan (2009) for lists of copula families. The most

popular are the Archimedean and elliptical families. When it comes to higher-dimensional dependence modeling, the number of available parametric models and their flexibility is rather limited. Consequently hierarchical copula models, like hierarchical Archimedean copulas and especially vine copulas turned out to be a break-though for high-dimensional dependence modeling. Vine copulas, also called pair-copula constructions, have been developed in Bedford and Cooke (2001), Bedford and Cooke (2002), Joe (1996), and Joe (1997). The article by Aas et al. (2009) about inference for vine copulas attracted much attention and has led to a lot of applied research in the area of high-dimensional dependence modeling with vine copulas. Vine copula theory and their applications are presented in great detail in the books of Joe (1997), Joe (2014), Kurowicka and Cooke (2006b) and Kurowicka and Joe (2011). In a recent review article, Aas (2016) gives an overview about vine copulas with a special focus on applications from financial econometrics, like in the areas capital asset pricing, credit risk, liquidity risk, market risk, systemic risk, option pricing and portfolio optimization.

The general idea of pair-copula constructions is to build up multivariate copulas using bivariate copulas (pair-copulas) as building blocks. A very general class, the so-called regular vine (R-vine) copulas, have been developed by Bedford and Cooke (2002) and Kurowicka and Cooke (2006b). Using graph-theoretical arguments, a so-called proximity condition as integral element of the definition of regular vines guarantees that the resulting regular vine copula model defines a valid multivariate copula. The commonly named *simplifying assumption* (Hobæk Haff et al. 2010) is the central model assumption rendering the model class tractable for statistical inference. The simplifying assumption is the assumption that every conditional copula (Definition 2) being a building block of the vine copula can be represented by an unconditional copula, i.e., that the conditional copula itself does not depend on the conditioning variables.

In the following, we describe the idea of vine copulas using a four-dimensional example, i.e., the random variables $U_{1:4}$ for which we assume w.l.o.g. that the marginals are uniformly distributed. From a graph-theoretic point of view, simplified vine copulas (SVCs) can be considered as an order sequence of trees, where bivariate unconditional copulas are assigned to each of the $d - j$ edges of tree j (Bedford and Cooke 2002). In the following, we will restrict ourselves to the subclass of drawable vines (D-vines) but the generalization to R-vines is straightforward and for example all implementations in Chapter IV have been done for R-vine copulas.

The density of a simplified D-vine copula (Joe (1996) and Aas et al. (2009)) for the four-dimensional example is given by the following product of six unconditional bivariate

(a) Simplified D-vine copula. **(b)** D-vine copula.
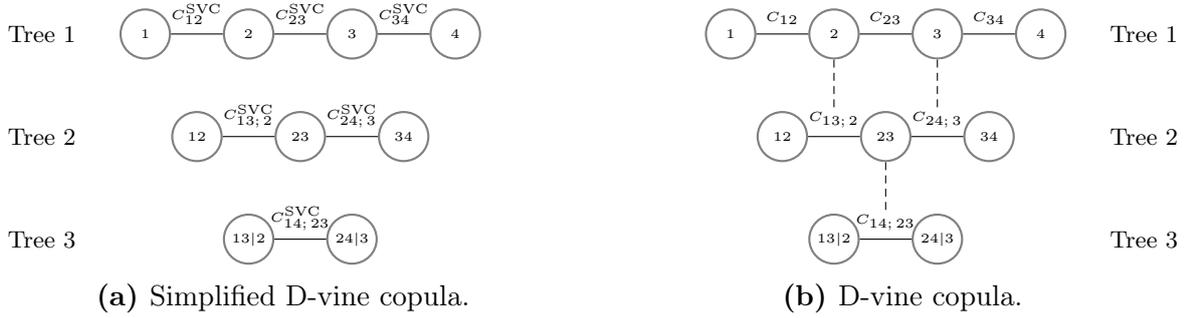
Figure 2: (Simplified) D-vine copula representation if $d = 4$. The influence of conditioning variables on the conditional copulas is indicated by dashed lines.

copula densities

$$c_{1:4}^{\text{SVC}}(u_{1:4}) = \underbrace{c_{12}^{\text{SVC}}(u_1, u_2) c_{23}^{\text{SVC}}(u_2, u_3) c_{34}^{\text{SVC}}(u_3, u_4)}_{\text{first tree}} \times \underbrace{c_{13;2}^{\text{SVC}}(u_{1|2}^{\text{SVC}}, u_{3|2}^{\text{SVC}}) c_{24;3}^{\text{SVC}}(u_{2|3}^{\text{SVC}}, u_{4|3}^{\text{SVC}})}_{\text{second tree}}$$

$$\times \underbrace{c_{14;23}^{\text{SVC}}(u_{1|23}^{\text{SVC}}, u_{4|23}^{\text{SVC}})}_{\text{third tree}},$$

where the arguments of the copulas in the second and third tree are exemplary obtained as

$$u_{1|2}^{\text{SVC}} = \partial_2 C_{12}^{\text{SVC}}(u_1, u_2), \qquad\qquad u_{4|23}^{\text{SVC}} = \partial_1 C_{24;3}^{\text{SVC}}(u_{2|3}^{\text{SVC}}, u_{4|3}^{\text{SVC}}).$$

A graphical representation of the four-dimensional simplified D-vine copula is given on the left of Figure 2, where the solid lines represent the edges to which the bivariate unconditional copulas are assigned.

Each of the involved $d(d-1)/2$ bivariate unconditional copulas, i.e., six in the four-dimensional case, can be chosen arbitrarily and the resulting SVC model $c_{1:4}^{\text{SVC}}(u_{1:4})$ always defines a valid four-dimensional copula density. Therefore, SVCs allow to construct a wide variety of flexible multivariate copulas. Pair-copula constructions or vine copulas are attractive for high-dimensional dependence modeling because they are solely constructed by bivariate unconditional copulas. This renders the multivariate density computationally feasible for model selection, parameter estimation and inference. It is obvious that not every copula density can be written as a SVC, see, for example, Hobæk Haff et al. (2010) for counterexamples.

In contrast, every copula can be represented by a (non-simplified) vine copula. The

four-dimensional D-vine copula is given by

$$c_{1:4}(u_{1:4}) = \underbrace{c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{34}(u_3, u_4)}_{\text{first tree}} \times \underbrace{c_{13;2}(u_{1|2}, u_{3|2}|u_2)c_{24;3}(u_{2|3}, u_{4|3}|u_3)}_{\text{second tree}}$$

$$\times \underbrace{c_{14;23}(u_{1|23}, u_{4|23}|u_{23})}_{\text{third tree}}, \tag{1.1}$$

where the building blocks assigned to the edges are not bivariate unconditional but conditional copulas (Definition 2) and the arguments of the copulas in the second and third tree are exemplary obtained as $u_{1|2} = \partial_2 C_{12}(u_1, u_2)$ and $u_{4|23} = \partial_1 C_{24;3}(u_{2|3}, u_{4|3}|u_3)$. The structure of the four-dimensional vine copula is illustrated on the right of Figure 2, where the influence of the conditioning variables on the conditional copulas is illustrated by dashed lines.

In applications, the simplifying assumption is typically imposed, i.e., it is assumed that all bivariate conditional copulas of the data generating vine copula degenerate to bivariate unconditional copulas.

Definition 3 (The simplifying assumption – Hobæk Haff et al. (2010)). *The four-dimensional D-vine copula* (1.1) *satisfies the simplifying assumption if the conditional copulas* $c_{13;2}(\cdot, \cdot|u_2)$, $c_{24;3}(\cdot, \cdot|u_3)$, $c_{14;23}(\cdot, \cdot|u_{23})$ *do not depend on the conditioning variables,* u_2, u_3 *and* u_{23}, *respectively.*

Usually it is assumed that the simplifying assumption is satisfied for the data generating copula. As a consequence, the copula can be represented by a SVC, which results in fast and simple statistical inference. Under this assumption, several methods for the consistent specification and estimation of vine copulas have been developed (Hobæk Haff (2012, 2013) and Dißmann et al. (2013)).

Hobæk Haff et al. (2010) show that the simplifying assumption is in general not satisfied. However, it can be easily verified that the multivariate Gaussian copula can be represented by a SVC, but the identified multivariate distributions for which the simplifying assumption is satisfied are very limited. Stöber et al. (2013) show that the only Archimedean copula that can be represented by a SVC is the Clayton copula and that the Student-t copula is the only copula arising from scale mixtures of normal distributions that is in line with the simplifying assumption. Acar et al. (2012) also study the simplifying assumption and suggest an estimator for conditional copulas with a one-dimensional conditioning variable which can be applied if the simplifying assumption is not satisfied.

Recently, the simplifying assumption has attracted more attention, especially in the context of partial and conditional copulas, but also for vine copulas. Statistical tests for

the simplifying assumption are studied in Acar et al. (2013), Derumigny and Fermanian (2017), Gijbels et al. (2017a, 2017b) and Kurz and Spanhel (2018). Graphical visualizations of the simplifying assumption and model distances have been analyzed in Killiches et al. (2016) and Killiches et al. (2017). Non- and semiparametric estimation of partial and conditional copulas under the simplifying assumption is studied in Gijbels et al. (2015a) and Portier and Segers (2018). Finally, different estimators for non-simplified vine copulas and conditional copulas have been proposed in Acar et al. (2012), Gijbels et al. (2011), Schellhase and Spanhel (2018), Vatter and Chavez-Demoulin (2015), Vatter and Nagler (2018) and Veraverbeke et al. (2011).

1.5   Higher-order partial copulas and the partial vine copula (PVC)

If one considers the conditional distribution of two random variables conditional on more than one random variable, i.e., a random vector, the partial correlation can be obtained in different ways. Using again the example, where we are interested in the partial correlation between Y_1 and Y_4 conditional on Y_{23}, we can determine the prediction errors after subtracting the best linear predictors $\mathcal{P}(Y_1|Y_{23})$ and $\mathcal{P}(Y_4|Y_{23})$. The partial correlation is then given by

$$\rho_{14;23} = \text{Corr}(\tilde{\mathcal{E}}_{1|23}, \tilde{\mathcal{E}}_{4|23}) = \text{Corr}(Y_1 - \mathcal{P}(Y_1|Y_{23}), Y_4 - \mathcal{P}(Y_4|Y_{23})).$$

As an alternative, the partial correlation can also be computed sequentially, see, for example, Yule and Kendall (1953). To obtain the sequential formula, we first consider four errors from linear predictions, each on a single random variable

$$\mathcal{E}_{1|2} = Y_1 - \mathcal{P}(Y_1|Y_2), \quad \mathcal{E}_{3|2} = Y_3 - \mathcal{P}(Y_3|Y_2), \quad \mathcal{E}_{2|3} = Y_2 - \mathcal{P}(Y_2|Y_3), \quad \mathcal{E}_{4|3} = Y_4 - \mathcal{P}(Y_4|Y_3).$$

In a second step, we consider the linear prediction of pairs of those first-order prediction errors, i.e.,

$$\mathcal{E}_{1|23} = \mathcal{E}_{1|2} - \mathcal{P}(\mathcal{E}_{1|2}|\mathcal{E}_{3|2}), \qquad\qquad \mathcal{E}_{4|23} = \mathcal{E}_{4|3} - \mathcal{P}(\mathcal{E}_{4|3}|\mathcal{E}_{2|3}).$$

One can then easily verify (Yule and Kendall 1953), that those prediction errors are equal to those obtained by directly correcting for the linear effect of the vector Y_{23}, i.e.,

$$\mathcal{E}_{1|23} = \mathcal{E}_{1|2} - \mathcal{P}(\mathcal{E}_{1|2}|\mathcal{E}_{3|2}) = Y_1 - \mathcal{P}(Y_1|Y_{23}) = \tilde{\mathcal{E}}_{1|23},$$
$$\mathcal{E}_{4|23} = \mathcal{E}_{4|3} - \mathcal{P}(\mathcal{E}_{4|3}|\mathcal{E}_{2|3}) = Y_4 - \mathcal{P}(Y_4|Y_{23}) = \tilde{\mathcal{E}}_{4|23}.$$
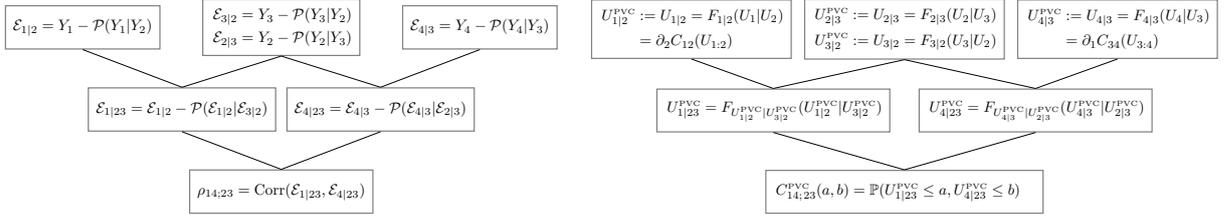
Figure 3: On the left, the sequential formula for deriving the partial correlation $\rho_{14;23}$ of Y_1 and Y_4 given Y_{23} is graphically illustrated. The right plot shows the corresponding sequential definition of the second-order partial copula $C_{14;23}^{\mathrm{PVC}}(a,b) = \mathbb{P}(U_{1|23}^{\mathrm{PVC}} \leq a, U_{4|23}^{\mathrm{PVC}} \leq b)$ as joint distribution of the PPITs $U_{1|23}^{\mathrm{PVC}}$ and $U_{4|23}^{\mathrm{PVC}}$.

As a result the partial correlation can also be computed sequentially via

$$\rho_{14;23} = \mathrm{Corr}(\mathcal{E}_{1|23}, \mathcal{E}_{4|23}).$$

This sequential approach is illustrated on the left of Figure 3.

When looking at the sequential formula for computing partial correlations, it is obvious that there is a close relation between partial correlations and vine copulas. In fact, many early papers and books about pair-copula constructions, vine copulas and so-called partial correlation vines point out, or work with this close relation, like for example Bedford and Cooke (2002) and Kurowicka and Cooke (2003, 2006a, 2006b).

Using the idea of proceeding sequentially, vine copulas can be applied to define a different generalization of partial correlations, the so-called higher-order partial copulas which constitute the partial vine copula (PVC). Higher-order partial copulas and the PVC are defined, explained as a concept for dependence modeling and analyzed in the second contribution (Chapter III: Spanhel and Kurz (2017)) of this cumulative dissertation.

> ### Chapter III:
> The partial vine copula: A dependence measure and approximation based on the simplifying assumption
>
> F. Spanhel and M. S. Kurz. 2017. "The partial vine copula: A dependence measure and approximation based on the simplifying assumption". ArXiv e-prints `arXiv: 1510.06971`.

To determine or estimate a partial copula for a high-dimensional conditioning variable, one needs to find a model for the corresponding CPITs $U_{1|23}$ and $U_{4|23}$. Usually, this involves the estimation of the joint distributions or copulas, i.e., $C_{1:3}$ and $C_{2:4}$, and then the CPITs

can be obtained as

$$U_{1|23} = F_{1|23}(U_1|U_{23}) = \frac{\partial_2 \partial_3 C_{1:3}(U_{1:3})}{c_{23}(U_2, U_3)} = \partial_2 C_{13;2}(U_{1|2}, U_{3|2}|U_2),$$

$$U_{4|23} = F_{4|23}(U_4|U_{23}) = \frac{\partial_1 \partial_2 C_{2:4}(U_{2:4})}{c_{23}(U_2, U_3)} = \partial_1 C_{24;3}(U_{2|3}, U_{4|3}|U_3).$$

Due to the lack of appropriate models for high-dimensional densities and the curse of dimensions, an attractive alternative is given by using a sequential approach. In this regard, we proceed similar to the sequential approach for computing partial correlations but generalized to copulas. It will turn out that a natural generalization in this regard is the PVC.

To understand this new measure of dependence, we start by considering the CPITs

$$U_{1|2} = F_{1|2}(U_1|U_2) =: U_{1|2}^{\text{PVC}}, \qquad\qquad U_{3|2} = F_{3|2}(U_3|U_2) =: U_{3|2}^{\text{PVC}}$$

and

$$U_{2|3} = F_{2|3}(U_2|U_3) =: U_{2|3}^{\text{PVC}}, \qquad\qquad U_{4|3} = F_{4|3}(U_4|U_3) =: U_{4|3}^{\text{PVC}}.$$

By Definition 2, the joint distributions of these CPITs are the partial copulas $C_{13;2}^{\text{P}}$ and $C_{24;3}^{\text{P}}$, respectively. If we continue sequentially, we can *remove* the effect of $U_{3|2}^{\text{PVC}}$ on $U_{1|2}^{\text{PVC}}$ by considering the corresponding CPIT through the partial copula, i.e,

$$U_{1|23}^{\text{PVC}} = F_{U_{1|2}^{\text{PVC}}|U_{3|2}^{\text{PVC}}}(U_{1|2}^{\text{PVC}}|U_{3|2}^{\text{PVC}}) = \partial_2 C_{13;2}^{\text{P}}(U_{1|2}^{\text{PVC}}, U_{3|2}^{\text{PVC}}),$$

$$U_{4|23}^{\text{PVC}} = F_{U_{4|3}^{\text{PVC}}|U_{2|3}^{\text{PVC}}}(U_{4|3}^{\text{PVC}}|U_{2|3}^{\text{PVC}}) = \partial_1 C_{24;3}^{\text{P}}(U_{2|3}^{\text{PVC}}, U_{4|3}^{\text{PVC}}),$$

which are defined as partial probability integral transforms (PPITs) in Spanhel and Kurz (2017). Finally, the corresponding sequential generalization of the partial correlation between U_1 and U_4 given U_{23} is the so-called second-order partial copula, which is the copula and joint distribution function of the PPITs, i.e.,

$$C_{14;23}^{\text{PVC}}(a, b) = \mathbb{P}(U_{1|23}^{\text{PVC}} \le a, U_{4|23}^{\text{PVC}} \le b).$$

The right graphic in Figure 3 summarizes the steps in the four-dimensional case and allows to easily compare it to the stepwise computation of the partial correlation coefficient which is illustrated on the left of Figure 3.

It is important to note that in general, the PPITs and CPITs do not coincide. As a consequence, the partial copula $C_{14;23}^{\text{P}}(a, b)$ is also in general not equal to the second-order partial copula $C_{14;23}^{\text{PVC}}(a, b)$. In Lemma 3.1 (Spanhel and Kurz 2017), it is shown that PPITs

14

and CPITs coincide if and only if the simplifying assumption holds. Therefore, partial copulas and higher-order partial copulas are equal if and only if the simplifying assumption holds for the corresponding vine copula.

Furthermore, every higher-order partial copula is not only determined by the index set representing the variables (e.g., $14;23$ for U_1 and U_4 given U_{23}), but also by the vine structure. In contrast to the D-vine structure discussed before, we can for example consider a C-vine with variable U_2 as root node in the first tree and variable U_3 as root node in the second tree (see Kurowicka and Cooke (2006b) for the definition of C-vines). The corresponding CPITs obtained from the bivariate copulas C_{21}, C_{23}, C_{24} are

$$U_{1|2} = F_{1|2}(U_1|U_2) = \partial_1 C_{21}(U_2, U_1) =: U_{1|2}^{\mathrm{PVC}},$$
$$U_{3|2} = F_{3|2}(U_3|U_2) = \partial_1 C_{23}(U_2, U_3) =: U_{3|2}^{\mathrm{PVC}},$$
$$U_{4|2} = F_{4|2}(U_4|U_2) = \partial_1 C_{24}(U_2, U_4) =: U_{4|2}^{\mathrm{PVC}},$$

and then using the partial copulas $C_{31;2}^{\mathrm{P}}$ and $C_{34;2}^{\mathrm{P}}$, we obtain the PPITs

$$U_{1|23}^{\mathrm{PVC}} = F_{U_{1|2}^{\mathrm{PVC}}|U_{3|2}^{\mathrm{PVC}}}(U_{1|2}^{\mathrm{PVC}}|U_{3|2}^{\mathrm{PVC}}) = \partial_1 C_{31;2}^{\mathrm{P}}(U_{3|2}^{\mathrm{PVC}}, U_{1|2}^{\mathrm{PVC}}),$$
$$U_{4|23}^{\mathrm{PVC}} = F_{U_{4|2}^{\mathrm{PVC}}|U_{3|2}^{\mathrm{PVC}}}(U_{4|2}^{\mathrm{PVC}}|U_{3|2}^{\mathrm{PVC}}) = \partial_1 C_{34;2}^{\mathrm{P}}(U_{3|2}^{\mathrm{PVC}}, U_{4|2}^{\mathrm{PVC}}),$$

and finally the second-order partial copula

$$C_{14;23}^{\mathrm{PVC}}(a, b) = \mathbb{P}(U_{1|23}^{\mathrm{PVC}} \leq a, U_{4|23}^{\mathrm{PVC}} \leq b).$$

This second-order partial copula describes the conditional dependence for the same variables as before, but it is not necessarily the same as the one obtained above with the D-vine structure. This illustrates that higher-order partial copulas for specific variable combinations are not unique and always depend on the involved random variables as well as the vine structure.

1.6 The partial vine copula, its properties and applications

If one continues this strategy and assigns higher-order partial copulas in every tree of a vine copula, one obtains the PVC which is introduced in Spanhel and Kurz (2017). In this regard, the PVC is a new multivariate dependence measure for d-dimensional random vectors in terms of $d(d-1)/2$ bivariate unconditional copulas and can be seen as a generalization of the partial correlation matrix which consists of $d(d-1)/2$ (partial) correlations.

Spanhel and Kurz (2017) further prove several properties of the PVC. For example, it

is shown that for non-simplified vine copulas the implicit bivariate margins of the PVC do not necessarily match the true bivariate margins. Moreover, it is proven that a pair of random variables may be considered as conditionally (in)dependent according to the PVC, although this is not the case for the data generating process.

Furthermore, the PVC plays a central role when considering SVCs as approximations for multivariate copulas. In Spanhel and Kurz (2017) it is shown that the PVC minimizes the KL divergence to the true copula in a stepwise fashion within the class of SVCs. In practice, this implies that the PVC is often the best feasible SVC approximation, because most model selection algorithms and estimators for SVCs proceed in a stepwise fashion. As a consequence, under regularity conditions, sequential estimators for vine copulas converge to the PVC irrespectively whether the data generating copula satisfies the simplifying assumption or not. However, it is also shown that in general the PVC does not globally minimize the KL divergence to the true copula in the space of SVCs. As a consequence, it might not be optimal to specify the true copulas in the first tree of a SVC approximation, which might be counter-intuitive. Additionally, joint and stepwise estimators of SVCs may not converge to the same probability limit if the simplifying assumption is not satisfied.

Nagler and Czado (2016) propose a non-parametric estimator of the PVC and derive its asymptotic properties. Their results show that the non-parametric estimator of the PVC often outperforms classical non-parametric approaches for multivariate density estimation. Additionally, the PVC turns out to be particularly useful for testing the simplifying assumption in high-dimensional vine copulas (Kurz and Spanhel 2018), which is the topic of the next section.

1.7 Testing the simplifying assumption in high-dimensional vine copulas

As pointed out earlier, the simplifying assumption is almost always used when developing methods for vine copulas or applying them for dependence modeling. A natural question that arises in this context is whether the simplifying assumption is valid and how this can be verified in practical applications. To answer this question, statistical tests for the simplifying assumption in high-dimensional vine copulas are developed in the third contribution (Chapter IV: Kurz and Spanhel (2018)) of this cumulative dissertation. Under the simplifying assumption, every conditional copula in the vine copula collapses to an unconditional copula. Kurz and Spanhel (2018) derive stochastic interpretations of the simplifying assumption, which form the basis for testing the simplifying assumption. It turns out that the simplifying assumption is equivalent to $(d-1)(d-2)/2$ vectorial independence assumptions.

> ## Chapter IV:
> ### Testing the simplifying assumption in high-dimensional vine copulas
>
> M. S. Kurz and F. Spanhel. 2018. "Testing the simplifying assumption in high-dimensional vine copulas". ArXiv e-prints `arXiv:1706.02338`.

These vectorial independencies are closely related to partial copulas, higher-order partial copulas and the PVC. In Kurz and Spanhel (2018) it is shown that especially the formulation based on the PVC, and the related partial probability integral transforms (PPITs), renders the test problem feasible even for high-dimensional vine copulas.

The key challenges of testing the simplifying assumption in high-dimensional vine copulas, are that the test problem involves checking restrictions on high-dimensional distributions and that tests have to be based on estimated observations. In the literature, tests for the simplifying assumption are so far restricted to the special case of testing a single bivariate conditional copula using non- or semiparametric estimators and tests (Acar et al. (2013) and Gijbels et al. (2017a, 2017b)). A recent survey is given by Derumigny and Fermanian (2017). In these studies, the simulations and applications are done for rather low-dimensional conditioning variables. In contrast, Kurz and Spanhel (2018) consider data sets of up to 49 dimensions in their applications and test the simplifying assumption for complete vine copula models and not only a single conditional copula.

In Kurz and Spanhel (2018), it is proposed to use a constant conditional correlation (CCC) test based on a discretization of the support of the conditioning variables. A decision tree algorithm is applied to detect the possibly largest deviation from the simplifying assumption. The decision tree approach results in a computationally feasible test even in high dimensions. A penalty is incorporated in the test statistic to derive its asymptotic distribution which also involves the application of the decision tree approach. A semi-parametric framework is used to test the simplifying assumption. Applying ranks, pseudo-observations from the PVC are obtained which are then used to compute stepwise maximum likelihood estimates of copula parameters. These estimates are plugged in in order to compute pseudo observations from the PPITs. The asymptotic distribution of the PPITs depending on estimated parameters can be obtained using the theory of (Hobæk Haff 2012, 2013) which is going back to Tsukahara (2005). Building on this, in Kurz and Spanhel (2018) the asymptotic distribution of the CCC test statistic under consideration of the estimation uncertainty in the PPITs is derived.

An extensive simulation study demonstrates the finite sample performance of the CCC test and particularly shows that the power of the test only slightly decreases with the

dimension of the conditioning variables. Moreover, the simulations show that vine structure selection and also misspecification of the parameters of the copulas in the lower trees only mildly affect the power properties of the CCC test. In the applications to 10 different data sets, mainly filtered financial returns as they are often modeled in financial econometrics, it is demonstrated that the simplifying assumption should be checked individually for each data set. The accompanying R-package `pacotest` (Kurz 2018c), available on CRAN, allows researchers to easily apply the proposed test procedure to their own vine copula models. The package is based on a C++-kernel, resulting in fast computations. Furthermore, the implementation contains an interface which allows researchers to vary all tuning parameters and graphical illustrations of the decision trees can be automatically produced.

The CCC test is not only useful as a specification test for SVC models, but it has also been applied to improve dependence models for high-dimensional data. In this regard, Schellhase and Spanhel (2018) used the CCC test to identify building blocks of vine copulas where the modeling of conditional copulas is more appropriate than relying on unconditional copulas. In other words, the test is applied to identify non-simplified building blocks in vine copulas. Additionally, Kraus and Czado (2017) apply the CCC test as part of new model selection algorithms which aim to find structures that are as simplified as possible. They show that their newly proposed model selection algorithms can outperform the popular Dißmann et al. (2013) algorithm.

2   Models for intertemporal dependence in financial econometrics: Weak white noise and GARCH processes

Econometric methods for modeling and predicting time series (Hamilton (1994), Tsay (2010) and Andersen et al. (2009)) are key to modern quantitative risk management. Risk managers, bankers, asset managers and regulators are often interested in estimating and predicting risk measures like value-at-risk or expected shortfall based on observed historic data. Usually, (log-)returns are studied and so-called stylized facts of financial time series are particularly helpful to model and predict risk measures, see, for example, Engle and Patton (2001) and Tsay (2010). For instance, many studies have reported that financial return series often only show little autocorrelation. In contrast, series of squared or absolute returns are often strongly autocorrelated and volatility is time-varying and appears in clusters. Therefore, volatility modeling became an integral part of quantitative risk management.

As a result of the ongoing technological progress, the amount and length of financial time series data available is constantly increasing. Prices and returns for assets can today be

recorded in very high frequencies, often even in a tick-by-tick frequency. A consequence of this data availability is the need to decide for a sampling frequency when applying methods for financial time series to implement quantitative risk management. A general analysis and discussion of the trade-off between sampling frequency and forecast horizon is given in Andersen and Bollerslev (1998) and Andersen et al. (1999).

Several frequencies and horizons have to be chosen, when modeling financial time series. The choice, however, is not always free and without restrictions. In risk management, for example, the frequencies and horizons are often fixed due to the specific use case. Think of an asset manager, or robo advisor, who is relying on a risk-based asset allocation approach. The strategy of the asset manager could for example be driven by a goal in terms of a value-at-risk (VaR) with a horizon of one week, one month or even one year. To achieve this goal, the asset manager has to predict risks over this horizon. However, often this risk assessment is not done on the same frequency as the prediction horizon, i.e., the asset manager is assessing it more frequently and for example estimates the risk of his asset allocation on a daily basis. This results in several frequencies and parameters which have to be chosen by the asset manager: The sampling frequency of the returns used for modeling risks, the forecast horizon of interest and the frequency for the risk assessments.

Regulators of banks and insurance companies face a similar situation. For instance, the assessment frequencies for market risks are rather high. To guarantee an ongoing risk assessment, daily reporting is often obligatory. At the same time the target horizon for the risk estimation can be much longer, like for example 10 days in Basel III or 259 days in Solvency II.

In this regard, several strands of research in financial econometrics are of particular interest. Three of them will be discussed in the following parts of the introduction. In particular, generalized autoregressive conditional heteroskedasticity (GARCH) processes, temporal aggregation of stochastic processes and consequences of the usage of overlapping (return) data for statistical inference.

2.1 Weak white noise processes: Gaussian white noise and GARCH processes

A key factor for sound risk management is the appropriate modeling of volatility. In the following, we will provide a short summary of weak white noise processes and the most popular volatility model, the GARCH model, introduced by Engle (1982) and Bollerslev (1986).

Weak white noise processes are determined by three properties: A constant mean, a constant and finite variance, and zero autocorrelation.

Definition 4 (Weak white noise). *A stochastic process, $(x_t)_{t\in\mathbb{Z}}$, is called weak white noise process, if $\forall t, t_1, t_2 \in \mathbb{Z}$, $t_1 \neq t_2$:*

(i) $\mathbb{E}(x_t) = \mu$, *with* $|\mu| < \infty$,

(ii) $\mathrm{Var}(x_t) = \sigma^2$, *with* $0 < \sigma^2 < \infty$,

(iii) $\mathrm{Cov}(x_{t_1}, x_{t_2}) = 0$.

Two special cases are of particular interest. The Gaussian white noise, which is the special case of a weak white noise process with independent and identically normally distributed random variables.

Example 1 (Gaussian white noise). *A stochastic process, $(x_t)_{t\in\mathbb{Z}}$, is called Gaussian white noise process, if $(x_t)_{t\in\mathbb{Z}}$ is a weak white noise process and $x_t \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$.*

The second weak white noise process that we will further study is the GARCH process.

Example 2 (GARCH(p,q)). *Let $(\epsilon_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. random variables and let $p \in \mathbb{N}$ and $q \in \mathbb{N}_0$. Further, let $\alpha_0 > 0$, $\alpha_1, \ldots, \alpha_q \geq 0$ and $\beta_1, \ldots, \beta_p \geq 0$ and assume $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$ such that the process is weakly stationary. Then, a GARCH(p,q) process, $(x_t)_{t\in\mathbb{Z}}$, with strictly positive volatility process, $(\sigma_t)_{t\in\mathbb{Z}}$, is defined by*

$$x_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2.$$

The GARCH process defines a model class that is widely used in academic research and in practice in order to model the volatility of financial returns. The novelty of the ARCH model by Engle (1982) is that for the first time volatility could be modeled parametrically and furthermore it is based on similar ideas as linear time series models (Andersen et al. 2009). The time varying conditional variance of an ARCH process is defined as a moving average of past squared returns. In this way, the ARCH model is capable to reproduce stylized facts that can be empirically observed for volatility time series. Bollerslev (1986) introduced an extension of the ARCH model, the GARCH model. In the GARCH model, the conditional variance can not only depend on past squared returns, but also on lagged conditional variances. This has the advantage that the ACF of squared returns often decays exponentially, but slower than inferred by fitted ARCH models, which can be reproduced

by fitted GARCH models. Moreover, GARCH models are often more parsimonious in empirical applications, as it is often sufficient to consider first orders, i.e., the GARCH(1,1) model, but for the ARCH model often high q-orders are suggested by model selection criteria. In the following, GARCH processes will primarily be used as data generating processes for financial returns. Two aspects render GARCH processes particularly useful in this case: First, as described before, GARCH processes are often well suited to describe financial returns and therefore provide a realistic data generating process. Second, GARCH processes have been extensively studied after their introduction. In particular, the existence of unconditional moments of GARCH process and closed form formulas for them (see, for example, He and Teräsvirta (1999b), He and Teräsvirta (1999a), Karanasos (1999), Ling and McAleer (2002a), Ling and McAleer (2002b) and Lindner (2009)) are of great interest when studying theoretical properties of risk estimates under the assumption of GARCH models as data generating processes.

2.2 Temporal aggregation of stochastic processes

Whenever one discusses the sampling strategy for return data and assumes a specific stochastic process for a specific sampling frequency, the theory about temporal aggregation of stochastic processes can help to understand the behavior of return series on a lower frequency which are obtained by aggregation. In a survey article, Silvestrini and Veredas (2008) provide a detailed overview about temporal aggregation of univariate as well as multivariate time series models and recently, for example, Kole et al. (2017) studied the impact of temporal and portfolio aggregation on the quality of 10-day ahead VaR forecasts. Silvestrini and Veredas (2008) point out that in economic studies the choice of the frequency is most of the times driven by fundamental arguments and that it clearly affects the estimation results.

In the following, we will restrict ourselves to the case of financial returns which are so-called flow variables. For flow variables, temporal aggregation to switch to a lower frequency is achieved by summing up the higher frequent observations. We denote the process of daily prices of an asset by $(P_t)_{t \in \mathbb{Z}}$, such that, with $r_t = \ln(P_t) - \ln(P_{t-1})$, the process of daily log-returns is given by $(r_t)_{t \in \mathbb{Z}}$. For $h > 1$, we can compute h-day returns by

$$r_{(h),t} = \ln(P_t) - \ln(P_{t-h}) = \sum_{j=0}^{h-1} r_{t-j}.$$

In the simplest case, we can assume a Gaussian white noise process (Example 1) for the

daily return process, $(r_t)_{t\in\mathbb{Z}}$. After aggregation, the series $(r_{(h),th})_{t\in\mathbb{Z}}$, where the aggregation horizon for the returns is the same as the sampling frequency, is again a Gaussian white noise process, but with a larger variance $\text{Var}(r_{(h),th}) = h\text{Var}(r_t)$. If the daily return process, $(r_t)_{t\in\mathbb{Z}}$, is not a Gaussian white noise process but a GARCH process, properties of the aggregated h-day return series, $(r_{(h),th})_{t\in\mathbb{Z}}$, are proven in Drost and Nijman (1993). For instance, Drost and Nijman (1993) show that weak GARCH processes are closed under temporal aggregation, i.e., the aggregated series is again a weak GARCH process.[2]

2.3 Overlapping samples and statistical implications

The usage of overlapping returns is often proposed, to obtain more observations on which risk assessments can be based. In this regard, the usage of overlapping returns is discussed for example by regulators of banks (Basel III) and insurance companies (Solvency II). Obviously, the usage of overlapping aggregation windows introduces a very special intertemporal dependence and can affect estimation results. In the following, we will discuss possible consequence of the usage of overlapping return data on risk estimation or, more general, statistical inference.

What most of the studies of temporal aggregation have in common is the fact that the aggregation horizon and the sampling frequency are synchronized. That means, that daily returns that have been aggregated to h-day returns are also sampled only every h-th day. The resulting time series of h-day returns, $(r_{(h),th})_{t\in\mathbb{Z}}$, is then a non-overlapping series in a sense that the h-day returns do not share any daily return observation. Alternatively, one can also consider overlapping time series of h-day returns. It is clear that the more observations are aggregated, the less aggregated non-overlapping observations one gets, but by allowing overlapping aggregation windows, the number of observations remains almost the same.

However, time series of overlapping returns are obviously dependent as neighboring h-day returns share daily returns and can therefore be strongly dependent. The effect of temporal aggregation on the serial dependence of overlapping samples can be nicely illustrated with the example of a Gaussian white noise process for the daily return series, $(r_t)_{t\in\mathbb{Z}}$, that we studied before. If we consider the daily series of overlapping h-day returns, $(r_{(h),t})_{t\in\mathbb{Z}}$, it is easy to show that the series follows a non-invertible $\text{MA}(h-1)$ process with parameters $\theta_j = 1$ for $1 \le j \le h-1$, i.e., $r_{(h),t} = \sum_{j=0}^{h-1} r_{t-j} = \sum_{j=1}^{h-1} \theta_j r_{t-j} + r_t$, where

---

[2] In their analysis of temporal aggregation of GARCH processes, Drost and Nijman (1993) distinguish between strong, semi-strong and weak GARCH processes and show that only weak GARCH processes are closed under temporal aggregation. For the exact definitions of the three different types of GARCH processes we refer to Drost and Nijman (1993).

$(r_t)_{t\in\mathbb{Z}}$ is the Gaussian white noise series of daily log-returns. The autocorrelation function (ACF) $\rho_{r_{(h),t}}(\ell)$ for the daily process of h-day returns, $(r_{(h),t})_{t\in\mathbb{Z}}$, is given by (cf. Mittnik (1988))

$$\rho_{r_{(h),t}}(\ell) = \text{Corr}(r_{(h),t}, r_{(h),t-\ell}) = \begin{cases} \frac{\sum_{j=0}^{h-1-|\ell|} \theta_j \theta_{j+|\ell|}}{\sum_{j=0}^{h-1} \theta_j^2} = \frac{h-|\ell|}{h} & , |\ell| \leq h-1, \\ 0 & , |\ell| \geq h, \end{cases}$$

where we set $\theta_0 = 1$ for notational simplicity. So, due to the temporal aggregation, there is obviously a very strong autocorrelation of $\rho_{r_{(h),t}}(\ell) = \frac{h-|\ell|}{h}$ for the first $h-1$ lags. Consequences of using overlapping data for statistical inference or, more specific, overlapping returns for risk estimation are analyzed in various studies.

One of the first authors explicitly suggesting to use overlapping data are Hansen and Hodrick (1980). They estimate regression models with overlapping data using OLS and suggest a standard error adjustment to correct for the autocorrelation introduced by the overlapping data. An estimator for the variance of overlapping returns has been proposed in Lo and MacKinlay (1988) and bias-corrected versions thereof for different data generating processes of the disaggregated returns have been analyzed in Bod et al. (2002) and Kluitman and Franses (2002). Mittnik (2011) illustrates and explains the introduced spurious dependence patterns that emerge when using overlapping yearly returns sampled on a daily frequency as it has been suggested in the Solvency II regulations. Hedegaard and Hodrick (2016) study the risk-return trade-off using overlapping monthly returns sampled on a daily frequency. In this regard, Hedegaard and Hodrick (2016) derive the asymptotic distribution of parameter estimates for GARCH models with overlapping data using the generalized method of moments (GMM, Hansen (1982)). More recently, Daníelsson et al. (2016) and Daníelsson and Zhou (2016) provide an empirical analysis of 10-day ahead value-at-risk and expected shortfall predictions in the context of Basel III. As part of their analysis Daníelsson et al. (2016) and Daníelsson and Zhou (2016) also studied the usage of overlapping return data.

Most referenced studies for the consequences of overlapping data on risk estimation have two things in common: First, the overlap in the data is explicit, i.e., neighboring observations of the series of h-day returns used for risk assessment have common higher-frequent return observations and are therefore strongly correlated. Second, the shown consequences and effects of the usage of overlapping data are most of the time presented statically, i.e., the properties of risk estimates for a specific point in time t are analyzed. The fourth contribution of this cumulative dissertation differs in two ways. The studied overlap is not explicit, but implicit, by assessing risks at a higher frequency from return
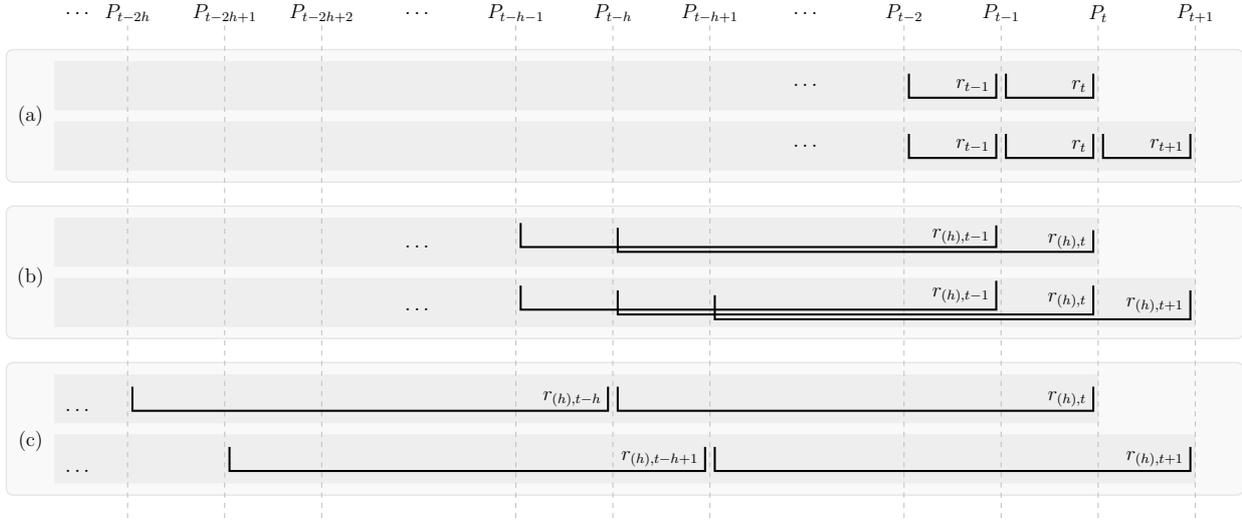
Figure 4: Illustration of different combinations of return intervals and sampling schemes for deriving h-day-ahead risk measures. Each panel consists of two rows: The first row sketches the data used for estimation at time t and the second row those at $t+1$. Panel (a) shows a scheme with daily sampling of daily returns. Here, risk estimates have to be scaled up to derive h-day-ahead risk estimates. Panel (b) illustrates the sampling scheme when using overlapping h-day returns. Panel (c) indicates the scheme when using non-overlapping h-day returns.

Figure source: Kurz and Mittnik (2018) / Chapter V.

data that have been aggregated to longer return interval. Moreover, we particularly analyze time series properties of risk estimates, which are of great importance for risk managers in financial institutions.

2.4 Risk assessment and spurious seasonality

Due to fundamental economic reasons and arguments, regulators and managers often specify the return interval, forecasting horizons and the assessment frequency in practical risk management applications. Despite that, the sampling scheme of the data is typically not fixed. Figure 4 summarizes three commonly applied strategies for assessing risk measures with a h-day horizon. To provide a better understanding of the time series properties, we show a panel, consisting of the sampling for a risk estimation at t in the first row and a second row showing the sampling for estimating in $t+1$, for each strategy, (a), (b) & (c). The first strategy, (a), consists of the usage of the most granular return data available, e.g., daily. To obtain h-day risk estimates, one then either has to (square-root) scale the one-day risk measures or needs to derive model-based multi-step forecasts. The goodness of square-root-of-time scaling for risk estimation and especially for volatilities has been extensively studied by Christoffersen et al. (1998), Daníelsson and Zigrand (2006), Diebold et al. (1997), Embrechts et al. (2005) and McNeil and Frey (2000). The second strategy, (b),

can be implemented by computing h-day returns via temporal aggregation and using a daily overlapping sample of h-day returns for risk estimation. In panel (b) of Figure 4, the explicit overlap is visible. For a discussion of the consequences of using overlapping data for risk estimation or, more general, statistical inference we refer to Section 2.3. The third strategy, (c), is also based on temporally aggregated return data, but this time every risk estimate is based on a non-overlapping sample of h-day returns. In panel (c) of Figure 4, one can clearly see that an overlap is no longer present, if a single risk estimate for a specific point in time is considered, but whenever the assessment frequency is higher, e.g., daily, an implicit overlap emerges. This implicit overlap and its consequences for the dynamic time series properties of risk estimates are the main topic of the fourth contribution (Chapter V: Kurz and Mittnik (2018)) of this cumulative dissertation. In this paper, we study the phenomenon *spurious seasonality* in sequentially estimated variances. Spurious seasonality arises when the non-overlapping return data used for estimation is sampled on, or aggregated to, a lower frequency than the assessment frequency. In Kurz and Mittnik (2018), spurious seasonality

> **Chapter V:**
> **Risk Assessment and Spurious Seasonality**
>
> M. S. Kurz and S. Mittnik. 2018. "Risk Assessment and Spurious Seasonality". Available at SSRN: `https://dx.doi.org/10.2139/ssrn.2990772`, *Center for Quantitative Risk Analysis (CEQURA), Working Paper Number 19.*

is introduced, explained and analyzed empirically using daily return data of the Dow Jones Industrial Average as well as theoretically based on weak white noise processes, like for example GARCH, as data generating processes for daily return data.

Kurz and Mittnik (2018) show that variance estimators, such as sample variances or the exponentially-weighted moving average (EWMA) variance estimator, are prone to exhibit spurious seasonality. Theoretically, spurious seasonality can be explained by deriving the autocorrelation function (ACF) of series of variance estimates. In Kurz and Mittnik (2018) this is done for weak white noise processes, satisfying specific moment conditions, as data generating process for the daily return series. To derive the ACF of sequences of variance estimates, Kurz and Mittnik (2018) show how many commonly used variance estimators can be written as quadratic forms in vectors of daily returns. For Gaussian white noise, the ACF can then be obtained as a corollary to well-known results about moments of products of quadratic forms in the very same vector of multivariate normally distributed random variables (Magnus 1978). Kurz and Mittnik (2018) generalize the result of Magnus (1978) to a richer class of weak white noise processes satisfying specific moment conditions. For

example the Gaussian white noise process (Example 1) as well as GARCH(p,q) processes (Example 2) with a symmetric innovation distribution satisfy these moment conditions. As a corollary the ACF of sequences of variances can be obtained.

To overcome spurious seasonality, Kurz and Mittnik (2018) suggest to rely on estimates using overlapping return data. In this regard, a exponentially-weighted moving average (EWMA) version of the two-scales variance estimator of Zhang et al. (2005) is introduced. Additionally, it is demonstrated how the two-scales estimator can be boundary-corrected. It is demonstrated that the boundary-corrected EWMA two-scales variance estimator overcomes spurious seasonality. Moreover, the estimator is compared to other estimators based on (non-)overlapping h-day returns in terms of bias, variance and mean squared error (MSE).

Recently, Daníelsson and Zhou (2016) also consider the three different sampling strategies (a), (b) & (c) to come up with h-day ahead risk estimates. They particularly focus on a comparison of estimation strategies based on square-root-of-time scaling with estimation based on overlapping return data and implications thereof for the bias and variance of risk estimates. Kurz and Mittnik (2018) differs in two aspects: First, only risk estimation strategies operating with h-day returns are considered. A main reason for that are the Basel III rules for which the Basel Committee on Banking Supervision (BCBS 2016) explicitly rules out any risk assessment based on scaling. Moreover, Kurz and Mittnik (2018) particularly analyze the time series properties of risk estimates which are of great practical importance for regulators, bankers, asset managers and risk managers who normally have to assess and report risks on a daily basis with longer than daily horizon.

The results in Kurz and Mittnik (2018) are exclusively derived and empirically illustrated for variance estimates. However, risk measures such as volatility, value-at-risk or expected shortfall are closely related to variance in a direct or indirect way. Empirically, one can therefore easily verify that spurious seasonality is also present in estimates for non-linear risk measures, like volatility, value-at-risk or expected shortfall. The phenomenon spurious seasonality is caused by the sampling scheme of the return data in combination with the assessment frequency. Therefore, the findings in Kurz and Mittnik (2018) also serve as an explanation for the rather large variation in GARCH parameter estimates observed by Hedegaard and Hodrick (2016). In their study of the risk-return trade-off, Hedegaard and Hodrick (2016) estimate GARCH models for 22 different shifted samples each consisting of non-overlapping monthly (22-day) returns and report substantial differences in the parameter estimates, but parameter equality could not be rejected with a formal test. In comparison to Kurz and Mittnik (2018), Hedegaard and Hodrick (2016) did not analyze longer time series of their estimates, but basically observed exactly one cycle of spurious

seasonality. Moreover, spurious seasonality is also present for other than daily frequencies like (ultra-)high-frequency data where realized-volatilities are often studied. The key factor for the presence of spurious seasonality is the setting consisting of the assessment frequency, the sampling frequency, the target horizon and whether overlapping returns are used or not.

3   State-space methods: The Kalman filter and smoother

Originating from engineering, state space methods are widely used in econometrics, especially in the context of time series analysis. References for the basics of state space models, the Kalman filter and its applications in econometrics are Durbin and Koopman (2012), Hamilton (1994) and Harvey (1990). The classical state space models are linear and assume a Gaussian distribution for the innovations. When it comes to prediction, smoothing and likelihood estimation, the Kalman filter (Kalman 1960) turns out to be particularly useful.

Consider the state space model[3]

$$X_t = A X_{t-1} + C u_t, \tag{3.1}$$

$$Z_t = D_1 X_t + R u_t, \tag{3.2}$$

where the measurement equation (3.2) relates the observable $p \times 1$ vector Z_t with the $n \times 1$ state vector X_t and the state equation (3.1) is of an autoregressive structure. For the $m \times 1$ vector of disturbances we assume $u_t \sim N(0, I_m)$ and the state space model is initialized via $X_0 \sim N(X_{0|0}, P_{0|0})$. We further use the notations $X_{t|t-s} = \mathbb{E}(X_t | Z_{1:t-s}, X_{0|0})$ for the conditional expectation, $P_{t|t} = \mathbb{E}((X_t - X_{t|t})(X_t - X_{t|t})')$ for the conditional variance and the short form $Z_{1:t} = (Z'_1, \ldots, Z'_t)'$ for vectors.

For applications in financial econometrics, non-Gaussian and nonlinear state space models have been developed, see, for example, Part II of Durbin and Koopman (2012) where methods like the extended and unscented Kalman filter and particle filter are discussed. However, for the sake of simplicity, we will restrict ourselves to the linear Gaussian models, which are also used in the fifth contribution of this cumulative dissertation (Chapter VI: Kurz (2018a)). In Kurz (2018a), a special focus will be on linear state space systems with a lagged state in the measurement equation, which we will introduce in the second part of this section.

---

[3] The fifth contribution of this cumulative dissertation (Chapter VI: Kurz (2018a)) is closely related to Nimark (2015). Therefore, in following the notation is chosen similar to Nimark (2015) and Kurz (2018a).

3.1 The Kalman filter for linear state space models

We now give a short introduction to the Kalman filter and smoother for linear state space models. The one-step ahead forecast error of Z_t given $Z_{1:t-1}$ is denoted by $\tilde{Z}_t$ and can be computed as

$$\tilde{Z}_t = Z_t - Z_{t|t-1} = Z_t - D_1 X_{t|t-1} = D_1(X_t - X_{t|t-1}) + Ru_t.$$

By the equation for updating linear projections (Hamilton 1994, Eq. [4.5.30]), the filtered state, $X_{t|t}$, is given by

$$X_{t|t} = X_{t|t-1} + \mathbb{E}(X_t \tilde{Z}_t)\mathbb{E}(\tilde{Z}_t \tilde{Z}_t')^{-1}\tilde{Z}_t.$$

As a result we obtain the forward Kalman filter updating equations

$$\tilde{Z}_t = Z_t - D_1 X_{t|t-1}, \qquad\qquad X_{t|t} = AX_{t-1|t-1} + K_t \tilde{Z}_t,$$

with the Kalman gain $K_t = \mathbb{E}(X_t \tilde{Z}_t)\mathbb{E}(\tilde{Z}_t \tilde{Z}_t')^{-1}$.

The smoothed state vector, $X_{t|T}$, is the estimate for the state at time $t \in \{1, \ldots, T\}$ if we condition on all observables, $Z_{1:T}$, i.e., past as well as the current and future observations. For the classical linear state space model, updating equations for the Kalman smoother that are purely based on filtered states and not on the observables are well known (Anderson and Moore (1979), Hamilton (1994) and Rauch et al. (1965)). The updating equations for the Kalman smoother are then given by (Hamilton 1994, Eq. [13.6.16])

$$X_{t|T} = X_{t|t} + J_t(X_{t+1|T} - X_{t+1|t}), \qquad\qquad J_t = P_{t|t}A'P_{t+1|t}^{-1}.$$

3.2 A modified low-dimensional Kalman filter for systems with lagged states in the measurement equation

Nimark (2015) considers state space systems with a lagged state in the measurement equation, given by

$$X_t = AX_{t-1} + Cu_t, \tag{3.3}$$

$$Z_t = D_1 X_t + D_2 X_{t-1} + Ru_t. \tag{3.4}$$

As pointed out by Nimark (2015), the standard approach in the literature to handle such a state space system is given by augmenting the state vector with lagged states. If one

considers the augmented state vector, $\bar{X}_t = (X_t', X_{t-1}')'$, one can rewrite the state space system with a lagged state in the measurement equation (3.3)-(3.4) as a standard linear state space system (3.1)-(3.2) and can apply the usual Kalman filter and Kalman smoother methods. A similar state space system was also studied by Qian (2014), who considers a more general case where a lagged observation and state can be present in the measurement as well as the state equation.

As an alternative to the augmentation, Nimark (2015) suggests a modified Kalman filter, which operates with a n-dimensional state vector and therefore is more parsimonious than the augmented system with a $2n$-dimensional state vector. To obtain the same filtered states as with the Kalman filter for the augmented system, Nimark (2015) derived modified filter equations. The modified filter can be summarized as

$$\tilde{Z}_t = Z_t - \tilde{D}X_{t-1|t-1}, \qquad\qquad P_{t|t-1} = AP_{t-1|t-1}A' + CC',$$
$$X_{t|t} = AX_{t-1|t-1} + K_t\tilde{Z}_t, \qquad\qquad P_{t|t} = P_{t|t-1} - K_t F_t K_t',$$

with $\tilde{D} = (D_1 A + D_2)$ and where the Kalman gain is given by $K_t = U_t F_t^{-1}$ with

$$U_t = \mathbb{E}(X_t \tilde{Z}_t') = AP_{t-1|t-1}\tilde{D}' + CC'D_1' + CR',$$
$$F_t = \mathbb{E}(\tilde{Z}_t \tilde{Z}_t') = \tilde{D}P_{t-1|t-1}\tilde{D}' + (D_1 C + R)(D_1 C + R)'.$$

3.3 On low-dimensional Kalman smoothers for systems with lagged states in the measurement equation

In Nimark (2015), the focus lies in the derivation of a modified low-dimensional Kalman filter for state space systems with a lagged state in the measurement equation. The fifth contribution of this cumulative dissertation (Chapter VI: Kurz (2018a)) is a study of Kalman smoothers for such state space systems. The main contribution of Kurz (2018a) is the

> Chapter VI:
> A note on low-dimensional Kalman smoothers for systems with lagged states in the measurement equation
>
> M. S. Kurz. 2018a. "A note on low-dimensional Kalman smoothers for systems with lagged states in the measurement equation". `https://doi.org/10.1016/j.econlet.2018.03.037`, *Economics Letters* 168: 42–45.

derivation of a MSE-minimizing low-dimensional Kalman smoother for state space systems

with a lagged state in the measurement equation in the lines of Anderson and Moore (1979), Hamilton (1994) and Rauch et al. (1965). Moreover, two computationally more efficient smoothing algorithms are presented, namely, a modified version of the de Jong (1988, 1989) and Kohn and Ansley (1989) smoother and a modified version of the disturbance-smoother-based Koopman (1993) smoother. The modified Kalman filter and all smoothing algorithms are implemented in an accompanying `MATLAB`-code package (Kurz 2018b). In Kurz (2018a), it is additionally proven that the conjecture in Nimark (2015) that one obtains a modified Kalman smoother by plugging the filtered states from the modified filter into the standard Kalman smoother formulas (Hamilton 1994) is in general not correct, i.e., it does not yield a MSE-minimizing smoothed state estimate. Finally, the smoother of Nimark (2015) is compared to the MSE-minimizing smoother using an ARMA(1,1) process with additive measurement error as data generating process. Similar processes have been used by Jacobs and van Norden (2011) for modeling data revisions which are a typical phenomenon of economic time series. The results indicate that the relative difference in the MSE between the Kalman smoother suggested in Nimark (2015) and the MSE-minimizing smoother derived in Kurz (2018a) can be rather large.

References for Chapter I

Aas, K., C. Czado, A. Frigessi, and H. Bakken. 2009. "Pair-copula constructions of multiple dependence". *Insurance: Mathematics and Economics* 44 (2): 182–198.

Aas, K. 2016. "Pair-Copula Constructions for Financial Applications: A Review". *Econometrics* 4 (4).

Acar, E. F., C. Genest, and J. Nešlehová. 2012. "Beyond simplified pair-copula constructions". *Journal of Multivariate Analysis* 110: 74–90.

Acar, E. F., R. V. Craiu, and F. Yao. 2013. "Statistical testing of covariate effects in conditional copula models". *Electronic Journal of Statistics* 7: 2822–2850.

Andersen, T. G., and T. Bollerslev. 1998. "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts". *International Economic Review* 39 (4): 885–905.

Andersen, T. G., T. Bollerslev, and S. Lange. 1999. "Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon". *Journal of Empirical Finance* 6 (5): 457–477.

Andersen, T. G., R. A. Davis, J.-P. Kreiß, and T. Mikosch, eds. 2009. *Handbook of Financial Time Series*. Berlin [u.a.]: Springer.

Anderson, B. D. O., and J. B. Moore. 1979. *Optimal Filtering*. Englewood Cliffs, N.J.: Prentice-Hall.

BCBS. 2016. "Minimum capital requirements for market risk". `http://www.bis.org/bcbs/publ/d352.htm`.

Balakrishnan, N. 2009. *Continuous Bivariate Distributions*. Ed. by C.-D. Lai. New York, NY: Springer.

Bedford, T., and R. M. Cooke. 2001. "Probability density decomposition for conditionally dependent random variables modeled by vines". *Annals of Mathematics and Artificial Intelligence* 32 (1): 245–268.

— . 2002. "Vines – A new graphical model for dependent random variables". *The Annals of Statistics* 30 (4): 1031–1068.

Bergsma, W. P. 2004. "Testing conditional independence for continuous random variables". URL: `https://www.eurandom.tue.nl/reports/2004/048-report.pdf`.

Bod, P., D. Blitz, P. H. Franses, and R. Kluitman. 2002. "An unbiased variance estimator for overlapping returns." *Applied Financial Economics* 12 (3): 155–158.

Bollerslev, T. 1986. "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics* 31 (3): 307–327.

Cherubini, U., E. Luciano, and W. Vecchiato. 2005. *Copula Methods in Finance*. Wiley Finance Series. Chichester, United Kingdom: John Wiley & Sons.

Cherubini, U., F. Gobbi, S. Mulinacci, and S. Romagnoli. 2012. *Dynamic Copula Methods in Finance*. Wiley Finance Series. Chichester, United Kingdom: John Wiley & Sons.

Christoffersen, P., F. X. Diebold, and T. Schuermann. 1998. "Horizon problems and extreme events in financial risk management". *Economic Policy Review* 4 (3): 98–16.

Daníelsson, J., and C. Zhou. 2016. "Why risk is so hard to measure". DNB Working Paper, No. 494 / January 2016, URL: `https://www.dnb.nl/en/news/dnb-publications/ dnb - working - papers - series / dnb - working - papers / working - papers - 2016 / dnb335907.jsp`.

Daníelsson, J., and J.-P. Zigrand. 2006. "On time-scaling of risk and the square-root-of-time rule". *Journal of Banking & Finance* 30 (10): 2701–2713.

Daníelsson, J., K. R. James, M. Valenzuela, and I. Zer. 2016. "Model risk of risk models". *Journal of Financial Stability* 23: 79–91.

Derumigny, A., and J.-D. Fermanian. 2017. "About tests of the "simplifying" assumption for conditional copulas". *Dependence Modeling* 5 (1): 154–197.

Diebold, F. X., A. Hickman, A. Inoue, and T. Schuermann. 1997. "Converting 1-day Volatility to h-day Volatility: Scaling by h is Worse than You Think". *Wharton Financial Institutions Center, Working Paper 97-34*.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka. 2013. "Selecting and estimating regular vine copulae and application to financial returns". *Computational Statistics & Data Analysis* 59: 52–69.

Drost, F. C., and T. E. Nijman. 1993. "Temporal Aggregation of GARCH Processes". *Econometrica* 61 (4): 909–927.

Durbin, J., and S. J. Koopman. 2012. *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series. OUP Oxford.

Embrechts, P., R. Kaufmann, and P. Patie. 2005. "Strategic Long-Term Financial Risks: Single Risk Factors". *Computational Optimization and Applications* 32 (1): 61–90.

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 A. J. Patton. 2001. "What good is a volatility model?" *Quantitative Finance* 1 (2): 237–245.

Genest, C., and J. MacKay. 1986. "The Joy of Copulas: Bivariate Distributions with Uniform Marginals". *The American Statistician* 40 (4): 280–283.

Genest, C., and J. G. Nešlehová. 2014. "Modeling dependence beyond correlation". In *Statistics in Action: A Canadian Outlook*, ed. by J. F. Lawless, 59–78. CRC Press.

Gijbels, I., N. Veraverbeke, and M. Omelka. 2011. "Conditional copulas, association measures and their applications". *Computational Statistics & Data Analysis* 55 (5): 1919–1932.

Gijbels, I., M. Omelka, and N. Veraverbeke. 2015a. "Estimation of a Copula when a Covariate Affects only Marginal Distributions". *Scandinavian Journal of Statistics* 42 (4): 1109–1126.

— . 2017a. "Nonparametric testing for no covariate effects in conditional copulas". *Statistics* 51 (3): 475–509.

— . 2015b. "Partial and average copulas and association measures". *Electronic Journal of Statistics* 9 (2): 2420–2474.

Gijbels, I., M. Omelka, M. Pešta, and N. Veraverbeke. 2017b. "Score tests for covariate effects in conditional copulas". *Journal of Multivariate Analysis* 159: 111–133.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press.

Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators". *Econometrica* 50 (4): 1029–1054.

Hansen, L. P., and R. J. Hodrick. 1980. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis". *Journal of Political Economy* 88 (5): 829–853.

Harvey, A. C. 1990. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

He, C., and T. Teräsvirta. 1999a. "Fourth Moment Structure of the GARCH(p,q) Process". *Econometric Theory* 15 (6): 824–846.

— . 1999b. "Properties of moments of a family of GARCH processes". *Journal of Econometrics* 92 (1): 173–192.

Hedegaard, E., and R. J. Hodrick. 2016. "Estimating the risk-return trade-off with overlapping data inference". *Journal of Banking & Finance* 67: 135–145.

Hobæk Haff, I. 2012. "Comparison of estimators for pair-copula constructions". *Journal of Multivariate Analysis* 110: 91–105.

— . 2013. "Parameter estimation for pair-copula constructions". *Bernoulli* 19 (2): 462–491.

Hobæk Haff, I., K. Aas, and A. Frigessi. 2010. "On the simplified pair-copula construction – Simply useful or too simplistic?" *Journal of Multivariate Analysis* 101 (5): 1296–1310.

Hoeffding, W. 1940. *Maßtabinvariante Korrelationstheorie*. Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik an der Universität Berlin, 5.

Jacobs, J. P. A. M., and S. van Norden. 2011. "Modeling data revisions: Measurement error and dynamics of "true" values". *Journal of Econometrics* 161 (2): 101–109.

Joe, H. 2014. *Dependence Modeling with Copulas*. New York: Chapman & Hall/CRC.

— . 1996. "Families of m-Variate Distributions with Given Margins and m(m-1)/2 Bivariate Dependence Parameters". *Lecture Notes-Monograph Series* 28: 120–141.

— . 1997. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Jondeau, E., and M. Rockinger. 2006. "The Copula-GARCH model of conditional dependencies: An international stock market application". *Journal of International Money and Finance* 25 (5): 827–853.

de Jong, P. 1988. "A Cross-Validation Filter for Time Series Models". *Biometrika* 75 (3): 594–600.

— . 1989. "Smoothing and Interpolation with the State-Space Model". *Journal of the American Statistical Association* 84 (408): 1085–1088.

Kalman, R. E. 1960. "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering* 82 (1): 35–45.

Karanasos, M. 1999. "The second moment and the autocovariance function of the squared errors of the GARCH model". *Journal of Econometrics* 90 (1): 63–76.

Kendall, M. G. 1938. "A New Measure of Rank Correlation". *Biometrika* 30 (1/2): 81–93.

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

— . 2016. "Using model distances to investigate the simplifying assumption, goodness-of-fit and truncation levels for vine copulas". ArXiv e-prints `arXiv:1610.08795`.

Kluitman, R., and P. H. Franses. 2002. "Estimating volatility on overlapping returns when returns are autocorrelated." *Applied Mathematical Finance* 9 (3): 179–188.

Kohn, R., and C. F. Ansley. 1989. "A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State Space Models". *Biometrika* 76 (1): 65–79.

Kole, E., T. Markwat, A. Opschoor, and D. van Dijk. 2017. "Forecasting Value-at-Risk under Temporal and Portfolio Aggregation". *Journal of Financial Econometrics* 15 (4): 649–677.

Koopman, S. J. 1993. "Disturbance Smoother for State Space Models". *Biometrika* 80 (1): 117–126.

Kraus, D., and C. Czado. 2017. "Growing simplified vine copula trees: improving Dißmann's algorithm". ArXiv e-prints `arXiv:1703.05203`.

Kurowicka, D., and R. Cooke. 2003. "A parameterization of positive definite matrices in terms of partial correlation vines". *Linear Algebra and its Applications* 372: 225–251.

— . 2006a. "Completion problem with partial correlation vines". *Linear Algebra and its Applications* 418 (1): 188–200.

— . 2006b. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Chichester: Wiley.

Kurowicka, D., and H. Joe, eds. 2011. *Dependence Modeling*. Singapore: World Scientific.

Kurz, M. S. 2018a. "A note on low-dimensional Kalman smoothers for systems with lagged states in the measurement equation". `https://doi.org/10.1016/j.econlet.2018.03.037`, *Economics Letters* 168: 42–45.

— . 2018b. *State Space Models with Lagged State (SSMwLS) in the measurement equation*. URL: `https://github.com/MalteKurz/SSMwLS`, `MATLAB`-code package.

— . 2018c. *pacotest: Testing for Partial Copulas and the Simplifying Assumption in Vine Copulas*. URL: `https://CRAN.R-project.org/package=pacotest`, `R`-Package version 0.3.

Kurz, M. S., and S. Mittnik. 2018. "Risk Assessment and Spurious Seasonality". Available at SSRN: `https://dx.doi.org/10.2139/ssrn.2990772`, *Center for Quantitative Risk Analysis (CEQURA), Working Paper Number 19*.

Kurz, M. S., and F. Spanhel. 2018. "Testing the simplifying assumption in high-dimensional vine copulas". ArXiv e-prints `arXiv:1706.02338`.

Lindner, A. M. 2009. "Stationarity, Mixing, Distributional Properties and Moments of GARCH(p, q)–Processes". In *Handbook of Financial Time Series*, ed. by T. Mikosch, J.-P. Kreiß, A. R. Davis, and G. T. Andersen, 43–69. Berlin, Heidelberg: Springer Berlin Heidelberg.

Ling, S., and M. McAleer. 2002a. "Necessary and Sufficient Moment Conditions for the GARCH(r,s) and Asymmetric Power GARCH(r,s) Models". *Econometric Theory* 18 (3): 722–729.

— . 2002b. "Stationarity and the existence of moments of a family of GARCH processes". *Journal of Econometrics* 106 (1): 109–117.

Lo, A. W., and A. C. MacKinlay. 1988. "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test". *Review of Financial Studies* 1 (1): 41–66.

Magnus, J. R. 1978. "The moments of products of quadratic forms in normal variables". *Statistica Neerlandica* 32 (4): 201–210.

McNeil, A. J., and R. Frey. 2000. "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach". Special issue on Risk Management, *Journal of Empirical Finance* 7 (3–4): 271–300.

McNeil, A. J., R. Frey, and P. Embrechts. 2005. *Quantitative Risk Management*. Princeton series in finance. Princeton NJ: Princeton Univ. Press.

Mesfioui, M., and J.-F. Quessy. 2008. "Dependence structure of conditional Archimedean copulas". *Journal of Multivariate Analysis* 99 (3): 372–385.

Mittnik, S. 1988. "Derivation of the theoretical autocovariance and autocorrelation function of autoregessive moving average processes". *Communications in Statistics - Theory and Methods* 17 (11): 3825–3831.

— . 2011. "Solvency II Calibrations: Where Curiosity Meets Spuriosity". Center for Quantitative Risk Analysis (CEQURA), Working Paper Number 04, URL: `http://www.cequra.uni-muenchen.de/download/cequra_wp_041.pdf`.

Nagler, T., and C. Czado. 2016. "Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas". *Journal of Multivariate Analysis* 151: 69–89.

Nelsen, R. B. 2006. *An Introduction to Copulas*. Springer series in statistics. New York, NY: Springer.

Nimark, K. P. 2015. "A low dimensional Kalman filter for systems with lagged states in the measurement equation". *Economics Letters* 127: 10–13.

Patton, A. J. 2006. "Modelling asymmetric exchange rate dependence". *International Economic Review* 47 (2): 527–556.

Patton, A. J. 2012. "A review of copula models for economic time series". *Journal of Multivariate Analysis* 110: 4–18.

Portier, F., and J. Segers. 2018. "On the weak convergence of the empirical conditional copula under a simplifying assumption". *Journal of Multivariate Analysis* 166: 160–181.

Qian, H. 2014. "A Flexible State Space Model and its Applications". *Journal of Time Series Analysis* 35 (2): 79–88.

Rauch, H. E., F Tung, and C. T. Striebel. 1965. "Maximum likelihood estimates of linear dynamic systems". *AIAA Journal* 3 (8): 1445–1450.

Remillard, B. 2013. *Statistical Methods for Financial Engineering*. Boca Raton, FL: CRC Press.

Schellhase, C., and F. Spanhel. 2018. "Estimating non-simplified vine copulas using penalized splines". *Statistics and Computing* 28: 387–409.

Schweizer, B., and E. Wolff. 1981. "On Nonparametric Measures of Dependence for Random Variables". *The Annals of Statistics* 9 (4): 879–885.

Sibuya, M. 1960. "Bivariate extreme statistics, I". *Annals of the Institute of Statistical Mathematics* 11 (3): 195–210.

Silvestrini, A., and D. Veredas. 2008. "Temporal aggregation of univariate and multivariate time series models: A survey". *Journal of Economic Surveys* 22 (3): 458–497.

Sklar, A. 1959. "Fonctions de répartition à n dimensions et leurs marges". *Publications de l'Institut de Statistique de L'Université de Paris* 8: 229–231.

Spanhel, F., and M. S. Kurz. 2016. "The partial copula: Properties and associated dependence measures". `https://doi.org/10.1016/j.spl.2016.07.014`, *Statistics & Probability Letters* 119: 76–83.

— . 2017. "The partial vine copula: A dependence measure and approximation based on the simplifying assumption". ArXiv e-prints `arXiv:1510.06971`.

Spearman, C. 1904. "The Proof and Measurement of Association between Two Things". *The American Journal of Psychology* 15 (1): 72–101.

Stöber, J., H. Joe, and C. Czado. 2013. "Simplified pair copula constructions – Limitations and extensions". *Journal of Multivariate Analysis* 119: 101–118.

Tsay, R. S. 2010. *Analysis of Financial Time Series*. Wiley Series in probability and statistics. Hoboken, NJ: Wiley.

Tsukahara, H. 2005. "Semiparametric estimation in copula models". *Canadian Journal of Statistics* 33 (3): 357–375.

Vatter, T., and V. Chavez-Demoulin. 2015. "Generalized additive models for conditional dependence structures". *Journal of Multivariate Analysis* 141: 147–167.

Vatter, T., and T. Nagler. 2018. "Generalized Additive Models for Pair-Copula Constructions". *Journal of Computational and Graphical Statistics (forthcoming)*.

Veraverbeke, N., M. Omelka, and I. Gijbels. 2011. "Estimation of a Conditional Copula and Association Measures". *Scandinavian Journal of Statistics* 38 (4): 766–780.

Yule, G. U., and M. G. Kendall. 1953. *An Introduction to the Theory of Statistics, 14th ed.* London: Charles Griffin & Company Limited.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia. 2005. "A Tale of Two Time Scales". *Journal of the American Statistical Association* 100 (472): 1394–1411.

Chapter II

The partial copula: Properties and associated dependence measures

This chapter is a reprint[*] **of:**

F. Spanhel and M. S. Kurz. 2016. "The partial copula: Properties and associated dependence measures". `https://doi.org/10.1016/j.spl.2016.07.014`, *Statistics & Probability Letters* 119: 76–83.

Copyright:

Author contributions:

The article was initiated by Fabian Spanhel and further developed jointly by Fabian Spanhel and Malte Kurz. Both authors contributed to the presented examples and properties of the partial copula. The writing was mainly done by Fabian Spanhel with contributions by Malte Kurz and the proofs have been worked out in close collaboration. The `MATLAB` graphics have been prepared by Malte Kurz.

---

[*]All reprints in this cumulative dissertation have been slightly adapted in order to harmonize the overall appearance, like the format, page margins, caption setup, citation style and bibliography style.

The partial copula: Properties and associated dependence measures

Fabian Spanhel [*†1] and Malte S. Kurz [‡1]

[1]Department of Statistics,
Ludwig-Maximilians-Universität München,
Akademiestr. 1, 80799 Munich, Germany.

Abstract

The partial correlation coefficient is a commonly used measure to assess the conditional dependence between two random variables. We provide a thorough explanation of the partial copula, which is a natural generalization of the partial correlation coefficient, and investigate several of its properties. In addition, properties of some associated partial dependence measures are examined.

Keywords: Partial copula, Conditional copula, Partial correlation, Partial Spearman's rho, Partial Kendall's tau.

---

[*]Corresponding author
[†]spanhel@stat.uni-muenchen.de
[‡]malte.kurz@stat.uni-muenchen.de

1 Introduction

Studying the dependence between two random variables Y_1 and Y_2 conditional on a random vector Z is an important topic in statistics. The partial correlation coefficient is often used to measure the conditional dependence between two random variables due to its simple computation and meaningful interpretation if the joint distribution of $(Y_{1:2}, Z)$ is given by an elliptical distribution. However, outside the elliptical world, the interpretation of the partial correlation coefficient as a measure of conditional dependence is less obvious and can be quite misleading. For instance, it can be zero if there is conditional dependence between two random variables and, even worse, its absolute value can be arbitrarily close to one if two random variables are conditionally independent. The partial copula is a natural generalization of the partial correlation coefficient and gives a meaningful measure of conditional dependence for general distributions. Moreover, there is a large class of distributions where the partial copula completely characterizes the conditional dependence.

We first motivate and define the partial copula in Section 2 and then turn to some examples in Section 3. In Section 4 we take a closer look at the properties of the partial copula. In particular, we examine the optimality of the partial copula as an approximation of the conditional copula and investigate its relation to conditional independence. Section 5 considers dependence measures of the partial copula and how they are related to expected dependence measures of the conditional copula.

2 The partial copula

For simplicity, we consider continuous real-valued random variables with a joint positive density and assume that $\mathbb{E}[Y_i] = 0$ for $i = 1, 2$. $Y_1 \perp Y_2 | Z$ means that Y_1 and Y_2 are independent given Z and $C^\perp$ denotes the bivariate product copula. For $i = 1, 2$, let $\beta_i := \mathbb{E}[Z'Z]^{-1}\mathbb{E}[Z'Y_i]$ so that $Z\beta_i$ denotes the best linear predictor of Y_i in terms of Z, and define

$$\mathcal{E}_1 := Y_1 - Z\beta_1, \quad \text{and} \quad \mathcal{E}_2 := Y_2 - Z\beta_2.$$

The partial correlation of Y_1 and Y_2 given Z can be expressed as

$$\rho_{Y_1, Y_2; Z} = \mathbb{C}\text{orr}[\mathcal{E}_1, \mathcal{E}_2].$$

Thus, $\rho_{Y_1, Y_2; Z}$ is the correlation of Y_1 and Y_2 when each random variable has been corrected for the linear influence of Z, i.e., $\mathbb{C}\text{orr}[\tilde{g}(\mathcal{E}_i), \tilde{h}(Z)] = 0$ for all linear functions $\tilde{g}$ and $\tilde{h}$. If $(Y_{1:2}, Z)$ is jointly elliptically distributed, it is well known that $\mathbb{E}[\mathcal{E}_1 | Z] = \mathbb{E}[\mathcal{E}_2 | Z] = 0$

41

(a.s.), implying that $\rho_{Y_1,Y_2;Z}$ describes the correlation of Y_1 and Y_2 when their expectation does not depend on Z anymore. Moreover, if the elliptical distribution is a Gaussian distribution, $\rho_{Y_1,Y_2;Z}$ completely characterizes the conditional dependence. One reason for this is that $\mathcal{E}_1 \perp Z$ and $\mathcal{E}_2 \perp Z$, which is equivalent to $\mathbb{C}\mathrm{orr}[h(\mathcal{E}_i), g(Z)] = 0$ for $i = 1, 2$, and all measurable functions h and g, so that $\rho_{Y_1,Y_2;Z}$ can be interpreted as the correlation of random variables which are individually independent of Z. In order to generalize this idea to the non-Gaussian case and to obtain random variables which are individually independent of Z, we define

$$U_1 := F_{Y_1|Z}(Y_1|Z), \quad \text{and} \quad U_2 := F_{Y_2|Z}(Y_2|Z).$$

U_i is called the conditional probability integral transform (CPIT) of Y_i wrt Z. If $(Y_{1:2}, Z)$ has a Gaussian distribution then $U_i = \Phi(\mathcal{E}_i/\sigma_{\mathcal{E}_i})$ so that $\mathbb{C}\mathrm{orr}[\Phi^{-1}(U_1), \Phi^{-1}(U_2)] = \rho_{Y_1,Y_2;Z}$. However, even when the joint distribution of $(Y_{1:2}, Z)$ is not Gaussian, we have that $U_1 \perp Z$ and $U_2 \perp Z$ (Proposition 2.1 in Spanhel and Kurz (2015)). Thus, dependence measures which are based on the distribution of (U_1, U_2) are also meaningful if the underlying distribution is not Gaussian. The joint distribution of the CPITs is the partial copula $C^{\mathrm{p}}_{Y_1,Y_2;Z}$ of $F_{Y_1,Y_2|Z}$ and has been introduced by Bergsma (2011), Gijbels et al. (2015a, 2015b) and Spanhel and Kurz (2015).[1]

In some special cases, e.g., the Gaussian or Student-t distribution (Stöber et al. 2013), the partial copula and the conditional cdfs $(C^{\mathrm{p}}_{Y_1,Y_2;Z}, F_{Y_1|Z}, F_{Y_2|Z})$ determine $F_{Y_1,Y_2|Z}$ via $F_{Y_1,Y_2|Z}(y_1, y_2|z) = C^{\mathrm{p}}_{Y_1,Y_2;Z}(u_1, u_2)$, where $u_i = F_{Y_i|Z}(y_i|z), i = 1, 2$. However, in general, we have that $F_{Y_1,Y_2|Z}(y_1, y_2|z) = C_{Y_1,Y_2|Z}(u_1, u_2|z) := \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2|Z = z)$, where $C_{Y_1,Y_2|Z}$ denotes the conditional copula of $F_{Y_1,Y_2|Z}$ (Patton 2006). Thus, the partial copula, which is identical to the expected conditional copula $\int C_{Y_1,Y_2|Z}(\cdot, \cdot|t)dF_Z(t)$, often acts as an approximation of $C_{Y_1,Y_2|Z}$ or $F_{Y_1,Y_2|Z}$. At first sight, it might appear that the partial copula is a rather rough approximation of $C_{Y_1,Y_2|Z}$ since one assumes that (U_1, U_2) are jointly independent of Z. However, by construction we have that $U_1 \perp Z$ and $U_2 \perp Z$, so these necessary conditions for joint independence are satisfied. In particular, $F_{Y_1,Y_2|Z}$ can be recovered from $(C^{\mathrm{p}}_{Y_1,Y_2;Z}, F_{Y_1|Z}, F_{Y_2|Z})$ if and only if Y_1 and Y_2 can depend on Z but the remaining dependence between U_1 and U_2, which are individually independent of Z, does not depend on Z. But also when $C^{\mathrm{p}}_{Y_1,Y_2;Z}$ just acts as an approximation of $C_{Y_1,Y_2|Z}$, it is an attractive dependence measure. First, $C^{\mathrm{p}}_{Y_1,Y_2;Z}$ is easier to estimate than $C_{Y_1,Y_2|Z}$

---

[1]Bergsma (2011) uses the partial copula to test for conditional independence. Gijbels et al. (2015a, 2015b) propose a non-parametric estimator of the partial copula. Spanhel and Kurz (2015) show that partial copulas are optimal in the second tree of simplified vine copula approximations regarding the stepwise Kullback-Leibler divergence minimization.

and, second, it measures conditional dependence by one bivariate unconditional copula and not by infinitely many bivariate unconditional copulas as it is the case for $C_{Y_1,Y_2|Z}$. In the following, we give some explicit examples of partial copulas. Moreover, we investigate to what extent the approximation of $C_{Y_1,Y_2|Z}$ by $C^{\mathrm{P}}_{Y_1,Y_2;Z}$ and the approximation of $F_{Y_1,Y_2|Z}$ by $(C^{\mathrm{P}}_{Y_1,Y_2;Z}, F_{Y_1|Z}, F_{Y_2|Z})$ is optimal and examine properties of the partial copula and related dependence measures.

3 Examples of partial copulas

Example 1 (Trivariate FGM copula)
The three-dimensional Farlie-Gumbel-Morgenstern (FGM) copula is given by

$$C_{1:3}(v_{1:3}; \theta) = \prod_{i=1}^{3} v_i + \theta \prod_{i=1}^{3} v_i(1 - v_i), \quad |\theta| \leq 1.$$

Since this copula is exchangeable, all three conditional copulas $C_{12|3}, C_{13|2}, C_{23|1}$ are identical. It is straightforward to show that

$$C_{12|3}(u_1, u_2|z) = \prod_{i=1,2} u_i + \theta(1 - 2z) \prod_{i=1,2} u_i(1 - u_i),$$

and

$$C^{\mathrm{P}}_{12;3} = C^{\perp}.$$

Example 2 (Trivariate Frank copula)
Consider the exchangeable three-dimensional Frank copula with dependence parameter $\theta > 0$,

$$C_{1:3}(v_{1:3}) = \log \left\{ 1 - (1 - \alpha) \prod_{i=1}^{3} \frac{1 - \alpha^{v_i}}{1 - \alpha} \right\} \Big/ \log(\alpha), \quad \alpha := \exp(-\theta).$$

The conditional copula (Mesfioui and Quessy 2008) belongs to the Ali-Mikhail-Haq (AMH) family with dependence parameter $\gamma(z; \theta) = 1 - \exp(-\theta z)$, i.e.,

$$C_{12|3}(u_1, u_2|z) = \frac{u_1 u_2}{1 - \gamma(z; \theta) \prod_{i=1,2}(1 - u_i)}.$$

In Appendix A the closed-form expression for the partial copula is derived, which is given

by

$$C_{12;3}^{\mathrm{P}}(u_1, u_2) = \frac{u_1 u_2}{\theta f(u_1, u_2)} \left[\log\left(1 - (1 - \exp(-\theta))(1 - f(u_1, u_2))\right) + \theta \right],$$

where $f(u_1, u_2) := u_1 + u_2 - u_1 u_2$. Figure 1 illustrates $C_{12|3}$ and $C_{12;3}^{\mathrm{P}}$.

Example 3 (Partial copulas of the Gaussian, Student-t, and Clayton copula)
For the Gaussian, Student-t, and Clayton copula, conditional and partial copulas coincide since the simplifying assumption holds (Stöber et al. 2013).
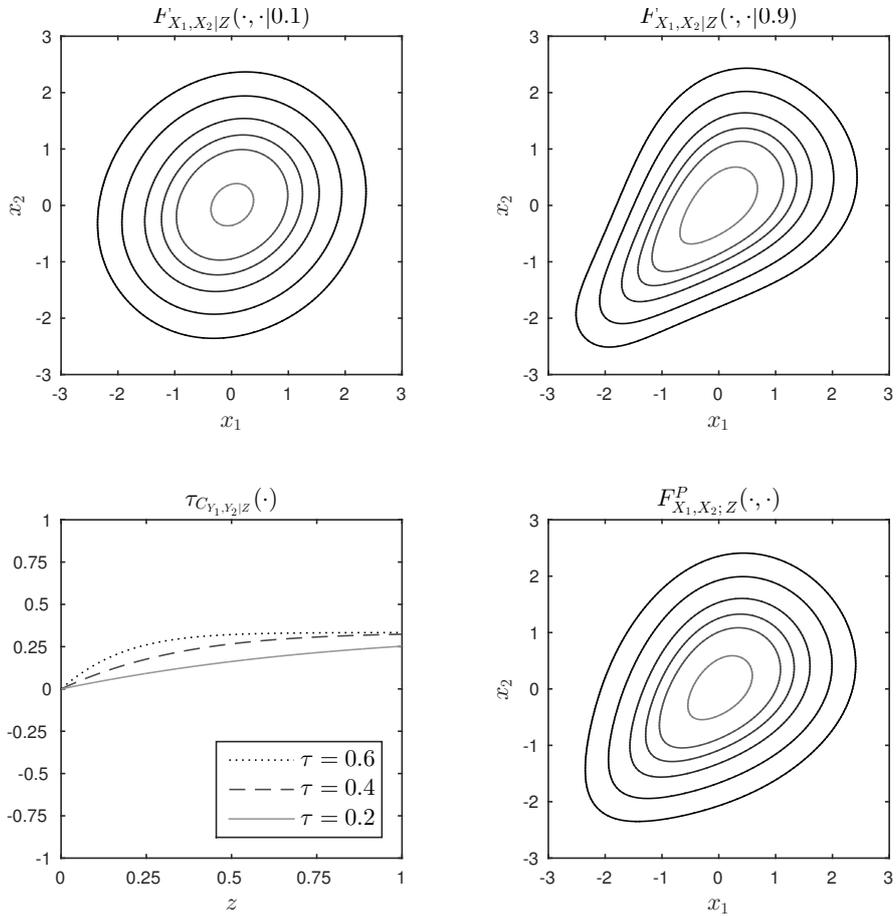
Figure 1: Illustration of the conditional and partial copula if the copula of $(Y_{1:2}, Z)$ is a trivariate Frank copula with pairwise values of Kendall's τ being 0.4. The upper panel shows contour plots of the density of $F_{X_1, X_2|Z}(\cdot, \cdot|z)$ where $X_i := \Phi^{-1}(U_i) = \Phi^{-1}(F_{Y_i|Z}(Y_i|Z))$, $i = 1, 2$, and Φ is the cdf of the standard normal distribution. On the left hand side in the lower panel the variation of Kendall's τ of $C_{Y_1, Y_2|Z}$ is depicted. The lower right figure shows contour plots of the density of $F_{X_1, X_2|Z}^{\mathrm{P}}$.

44

4 Properties of the partial copula

As stated in Section 2, the partial copula can be considered as an approximation of the conditional copula. The first property shows that the partial copula minimizes the Kullback-Leibler (KL) divergence from the conditional copula.

Property 1 (KL divergence minimization)
The partial copula $C^P_{Y_1,Y_2;Z}$ minimizes the KL divergence from the conditional copula $C_{Y_1,Y_2|Z}$ in the space of absolutely continuous bivariate distribution functions.
Proof: This follows from equation (3.3) in Theorem 3.1 in Spanhel and Kurz (2015).

However, the partial copula does not always minimize the L^2 distance to the conditional copula.

Property 2 (L^2 distance minimization)
Let $C_{Y_1,Y_2|Z}(U_1, U_2|Z)$ have finite variance and $\mathcal{F}_2$ be the space of bivariate cdfs so that each $C_{Y_1,Y_2;Z}(U_1, U_2) \in \mathcal{F}_2$ is a $C_{Y_1,Y_2|Z}$-measurable random variable with finite variance. Let $C^{L^2}_{Y_1,Y_2;Z} \in \mathcal{F}_2$ denote the bivariate cdf which minimizes the L^2 distance to $C_{Y_1,Y_2|Z}$. In general,

$$C^{L^2}_{Y_1,Y_2;Z} \neq C^P_{Y_1,Y_2;Z}.$$

Proof: See Appendix B.

The first two properties address the optimality of the partial copula when it comes to approximating conditional copulas. But partial copulas, together with univariate conditional cdfs, can also be used to provide a model for a general bivariate conditional distribution. For instance, Song et al. (2009) use generalized linear models for univariate conditional cdfs and join these conditional cdfs with an unconditional copula. Also, the conditional cdfs of financial returns are often filtered with ARMA-GARCH models and the remaining dependence is then modeled by an unconditional copula (Chen and Fan 2006, Liu and Luger 2009, Min and Czado 2014, Nikoloulopoulos et al. 2012). Therefore, the next property is of interest if conditional cdfs are linked with the partial copula.

Property 3 (Non-optimality of partial copula-induced approximations)
Let

$$F^P_{Y_1,Y_2|Z}(y_1, y_2|z) := C^P_{Y_1,Y_2;Z}\big(F_{Y_1|Z}(y_1|z), F_{Y_2|Z}(y_2|z)\big),$$

be the approximation of $F_{Y_1,Y_2|Z}$ that emerges if the conditional copula is approximated by the partial copula and the true univariate conditional cdfs are used. In general, $F^P_{Y_1,Y_2|Z}$ does neither minimize the KL divergence nor the L^2 distance from $F_{Y_1,Y_2|Z}$.

Proof: See Appendix C.

Although the partial copula $C^{\mathrm{P}}_{Y_1,Y_2;Z}$ minimizes the KL divergence from $C_{Y_1,Y_2|Z}$ (Property 1), Property 3 reveals the surprising result that this does not imply that the induced approximation $F^{\mathrm{P}}_{Y_1,Y_2|Z}$ also minimizes the KL divergence from $F_{Y_1,Y_2|Z}$. Note that Property 1 implies that $C^{\mathrm{P}}_{Y_1,Y_2;Z}$ is the bivariate copula that minimizes the KL divergence if the true conditional cdfs $(F_{Y_1|Z}, F_{Y_2|Z})$ are specified. Thus, Property 3 implies that, in general, one can obtain a better approximation by an adequate misspecification of the conditional cdfs. That is, there are conditional cdfs $(\tilde{F}_{Y_1|Z}, \tilde{F}_{Y_2|Z})$ such that $(\tilde{F}_{Y_1|Z}, \tilde{F}_{Y_2|Z}) \neq (F_{Y_1|Z}, F_{Y_2|Z})$ and $\mathbb{E}[C^{\mathrm{P}}_{Y_1,Y_2;Z}(\tilde{F}_{Y_1|Z}(Y_1|Z), \tilde{F}_{Y_2|Z}(Y_2|Z))] > \mathbb{E}[F^{\mathrm{P}}_{Y_1,Y_2|Z}(Y_1, Y_2|Z)]$. Because the marginal distributions of $(\tilde{F}_{Y_1|Z}(Y_1|Z), \tilde{F}_{Y_2|Z}(Y_2|Z))$ are not uniform in general, one can further improve the approximation if one specifies a pseudo-copula (Fermanian and Wegkamp 2012). Another interesting implication of Property 3 is that, if the conditional cdfs and the partial copula are estimated, the joint and stepwise ML estimator may have a different probability limit if the partial and conditional copula do not coincide.

Property 4 (Joint and stepwise ML estimation)

Wlog assume that the following random variables and parameters are scalars. Let $(F_{Y_1|Z}, F_{Y_2|Z})$ be the true conditional cdfs and $C^P_{Y_1,Y_2;Z}$ be the true partial copula of $C_{Y_1,Y_2|Z}$. Assume that we observe n independent samples from $(Y_{1:2}, Z)$. For $i = 1, 2$, let $\tilde{F}_{Y_i|Z}(\tilde{\theta}_i)$ be a parametric conditional cdf and assume that $\exists \theta_i \in \Theta_i : \tilde{F}_{Y_i|Z}(\theta_i) = F_{Y_i|Z}$. Let $\tilde{C}_{12}(\tilde{\theta}_3)$ be a parametric bivariate copula family so that $\exists \theta_3 \in \Theta_3 : \tilde{C}_{12}(\theta_3) = C^P_{Y_1,Y_2;Z}$. Let

$$\theta^J_n := \underset{\tilde{\theta}_{1:3} \in \Theta_{1:3}}{\arg\max} \sum_{i=1}^{n} \log\left(\tilde{c}_{12}\big(\tilde{F}_{Y_1|Z}(Y_{1,i}|Z_i; \tilde{\theta}_1), \tilde{F}_{Y_2|Z}(Y_{2,i}|Z_i; \tilde{\theta}_2); \tilde{\theta}_3\big) \prod_{j=1}^{2} \tilde{f}_{Y_j|Z}(Y_{j,i}|Z_i; \tilde{\theta}_j)\right)$$

be the joint ML estimator and

$$\theta^S_n := \begin{pmatrix} \theta^S_{1:2} \\ \theta^S_3 \end{pmatrix} = \begin{pmatrix} \underset{\tilde{\theta}_{1:2} \in \Theta_{1:2}}{\arg\max} \sum_{j=1}^{2} \sum_{i=1}^{n} \log \tilde{f}_{Y_j|Z}(Y_{j,i}|Z_i; \tilde{\theta}_j) \\ \underset{\tilde{\theta}_3 \in \Theta_3}{\arg\max} \sum_{i=1}^{n} \log \tilde{c}_{12}\big(\tilde{F}_{Y_1|Z}(Y_{1,i}|Z_i; \theta^S_1), \tilde{F}_{Y_2|Z}(Y_{2,i}|Z_i; \theta^S_2); \tilde{\theta}_3\big) \end{pmatrix}$$

be the stepwise ML estimator. Assume that the regularity conditions stated in Joe (2005) hold and that

$$\gamma := \underset{\tilde{\theta}_{1:3} \in \Theta_{1:3}}{\arg\max} \mathbb{E}\left[\log\left(\tilde{c}_{12}\big(\tilde{F}_{Y_1|Z}(Y_1|Z; \tilde{\theta}_1), \tilde{F}_{Y_2|Z}(Y_2|Z; \tilde{\theta}_2); \tilde{\theta}_3\big) \prod_{j=1}^{2} \tilde{f}_{Y_j|Z}(Y_j|Z; \tilde{\theta}_j)\right)\right]$$

exists. If $C_{Y_1,Y_2|Z} = C^P_{Y_1,Y_2;Z}$ (a.s.), then $\theta^J_n \overset{p}{\to} \theta$ and $\theta^S_n \overset{p}{\to} \theta$ for $n \to \infty$. However, if $C_{Y_1,Y_2|Z} \neq C^P_{Y_1,Y_2;Z}$ (a.s.), then $\theta^J_n \overset{p}{\to} \gamma$ and $\theta^S_n \overset{p}{\to} \theta$ for $n \to \infty$. In particular, $\theta^J_{i,n} - \theta^S_{i,n} \overset{p}{\to} 0$ may <u>not</u> hold for all $i = 1, 2, 3$.

Proof: See Appendix D.

Thus, the well known result that the joint and stepwise ML estimator of conditional cdfs and an unconditional copula converge to the same probability limit may not hold if the partial and conditional copula do not match. The next two properties are related to the parametric family of the partial copula.

Property 5 (Archimedean copulas)

Let the copula of $(Y_{1:2}, Z)$ be Archimedean. Then $C^P_{Y_1,Y_2;Z}$ might not be Archimedean.
Proof: See Appendix E.

Note that, if the copula of $(Y_{1:2}, Z)$ is Archimedean, $C_{Y_1,Y_2|Z}$ is always Archimedean (Mesfioui and Quessy 2008).

Property 6 (Family of the partial copula)

Assume that there is a bivariate parametric copula family $C^F(\cdot, \cdot; \theta)$ with parameter $\theta \in \Theta$ and a measurable function g such that $C_{Y_1,Y_2|Z}(\cdot, \cdot, |z) = C^F(\cdot, \cdot; g(z))$ for almost all z. In general, it does not hold that $\exists \theta \in \Theta : C^P_{Y_1,Y_2;Z}(\cdot, \cdot) = C^F(\cdot, \cdot; \theta)$.
Proof: This follows from Property 5.

Thus, partial copulas can also be used to obtain new (unconditional) copulas (see Example 2). One deficiency of the partial correlation coefficient is that its absolute value can be arbitrarily close to one if we have conditional independence. The partial copula is more attractive in this regard as the following property demonstrates.

Property 7 (Conditional (in)dependence)

Let $Y_1 \perp Y_2|Z$. The smallest upper bound for the absolute value of $\rho_{Y_1,Y_2;Z}$ is one. However, we always have that $C^P_{Y_1,Y_2;Z} = C^\perp$. On the other side, $C^P_{Y_1,Y_2;Z} = C^\perp$ or $\rho_{Y_1,Y_2;Z} = 0$ does not imply that $Y_1 \perp Y_2|Z$.
Proof: See Appendix F.

The next property points out that a varying conditional correlation is not sufficient for the non-equality of the partial and conditional copula.

Property 8 (Conditional correlation)

If $\rho_{Y_1,Y_2|Z}(Z) = (\mathbb{E}[Y_1 Y_2|Z] - \mathbb{E}[Y_1|Z]\mathbb{E}[Y_2|Z])/\sqrt{Var[Y_1|Z]Var[Y_2|Z]}$ is not almost surely a constant this does not imply that $\mathbb{P}(C_{Y_1,Y_2|Z} = C^P_{Y_1,Y_2;Z}) < 1$.
Proof: See Appendix G.

5   Properties of partial dependence measures

Let $\rho_{C_{Y_1,Y_2;Z}^{\mathrm{P}}}, \tau_{C_{Y_1,Y_2;Z}^{\mathrm{P}}}, \lambda_{C_{Y_1,Y_2;Z}^{\mathrm{P}}}^{l}$, and $\lambda_{C_{Y_1,Y_2;Z}^{\mathrm{P}}}^{u}$, denote Spearman's ρ, Kendall's τ, and the corresponding lower and upper tail dependence coefficients of $C_{Y_1,Y_2;Z}^{\mathrm{P}}$. We refer to these dependence measures as partial dependence measures, see also Gijbels et al. (2015b). The next proposition summarizes that partial Spearman's ρ and the partial tail dependence coefficients are equal to the expected Spearman's ρ and the expected tail dependence coefficients of the conditional copula.

Property 9 (Partial Spearman's ρ and partial tail dependence)
It holds that

$$\rho_{C_{Y_1,Y_2;Z}^{P}} = \mathbb{E}[\rho_{C_{Y_1,Y_2|Z}}(Z)]$$
$$\lambda_{C_{Y_1,Y_2;Z}^{P}}^{l} = \mathbb{E}[\lambda_{C_{Y_1,Y_2|Z}}^{l}(Z)]$$
$$\lambda_{C_{Y_1,Y_2;Z}^{P}}^{u} = \mathbb{E}[\lambda_{C_{Y_1,Y_2|Z}}^{u}(Z)]$$

Proof: These statements are easily verified by computing the expectations.

However, the expected conditional Kendall's τ is in general not equal to partial Kendall's τ (Gijbels et al. 2015b).

Property 10 (Partial Kendall's τ)
In general, $\tau_{C_{Y_1,Y_2;Z}^{P}} \neq \mathbb{E}[\tau_{C_{Y_1,Y_2|Z}}(Z)]$.
Proof: See Appendix H.

Unless $|\rho_{C_{Y_1,Y_2;Z}^{\mathrm{P}}}| = 1$, the value of $\rho_{C_{Y_1,Y_2;Z}^{\mathrm{P}}}$ does not provide any information about the value of $\rho_{C_{Y_1,Y_2|Z}}$. E.g., $\rho_{C_{Y_1,Y_2;Z}^{\mathrm{P}}} = 0$ does not imply that $\rho_{C_{Y_1,Y_2|Z}} = 0$ (a.s.) (see Example 1). However, for the coefficients of tail dependence we obtain the following relation.

Property 11 (Tail dependence)
$\lambda_{C_{Y_1,Y_2;Z}^{P}}^{l} = 0 \Leftrightarrow \lambda_{C_{Y_1,Y_2|Z}}^{l}(Z) = 0$ *(a.s)* and $\lambda_{C_{Y_1,Y_2;Z}^{P}}^{u} = 0 \Leftrightarrow \lambda_{C_{Y_1,Y_2|Z}}^{u}(Z) = 0$ *(a.s.)*.
Proof: Follows by Property 9 and because the coefficients of tail dependence are non-negative.

Thus, the partial copula has no tail dependence if and only if the conditional copula has no tail dependence (a.s.). This is a useful result because we can test for tail dependence of the conditional copula by testing for tail dependence of the partial copula. For instance, when modeling a time-varying conditional copula it is important to know whether the time-varying conditional copula should exhibit tail dependence.

6 Concluding remarks

The partial copula is a natural generalization of the partial correlation coefficient which does not share some of its drawbacks. We presented examples of the partial copula and investigated several of its properties. The bivariate partial copula minimizes the KL divergence to a bivariate conditional copula but the resulting approximation of a general bivariate conditional cdf does in general not minimize the KL divergence. As a result, the joint and stepwise ML estimator may converge to different probability limits.

While the partial copula has attractive theoretical properties, its estimation is much more involved than the estimation of the partial correlation coefficient. Non-parametric estimation of the partial copula has been proposed by Gijbels et al. (2015a, 2015b) if there is only one conditioning variable. However, further investigation is required to determine whether such a non-parametric estimator is reasonable if the set of conditioning variables is not very small and we have finite sample sizes. Alternatively, one could use higher-order partial copulas (Spanhel and Kurz 2015) which provide a different generalization of the partial correlation coefficient and are derived from the partial vine copula (PVC). While higher-order partial copulas do not share all properties of partial copulas, they can be efficiently estimated for a very large set of conditioning variables.

References for Chapter II

Bergsma, W. 2011. "Nonparametric testing of conditional independence by means of the partial copula". ArXiv e-prints `arXiv:1101.4607`.

Chen, X., and Y. Fan. 2006. "Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification". *Journal of Econometrics* 135 (1-2): 125–154.

Fermanian, J.-D., and M. H. Wegkamp. 2012. "Time-dependent copulas". *Journal of Multivariate Analysis* 110: 19–29.

Gijbels, I., M. Omelka, and N. Veraverbeke. 2015a. "Estimation of a copula when a covariate affects only marginal distributions". *Scandinavian Journal of Statistics* 42 (4): 1109–1126.

— . 2015b. "Partial and average copulas and association measures". *Electronic Journal of Statistics* 9 (2): 2420–2474.

Joe, H. 2005. "Asymptotic efficiency of the two-stage estimation method for copula-based models". *Journal of Multivariate Analysis* 94 (2): 401–419.

Liu, Y., and R. Luger. 2009. "Efficient estimation of copula-GARCH models". *Computational Statistics & Data Analysis* 53 (6): 2284–2297.

McNeil, A. J., R. Frey, and P. Embrechts. 2005. *Quantitative Risk Management*. Princeton series in finance. Princeton NJ: Princeton Univ. Press.

Mesfioui, M., and J.-F. Quessy. 2008. "Dependence structure of conditional Archimedean copulas". *Journal of Multivariate Analysis* 99 (3): 372–385.

Min, A., and C. Czado. 2014. "SCOMDY models based on pair-copula constructions with application to exchange rates". *Computational Statistics & Data Analysis* 76: 523–535.

Nelsen, R. B. 2006. *An Introduction to Copulas*. Springer series in statistics. New York, NY: Springer.

Nikoloulopoulos, A. K., H. Joe, and H. Li. 2012. "Vine copulas with asymmetric tail dependence and applications to financial return data". *Computational Statistics & Data Analysis* 56 (11): 3659–3673.

Patton, A. J. 2006. "Modelling Asymmetric Exchange Rate Dependence". *International Economic Review* 47 (2): 527–556.

Song, P. X.-K., M. Li, and Y. Yuan. 2009. "Joint Regression Analysis of Correlated Data Using Gaussian Copulas". *Biometrics* 65 (1): 60–68.

Spanhel, F., and M. S. Kurz. 2015. "Simplified vine copula models: Approximations based on the simplifying assumption". ArXiv e-prints `arXiv:1510.06971`.

Stöber, J., H. Joe, and C. Czado. 2013. "Simplified pair copula constructions—Limitations and extensions". *Journal of Multivariate Analysis* 119: 101–118.

Appendix

A   Derivation of the partial Frank copula (Example 2)

Let $g(u_1, u_2, \theta) := 1 - (1 - \exp(-\theta))(1 - u_1)(1 - u_2)$ and $f(u_1, u_2) := u_1 + u_2 - u_1 u_2$. The partial copula is given by

$$
\begin{aligned}
C_{12;3}^{\mathrm{P}}(u_1, u_2) &= \int_0^1 C_{12|3}(u_1, u_2|z)\,\mathrm{d}z = \int_0^1 \frac{u_1 u_2}{1 - (1 - \exp(-\theta z))(1 - u_1)(1 - u_2)}\mathrm{d}z \\
&= \int_1^{g(u_1,u_2,\theta)} \frac{u_1 u_2}{x\theta(f(u_1, u_2) - x)}\mathrm{d}x = \frac{u_1 u_2}{\theta f(u_1, u_2)} \int_1^{g(u_1,u_2,\theta)} \frac{1}{x} - \frac{1}{x - f(u_1, u_2)}\mathrm{d}x \\
&= \frac{u_1 u_2}{\theta f(u_1, u_2)} \left[\log(x) - \log(x - f(u_1, u_2))\right]_{x=1}^{x=g(u_1,u_2,\theta)} \\
&= \frac{u_1 u_2}{\theta f(u_1, u_2)} \left[\log g(u_1, u_2, \theta) + \theta\right].
\end{aligned}
$$

B   Proof of Property 2

It is well known that, if the variance of Y exists, $\mathbb{E}[Y|X]$ minimizes the L^2 distance to Y over all X-measurable random variables with finite variance. Thus

$$
\begin{aligned}
C_{Y_1,Y_2;Z}^{L2}(U_1, U_2) &= \underset{C_{Y_1,Y_2;Z} \in \mathcal{F}_2}{\arg\min} \ \mathbb{E}\left[(C_{Y_1,Y_2|Z}(U_1, U_2|Z) - C_{Y_1,Y_2;Z}(U_1, U_2))^2\right] \\
&= \mathbb{E}\left[C_{Y_1,Y_2|Z}(U_1, U_2|Z)|U_1, U_2\right].
\end{aligned}
$$

If $C_{Y_1,Y_2,Z}$ is the FGM copula given in Example 1 then

$$
\begin{aligned}
C_{Y_1,Y_2;Z}^{L^2}(u_1, u_2) &= \int_0^1 C_{Y_1,Y_2|Z}(u_1, u_2|z) f_{Z|U_1,U_2}(z|u_1, u_2)dz \\
&= u_1 u_2 \left(1 + \frac{\theta^2(4u_1^2 u_2^2 - 6(u_1^2 u_2 + u_1 u_2^2) + 2(u_1^2 + u_2^2) + 9u_1 u_2 - 3(u_1 + u_2) + 1)}{3}\right) \\
&\neq u_1 u_2 = C_{Y_1,Y_2;Z}^{\mathrm{P}}(u_1, u_2).
\end{aligned}
$$

C   Proof of Property 3

Wlog assume that Z is a scalar and that (Y_1, Y_2, Z) is a uniform random vector with $(Y_1, Z) \sim C_{12}$ and $(Z, Y_2) \sim C_{23}$. The KL divergence of $\tilde{F}_{Y_1,Y_2|Z}$ from $F_{Y_1,Y_2|Z}$ is given by

$$
D_{KL}(F_{Y_1,Y_2|Z}, \tilde{F}_{Y_1,Y_2|Z}) = \mathbb{E}\left[\log \frac{f_{Y_1,Y_2|Z}(Y_1, Y_2|Z)}{\tilde{f}_{Y_1,Y_2|Z}(Y_1, Y_2|Z)}\right]
$$

and is identical to the KL divergence given in equation (3.1) in Spanhel and Kurz (2015). From equation (3.7) in Theorem 3.2 in Spanhel and Kurz (2015) it follows that $D_{KL}(F_{Y_1,Y_2|Z}, \tilde{F}_{Y_1,Y_2|Z})$ does, in general, not attain a minimum at $\tilde{F}_{Y_1,Y_2|Z} = F^{\mathrm{P}}_{Y_1,Y_2|Z}$. That $F^{\mathrm{P}}_{Y_1,Y_2|Z}$ may not minimize the L^2 distance to $F_{Y_1,Y_2|Z}$ follows from Property 2.

D   Proof of Property 4

Joe (2005) establishes regularity conditions so that, if $C_{Y_1,Y_2|Z} = C^{\mathrm{P}}_{Y_1,Y_2;Z}$ (a.s.), then $\theta^S_n \xrightarrow{p} \theta$ and $\theta^J_n \xrightarrow{p} \theta$, which implies that $\theta^J_n - \theta^S_n \xrightarrow{p} 0$ holds. Now, let $C_{Y_1,Y_2|Z} \neq C^{\mathrm{P}}_{Y_1,Y_2;Z}$ and assume that $\gamma = \arg\max_{\tilde{\theta}_{1:3}\in\Theta_{1:3}} \mathbb{E}\big[\log\big(\tilde{c}_{12}\big(\tilde{F}_{Y_1|Z}(Y_1|Z;\tilde{\theta}_1), \tilde{F}_{Y_2|Z}(Y_2|Z;\tilde{\theta}_2);\tilde{\theta}_3\big)\prod_{j=1}^2 \tilde{f}_{Y_j|Z}(Y_j|Z;\tilde{\theta}_j)\big)\big]$ exists. Note that,

$$\theta_{1:2} = \arg\max_{\tilde{\theta}_{12}\in\Theta_{12}} \mathbb{E}\Big[\sum_{j=1}^2 \log\tilde{f}_{Y_j|Z}(Y_j|Z;\tilde{\theta}_j)\Big]$$

and

$$\theta_3 = \arg\max_{\tilde{\theta}_3\in\Theta_3} \mathbb{E}\big[\log\tilde{c}_{12}\big(\tilde{F}_{Y_1|Z}(Y_1|Z;\theta_1), \tilde{F}_{Y_2|Z}(Y_2|Z;\theta_2);\tilde{\theta}_3\big)\big].$$

Thus, under the regularity conditions in Joe (2005), it follows that $\theta^S_n \xrightarrow{p} \theta$. Property 3 and the subsequent remarks imply that, in general, $\gamma_i \neq \theta_i$ for all $i = 1, 2, 3$. Provided the regularity conditions in Joe (2005) are satisfied, it follows that $\theta^J_{i,n} \xrightarrow{p} \gamma_i$ for all $i = 1, 2, 3$, which finishes the proof.

E   Proof of Property 5

Let $C^{\mathrm{P}}_{Y_1,Y_2;Z}$ be given as in Example 2. We observe that

$$C^{\mathrm{P}}_{12;3}\big(C^{\mathrm{P}}_{12;3}(0.25, 0.5), 0.5\big) \neq C^{\mathrm{P}}_{12;3}\big(0.25, C^{\mathrm{P}}_{12;3}(0.5, 0.5)\big),$$

which shows that $C^{\mathrm{P}}_{12;3}$ is not associative, which is sufficient for the copula not to be Archimedean (Nelsen 2006).

F   Proof of Property 7

Let $Z \sim N(0,1)$ and $Y_1 = -1+Z^2+\mathcal{E}_1$, $Y_2 = -1+Z^2+\mathcal{E}_2$, where $\mathcal{E}_1, \mathcal{E}_2$, and Z are mutually independent, $\mathbb{E}[\mathcal{E}_i] = 0$ and $\mathbb{V}\mathrm{ar}[\mathcal{E}_i] = \sigma$ for $i = 1, 2$.[2] It is easy to show that $Y_1 \perp Y_2|Z$. Note that $\mathbb{C}\mathrm{ov}[Y_i, Z] = \mathbb{E}[Z^3] = 0$ so that $Y_i - (1, Z)\beta_i = Z^2 + \mathcal{E}_i$, where β_i is the best

---

[2]If $\sigma = 0$, then $(Y_{1:2}, Z)$ does not have a Lebesgue density, which we assume throughout the paper. However, if we allow that $\mathcal{E}_i$ is almost surely a constant, it follows that the maximal absolute value of $\rho_{Y_1,Y_2;Z}$ is one.

linear predictor of Y_i in terms of Z. Thus, $\rho_{Y_1,Y_2;Z} = \mathbb{C}\mathrm{orr}[Z^2 + \mathcal{E}_1, Z^2 + \mathcal{E}_2] = \frac{\mathbb{V}\mathrm{ar}[Z^2]}{\mathbb{V}\mathrm{ar}[Z^2]+\sigma}$, and $\lim_{\sigma \to 0} \rho_{Y_1,Y_2;Z} = 1$. Setting $Y_2 = 1 - Z^2 + \mathcal{E}_2$ shows that $\lim_{\sigma \to 0} \rho_{Y_1,Y_2;Z} = -1$. If $Y_1 \perp Y_2|Z$ then $C_{Y_1,Y_2|Z}(U_1, U_2|Z) = U_1 U_2$, i.e., $C^{\mathrm{P}}_{Y_1,Y_2;Z} = C^{\perp}$. From Example 1 it follows that $C^{\mathrm{P}}_{Y_1,Y_2;Z} = C^{\perp}$ does not imply that $Y_1 \perp Y_2|Z$.

G   Proof of Property 8

Let Z be exponentially distributed with unit mean, $Y_1|Z \sim \log\mathcal{N}(0,1)$ and $Y_2|Z = z \sim \log\mathcal{N}(0,z)$, where $\log\mathcal{N}(0,\sigma)$ denotes the log-normal distribution with zero location parameter and scale parameter σ. Using the same arguments as in Example 5.26 in McNeil et al. 2005, we obtain that, if $C_{Y_1,Y_2|Z}(U_1, U_2|Z) = \min(U_1, U_2) = C^{\mathrm{P}}_{Y_1,Y_2;Z}$, the support of the random variable

$$\rho_{Y_1,Y_2|Z}(Z) = \frac{\exp(Z) - 1}{\sqrt{(\exp(1) - 1)(\exp(Z^2) - 1)}}$$

is the unit interval $[0, 1]$.

H   Proof of Property 10

Let Z be uniformly distributed and consider the following conditional copula

$$C_{Y_1,Y_2|Z}(u_1, u_2|z) = u_1 u_2 + z u_1 u_2 (1 - u_1)(1 - u_2)(1 + u_1 u_2).$$

The corresponding partial copula is

$$C^{\mathrm{P}}_{Y_1,Y_2;Z}(u_1, u_2) = u_1 u_2 + \frac{1}{2} u_1 u_2 (1 - u_1)(1 - u_2)(1 + u_1 u_2).$$

Elementary integration yields that Kendall's τ for the conditional copula is given by

$$\tau_{C_{Y_1,Y_2|Z}}(Z) = 4 \int_{[0,1]^2} C_{Y_1,Y_2|Z}(u_1, u_2|Z)\, dC_{Y_1,Y_2|Z}(u_1, u_2|Z) - 1 = Z^2/450 + 5Z/18$$

Thus,

$$\mathbb{E}[\tau_{C_{Y_1,Y_2|Z}}(Z)] = \frac{377}{2700} \neq \frac{251}{1800} = \tau_{C^{\mathrm{P}}_{Y_1,Y_2;Z}}.$$

Chapter III

The partial vine copula: A dependence measure and approximation based on the simplifying assumption

This chapter is a reprint[*] of:

F. Spanhel and M. S. Kurz. 2017. "The partial vine copula: A dependence measure and approximation based on the simplifying assumption". ArXiv e-prints `arXiv:1510.06971`.

Revision note:

This chapter is a reprint of the revised manuscript available on arXiv since June 11, 2017. On October 23, 2015, the first version was made publicly available on arXiv under the title "Simplified vine copula models: Approximations based on the simplifying assumption".

Author contributions:

The project was initiated jointly by Fabian Spanhel and Malte Kurz and further developed in close collaboration. Both authors contributed to the manuscript. The terminology of higher-order partial copulas and the partial vine copulas was suggested by Fabian Spanhel and further developed in close collaboration by Fabian Spanhel and Malte Kurz. All proofs have been developed in close collaboration and have been improved and simplified in several rounds of revision. Both authors contributed equally to the proofs. In 2016, a major restructuring of the paper was drafted by Fabian Spanhel and implemented jointly. Several revisions have been done jointly by both authors. Most implementations, graphics and simulations have been done by Malte Kurz with support by Fabian Spanhel.

---

[*]All reprints in this cumulative dissertation have been slightly adapted in order to harmonize the overall appearance, like the format, page margins, caption setup, citation style and bibliography style.

The partial vine copula: A dependence measure and approximation based on the simplifying assumption [*]

Fabian Spanhel [†‡1] and Malte S. Kurz [§1]

[1]Department of Statistics,
Ludwig-Maximilians-Universität München,
Akademiestr. 1, 80799 Munich, Germany.

Abstract

Simplified vine copulas (SVCs), or pair-copula constructions, have become an important tool in high-dimensional dependence modeling. So far, specification and estimation of SVCs has been conducted under the simplifying assumption, i.e., all bivariate conditional copulas of the vine are assumed to be bivariate unconditional copulas. We introduce the partial vine copula (PVC) which provides a new multivariate dependence measure and which plays a major role in the approximation of multivariate distributions by SVCs. The PVC is a particular SVC where to any edge a j-th order partial copula is assigned and constitutes a multivariate analogue of the bivariate partial copula. We investigate to what extent the PVC describes the dependence structure of the underlying copula. We show that the PVC does not minimize the Kullback-Leibler divergence from the true copula and that the best approximation satisfying the simplifying assumption is given by a vine pseudo-copula. However, under regularity conditions, stepwise estimators of pair-copula constructions converge to the PVC irrespective of whether the simplifying assumption holds or not. Moreover, we elucidate why the PVC is the best feasible SVC approximation in practice.

Keywords: Vine copula, Pair-copula construction, Simplifying assumption, Conditional copula, Approximation.

---

1 Introduction

Copulas constitute an important tool to model dependence (Nelsen 2006, Joe 1997, McNeil et al. 2005). While it is easy to construct bivariate copulas, the construction of flexible high-dimensional copulas is a sophisticated problem. The introduction of simplified vine copulas (Joe (1996)), or pair-copula constructions (Aas et al. (2009)), has been an enormous advance for high-dimensional dependence modeling. Simplified vine copulas are hierarchical structures, constructed upon a sequence of bivariate unconditional copulas, which capture the conditional dependence between pairs of random variables if the data generating process satisfies the simplifying assumption. In this case, all conditional copulas of the data generating vine collapse to unconditional copulas and the true copula can be represented in terms of a simplified vine copula. Vine copula methodology and application have been extensively developed under the simplifying assumption (Dißmann et al. 2013, Grothe and Nicklas 2013, Joe et al. 2010, Kauermann and Schellhase 2014, Nikoloulopoulos et al. 2012), with studies showing the superiority of simplified vine copula models over elliptical copulas and nested Archimedean copulas (Aas and Berg (2009) and Fischer et al. (2009)).

Although some copulas can be expressed as a simplified vine copula, the simplifying assumption is not true in general. Hobæk Haff et al. (2010) point out that the simplifying assumption is in general not valid and provide examples of multivariate distributions which do not satisfy the simplifying assumption. Stöber et al. (2013) show that the Clayton copula is the only Archimedean copula for which the simplifying assumption holds, while the Student-t copula is the only simplified vine copula arising from a scale mixture of normal distributions. In fact, it is very unlikely that the unknown data generating process satisfies the simplifying assumption in a strict mathematical sense. As a result, researchers have recently started to investigate new dependence concepts that are related to the simplifying assumption and arise if it does not hold. In particular, studies on the bivariate partial copula, a generalization of the partial correlation coefficient, have (re-)emerged lately (Bergsma 2004, Gijbels et al. 2015a, 2015b, Spanhel and Kurz 2016b, Portier and Segers 2015).

We introduce the partial vine copula (PVC) which constitutes a multivariate analogue of the bivariate partial copula and which generalizes the partial correlation matrix. The PVC is a particular simplified vine copula where to any edge a j-th order partial copula is assigned. It provides a new multivariate dependence measure for a d-dimensional random vector in terms of $d(d-1)/2$ bivariate unconditional copulas and can be readily estimated for high-dimensional data (Nagler and Czado 2016). We investigate several properties of the PVC and show to what extent the dependence structure of the underlying distribution is captured. The PVC plays a crucial role in terms of approximating a multivariate distri-

bution by a simplified vine copula (SVC). We show that many estimators of SVCs converge to the PVC if the simplifying assumption does not hold. However, we also prove that the PVC may not minimize the Kullback-Leibler divergence from the true copula and thus may not be the best approximation in the space of simplified vine copulas. This result is rather surprising, because it implies that it may not be optimal to specify the true copulas in the first tree of a simplified vine copula approximation. Moreover, joint and stepwise estimators of SVCs may not converge to the same probability limit any more if the simplifying assumption does not hold. Nevertheless, due to the prohibitive computational burden or simply because only a stepwise model selection and estimation is possible, the PVC is the best feasible SVC approximation in practice. Moreover, the PVC is used by Nagler and Czado (2016) to construct a new non-parametric estimator of a multivariate distribution that can outperform classical non-parametric approaches and by Kurz and Spanhel (2017) to test the simplifying assumption in high-dimensional vine copulas. All in all, these facts highlight the great practical importance of the PVC for multivariate dependence modeling.

The rest of this paper is organized as follows. (Simplified) vine copulas, the simplifying assumption, conditional and partial copulas, are discussed in Section 2. The PVC and j-th order partial copulas are introduced in Section 3. Properties of the PVC and some examples are presented in Section 4. In Section 5 we analyze the role of the PVC for simplified vine copula approximations and explain why the PVC is the best feasible approximation in practical applications. A parametric estimator for the PVC is presented in Section 6 and implications for the stepwise and joint maximum likelihood estimator of simplified vine copulas are illustrated. Section 7 contains some concluding remarks.

The following notation and assumptions are used throughout the paper. We write $X_{1:d} := (X_1, \ldots, X_d)$, so that $F_{X_{1:d}}(x_{1:d}) := \mathbb{P}(\forall i = 1, \ldots, d \colon X_i \leq x_i)$, and $\mathrm{d}x_{1:d} := \mathrm{d}x_1 \ldots \mathrm{d}x_d$ to denote the variables of integration in $\int f_{X_{1:d}}(x_{1:d})\mathrm{d}x_{1:d}$. $C^\perp$ refers to the independence copula. $X \perp Y$ means that X and Y are stochastically independent. For $1 \leq k \leq d$, the partial derivative of g w.r.t. the k-th argument is denoted by $\partial_k g(x_{1:d})$. We write $\mathbb{1}_{\{A\}} = 1$ if A is true, and $\mathbb{1}_{\{A\}} = 0$ otherwise. For simplicity, we assume that all random variables are real-valued and continuous. In the following, let $d \geq 3$, if not otherwise specified, and $\mathcal{C}_d$ be the space of absolutely continuous d-dimensional copulas with positive density (a.s.). The distribution function of a random vector $U_{1:d}$ with uniform margins is denoted by $F_{1:d} = C_{1:d} \in \mathcal{C}_d$. We set $\mathcal{I}_l^d := \{(i,j) \colon j = l, \ldots, d-1, i = 1, \ldots, d-j\}$ and $S_{ij} := i+1 : i+j-1 := i+1, \ldots, i+j-1$. We focus on D-vine copulas, but all results carry over to regular vine copulas (Bedford and Cooke (2002), Kurowicka and Joe (2011)). An overview of the used notation can be found in Table 1. All proofs are deferred to the appendix.

Table 1: Notation for simplified D-vine copulas. $U_{1:d}$ has standard uniform margins, $d \geq 3, (i,j) \in \mathcal{I}_1^d, k = i, i+j$.

Notation	Explanation
$F_{1:d}$ or $C_{1:d}$	cdf and copula of $U_{1:d}$
$\mathcal{C}_d$	space of d-dimensional copulas with positive density
$\mathcal{C}_d^{\mathrm{SVC}}$	space of d-dimensional simplified D-vine copulas with positive density
$\mathcal{I}_1^d$	$\mathcal{I}_1^d := \{(i,j) \colon j = 1, \ldots, d-1, i = 1, \ldots, d-j\}$, the conditioned set of a D-vine copula density
S_{ij}	$S_{ij} := i+1 : i+j-1 := i+1, \ldots, i+j-1$, the conditioning set of an edge in a D-vine
$U_{k\mid S_{ij}}$	$F_{k\mid S_{ij}}(U_k\mid U_{S_{ij}})$, conditional probability integral transform (CPIT) of U_k w.r.t. $U_{S_{ij}}$
$C_{i,i+j;S_{ij}}$	bivariate conditional copula of $F_{i,i+j\mid S_{ij}}$, i.e., $C_{i,i+j;S_{ij}} = F_{U_{i\mid S_{ij}}, U_{i+j\mid S_{ij}}\mid U_{S_{ij}}}$
$C_{i,i+j;S_{ij}}^{\mathrm{SVC}}$	arbitrary bivariate (unconditional) copula that is used to model $C_{i,i+j;S_{ij}}$
$C_{i,i+j;S_{ij}}^{\mathrm{P}}$	partial copula of $C_{i,i+j;S_{ij}}$, i.e., $C_{i,i+j;S_{ij}}^{\mathrm{P}} = F_{U_{i\mid S_{ij}}, U_{i+j\mid S_{ij}}}$
$C_{i,i+j;S_{ij}}^{\mathrm{PVC}}$	$(j-1)$-th order partial copula of $C_{i,i+j;S_{ij}}$
$U_{k\mid S_{ij}}^{\mathrm{PVC}}$	$F_{k\mid S_{ij}}^{\mathrm{PVC}}(U_k\mid U_{S_{ij}})$, $(j-2)$-th order partial probability integral transform (PPIT) of U_k w.r.t. $U_{S_{ij}}$
$C_{1:d}^{\mathrm{PVC}}$	Partial vine copula (PVC) of $C_{1:d}$, if $d = 3$, then $c_{1:3}^{\mathrm{PVC}}(u_{1:3}) = c_{12}(u_1, u_2)\, c_{23}(u_2, u_3)\, c_{13;2}^{\mathrm{PVC}}(u_{1\mid 2}, u_{3\mid 2})$

2 Simplified vine copulas, conditional copulas, and higher-order partial copulas

In this section, we discuss (simplified) vine copulas and the simplifying assumption. Thereafter, we introduce the partial copula which can be considered as a generalization of the partial correlation coefficient and as an approximation of a bivariate conditional copula.

Definition 2.1 (Simplified D-vine copula or pair-copula construction – Joe (1996) and Aas et al. (2009))
For $(i,j) \in \mathcal{I}_1^d$, let $C_{i,i+j;S_{ij}}^{SVC} \in \mathcal{C}_2$ with density $c_{i,i+j;S_{ij}}^{SVC}$. For $j = 1$ and $i = 1, \ldots, d-j$, we set $C_{i,i+j;S_{ij}}^{SVC} = C_{i,i+1}^{SVC}$ and $u_{k\mid S_{ij}}^{SVC} = u_k$ for $k = i, i+j$. For $(i,j) \in \mathcal{I}_2^d$, define

$$u_{i\mid S_{ij}}^{SVC} := F_{i\mid S_{ij}}^{SVC}(u_i\mid u_{S_{ij}}) = \partial_2 C_{i,i+j-1;S_{i,j-1}}^{SVC}(u_{i\mid S_{i,j-1}}^{SVC}, u_{i+j-1\mid S_{i,j-1}}^{SVC}),$$
$$u_{i+j\mid S_{ij}}^{SVC} := F_{i+j\mid S_{ij}}^{SVC}(u_{i+j}\mid u_{S_{ij}}) = \partial_1 C_{i+1,i+j;S_{i+1,j-1}}^{SVC}(u_{i+1\mid S_{i+1,j-1}}^{SVC}, u_{i+j\mid S_{i+1,j-1}}^{SVC}).$$

(a) Simplified D-vine copula. **(b)** D-vine copula.
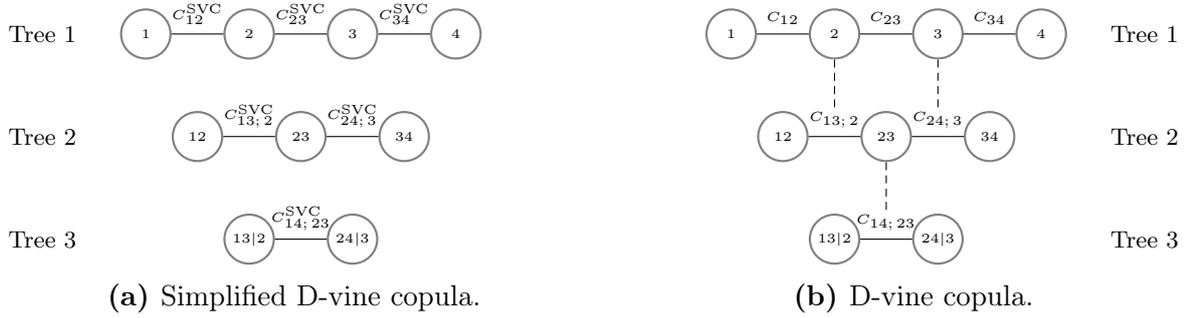
Figure 1: (Simplified) D-vine copula representation if $d = 4$. The influence of conditioning variables on the conditional copulas is indicated by dashed lines.

Then

$$c_{1:d}^{SVC}(u_{1:d}) = \prod_{(i,j) \in \mathcal{I}_1^d} c_{i,i+j;S_{ij}}^{SVC}\big(u_{i|S_{ij}}^{SVC}, u_{i+j|S_{ij}}^{SVC}\big)$$

is the density of a d-dimensional simplified D-vine copula $C_{1:d}^{SVC}$. We denote the space of d-dimensional simplified D-vine copulas by $\mathcal{C}_d^{SVC}$.

From a graph-theoretic point of view, simplified (regular) vine copulas can be considered as an ordered sequence of trees, where j refers to the number of the tree and a bivariate unconditional copula $C_{i,i+j;S_{ij}}^{\text{SVC}}$ is assigned to each of the $d - j$ edges of tree j (Bedford and Cooke (2002)). The left hand side of Figure 1 shows the graphical representation of a simplified D-vine copula for $d = 4$, i.e.,

$$c_{1:4}^{\text{SVC}}(u_{1:4}) = \underbrace{c_{12}^{\text{SVC}}(u_1, u_2)c_{23}^{\text{SVC}}(u_2, u_3)c_{34}^{\text{SVC}}(u_3, u_4)}_{\text{first tree}} \times \underbrace{c_{13;2}^{\text{SVC}}(u_{1|2}^{\text{SVC}}, u_{3|2}^{\text{SVC}})c_{24;3}^{\text{SVC}}(u_{2|3}^{\text{SVC}}, u_{4|3}^{\text{SVC}})}_{\text{second tree}}$$
$$\times \underbrace{c_{14;2:3}^{\text{SVC}}(u_{1|2:3}^{\text{SVC}}, u_{4|2:3}^{\text{SVC}})}_{\text{third tree}}.$$

The bivariate unconditional copulas $C_{i,i+j;S_{ij}}^{\text{SVC}}$ are also called pair-copulas, so that the resulting model is often termed a pair-copula construction (PCC). By means of simplified vine copula models one can construct a wide variety of flexible multivariate copulas because each of the $d(d-1)/2$ bivariate unconditional copulas $C_{i,i+j;S_{ij}}^{\text{SVC}}$ can be chosen arbitrarily and the resulting model is always a valid d-dimensional copula. Moreover, a pair-copula construction does not suffer from the curse of dimensions because it is build upon a sequence of bivariate unconditional copulas which renders it very attractive for high-dimensional applications. Obviously, not every multivariate copula can be represented by a simplified vine copula. However, every copula can be represented by the following (non-simplified) D-vine

copula.

Definition 2.2 (D-vine copula – Kurowicka and Cooke (2006))

Let $U_{1:d}$ be a random vector with cdf $F_{1:d} = C_{1:d} \in \mathcal{C}_d$. For $j = 1$ and $i = 1, \ldots, d - j$, we set $C_{i,i+j;S_{ij}} = C_{i,i+1}$ and $u_{k|S_{ij}} = u_k$ for $k = i, i + j$. For $(i,j) \in \mathcal{I}_2^d$, let $C_{i,i+j;S_{ij}}$ denote the conditional copula of $F_{i,i+j|S_{ij}}$ (Definition 2.5) and let $u_{k|S_{ij}} := F_{k|S_{ij}}(u_k|u_{S_{ij}})$ for $k = i, i + j$. The density of a D-vine copula decomposes the copula density of $U_{1:d}$ into $d(d-1)/2$ bivariate conditional copula densities $c_{i,i+j;S_{ij}}$ according to the following factorization:

$$c_{1:d}(u_{1:d}) = \prod_{(i,j)\in\mathcal{I}_1^d} c_{i,i+j;S_{ij}}(u_{i|S_{ij}}, u_{i+j|S_{ij}}|u_{S_{ij}}).$$

Contrary to a simplified D-vine copula in Definition 2.1, a bivariate conditional copula $C_{i,i+j;S_{ij}}$, which is in general a function of $j + 1$ variables, is assigned to each edge of a D-vine copula in Definition 2.2. The influence of the conditioning variables on the conditional copulas is illustrated by dashed lines in the right hand side of Figure 1. In applications, the simplifying assumption is typically imposed, i.e., it is assumed that all bivariate conditional copulas of the data generating vine copula degenerate to bivariate unconditional copulas.

Definition 2.3 (The simplifying assumption – Hobæk Haff et al. (2010))

The D-vine copula in Definition 2.2 satisfies the simplifying assumption if $c_{i,i+j;S_{ij}}(\cdot, \cdot|u_{S_{ij}})$ does not depend on $u_{S_{ij}}$ for all $(i,j)\in\mathcal{I}_2^d$.

If the data generating copula satisfies the simplifying assumption, it can be represented by a simplified vine copula, resulting in fast and simple statistical inference. Several methods for the consistent specification and estimation of pair-copula constructions have been developed under this assumption (Hobæk Haff (2013), Dißmann et al. (2013)). However, in view of Definition 2.2 and Definition 2.1 it is evident that it is extremely unlikely that the data generating vine copula strictly satisfies the simplifying assumption in practical applications.

Several questions arise if the data generating process does not satisfy the simplifying assumption and a simplified D-vine copula model (Definition 2.1) is used to approximate a general D-vine copula (Definition 2.2). First of all, what bivariate unconditional copulas $C_{i,i+j;S_{ij}}^{\text{svc}}$ should be chosen in Definition 2.1 to model the bivariate conditional copulas $C_{i,i+j;S_{ij}}$ in Definition 2.2 so that the best approximation w.r.t. a certain criterion is obtained? What simplified vine copula model do established stepwise procedures (asymptotically) specify and estimate if the simplifying assumption does not hold for the data generating vine copula? What are the properties of an optimal approximation? Before we

address these questions in Section 5, it is useful to recall the definition of the conditional and partial copula in the remainder of this section and to introduce and investigate the partial vine copula in Section 3 and Section 4 because it plays a major role in the approximation of copulas by simplified vine copulas.

Definition 2.4 (Conditional probability integral transform (CPIT))
Let $U_{1:d} \sim F_{1:d} \in \mathcal{C}_d$, $(i,j) \in \mathcal{I}_2^d$ and $k = i, i+j$. We call $U_{k|S_{ij}} := F_{k|S_{ij}}(U_k|U_{S_{ij}})$ the conditional probability integral transform of U_k w.r.t. $U_{S_{ij}}$.

It can be readily verified that, under the assumptions in Definition 2.4, $U_{k|S_{ij}} \sim \mathcal{U}(0,1)$ and $U_{k|S_{ij}} \perp U_{S_{ij}}$. Thus, applying the random transformation $F_{k|S_{ij}}(\cdot|U_{S_{ij}})$ to U_k removes possible dependencies between U_k and $U_{S_{ij}}$ and $U_{k|S_{ij}}$ can be interpreted as the remaining variation in U_k that can not be explained by $U_{S_{ij}}$. This interpretation of the CPIT is crucial for understanding the conditional and partial copula which are related to the (conditional) joint distribution of CPITs. The conditional copula has been introduced by Patton (2006) and we restate its definition here.[1]

Definition 2.5 (Bivariate conditional copula – Patton (2006))
Let $U_{1:d} \sim F_{1:d} \in \mathcal{C}_d$ and $(i,j) \in \mathcal{I}_2^d$. The (a.s.) unique conditional copula $C_{i,i+j;S_{ij}}$ of the conditional distribution $F_{i,i+j|S_{ij}}$ is defined by

$$
\begin{aligned}
C_{i,i+j;S_{ij}}(a,b|u_{S_{ij}}) &:= \mathbb{P}(U_{i|S_{ij}} \le a, U_{i+j|S_{ij}} \le b | U_{S_{ij}} = u_{S_{ij}}) \\
&= F_{i,i+j|S_{ij}}(F_{i|S_{ij}}^{-1}(a|u_{S_{ij}}), F_{i+j|S_{ij}}^{-1}(b|u_{S_{ij}})|u_{S_{ij}}).
\end{aligned}
$$

Equivalently, we have that

$$
F_{i,i+j|S_{ij}}(u_i, u_{i+j}|u_{S_{ij}}) = C_{i,i+j;S_{ij}}(F_{i|S_{ij}}(u_i|u_{S_{ij}}), F_{i+j|S_{ij}}(u_{i+j}|u_{S_{ij}})|u_{S_{ij}}),
$$

so that the effect of a change in $u_{S_{ij}}$ on the conditional distribution $F_{i,i+j|S_{ij}}(u_i, u_{i+j}|u_{S_{ij}})$ can be separated into two effects. First, the values of the CPITs, $(F_{i|S_{ij}}(u_i|u_{S_{ij}}), F_{i+j|S_{ij}}(u_{i+j}|u_{S_{ij}}))$, at which the conditional copula is evaluated, may change. Second, the functional form of the conditional copula $C_{i,i+j;S_{ij}}(\cdot,\cdot|u_{S_{ij}})$ may vary. In comparison to the conditional copula, which is the conditional distribution of two CPITs, the partial copula is the unconditional distribution and copula of two CPITs.

---

[1] Patton's notation for the conditional copula is given by $C_{i,i+j|S_{ij}}$. Originally, this notation has also been used in the vine copula literature (Aas et al. 2009, Kurowicka and Joe 2011, Acar et al. 2012). However, the current notation for a(n) (un)conditional copula that is assigned to an edge of a vine is given by $C_{i,i+j;S_{ij}}$ and $C_{i,i+j|S_{ij}}$ is used to denote $F_{U_i,U_{i+j}|U_{S_{ij}}}$ (Joe et al. 2010, Stöber et al. 2013, Krupskii and Joe 2013). In order to avoid possible confusions, we use $C_{i,i+j;S_{ij}}$ to denote a conditional copula and $C_{i,i+j;S_{ij}}^{\text{SVC}}$ to denote an unconditional copula.

Definition 2.6 (Bivariate partial copula - Bergsma (2004))

Let $U_{1:d} \sim F_{1:d} \in \mathcal{C}_d$ and $(i, j) \in \mathcal{I}_2^d$. The partial copula $C_{i,i+j;S_{ij}}^P$ of the distribution $F_{i,i+j|S_{ij}}$ is defined by

$$C_{i,i+j;S_{ij}}^P(a, b) := \mathbb{P}(U_{i|S_{ij}} \leq a, U_{i+j|S_{ij}} \leq b).$$

Since $U_{i|S_{ij}} \perp U_{S_{ij}}$ and $U_{i+j|S_{ij}} \perp U_{S_{ij}}$, the partial copula represents the distribution of random variables which are individually independent of the conditioning vector $U_{S_{ij}}$. This is similar to the partial correlation coefficient, which is the correlation of two random variables from which the linear influence of the conditioning vector has been removed. The partial copula can also be interpreted as the expected conditional copula,

$$C_{i,i+j;S_{ij}}^P(a, b) = \int_{\mathbb{R}^{j-1}} C_{i,i+j;S_{ij}}(a, b|u_{S_{ij}}) \mathrm{d}F_{S_{ij}}(u_{S_{ij}}),$$

and be considered as an approximation of the conditional copula. Indeed, it is easy to show that the partial copula $C_{i,i+j;S_{ij}}^P$ minimizes the Kullback-Leibler divergence from the conditional copula $C_{i,i+j;S_{ij}}$ in the space of absolutely continuous bivariate distribution functions. The partial copula is first mentioned by Bergsma (2004) who applies the partial copula to test for conditional independence. Recently, there has been a renewed interest in the partial copula. Spanhel and Kurz (2016b) investigate properties of the partial copula and mention some explicit examples whereas Gijbels et al. (2015a, 2015b) and Portier and Segers (2015) focus on the non-parametric estimation of the partial copula.

3 Higher-order partial copulas and the partial vine copula

A generalization of the partial correlation coefficient that is different from the partial copula is given by the higher-order partial copula. To illustrate this relation, let us recall the common definition of the partial correlation coefficient. Assume that all univariate margins of $Y_{1:d}$ have zero mean and finite variance. For $k = i, i + j$, let $\mathcal{P}(Y_k|Y_{S_{ij}})$ denote the best linear predictor of Y_k w.r.t $Y_{S_{ij}}$ which minimizes the mean squared error so that $\tilde{\mathcal{E}}_{k|S_{ij}} = Y_k - \mathcal{P}(Y_k|Y_{S_{ij}})$ is the corresponding prediction error. The partial correlation coefficient of Y_i and Y_{i+j} given $Y_{S_{ij}}$ is then defined by $\rho_{i,i+j;S_{ij}} = \mathbb{C}\mathrm{orr}[\tilde{\mathcal{E}}_{i|S_{ij}}, \tilde{\mathcal{E}}_{i+j|S_{ij}}]$. An

equivalent definition is given as follows. For $i = 1, \ldots, d-2$, let

$$\mathcal{E}_{i|i+1} := Y_i - \mathcal{P}(Y_i|Y_{i+1}), \quad \text{and} \quad \mathcal{E}_{i+2|i+1} := Y_{i+2} - \mathcal{P}(Y_{i+2}|Y_{i+1}). \tag{3.1}$$

Moreover, for $j = 3, \ldots, d-1$, and $i = 1, \ldots, d-j$, define

$$\begin{aligned}
\mathcal{E}_{i|S_{ij}} &:= \mathcal{E}_{i|S_{i,j-1}} - \mathcal{P}(\mathcal{E}_{i|S_{i,j-1}}|\mathcal{E}_{i+j-1|S_{i,j-1}}), \\
\mathcal{E}_{i+j|S_{ij}} &:= \mathcal{E}_{i+j|S_{i+1,j-1}} - \mathcal{P}(\mathcal{E}_{i+j|S_{i+1,j-1}}|\mathcal{E}_{i+1|S_{i+1,j-1}}).
\end{aligned} \tag{3.2}$$

It is easy to show that $\mathcal{E}_{k|S_{ij}} = \tilde{\mathcal{E}}_{k|S_{ij}}$ for all $k = i, i+j$ and $(i,j) \in \mathcal{I}_2^d$. That is, $\mathcal{E}_{k|S_{ij}}$ is the error of the best linear prediction of Y_k in terms of $Y_{S_{ij}}$. Thus, $\rho_{i,i+j;S_{ij}} = \mathbb{C}\mathrm{orr}[\mathcal{E}_{i|S_{ij}}, \mathcal{E}_{i+j|S_{ij}}]$. However, the interpretation of the partial correlation coefficient as a measure of conditional dependence is different depending on whether one considers it as the correlation of $(\tilde{\mathcal{E}}_{i|S_{ij}}, \tilde{\mathcal{E}}_{i+j|S_{ij}})$ or $(\mathcal{E}_{i|S_{ij}}, \mathcal{E}_{i+j|S_{ij}})$. For instance, $\rho_{14;23} = \mathbb{C}\mathrm{orr}[\tilde{\mathcal{E}}_{1|23}, \tilde{\mathcal{E}}_{4|23}]$ can be interpreted as the correlation between Y_1 and Y_4 after each variable has been corrected for the linear influence of $Y_{2:3}$, i.e., $\mathbb{C}\mathrm{orr}[g(\tilde{\mathcal{E}}_{k|23}), h(Y_{2:3})] = 0$ for all linear functions g and h. The idea of the partial copula is to replace the prediction errors $\mathcal{E}_{1|23}$ and $\mathcal{E}_{4|23}$ by the CPITS $U_{1|23}$ and $U_{4|23}$ which are independent of $Y_{2:3}$. On the other side, $\rho_{14;23} = \mathbb{C}\mathrm{orr}\left[\mathcal{E}_{1|23}, \mathcal{E}_{4|23}\right]$ 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 $\mathcal{E}_{2|3}$. Consequently, a different generalization of the partial correlation coefficient emerges if we do not only decorrelate the involved random variables in (3.1) and (3.2) but render them independent by replacing each expression of the form $X - \mathcal{P}(X|Z)$ in (3.1) and (3.2) by the corresponding CPIT $F_{X|Z}(X|Z)$. The joint distribution of a resulting pair of random variables is given by the j-th order partial copula and the set of these copulas together with a vine structure constitute the partial vine copula.

Definition 3.1 (Partial vine copula (PVC) and j-th order partial copulas)
Consider the D-vine copula $C_{1:d} \in \mathcal{C}_d$ stated in Definition 2.2. In the first tree, we set for $i = 1, \ldots, d-1$: $C_{i,i+1}^{PVC} = C_{i,i+1}$, while in the second tree, we denote for $i = 1, \ldots, d-2$, $k = i, i+2$: $C_{i,i+2;i+1}^{PVC} = C_{i,i+2;i+1}^{P}$ and $U_{k|i+1}^{PVC} = U_{k|i+1} = F_{k|i+1}(U_k|U_{i+1})$. In the remaining trees $j = 3, \ldots, d-1$, for $i = 1, \ldots, d-j$, we define

$$\begin{aligned}
U_{i|S_{ij}}^{PVC} &:= F_{i|S_{ij}}^{PVC}(U_i|U_{S_{ij}}) := \partial_2 C_{i,i+j-1;S_{i,j-1}}^{PVC}(U_{i|S_{i,j-1}}^{PVC}, U_{i+j-1|S_{i,j-1}}^{PVC}), \\
U_{i+j|S_{ij}}^{PVC} &:= F_{i+j|S_{ij}}^{PVC}(U_{i+j}|U_{S_{ij}}) := \partial_1 C_{i+1,i+j;S_{i+1,j-1}}^{PVC}(U_{i+1|S_{i+1,j-1}}^{PVC}, U_{i+j|S_{i+1,j-1}}^{PVC}),
\end{aligned}$$

and

$$C_{i,i+j;S_{ij}}^{PVC}(a, b) := \mathbb{P}(U_{i|S_{ij}}^{PVC} \leq a, U_{i+j|S_{ij}}^{PVC} \leq b).$$

We call the resulting simplified vine copula $C_{1:d}^{PVC}$ the partial vine copula (PVC) of $C_{1:d}$. Its density is given by

$$c_{1:d}^{PVC}(u_{1:d}) := \prod_{(i,j)\in\mathcal{I}_1^d} c_{i,i+j;S_{ij}}^{PVC}(u_{i|S_{ij}}^{PVC}, u_{i+j|S_{ij}}^{PVC}).$$

For $k = i, i + j$, we call $U_{k|S_{ij}}^{PVC}$ the $(j-2)$-th order partial probability integral transform (PPIT) of U_k w.r.t. $U_{S_{ij}}$ and $C_{i,i+j;S_{ij}}^{PVC}$ the $(j-1)$-th order partial copula of $F_{i,i+j|S_{ij}}$ that is induced by $C_{1:d}^{PVC}$.

Note that the first-order partial copula coincides with the partial copula of a conditional distribution with one conditioning variable. If $j \geq 3$, we call $C_{i,i+j;S_{ij}}^{\mathrm{PVC}}$ a higher-order partial copula. It is easy to show that, for all $(i,j) \in \mathcal{I}_1^d$, $U_{i|S_{ij}}^{\mathrm{PVC}}$ is the CPIT of $U_{i|S_{i,j-1}}^{\mathrm{PVC}}$ w.r.t. $U_{i+j-1|S_{i,j-1}}^{\mathrm{PVC}}$ and $U_{i+j|S_{ij}}^{\mathrm{PVC}}$ is the CPIT of $U_{i+j|S_{i+1,j-1}}^{\mathrm{PVC}}$ w.r.t. $U_{i+1|S_{i+1,j-1}}^{\mathrm{PVC}}$. Thus, PPITs are uniformly distributed and higher-order partial copulas are indeed copulas. Since $U_{i|S_{ij}}^{\mathrm{PVC}}$ is the CPIT of $U_{i|S_{i,j-1}}^{\mathrm{PVC}}$ w.r.t. $U_{i+j-1|S_{i,j-1}}^{\mathrm{PVC}}$, it is independent of $U_{i+j-1|S_{i,j-1}}^{\mathrm{PVC}}$. However, in general it is not true that $U_{i|S_{ij}}^{\mathrm{PVC}} \perp U_{S_{ij}}$ as the following proposition clarifies.

Lemma 3.1 (Relation between PPITs and CPITs)
For $(i,j) \in \mathcal{I}_2^d$ and $k = i, i + j$, it holds:

$$U_{k|S_{ij}}^{PVC} \perp U_{S_{ij}} \Leftrightarrow U_{k|S_{ij}}^{PVC} = U_{k|S_{ij}} \ (a.s.).$$

Note that $(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}) = (U_{i|S_{ij}}, U_{i+j|S_{ij}})$ (a.s.) if and only if $C_{i,i+j;S_{ij}}^{\mathrm{PVC}} = C_{i,i+j;S_{ij}}^{\mathrm{P}}$. Consequently, if a higher-order partial copula does not coincide with the partial copula, it describes the distribution of a pair of uniformly distributed random variables which are neither jointly nor individually independent of the conditioning variables of the corresponding conditional copula. Thus, if the simplifying assumption holds, then $C_{1:d} = C_{1:d}^{\mathrm{PVC}}$, i.e., higher-order partial copulas, partial copulas and conditional copulas coincide. This insight is used by Kurz and Spanhel (2017) to develop tests for the simplifying assumption in high-dimensional vine copulas.

Let $k = i, i + j$, and $G_{k|S_{ij}}^{\mathrm{PVC}}(t_k|t_{S_{ij}}) = (F_{k|S_{ij}}^{\mathrm{PVC}})^{-1}(t_k|t_{S_{ij}})$ denote the inverse of $F_{k|S_{ij}}^{\mathrm{PVC}}(\cdot|t_{S_{ij}})$ w.r.t. the first argument. A $(j-1)$-th order partial copula is then given by

$$C_{i,i+j;S_{ij}}^{\mathrm{PVC}}(a,b) = \mathbb{P}(U_{i|S_{ij}}^{\mathrm{PVC}} \leq a, U_{i+j|S_{ij}}^{\mathrm{PVC}} \leq b) = \mathbb{E}\big[\mathbb{P}(U_{i|S_{ij}}^{\mathrm{PVC}} \leq a, U_{i+j|S_{ij}}^{\mathrm{PVC}} \leq b|U_{S_{ij}})\big]$$

$$= \int_{[0,1]^{j-1}} C_{i,i+j;S_{ij}}\Big(F_{i|S_{ij}}\big(G_{i|S_{ij}}^{\mathrm{PVC}}(a|t_{S_{ij}})\big|t_{S_{ij}}\big), F_{i+j|S_{ij}}\big(G_{i+j|S_{ij}}^{\mathrm{PVC}}(b|t_{S_{ij}})\big|t_{S_{ij}}\big)\Big|t_{S_{ij}}\Big)\mathrm{d}F_{S_{ij}}(t_{S_{ij}}).$$

If $j \geq 3$, $C_{i,i+j;S_{ij}}^{\mathrm{PVC}}$ depends on $F_{i|S_{ij}}, F_{i+j|S_{ij}}, C_{i,i+j;S_{ij}}$, and $F_{S_{ij}}$, i.e., it depends on $C_{i:i+j}$. Moreover, $C_{i,i+j;S_{ij}}^{\mathrm{PVC}}$ also depends on $G_{i|S_{ij}}^{\mathrm{PVC}}$ and $G_{i+j|S_{ij}}^{\mathrm{PVC}}$, which are determined by the regular vine structure. Thus, the corresponding PVCs of different regular vines may be different. In particular, if the simplifying assumption does not hold, higher-order partial copulas of different PVCs which refer to the same conditional distribution may not be identical. This is different from the partial correlation coefficient or the partial copula which do not depend on the structure of the regular vine.

In general, higher-order partial copulas do not share the simple interpretation of the partial copula because they can not be considered as expected conditional copulas. However, higher-order partial copulas can be more attractive from a practical point of view. The estimation of the partial copula of $C_{i,i+j;S_{ij}}$ requires the estimation of the two j-dimensional conditional cdfs $F_{i|S_{ij}}$ and $F_{i+j|S_{ij}}$ to construct pseudo-observations from the CPITs $(U_{i|S_{ij}}, U_{i+j|S_{ij}})$. As a result, a non-parametric estimation of the partial copula is only sensible if j is very small. In contrast, a higher-order partial copula is the distribution of two PPITs $(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}})$ which are made up of only two-dimensional functions (Definition 3.1). Thus, the non-parametric estimation of a higher-order partial copula does not suffer from the curse of dimensionality and is also sensible for large j (Nagler and Czado 2016). But also in a parametric framework the specification of the model family is much easier for a higher-order partial copula than for a conditional copula. This renders higher-order partial copulas very attractive from a modeling point of view to analyze and estimate bivariate conditional dependencies. As we show in Section 6, the PVC is also the probability limit of many estimators of pair-copula constructions and thus of great practical importance.

4   Properties of the partial vine copula and examples

In this section, we analyze to what extent the PVC describes the dependence structure of the data generating copula if the simplifying assumption does not hold. We first investigate whether the bivariate margins of $C_{1:d}^{\mathrm{PVC}}$ match the bivariate margins of $C_{1:d}$ and then take a closer look at conditional independence relations. By construction, the bivariate margins $C_{i,i+1}^{\mathrm{PVC}}, i = 1, \ldots, d-1$, of the PVC given in Definition 3.1 are identical to the corresponding margins $C_{i,i+1}, i = 1, \ldots, d-1$, of $C_{1:d}$. That is because the PVC explicitly specifies these $d-1$ margins in the first tree of the vine. The other bivariate margins $C_{i,i+j}^{\mathrm{PVC}}$, where

$(i, j) \in \mathcal{I}_2^d$, are implicitly specified and given by

$$C_{i,i+j}^{\mathrm{PVC}}(u_i, u_{i+j}) = \int_{[0,1]^{j-1}} C_{i,i+j;S_{ij}}^{\mathrm{PVC}}(F_{i|S_{ij}}^{\mathrm{PVC}}(u_i|u_{S_{ij}}), F_{i+j}^{\mathrm{PVC}}(u_{i+j}|u_{S_{ij}})) \mathrm{d}C_{S_{ij}}^{\mathrm{PVC}}(u_{S_{ij}}).$$

The relation between the implicitly given bivariate margins of the PVC and the underlying copula are summarized in the following lemma.

Lemma 4.1 (Implicitly specified margins of the PVC)
Let $C_{1:d} \in \mathcal{C}_d \backslash \mathcal{C}_d^{SVC}$, $(i, j) \in \mathcal{I}_2^d$, and τ_E and ρ_E denote Kendall's τ and Spearman's ρ of the copula $E \in \mathcal{C}_2$. In general, it holds that $C_{i,i+j}^{PVC} \neq C_{i,i+j}, \rho_{C_{i,i+j}^{PVC}} \neq \rho_{C_{i,i+j}}$, and $\tau_{C_{i,i+j}^{PVC}} \neq \tau_{C_{i,i+j}}$.

The next example provides an example of a three-dimensional PVC and illustrates the results of Lemma 4.1. Other examples of PVCs in three dimensions are given in Spanhel and Kurz (2016b).

Example 4.1
Let $C^{FGM_2}(\theta)$ denote the bivariate FGM copula

$$C^{FGM_2}(u_1, u_2; \theta) = u_1 u_2 [1 + \theta(1 - u_1)(1 - u_2)], \quad |\theta| \leq 1,$$

and $C^A(\gamma)$ denote the following asymmetric version of the FGM copula (Nelsen 2006, Example 3.16)

$$C^A(u_1, u_2; \gamma) = u_1 u_2 [1 + \gamma u_1 (1 - u_1)(1 - u_2)], \quad |\gamma| \leq 1. \tag{4.1}$$

Assume that $C_{12} = C^A(\gamma), C_{23} = C^\perp, C_{13;2}(\cdot, \cdot; u_2) = C^{FGM_2}(\cdot, \cdot; 1 - 2u_2)$ for all u_2, so that

$$C_{1:3}(u_{1:3}) = \int_0^{u_2} C^{FGM_2}(\partial_2 C^A(u_1, t_2), u_3; 1 - 2t_2) dt_2.$$

Elementary computations show that the implicit margin is given by

$$C_{13}(u_1, u_3) = u_1 u_3 [\gamma(u_1 - 3u_1^2 + 2u_1^3)(1 - u_3) + 3]/3,$$

which is a copula with quartic sections in u_1 and square sections in u_3 if $\gamma \neq 0$. The corresponding PVC is

$$C_{1:3}^{PVC}(u_{1:3}) = \int_0^{u_2} C_{13;2}^{PVC}(F_{1|2}(u_1|t_2), F_{3|2}(u_3|t_2)) dt_2 \overset{C_{13;2}^{PVC}=C^\perp}{=} u_3 \int_0^{u_2} \partial_2 C^A(u_1, t_2) dt_2$$

and the implicit margin of $C_{1:3}^{PVC}$ is

$$C_{13}^{PVC}(u_1, u_3) = C_{1:3}^{PVC}(u_1, 1, u_3) = u_1 u_3.$$

Moreover, $\rho_{C_{13}} = -\gamma/1080, \tau_{C_{13}} = -\gamma/135$, but $\rho_{C_{13}^{PVC}} = \tau_{C_{13}^{PVC}} = 0$.

Higher-order partial copulas can also be used to construct new measures of conditional dependence. For instance, if $X_{1:d}$ is a random vector with copula $C_{1:d} \in \mathcal{C}_d$, higher-order partial Spearman's ρ and Kendall's τ of X_i and X_{i+j} given $X_{S_{ij}}$ are defined by

$$\tau_{C_{i,i+j;S_{ij}}^{\mathrm{PVC}}} = 4 \int_{[0,1]^2} C_{i,i+j;S_{ij}}^{\mathrm{PVC}}(a,b) \mathrm{d}C_{i,i+j;S_{ij}}^{\mathrm{PVC}}(a,b) - 1,$$

$$\rho_{C_{i,i+j;S_{ij}}^{\mathrm{PVC}}} = 12 \int_{[0,1]^2} C_{i,i+j;S_{ij}}^{\mathrm{PVC}}(a,b) \mathrm{d}a \mathrm{d}b - 3.$$

Note that all dependence measures that are derived from a higher-order partial copula are defined w.r.t. a regular vine structure and that they coincide with their conditional analogues if the simplifying assumption holds. A partial correlation coefficient of zero is commonly interpreted as an indication of conditional independence, although this can be quite misleading if the underlying distribution is not close to a Normal distribution (Spanhel and Kurz (2016b)). Therefore, one might wonder to what extent higher-order partial copulas can be used to check for conditional independencies. If $C_{i,i+j;S_{ij}}^{\mathrm{PVC}}$ equals the independence copula, we say that X_i and X_{i+j} are (j-th order) partially independent given $X_{S_{ij}}$ and write $X_i \overset{\mathrm{PVC}}{\perp} X_{i+j}|X_{S_{ij}}$. The following theorem establishes that there is in general no relation between conditional independence and higher-order partial independence.

Theorem 4.1 (Conditional independence and j-th order partial independence)
Let $d \geq 4$, $(i,j) \in \mathcal{I}_1^d$, and $C_{1:d} \in \mathcal{C}_d \backslash \mathcal{C}_d^{svc}$ be the copula of $X_{1:d}$. It holds that

$$X_i \perp X_{i+2}|X_{i+1} \Rightarrow X_i \overset{PVC}{\perp} X_{i+2}|X_{i+1},$$

$$\forall j \geq 3 : X_i \perp X_{i+j}|X_{S_{ij}} \not\Rightarrow X_i \overset{PVC}{\perp} X_{i+j}|X_{S_{ij}},$$

and

$$\forall j \geq 2 : X_i \perp X_{i+j}|X_{S_{ij}} \not\Leftarrow X_i \overset{PVC}{\perp} X_{i+j}|X_{S_{ij}}.$$

The next five-dimensional example illustrates higher-order partial copulas, higher-order PPITs, and the relation between partial independence and conditional independence.
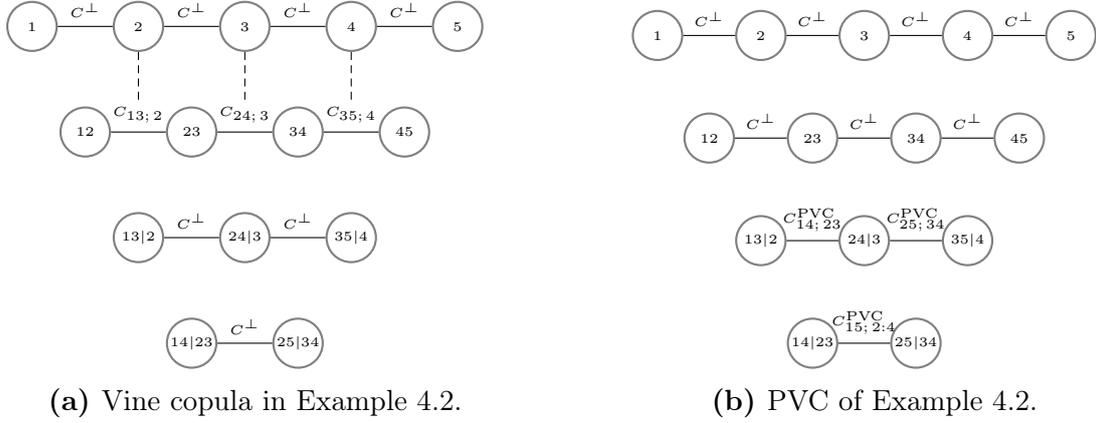
(a) Vine copula in Example 4.2.   **(b)** PVC of Example 4.2.

Figure 2: The non-simplified D-vine copula given in Example 4.2 and its PVC. The influence of conditioning variables on the conditional copulas is indicated by dashed lines.

Example 4.2

Consider the following exchangeable D-vine copula $C_{1:5}$ which does not satisfy the simplifying assumption:

$$C_{12} = C_{23} = C_{34} = C_{45}, \quad C_{13;2} = C_{24;3} = C_{35;4}, \quad C_{14;2:3} = C_{25;3:4},$$

$$C_{12} = C^{\perp}, \tag{4.2}$$

$$C_{13;2}(a, b | u_2) = C^{FGM_2}(a, b\,; 1 - 2u_2), \quad \forall (a, b, u_2) \in [0, 1]^3 \tag{4.3}$$

$$C_{14;2:3} = C^{\perp}, \tag{4.4}$$

$$C_{15;2:4} = C^{\perp}, \tag{4.5}$$

where $C_{i,i+j;S_{ij}} = C^{\perp}$ means that $C_{i,i+j;S_{ij}}(a, b | u_{S_{ij}}) = ab$ for all $(a, b, u_{S_{ij}}) \in [0, 1]^{j+1}$.

All conditional copulas of the vine copula in Example 4.2 correspond to the independence copula except for the second tree. Note that for all $i = 1, 2, 3$, $(U_i, U_{i+1}, U_{i+2}) \sim C^{FGM_3}(1)$, where $C^{FGM_3}(u_{1:3}; \theta) = \prod_{i=1}^{3} u_i + \theta \prod_{i=1}^{3} u_i(1 - u_i), |\theta| \le 1$, is the three-dimensional FGM copula. The left panel of Figure 2 illustrates the D-vine copula of the data generating process. We now investigate the PVC of $C_{1:5}$ which is illustrated in the right panel of Figure 2. Since $C_{1:5}$ and $C_{1:5}^{\text{PVC}}$ are exchangeable copulas, we only report the PPITs $U_{1|2}^{\text{PVC}}, U_{1|2:3}^{\text{PVC}}$ and $U_{1|2:4}^{\text{PVC}}$ in the following lemma.

Lemma 4.2 (The PVC of Example 4.2)

Let $C_{1:5}$ be defined as in Example 4.2. Then

$$C_{12}^{PVC} = C_{23}^{PVC} = C_{34}^{PVC} = C_{45}^{PVC}, \quad C_{13;2}^{PVC} = C_{24;3}^{PVC} = C_{35;4}^{PVC}, \quad C_{14;2:3}^{PVC} = C_{25;3:4}^{PVC},$$

$$C_{12}^{PVC} = C^{\perp},$$

$$C^{PVC}_{13;2} = C^{\perp},$$
$$C^{PVC}_{14;2:3}(a,b) = C^{FGM_2}(a,b\,;1/9), \quad \forall (a,b) \in [0,1]^2$$
$$C^{PVC}_{15;2:4} \neq C^{\perp},$$

and

$$U^{PVC}_{1|2} = U_1 = U_{1|2},$$
$$U^{PVC}_{1|2:3} = U_1 \neq U_{1|2:3} = U_1[1 + (1 - U_1)(1 - 2U_2)(1 - 2U_3)],$$
$$U^{PVC}_{1|2:4} = U_1[1 + \theta(1 - U_1)(1 - 2U_4)] \neq U_{1|2:4} = U_{1|2:3}.$$

Lemma 4.2 demonstrates that j-th order partial copulas may not be independence copulas, although the corresponding conditional copulas are independence copulas. In particular, under the data generating process the edges of the third tree of $C_{1:5}$ are independence copulas. Neglecting the conditional copulas in the second tree and replacing them with first-order partial copulas induces spurious dependencies in the third tree of $C^{PVC}_{1:5}$. The introduced spurious dependence also carries over to the fourth tree where we have (conditional) independence in fact. Nevertheless, the PVC reproduces the bivariate margins of $C_{1:5}$ pretty well. It can be readily verified that $(C^{PVC}_{13}, C^{PVC}_{14}, C^{PVC}_{24}, C^{PVC}_{25}, C^{PVC}_{35}) = (C_{13}, C_{14}, C_{24}, C_{25}, C_{35})$, i.e., except for C^{PVC}_{15}, all bivariate margins of $C^{PVC}_{1:5}$ match the bivariate margins of $C_{1:5}$ in Example 4.2. Moreover, the mutual information in the third and fourth tree are larger if higher-order partial copulas are used instead of the true conditional copulas. Thus, the spurious dependence in the third and fourth tree decreases the Kullback-Leibler divergence from $C_{1:5}$ and therefore acts as a countermeasure for the spurious (conditional) independence in the second tree. Lemma 4.2 also reveals that $U_{1|2:4}$ is a function of U_2 and U_3, i.e. the true conditional distribution function $F_{1|2:4}$ depends on u_2 and u_3. In contrast, $F^{PVC}_{1|2:4}$, the resulting model for $F_{1|2:4}$ which is implied by the PVC, depends only on u_4. That is, the implied conditional distribution function of the PVC depends on the conditioning variable which actually has no effect.

5   Approximations based on the partial vine copula

The specification and estimation of SVCs is commonly based on procedures that asymptotically minimize the Kullback-Leibler divergence (KLD) in a stepwise fashion. For instance, if a parametric vine copula model is used, the step-by-step ML estimator (Hobæk Haff (2012, 2013)), where one estimates tree after tree and sequentially minimizes the estimated KLD conditional on the estimates from the previous trees, is often employed in order to select and estimate the parametric pair-copula families of the vine. But also the non-parametric

methods of Kauermann and Schellhase (2014) and Nagler and Czado (2016) proceed in a stepwise manner and asymptotically minimize the KLD of each pair-copula separately under appropriate conditions. In this section, we investigate the role of the PVC when it comes to approximating non-simplified vine copulas.

Let $C_{1:d} \in \mathcal{C}_d$ and $C_{1:d}^{\text{SVC}} \in \mathcal{C}_d^{\text{SVC}}$. The KLD of $C_{1:d}^{\text{SVC}}$ from the true copula $C_{1:d}$ is given by

$$D_{KL}(C_{1:d}||C_{1:d}^{\text{SVC}}) = \mathbb{E}\left[\log \frac{c_{1:d}(U_{1:d})}{c_{1:d}^{\text{SVC}}(U_{1:d})}\right],$$

where the expectation is taken w.r.t. the true distribution $C_{1:d}$. We now decompose the KLD into the Kullback-Leibler divergences related to each of the $d-1$ trees. For this purpose, let $j = 1, \ldots, d-1$ and define

$$\mathbf{T}_j := \left\{(C_{i,i+j;S_{ij}}^{\text{SVC}})_{i=1,\ldots,d-j} \colon C_{i,i+j;S_{ij}}^{\text{SVC}} \in \mathcal{C}_2 \text{ for } 1 \leq i \leq d-j\right\},$$

so that $\mathbf{T}_{1:j} = \times_{k=1}^{j}\mathbf{T}_k$ represents all possible SVCs up to and including the j-th tree. Let $\mathcal{T}_j \in \mathbf{T}_j, \mathcal{T}_{1:j-1} \in \mathbf{T}_{1:j-1}$. The KLD of the SVC associated with $\mathcal{T}_{1:d-1}$ is given by

$$D_{KL}(C_{1:d}||\mathcal{T}_{1:d-1}) = \sum_{j=1}^{d-1} D_{KL}^{(j)}(\mathcal{T}_j(\mathcal{T}_{1:j-1})), \tag{5.1}$$

where

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

denotes the KLD related to the first tree, and for the remaining trees $j = 2, \ldots, d-1$, the related KLD is

$$D_{KL}^{(j)}(\mathcal{T}_j(\mathcal{T}_{1:j-1})) := \sum_{i=1}^{d-j} \mathbb{E}\left[\log \frac{c_{i,i+j;S_{ij}}(U_{i|S_{ij}}, U_{i+j|S_{ij}}|U_{S_{ij}})}{c_{i,i+j;S_{ij}}^{\text{SVC}}(U_{i|S_{ij}}^{\text{SVC}}, U_{i+j|S_{ij}}^{\text{SVC}})}\right].$$

For instance, if $d = 3$, the KLD can be decomposed into the KLD related to the first tree $D_{KL}^{(1)}$ and to the second tree $D_{KL}^{(2)}$ as follows

$$D_{KL}(C_{1:3}||\mathcal{T}_{1:2}) = D_{KL}(C_{1:3}||(\mathcal{T}_1, \mathcal{T}_2)) = D_{KL}^{(1)}(\mathcal{T}_1) + D_{KL}^{(2)}(\mathcal{T}_2(\mathcal{T}_1))$$

$$= \mathbb{E}\left[\log \frac{c_{12}(U_{1:2})c_{23}(U_{2:3})}{c_{12}^{\text{SVC}}(U_{1:2})c_{23}^{\text{SVC}}(U_{2:3})}\right] + \mathbb{E}\left[\log \frac{c_{13;2}\big(\partial_2 C_{12}(U_{1:2}), \partial_1 C_{23}(U_{2:3})|U_2\big)}{c_{13;2}^{\text{SVC}}\big(\partial_2 C_{12}^{\text{SVC}}(U_{1:2}), \partial_1 C_{23}^{\text{SVC}}(U_{2:3})\big)}\right].$$

Note that the KLD related to tree j depends on the specified copulas in the lower trees because they determine at which values the copulas in tree j are evaluated. The following

71

theorem shows that, if one sequentially minimizes the KLD related to each tree, then the optimal SVC is the PVC.

Theorem 5.1 (Tree-by-tree KLD minimization using the PVC)

Let $C_{1:d} \in \mathcal{C}_d$ be the data generating copula and $\mathcal{T}_j^{PVC} := (C_{i,i+j;S_{ij}}^{PVC})_{i=1,\ldots,d-j}$, so that $\mathcal{T}_{1:j}^{PVC} := \times_{k=1}^{j} \mathcal{T}_k^{PVC}$ collects all copulas of the PVC up to and including the j-th tree. It holds that

$$\forall j = 1, \ldots, d-1: \quad \underset{\mathcal{T}_j \in \mathbf{T}_j}{\arg \min} \; D_{KL}^{(j)}(\mathcal{T}_j(\mathcal{T}_{1:j-1}^{PVC})) = \mathcal{T}_j^{PVC}. \tag{5.2}$$

According to Theorem 5.1, if the true copulas are specified in the first tree, one should choose the first-order partial copulas in the second tree, the second-order partial copulas in the third tree etc. to minimize the KLD tree-by-tree. Theorem 5.1 also remains true if we replace $\mathcal{C}_2$ in the definition of $\mathbf{T}_j$ by the space of absolutely continuous bivariate cdfs. The PVC ensures that random variables in higher trees are uniformly distributed since the resulting random variables in higher trees are higher-order PPITs. If one uses a different approximation, such as the one used by Hobæk Haff et al. (2010) and Stöber et al. (2013), then the random variables in higher trees are not necessarily uniformly distributed and pseudo-copulas (Fermanian and Wegkamp (2012)) can be used to further minimize the KLD. Stöber et al. (2013) note in their appendix that if $C_{1:3}$ is a FGM copula and the copulas in the first tree are correctly specified, then the KLD from the true distribution has an extremum at $C_{13;2}^{\mathrm{svc}} = C^{\perp} = C_{13;2}^{\mathrm{PVC}}$. If $C_{13;2}$ belongs to a parametric family of bivariate copulas whose parameter depends on u_2, then $C_{13;2}^{\mathrm{PVC}}$ is in general not a member of the same copula family with a constant parameter, see Spanhel and Kurz (2016b). Together with Theorem 5.1 it follows that the proposed simplified vine copula approximations of Hobæk Haff et al. (2010) and Stöber et al. (2013) can be improved if the first-order partial copula is chosen in the second tree, and not a copula of the same parametric family as the conditional copula but with a constant dependence parameter such that the KLD is minimized.

Besides its interpretation as generalization of the partial correlation matrix, the PVC can also be interpreted as the SVC that minimizes the KLD tree-by-tree. This sequential minimization neglects that the KLD related to a tree depends on the copulas that are specified in the former trees. For instance, if $d = 3$, the KLD of the first tree $D_{KL}^{(1)}(\mathcal{T}_1)$ is minimized over the copulas $(C_{12}^{\mathrm{svc}}, C_{23}^{\mathrm{svc}})$ in the first tree $\mathcal{T}_1$, but the effect of the chosen copulas in the first tree $\mathcal{T}_1$ on the KLD related to the second tree $D_{KL}^{(2)}(\mathcal{T}_2(\mathcal{T}_1))$ is not taken into account. Therefore, we now analyze whether the PVC also globally minimizes the KLD. Note that specifying the wrong margins in the first tree $\mathcal{T}_1$, e.g., $(C_{12}^{\mathrm{svc}}, C_{23}^{\mathrm{svc}}) \neq (C_{12}, C_{23})$, increases $D_{KL}^{(1)}(\mathcal{T}_1)$ in any case. Thus, without any further investigation, it is

absolutely indeterminate whether the definite increase in $D_{KL}^{(1)}(\mathcal{T}_1)$ can be overcompensated by a possible decrease in $D_{KL}^{(2)}(\mathcal{T}_2(\mathcal{T}_1))$ if another approximation is chosen. The next theorem shows that the PVC is in general not the global minimizer of the KLD.

Theorem 5.2 (Global KLD minimization if $C_{1:d} \in \mathcal{C}_d^{\text{svc}}$ or $C_{1:d} \in \mathcal{C}_d \backslash \mathcal{C}_d^{\text{svc}}$)
If $C_{1:d} \in \mathcal{C}_d^{SVC}$, i.e., the simplifying assumption holds for $C_{1:d}$, then

$$\operatorname*{arg\,min}_{C_{1:d}^{SVC} \in \mathcal{C}_d^{SVC}} D_{KL}(C_{1:d}||C_{1:d}^{SVC}) = C_{1:d}^{PVC}. \tag{5.3}$$

If the simplifying assumption does not hold for $C_{1:d}$, then $C_{1:d}^{PVC}$ might not be a global minimum. That is, $\exists C_{1:d} \in \mathcal{C}_d \backslash \mathcal{C}_d^{SVC}$ such that

$$\operatorname*{arg\,min}_{C_{1:d}^{SVC} \in \mathcal{C}_d^{SVC}} D_{KL}(C_{1:d}||C_{1:d}^{SVC}) \neq C_{1:d}^{PVC}, \tag{5.4}$$

and $\forall (\mathcal{T}_2, \ldots, \mathcal{T}_{d-1}) \in \times_{k=2}^{d-1} \mathbf{T}_k$

$$\operatorname*{arg\,min}_{\tilde{\mathcal{T}}_{1:d-1} \in \mathbf{T}_{1:d-1}} D_{KL}(C_{1:d}||\tilde{\mathcal{T}}_{1:d-1}) \neq (\mathcal{T}_1^{PVC}, \mathcal{T}_2, \ldots, \mathcal{T}_{d-1}). \tag{5.5}$$

Theorem 5.2 states that, if the simplifying assumption does not hold, the KLD may not be minimized by choosing the true copulas in the first tree, first-order partial copulas in the second tree and higher-order partial copulas in the remaining trees (see (5.4)). It follows that, if the objective is the minimization of the KLD, it may not be optimal to specify the true copulas in the first tree, no matter what bivariate copulas are specified in the other trees (see (5.5)). This rather puzzling result can be explained by the fact that, if the simplifying assumption does not hold, then the approximation error of the implicitly modeled bivariate margins is not minimized (see Lemma 4.1). For instance, if $d = 3$, a departure from the true copulas (C_{12}, C_{23}) in the first tree increases the KLD related to the first tree, but it can decrease the KLD of the implicitly modeled margin C_{13}^{svc} from C_{13}. As a result, the increase in $D_{KL}^{(1)}$ can be overcompensated by a larger decrease in $D_{KL}^{(2)}$, so that the KLD can be decreased.

Theorem 5.2 does not imply that the PVC never minimizes the KLD from the true copula. For instance, if $d = 3$ and if $C_{13;2}^{\text{PVC}} = C^{\perp}$, then $D_{KL}(C_{1:3}||C_{1:3}^{\text{PVC}})$ is an extremum, which directly follows from equation (5.2) since

$$\operatorname*{arg\,min}_{\mathcal{T}_1 \in \mathbf{T}_1} D_{KL}(C_{1:3}||(\mathcal{T}_1, (C^{\perp}))) = \operatorname*{arg\,min}_{\mathcal{T}_1 \in \mathbf{T}_1} D_{KL}^{(1)}(\mathcal{T}_1).$$

It is an open problem whether and when the PVC can be the global minimizer of the

KLD. Unfortunately, the simplified vine copula approximation that globally minimizes the KLD is not tractable. However, if the simplified vine copula approximation that minimizes the KLD does not specify the true copulas in the first tree, the random variables in the higher tree are not CPITs. Thus, it is not guaranteed that these random variables are uniformly distributed and we could further decrease the KLD by assigning pseudo-copulas (Fermanian and Wegkamp (2012)) to the edges in the higher trees. It can be easily shown that the resulting best approximation is then a pseudo-copula. Consequently, the best approximation satisfying the simplifying assumption is in general not a SVC but a simplified vine pseudo-copula if one considers the space of regular vines where each edge corresponds to a bivariate cdf.

While the PVC may not be the best approximation in the space of SVCs, it is the best feasible SVC approximation in practical applications. That is because the stepwise specification and estimation of a SVC is also feasible for (very) large dimensions which is not true for a joint specification and estimation. For instance, if all pair-copula families of a parametric vine copula are chosen simultaneously and the selection is done by means of information criteria, we have to estimate $K^{d(d-1)/2}$ different models, where d is the dimension and K the number of possible pair-copula families that can be assigned to each edge. On the contrary, a stepwise procedure only requires the estimation of $Kd(d-1)/2$ models. To illustrate the computational burden, consider the `R`-package `VineCopula` (Schepsmeier et al. 2016) where $K = 40$. For this number of pair-copula families, a joint specification requires the estimation of 64,000 ($d = 3$) or more than four billion ($d = 4$) models whereas only 120 ($d = 3$) or 240 ($d = 4$) models are needed for a stepwise specification. For many non-parametric estimation approaches (kernels (Nagler and Czado 2016), empirical distributions (Hobæk Haff and Segers 2015)), only the sequential estimation of a SVC is possible. The only exception is the spline-based approach of Kauermann and Schellhase (2014). However, due to the large number of parameters and the resulting computational burden, a joint estimation is only feasible for $d \leq 5$ (Kauermann et al. 2013).

6 Convergence to the partial vine copula

If the data generating process satisfies the simplifying assumption, consistent stepwise procedures for the specification and estimation of parametric and non-parametric simplified vine copula models asymptotically minimize the KLD from the true copula. Theorem 5.1 implies that this is not true in general if the data generating process does not satisfy the simplifying assumption. An implication of this result for the application of SVCs is pointed out in the next corollary.

Corollary 6.1

Denote the sample size by N. Let $C_{1:d} \in \mathcal{C}_d$ be the data generating copula and $C_{1:d}^{SVC}(\theta) \in \mathcal{C}_d^{SVC}$, $\theta \in \Theta$, be a parametric SVC so that $\exists_1 \theta_{PVC} \in \Theta : C_{1:d}^{SVC}(\theta_{PVC}) = C_{1:d}^{PVC}$. The pseudo-true parameters which minimize the KLD from the true distribution are assumed to exist (see White (1982) for sufficient conditions) and denoted by

$$\theta^\star = \arg\min_{\theta \in \Theta} D_{KL}(C_{1:d} || C_{1:d}^{SVC}(\theta)).$$

Let $\hat{\theta}_S$ denote the (semi-parametric) step-by-step ML estimator and $\hat{\theta}_J$ denote the (semi-parametric) joint ML estimator defined in Hobæk Haff (2012, 2013). Under regularity conditions (e.g., Condition 1 and Condition 2 in (Spanhel and Kurz 2016a)) and for $N \to \infty$, it holds that:

(i) $\hat{\theta}^S \xrightarrow{p} \theta^{PVC}$.

(ii) $\hat{\theta}^J \xrightarrow{p} \theta^\star$.

(iii) $\exists C_{1:d} \in \mathcal{C}_d \backslash \mathcal{C}_d^{SVC}$ such that $\hat{\theta}^S \not\xrightarrow{p} \theta^\star$.

Corollary 6.1 shows that the step-by-step and joint ML estimator may not converge to the same limit (in probability) if the simplifying assumption does not hold for the data generating vine copula. For this reason, we investigate in the following the difference between the step-by-step and joint ML estimator in finite samples. Note that the convergence of kernel-density estimators to the PVC has been recently established by Nagler and Czado (2016). However, in this case, only a sequential estimation of a simplified vine copula is possible and thus the best feasible approximation in the space of simplified vine copulas is given by the PVC.

6.1 Difference between step-by-step and joint ML estimates

We compare the step-by-step and the joint ML estimator under the assumption that the pair-copula families of the PVC are specified for the parametric vine copula model. For this purpose, we simulate data from two three-dimensional copulas $C_{1:3}$ with sample sizes $N = 500, 2500, 25000$, perform a step-by-step and joint ML estimation, and repeat this 1000 times. For ease of exposition and because the qualitative results are not different, we consider copulas where $C_{12} = C_{23}$ and only present the estimates for $(\theta_{12}, \theta_{13;2})$.

Example 6.1 (PVC of the Frank copula)

Let $C^{Fr}(\theta)$ denote the bivariate Frank copula with dependence parameter θ and $C^{P\text{-}Fr}(\theta)$ be the partial Frank copula (Spanhel and Kurz 2016b) with dependence parameter θ. Let

$C_{1:3}$ be the true copula with $(C_{12}, C_{23}, C_{13;2}) = (C^{Fr}(5.74), C^{Fr}(5.74), C^{P\text{-}Fr}(5.74))$, i.e., $C_{1:3} = C_{1:3}^{PVC}$, and $C_{1:3}^{SVC}(\theta) = (C^{Fr}(\theta_{12}), C^{Fr}(\theta_{23}), C^{P\text{-}Fr}(\theta_{13;2}))$ be the parametric SVC that is fitted to data generated from $C_{1:3}$.
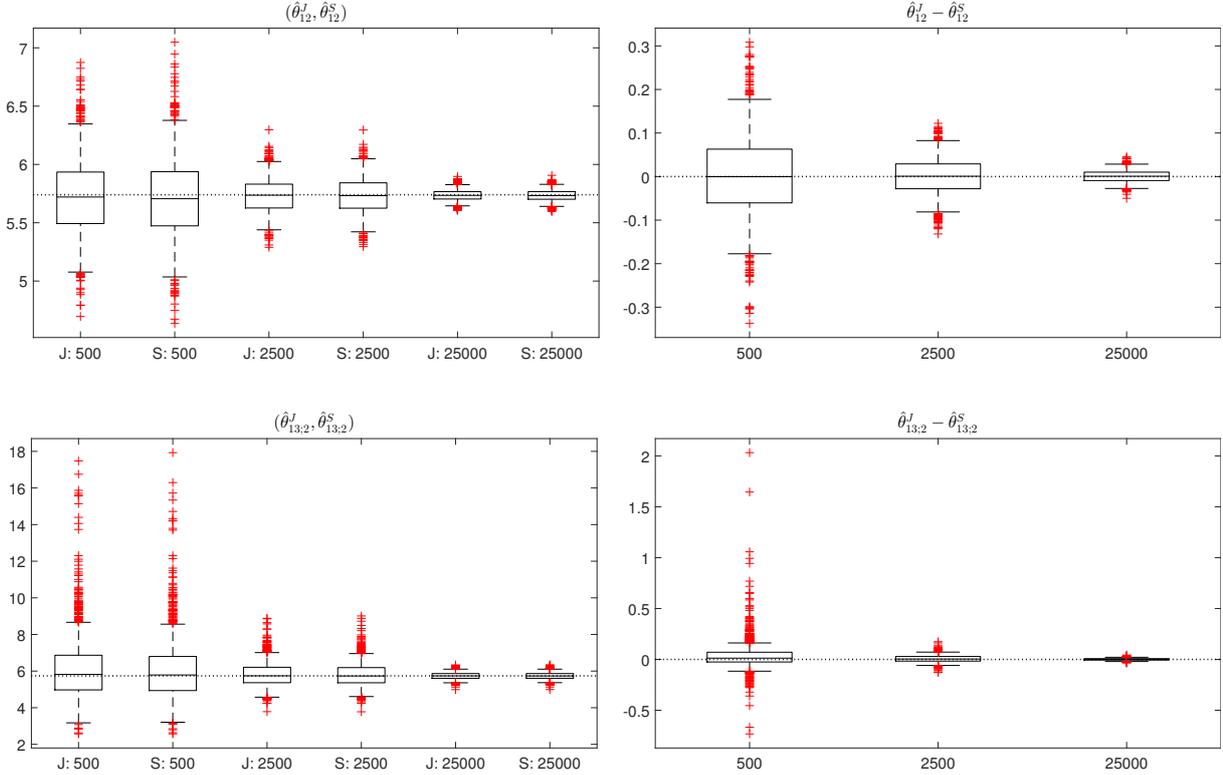
Figure 3: Box plots of joint (J) and sequential (S) ML estimates and their difference for sample sizes $N = 500, 2500, 25000$, if the data is generated from $C_{1:3}$ in Example 6.1 and the pair-copula families of the SVC are given by the corresponding PVC. The dotted line indicates the pseudo-true parameter and zero, respectively. The end of the whiskers is 0.953 times the inter-quartile range, corresponding to approximately 95% coverage if the data is generated by a normal distribution.

Example 6.1 presents a data generating process which satisfies the simplifying assumption, implying $\theta^{\text{PVC}} = \theta^\star$. It is the PVC of the three-dimensional Frank copula with Kendall's τ approximately equal to 0.5. Figure 3 shows the corresponding box plots of joint and step-by-step ML estimates and their difference. The left panel confirms the results of Hobæk Haff (2012, 2013). Although the joint ML estimator is more efficient, the loss in efficiency for the step-by-step ML estimator is negligible and both estimators converge to the true parameter value. Moreover, the right panel of Figure 3 shows that the difference between joint and step-by-step ML estimates is never statistically significant at a 5% level. Since the computational time for a step-by-step ML estimation is much lower than for a joint ML estimation (Hobæk Haff 2012), the step-by-step ML estimator is very attractive for estimating high-dimensional vine copulas that satisfy the simplifying assump-

tion. Moreover, the step-by-step ML estimator is then inherently suited for selecting the pair-copula families in a stepwise manner. However, if the simplifying assumption does not hold for the data generating vine copula, the step-by-step and joint ML estimator can converge to different limits (Corollary 6.1), as the next example demonstrates.

Example 6.2 (Frank copula)
Let $C_{1:3}$ be the Frank copula with dependence parameter $\theta = 5.74$, i.e., $C_{1:3} \neq C_{1:3}^{PVC}$, and $C_{1:3}^{SVC} = (C^{Fr}(\theta_{12}), C^{Fr}(\theta_{23}), C^{P\text{-}Fr}(\theta_{13;2}))$ be the parametric SVC that is fitted to data generated from $C_{1:3}$.
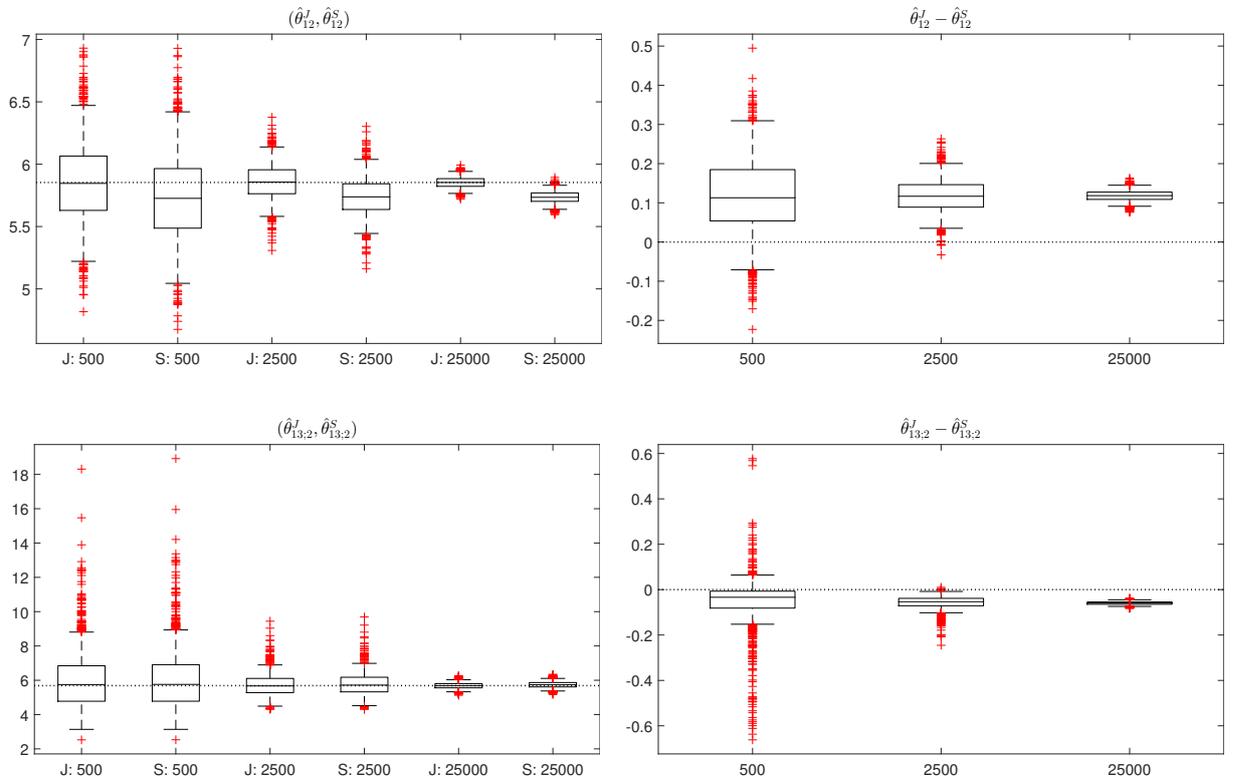
Figure 4: Box plots of joint (J) and sequential (S) ML estimates and their difference for sample sizes $N = 500, 2500, 25000$, if the data is generated from $C_{1:3}$ in Example 6.2 and the pair-copula families of the SVC are given by the corresponding PVC. The dotted line indicates the pseudo-true parameter and zero, respectively. The end of the whiskers is 0.953 times the inter-quartile range, corresponding to approximately 95% coverage if the data is generated by a normal distribution.

Example 6.2 is identical to Example 6.1, with the only difference that the conditional copula is varying in such a way that the resulting three-dimensional copula is a Frank copula. Although the Frank copula does not satisfy the simplifying assumption, it is pretty close to a copula for which the simplifying assumption holds, because the variation in the conditional copula is strongly limited for many Archimedean copulas (Mesfioui and Quessy (2008)). Nevertheless, the right panel of Figure 4 shows that the step-by-step and joint ML estimates

for θ_{12} are significantly different at the 5% level if the sample size is 2500 observations. The difference between step-by-step and joint ML estimates for $\theta_{13;2}$ is less pronounced, but also highly significant for sample sizes with 2500 observations or more. Thus, only in Example 6.1 the step-by-step ML estimator is a consistent estimator of a simplified vine copula model that minimizes the KLD from the underlying copula, whereas the joint ML estimator is a consistent minimizer in both examples. A third example where the distance between the data generating copula and the PVC and thus the difference between the step-by-step and joint ML estimates is more pronounced is given in Appendix A.9.

7 Conclusion

We introduced the partial vine copula (PVC) which is a particular simplified vine copula that coincides with the data generating copula if the simplifying assumption holds. The PVC can be regarded as a generalization of the partial correlation matrix where partial correlations are replaced by j-th order partial copulas. Consequently, it provides a new dependence measure of a d-dimensional distribution in terms of $d(d-1)/2$ bivariate unconditional copulas. While a higher-order partial copula of the PVC is related to the partial copula, it does not suffer from the curse of dimensionality and can be estimated for high-dimensional data (Nagler and Czado 2016). We analyzed to what extent the dependence structure of the underlying distribution is reproduced by the PVC. In particular, we showed that a pair of random variables may be considered as conditionally (in)dependent according to the PVC although this is not the case for the data generating process.

We also revealed the importance of the PVC for the modeling of high-dimensional distributions by means of simplified vine copulas (SVCs). Up to now, the estimation of SVCs has almost always been based on the assumption that the data generating process satisfies the simplifying assumption. Moreover, the implications that follow if the simplifying assumption is not true have not been investigated. We showed that the PVC is the SVC approximation that minimizes the Kullback-Leibler divergence in a stepwise fashion. Since almost all estimators of SVCs proceed sequentially, it follows that, under regularity conditions, many estimators of SVCs converge to the PVC also if the simplifying assumption does not hold. However, we also proved that the PVC may not minimize the Kullback-Leibler divergence from the true copula and thus may not be the best SVC approximation in theory. Nevertheless, due to the prohibitive computational burden or simply because only a stepwise model specification and estimation is possible, the PVC is the best feasible SVC approximation in practice.

The analysis in this paper showed the relative optimality of the PVC when it comes

to approximating multivariate distributions by SVCs. Obviously, it is easy to construct (theoretical) examples where the PVC does not provide a good approximation in absolute terms. But such examples do not provide any information about the appropriateness of the simplifying assumption in practice. To investigate whether the simplifying assumption is true and the PVC is a good approximation in applications, one can use Lemma 3.1 to develop tests for the simplifying assumption, see Kurz and Spanhel (2017). Moreover, even in cases where the simplifying assumption is strongly violated, an estimator of the PVC can yield an approximation that is superior to competing approaches. Recently, it has been demonstrated in Nagler and Czado (2016) that the structure of the PVC can be used to obtain a constrained kernel-density estimator that can be much closer to the data generating process than the classical unconstrained kernel-density estimator, even if the distance between the PVC and the data generating copula is large.

Acknowledgements

References for Chapter III

Aas, K., and D. Berg. 2009. "Models for construction of multivariate dependence – a comparison study". *The European Journal of Finance* 15 (7-8): 639–659.

Aas, K., C. Czado, A. Frigessi, and H. Bakken. 2009. "Pair-Copula Constructions of Multiple Dependence". *Insurance: Mathematics and Economics* 44 (2): 182–198.

Acar, E. F., C. Genest, and J. Nešlehová. 2012. "Beyond simplified pair-copula constructions". *Journal of Multivariate Analysis* 110: 74–90.

Bedford, T., and R. M. Cooke. 2002. "Vines – A New Graphical Model for Dependent Random Variables". *The Annals of Statistics* 30 (4): 1031–1068.

Bergsma, W. P. 2004. "Testing conditional independence for continuous random variables". URL: https://www.eurandom.tue.nl/reports/2004/048-report.pdf.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka. 2013. "Selecting and estimating regular vine copulae and application to financial returns". *Computational Statistics & Data Analysis* 59: 52–69.

Fermanian, J.-D., and M. H. Wegkamp. 2012. "Time-dependent copulas". *Journal of Multivariate Analysis* 110: 19–29.

Fischer, M., C. Köck, S. Schlüter, and F. Weigert. 2009. "An empirical analysis of multivariate copula models". *Quantitative Finance* 9 (7): 839–854.

Gijbels, I., M. Omelka, and N. Veraverbeke. 2015a. "Estimation of a Copula when a Covariate Affects only Marginal Distributions". *Scandinavian Journal of Statistics* 42 (4): 1109–1126.

— . 2015b. "Partial and average copulas and association measures". *Electronic Journal of Statistics* 9 (2): 2420–2474.

Grothe, O., and S. Nicklas. 2013. "Vine constructions of Lévy copulas". *Journal of Multivariate Analysis* 119: 1–15.

Hobæk Haff, I. 2012. "Comparison of estimators for pair-copula constructions". *Journal of Multivariate Analysis* 110: 91–105.

— . 2013. "Parameter estimation for pair-copula constructions". *Bernoulli* 19 (2): 462–491.

Hobæk Haff, I., and J. Segers. 2015. "Nonparametric estimation of pair-copula constructions with the empirical pair-copula". *Computational Statistics & Data Analysis* 84: 1–13.

Hobæk Haff, I., K. Aas, and A. Frigessi. 2010. "On the simplified pair-copula construction – Simply useful or too simplistic?" *Journal of Multivariate Analysis* 101 (5): 1296–1310.

Joe, H. 1996. "Families of m-Variate Distributions with Given Margins and m(m-1)/2 Bivariate Dependence Parameters". *Lecture Notes-Monograph Series* 28: 120–141.

— . 1997. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Joe, H., H. Li, and A. K. Nikoloulopoulos. 2010. "Tail dependence functions and vine copulas". *Journal of Multivariate Analysis* 101 (1): 252–270.

Kauermann, G., and C. Schellhase. 2014. "Flexible pair-copula estimation in D-vines using bivariate penalized splines". *Statistics and Computing* 24 (6): 1081–1100.

Kauermann, G., C. Schellhase, and D. Ruppert. 2013. "Flexible Copula Density Estimation with Penalized Hierarchical B-splines". *Scandinavian Journal of Statistics* 40 (4): 685–705.

Krupskii, P., and H. Joe. 2013. "Factor copula models for multivariate data". *Journal of Multivariate Analysis* 120: 85–101.

Kurowicka, D., and R. Cooke. 2006. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Chichester: Wiley.

Kurowicka, D., and H. Joe, eds. 2011. *Dependence Modeling*. Singapore: World Scientific.

Kurz, M. S., and F. Spanhel. 2017. "Testing the simplifying assumption in high-dimensional vine copulas". ArXiv e-prints `arXiv:1706.02338`.

McNeil, A. J., R. Frey, and P. Embrechts. 2005. *Quantitative Risk Management*. Princeton series in finance. Princeton NJ: Princeton Univ. Press.

Mesfioui, M., and J.-F. Quessy. 2008. "Dependence structure of conditional Archimedean copulas". *Journal of Multivariate Analysis* 99 (3): 372–385.

Nagler, T., and C. Czado. 2016. "Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas". *Journal of Multivariate Analysis* 151: 69–89.

Nelsen, R. B. 2006. *An Introduction to Copulas*. Springer series in statistics. New York, NY: Springer.

Nikoloulopoulos, A. K., H. Joe, and H. Li. 2012. "Vine copulas with asymmetric tail dependence and applications to financial return data". *Computational Statistics & Data Analysis* 56 (11): 3659–3673.

Patton, A. J. 2006. "Modelling Asymmetric Exchange Rate Dependence". *International Economic Review* 47 (2): 527–556.

Portier, F., and J. Segers. 2015. "On the weak convergence of the empirical conditional copula under a simplifying assumption". ArXiv e-prints `arXiv:1511.06544`.

Remillard, B. 2013. *Statistical Methods for Financial Engineering*. Boca Raton, FL: CRC Press.

Schepsmeier, U., J. Stoeber, E. C. Brechmann, B. Graeler, T. Nagler, and T. Erhardt. 2016. *VineCopula: Statistical Inference of Vine Copulas*. URL: `https://CRAN.R-project.org/package=VineCopula`, R package version 2.0.5.

Spanhel, F., and M. S. Kurz. 2016a. "Estimating parametric simplified vine copulas under general conditions". Unpublished working paper.

— . 2016b. "The partial copula: Properties and associated dependence measures". *Statistics & Probability Letters* 119: 76–83.

Stöber, J., H. Joe, and C. Czado. 2013. "Simplified pair copula constructions – Limitations and extensions". *Journal of Multivariate Analysis* 119: 101–118.

White, H. 1982. "Maximum Likelihood Estimation of Misspecified Models". *Econometrica* 50 (1): 1–25.

Appendix

A.1 Proof of Lemma 3.1

$U_{k|S_{ij}}^{\text{PVC}} = U_{k|S_{ij}}$ (a.s.) $\Rightarrow U_{k|S_{ij}}^{\text{PVC}} \perp U_{S_{ij}}$ is true because $U_{k|S_{ij}}$ is a CPIT. For the converse, let $A := \times_{k=i+1}^{i+j-1}[0, u_k]$ and consider

$$\mathbb{P}(U_{k|S_{ij}}^{\text{PVC}} \leq a, U_{S_{ij}} \leq u_{S_{ij}}) = \int_A F_{k|S_{ij}}\big((F_{k|S_{ij}}^{\text{PVC}})^{-1}(a|t_{S_{ij}})|t_{S_{ij}}\big)\mathrm{d}C_{S_{ij}}(t_{S_{ij}}). \quad (\text{A.1})$$

Since $U_{k|S_{ij}}^{\text{PVC}} \sim U(0, 1)$ it follows that if $U_{k|S_{ij}}^{\text{PVC}} \perp U_{S_{ij}}$ then $\mathbb{P}(U_{k|S_{ij}}^{\text{PVC}} \leq a, U_{S_{ij}} \leq u_{S_{ij}}) = aC_{S_{ij}}(u_{S_{ij}})$ for all $(a, u_{S_{ij}}) \in [0, 1]^j$. This implies that

$$\mathbb{P}(U_{k|S_{ij}}^{\text{PVC}} \leq a, U_{S_{ij}} \leq u_{S_{ij}}) = \int_A a\, \mathrm{d}C_{S_{ij}}(t_{S_{ij}})$$

equals the right hand side of (A.1) for all $(a, u_{S_{ij}}) \in [0, 1]^j$. It follows that the integrands must be identical (a.s.) as well and $F_{k|S_{ij}}(F_{k|S_{ij}}^{\text{PVC}})^{-1}(a|t_{S_{ij}}) = a$ for all $a \in [0, 1]$ and almost every $t_{S_{ij}} \in [0, 1]^{j-1}$. Thus $F_{k|S_{ij}} = F_{k|S_{ij}}^{\text{PVC}}$ (a.s.) which is equivalent to $U_{k|S_{ij}}^{\text{PVC}} = U_{k|S_{ij}}$ (a.s.).

A.2 Proof of Lemma 4.1

Let $C_{1:3}^{\star} \in \mathcal{C}_3^{\text{SVC}}$ be the SVC given in Example 4.1. We define $C_{1:d}$ as follows. Let $C_{1,d-1;2:d-2} = C_{12}^{\star}, C_{2,d;3:d-1} = C_{23}^{\star}, C_{1,d;2:d-1} = D_{1,3;2}^{\star}$, where $D_{1,3;2}^{\star}$ is the corresponding conditional copula in Example 4.1 and $C_{i,i+j;S_{ij}} = E_{k,l} \in \mathcal{C}_2, (k, l) \in \mathcal{I}_1^d$ means that $C_{i,i+j;S_{ij}}(a, b|u_{S_{ij}}) = E_{k,l}(a, b)$ for all $(a, b, u_{S_{ij}}) \in [0, 1]^{j+1}$. Moreover, let $C_{i,i+j;S_{ij}} = C^{\perp}$ for $(i, j) \in \mathcal{I}_1^d \backslash \{(1, d-2), (2, d-2), (1, d-1)\}$. The conclusion now follows from Example 4.1.

A.3 Proof of Theorem 4.1

W.l.o.g. assume that the margins of $X_{1:d}$ are uniform. Let $C^{FGM_3}(u_{1:3}; \theta) = \prod_{i=1}^3 u_i + \theta \prod_{i=1}^3 u_i(1 - u_i), |\theta| \leq 1$, be the three-dimensional FGM copula, $d \geq 4$, and $(i, j) \in \mathcal{I}_1^d$. It is obvious that $C_{i,i+2;i+1} = C^{\perp} \Rightarrow C_{i,i+2;i+1}^{\text{PVC}} = C^{\perp}$ is true. Let $J \in \{2, \ldots, d-2\}$ be fixed. Assume that $C_{1:d}$ has the following D-vine copula representation of the non-simplified form

$$C_{1,1+J;2:J} = \partial_3 C^{FGM_3}(u_1, u_{1+J}, u_2; 1)$$
$$C_{2,2+J;3:J+1} = \partial_3 C^{FGM_3}(u_2, u_{2+J}, u_{1+J}; 1)$$

and $C_{i,i+j;S_{i,j}} = C^\perp$ for all other $(i,j) \in \mathcal{I}_1^d$. Using the same arguments as in the proof of Lemma 4.2 we obtain

$$
\begin{aligned}
C_{i,i+J;i+1:J-1}^{\text{PVC}} &= C^\perp, \quad i = 1, 2, \\
C_{1,2+J;2:J+1}^{\text{PVC}} &= C^{FGM_2}(1/9).
\end{aligned}
$$

This proves that $C_{i,i+2;i+1} = C^\perp \Leftarrow C_{i,i+2;i+1}^{\text{PVC}} = C^\perp$ is not true in general and that, for $j \geq 3$, neither the statement $C_{i,i+j;S_{ij}} = C^\perp \Rightarrow C_{i,i+j;S_{ij}}^{\text{PVC}} = C^\perp$ nor the statement $C_{i,i+j;S_{ij}} = C^\perp \Leftarrow C_{i,i+j;S_{ij}}^{\text{PVC}} = C^\perp$ is true in general.

A.4 Proof of Lemma 4.2

We show a more general result and set $C_{i,i+2;i+1}(u_i, u_{i+2}|u_{i+1}) = C^{FGM_2}(u_i, u_{i+2}; g(u_{i+1}))$ in (4.3) where $g: [0,1] \to [-1,1]$ is a non-constant measurable function such that

$$
\forall u \in [0.5, 1]: g(0.5 + u) = -g(0.5 - u). \tag{A.2}
$$

For $i = 1, 2, 3$, the copula in the second tree of the PVC is given by

$$
\begin{aligned}
C_{i,i+2;i+1}^{\text{PVC}}(a, b) &= \mathbb{P}(U_{i|i+1} \leq a, U_{i+2|i+1} \leq b) = \int_{[0,1]} C_{i,i+2;i+1}(a, b|u_{i+1}) du_{i+1} \\
&\overset{(4.3)}{=} ab\big(1 + (1-a)(1-b) \int_{[0,1]} g(u_{i+1}) du_{i+1}\big) \overset{(A.2)}{=} ab, \tag{A.3}
\end{aligned}
$$

which is the independence copula. For $i = 1, 2, k = i, i + 3$, the true CPIT of U_k w.r.t. $U_{i+1:i+2}$ is a function of $U_{i+1:i+2}$ because

$$
U_{i|i+1:i+2} = U_i[1 + g(U_{i+1})(1 - U_i)(1 - 2U_{i+2})], \tag{A.4}
$$

$$
U_{i+3|i+1:i+2} = U_{i+3}[1 + g(U_{i+2})(1 - U_{i+3})(1 - 2U_{i+1})]. \tag{A.5}
$$

However, for $i = 1, 2, k = i, i + 3$, the PPIT of U_k w.r.t. $U_{i+1:i+2}$ is not a function of $U_{i+1:i+2}$ because

$$
\begin{aligned}
U_{i|i+1:i+2}^{\text{PVC}} = F_{i|i+1:i+2}^{\text{PVC}}(U_i|U_{i+1:i+2}) &= F_{U_{i|i+1}|U_{i+2|i+1}}(U_{i|i+1}|U_{i+2|i+1}) \\
&= \partial_2 C_{i,i+2;i+1}^{\text{PVC}}(U_{i|i+1}, U_{i+2|i+1}) \overset{(A.3)}{=} U_{i|i+1} \overset{(4.2)}{=} U_i, \tag{A.6}
\end{aligned}
$$

and, by symmetry,

$$
U_{i+3|i+1:i+2}^{\text{PVC}} = U_{i+3}. \tag{A.7}
$$

For $i = 1, 2$, the joint distribution of these first-order PPITs is a copula in the third tree of the PVC which is given by

$$
\begin{aligned}
C_{i,i+3;i+1:i+2}^{\text{PVC}}(a, b) &= \mathbb{P}(U_{i|i+1:i+2}^{\text{PVC}} \leq a, U_{i+3|i+1:i+2}^{\text{PVC}} \leq b) \\
&\overset{\text{(A.6),(A.7)}}{=} \quad \mathbb{P}(U_i \leq a, U_{i+3} \leq b) = C_{i,i+3}(a, b) \\
&\overset{\text{(4.4)}}{=} \quad \int_{[0,1]^2} F_{i|i+1:i+2}(a|u_{i+1:i+2}) F_{i+3|i+1:i+2}(b|u_{i+1:i+2}) \mathrm{d}u_{i+1:i+2} \\
&\overset{\text{(A.4),(A.5)}}{=} \quad ab\bigg[1 + (1-a)(1-b) \int_{[0,1]} g(u_{i+1})(1 - 2u_{i+1}) \mathrm{d}u_{i+1} \\
&\qquad \times \int_{[0,1]} g(u_{i+2})(1 - 2u_{i+2}) \mathrm{d}u_{i+2}\bigg] \\
&= ab[1 + \theta(1-a)(1-b)] = C^{FGM_2}(\theta),
\end{aligned}
\tag{A.8}
$$

where $\theta := 4(\int_{[0,1]} u g(u) \mathrm{d}u)^2 > 0$, by the properties of g. Thus, a copula in the third tree of the PVC is a bivariate FGM copula whereas the true conditional copula is the independence copula.

The CPITs of U_1 or U_5 w.r.t. $U_{2:4}$ are given by

$$
\begin{aligned}
U_{1|2:4} = F_{1|2:4}(U_1|U_{2:4}) = \partial_2 C_{14;2:3}(U_{1|2:3}, U_{4|2:3}|U_{2:3}) &\overset{\text{(4.4)}}{=} U_{1|2:3} \\
&\overset{\text{(A.4)}}{=} U_1[1 + g(U_2)(1 - U_1)(1 - 2U_3)],
\end{aligned}
\tag{A.9}
$$

$$
U_{5|2:4} = U_5[1 + g(U_4)(1 - U_5)(1 - 2U_3)],
\tag{A.10}
$$

whereas the corresponding second-order PPITs are given by

$$
\begin{aligned}
U_{1|2:4}^{\text{PVC}} = F_{1|2:4}^{\text{PVC}}(U_1|U_{2:4}) = F_{U_{1|2:3}^{\text{PVC}}|U_{4|2:3}^{\text{PVC}}}(U_{1|2:3}^{\text{PVC}}|U_{4|2:3}^{\text{PVC}}) &\overset{\text{(A.6),(A.7)}}{=} F_{U_1|U_4}(U_1|U_4) \\
= U_{1|4} = \partial_2 C_{14}(U_1, U_4) &\overset{\text{(A.8)}}{=} U_1[1 + \theta(1 - U_1)(1 - 2U_4)],
\end{aligned}
\tag{A.11}
$$

$$
U_{5|2:4}^{\text{PVC}} = U_5[1 + \theta(1 - U_5)(1 - 2U_2)].
\tag{A.12}
$$

For the copula in the fourth tree of the PVC it holds

$$
\begin{aligned}
C_{15;2:4}^{\text{PVC}}(a, b) &= \mathbb{P}(U_{1|2:4}^{\text{PVC}} \leq a, U_{5|2:4}^{\text{PVC}} \leq b) \overset{\text{(A.11),(A.12)}}{=} \mathbb{P}(U_{1|4} \leq a, U_{5|2} \leq b) \\
&= \mathbb{P}(U_1 \leq F_{1|4}^{-1}(a|U_4), U_5 \leq F_{5|2}^{-1}(b|U_2)) \\
&= \int_{[0,1]^3} F_{15|2:4}(F_{1|4}^{-1}(a|u_4), F_{5|2}^{-1}(b|u_2)|u_{2:4}) c_{2:4}(u_{2:4}) \mathrm{d}u_{2:4} \\
&= \int_{[0,1]^3} C_{15;2:4}(F_{1|2:4}(F_{1|4}^{-1}(a|u_4)|u_{2:4}), F_{5|2:4}(F_{5|2}^{-1}(b|u_2)|u_{2:4})|u_{2:4}) c_{2:4}(u_{2:4}) \mathrm{d}u_{2:4}
\end{aligned}
$$

$$\overset{(4.5)}{=} \int_{[0,1]^3} F_{1|2:4}(F_{1|4}^{-1}(a|u_4)|u_{2:4}) F_{5|2:4}(F_{5|2}^{-1}(b|u_2)|u_{2:4}) c_{2:4}(u_{2:4}) du_{2:4}$$

$$\overset{(A.9),(A.10)}{=} \int_{[0,1]^3} F_{1|2:3}(F_{1|4}^{-1}(a|u_4)|u_{2:3}) F_{5|3:4}(F_{5|2}^{-1}(b|u_2)|u_{3:4}) c_{2:4}(u_{2:4}) du_{2:4}$$

$$\overset{(A.4),(A.5)}{=} \int_{[0,1]^3} F_{1|4}^{-1}(a|u_4)[1 + g(u_2)(1 - F_{1|4}^{-1}(a|u_4))(1 - 2u_3)]$$

$$\times F_{5|2}^{-1}(b|u_2)[1 + g(u_4)(1 - F_{5|2}^{-1}(b|u_2))(1 - 2u_3)]$$

$$\times [1 + g(u_3)(1 - 2u_2)(1 - 2u_4)] du_{2:4}$$

$$= \int_{[0,1]^2} F_{1|4}^{-1}(a|u_4) F_{5|2}^{-1}(b|u_2)$$

$$\times \Bigg[1 + \int_{[0,1]} (1 - 2u_3)^2 du_3 (1 - F_{1|4}^{-1}(a|u_4))(1 - F_{5|2}^{-1}(b|u_2)) g(u_4) g(u_2)$$

$$+ \int_{[0,1]} (1 - 2u_3) g(u_3) du_3 (1 - F_{5|2}^{-1}(b|u_2))(1 - 2u_2)(1 - 2u_4) g(u_4)$$

$$+ \int_{[0,1]} (1 - 2u_3) g(u_3) du_3 (1 - F_{1|4}^{-1}(a|u_4))(1 - 2u_4)(1 - 2u_2) g(u_2) \Bigg] du_2 du_4,$$

where we used that $\int_{[0,1]}(1-2u_3)du_3 = 0$, $\int_{[0,1]} g(u_3)du_3 \overset{(A.2)}{=} 0$ and $\int_{[0,1]}(1-2u_3)^2 g(u_3)du_3 \overset{(A.2)}{=} 0$. By setting $\gamma := -2\int_{[0,1]} u g(u) du$ we can write the copula function as

$$C_{15;2:4}^{\text{PVC}}(a,b) = \int_{[0,1]^2} F_{1|4}^{-1}(a|u_4) F_{5|2}^{-1}(b|u_2) \Bigg[1 + \frac{1}{3}(1 - F_{1|4}^{-1}(a|u_4))(1 - F_{5|2}^{-1}(b|u_2)) g(u_4) g(u_2)$$

$$+ \gamma(1 - F_{5|2}^{-1}(b|u_2))(1 - 2u_2)(1 - 2u_4) g(u_4)$$

$$+ \gamma(1 - F_{1|4}^{-1}(a|u_4))(1 - 2u_4)(1 - 2u_2) g(u_2) \Bigg] du_2 du_4$$

$$= \left(\int_0^1 F_{1|4}^{-1}(a|u_4) du_4 \right) \left(\int_0^1 F_{5|2}^{-1}(b|u_2) du_2 \right)$$

$$+ \gamma \left(\int_0^1 (1 - 2u_2)(F_{5|2}^{-1}(b|u_2) - (F_{5|2}^{-1}(b|u_2))^2) du_2 \right)$$

$$\times \left(\int_0^1 (1 - 2u_4) g(u_4) F_{1|4}^{-1}(a|u_4) du_4 \right)$$

$$+ \gamma \left(\int_0^1 (1 - 2u_4)(F_{1|4}^{-1}(a|u_4) - (F_{1|4}^{-1}(a|u_4))^2) du_4 \right)$$

$$\times \left(\int_0^1 (1 - 2u_2) g(u_2) F_{5|2}^{-1}(b|u_2) du_2 \right)$$

$$+ \frac{1}{3} \left(\int_0^1 g(u_4)(F_{1|4}^{-1}(a|u_4) - (F_{1|4}^{-1}(a|u_4))^2) du_4 \right)$$

$$\times \left(\int_0^1 g(u_2)(F_{5|2}^{-1}(b|u_2) - (F_{5|2}^{-1}(b|u_2))^2) \mathrm{d}u_2 \right).$$

If $(U, V) \sim C^{FGM_2}(\theta)$, the quantile function is given by (cf. Remillard (2013))

$$F_{U|V}^{-1}(u|v) = \frac{1 + h(v) - \sqrt{(1 + h(v))^2 - 4h(v)u}}{2h(v)},$$

with $h(v) := \theta(1 - 2v)$, which implies

$$\frac{\partial}{\partial u} F_{U|V}^{-1}(u|v) = \frac{1}{\sqrt{(1 + h(v))^2 - 4h(v)u}} =: G(u, v), \tag{A.13}$$

and

$$\frac{\partial}{\partial u} (F_{U|V}^{-1}(u|v))^2 = \frac{1}{h(v)} \left[(1 + h(v))G(u, v) - 1 \right]. \tag{A.14}$$

For the density of the copula in the fourth tree of the PVC it follows

$$
\begin{aligned}
c_{15;2:4}^{\text{PVC}}(a, b) &= \frac{\partial^2}{\partial a \partial b} C_{15;2:4}^{\text{PVC}}(a, b) \\
&\overset{(A.13),(A.14)}{=} \left(\int_0^1 G(a, u_4) \mathrm{d}u_4 \right) \left(\int_0^1 G(b, u_2) \mathrm{d}u_2 \right) \\
&\quad + \frac{1}{\gamma} \left(1 - \int_0^1 G(b, u_2) \mathrm{d}u_2 \right) \left(\int_0^1 (1 - 2u_4)g(u_4)G(a, u_4) \mathrm{d}u_4 \right) \\
&\quad + \frac{1}{\gamma} \left(1 - \int_0^1 G(a, u_4) \mathrm{d}u_4 \right) \left(\int_0^1 (1 - 2u_2)g(u_2)G(b, u_2) \mathrm{d}u_2 \right) \\
&\quad + \frac{1}{3} \left(\int_0^1 \frac{g(u_4)}{h(u_4)} \left[1 - G(a, u_4) \right] \mathrm{d}u_4 \right) \left(\int_0^1 \frac{g(u_2)}{h(u_2)} \left[1 - G(b, u_2) \right] \mathrm{d}u_2 \right) \\
&= \frac{1}{4\theta^2} \log(\sigma(a)) \log(\sigma(b)) \\
&\quad + \frac{1}{\gamma} \left(1 - \frac{1}{2\theta} \log(\sigma(b)) \right) \left(\int_0^1 (1 - 2u_4)g(u_4)G(a, u_4) \mathrm{d}u_4 \right) \\
&\quad + \frac{1}{\gamma} \left(1 - \frac{1}{2\theta} \log(\sigma(a)) \right) \left(\int_0^1 (1 - 2u_2)g(u_2)G(b, u_2) \mathrm{d}u_2 \right) \\
&\quad + \frac{1}{3} \left(\int_0^1 \frac{g(u_4)}{h(u_4)} \left[1 - G(a, u_4) \right] \mathrm{d}u_4 \right) \left(\int_0^1 \frac{g(u_2)}{h(u_2)} \left[1 - G(b, u_2) \right] \mathrm{d}u_2 \right),
\end{aligned}
$$

where

$$\sigma(i) = \left(\frac{\sqrt{(1+\theta)^2 - 4\theta i} + 1 - 2i + \theta}{\sqrt{(1-\theta)^2 + 4\theta i} + 1 - 2i - \theta} \right) \quad \text{for } i \in \{a, b\}.$$

If we set $g(u) := 1 - 2u$, then $\theta = 1/9$ and $\gamma = 1/3$, and we get

$$c_{15;2:4}^{\text{PVC}}(a, b) = \frac{81}{4} \prod_{i=a,b} \log(s(i)) + 27 \prod_{i=a,b} \left(1 - \frac{81}{4} \log(s(i)) \right)$$

$$+ \frac{2187}{2} \sum_{(i,j) \in I_{a,b}} \left(1 - \frac{9}{2} \log(s(i)) \right) \left[(6j^2 - 6j + 1) \log(s(j)) \right.$$

$$\left. + \frac{1}{9} \left((6j - \frac{26}{9}) \sqrt{25 - 9j} - (6j - \frac{28}{9}) \sqrt{16 + 9j} \right) \right]$$

where

$$s(i) = \frac{\sqrt{25 - 9i} + 5 - 9i}{\sqrt{16 + 9i} + 4 - 9i} \quad \text{for } i \in \{a, b\} \text{ and } I_{a,b} := \{(a, b), (b, a)\}.$$

Evaluating the density shows that $C_{15;2:4}^{\text{PVC}}$ is not the independence copula.

A.5 Proof of Theorem 5.1

The KLD related to tree j, $D_{KL}^{(j)}(\mathcal{T}_j(\mathcal{T}_{1:j-1}))$, is minimized when the negative cross entropy related to tree j is maximized. The negative cross entropy related to tree j is given by

$$H^{(j)}(\mathcal{T}_j(\mathcal{T}_{1:j-1})) := \sum_{i=1}^{d-j} \mathbb{E} \left[\log c_{i,i+j;S_{ij}}^{\text{SVC}}(F_{i|S_{ij}}^{\text{SVC}}(U_i | U_{S_{ij}}), F_{i+j|S_{ij}}^{\text{SVC}}(U_{i+j} | U_{S_{ij}})) \right]$$

$$=: \sum_{i=1}^{d-j} H_i^{(j)}(c_{i,i+j;S_{ij}}^{\text{SVC}}, F_{i|S_{ij}}^{\text{SVC}}, F_{i+j|S_{ij}}^{\text{SVC}}).$$

Obviously, to maximize $H^{(j)}(\mathcal{T}_j(\mathcal{T}_{1:j-1}))$ w.r.t. $\mathcal{T}_j$ we can maximize each $H_i^{(j)}(c_{i,i+j;S_{ij}}^{\text{SVC}}, F_{i|S_{ij}}^{\text{SVC}}, F_{i+j|S_{ij}}^{\text{SVC}})$ individually for all $i = 1, \ldots, d - j$. If $j = 1$, then

$$H_i^{(j)}(c_{i,i+j;S_{ij}}^{\text{SVC}}, F_{i|S_{ij}}^{\text{PVC}}, F_{i+j|S_{ij}}^{\text{PVC}}) = \mathbb{E} \left[\log \frac{c_{i,i+1}(U_i, U_{i+1})}{c_{i,i+1}^{\text{SVC}}(U_i, U_{i+1})} \right]$$

which is maximized for $C_{i,i+1}^{\text{SVC}} = C_{i,i+1}$ by Gibbs' inequality. Thus, if $j = 1$, then

$$\arg\min_{\mathcal{T}_j \in \mathbf{T}_j} D_{KL}^{(j)}(\mathcal{T}_j(\mathcal{T}_{1:j-1}^{\text{PVC}})) = \mathcal{T}_j^{\text{PVC}}. \tag{A.15}$$

To show that (A.15) holds for $j \geq 2$ we use induction. Assume that

$$\arg\min_{\mathcal{T}_j \in \mathbf{T}_j} D_{KL}^{(j)}(\mathcal{T}_j(\mathcal{T}_{1:j-1}^{\text{PVC}})) = \mathcal{T}_j^{\text{PVC}}$$

holds for $1 \leq j \leq d - 2$. To minimize the KLD related to tree $j + 1 =: n$ w.r.t. $\mathcal{T}_n$, conditional on $\mathcal{T}_{1:n-1} = \mathcal{T}_{1:n-1}^{\mathrm{PVC}}$, we have to maximize the negative cross entropy which is maximized if

$$
\begin{aligned}
&H_i^{(n)}(c_{i,i+n;S_{i,n}}^{\mathrm{SVC}}, F_{i|S_{i,n}}^{\mathrm{PVC}}, F_{i+n|S_{i,n}}^{\mathrm{PVC}}) \\
&= \mathbb{E}\left[\log c_{i,i+n;S_{i,n}}^{\mathrm{SVC}}\left(F_{i|S_{i,n}}^{\mathrm{PVC}}(U_i|U_{S_{i,n}}), F_{i+n|S_{i,n}}^{\mathrm{PVC}}(U_{i+n}|U_{S_{i,n}})\right)\right]
\end{aligned}
$$

is maximized for all $i = 1, \ldots, d - n$. Using the substitution $u_i = (F_{i|S_{i,n}}^{\mathrm{PVC}})^{-1}(t_i|u_{S_{i,n}}) = G_{i|S_{i,n}}(t_i|u_{S_{i,n}})$ and $u_{i+n} = (F_{i+n|S_{i,n}}^{\mathrm{PVC}})^{-1}(t_{i+n}|u_{S_{i,n}}) = G_{i+n|S_{i,n}}(t_{i+n}|u_{S_{i,n}})$, we obtain

$$
\begin{aligned}
&H_i^{(n)}(c_{i,i+n;S_{i,n}}^{\mathrm{SVC}}, F_{i|S_{i,n}}^{\mathrm{PVC}}, F_{i+n|S_{i,n}}^{\mathrm{PVC}}) = \int_{[0,1]^{n+1}} \log c_{i,i+n;S_{i,n}}^{\mathrm{SVC}}(t_i, t_{i+n}) \\
&\quad \times c_{i,i+n;S_{i,n}}\left(F_{i|S_{i,n}}\left(G_{i|S_{i,n}}(t_i|u_{S_{i,n}})\big|u_{S_{i,n}}\right), F_{i|S_{i,n}}\left(G_{i+n|S_{i,n}}(t_{i+n}|u_{S_{i,n}})\big|u_{S_{i,n}}\right)\Big|u_{S_{i,n}}\right) \\
&\quad \times \frac{\prod_{k=i,i+n} f_{k|S_{i,n}}\left(G_{k|S_{i,n}}(t_k|u_{S_{i,n}})\big|u_{S_{i,n}}\right)}{\prod_{k=i,i+n} f_{k|S_{i,n}}^{\mathrm{PVC}}\left(G_{k|S_{i,n}}(t_k|u_{S_{i,n}})\big|u_{S_{i,n}}\right)} c_{S_{i,n}}(u_{S_{i,n}}) \mathrm{d}u_{S_{i,n}} \mathrm{d}t_i \mathrm{d}t_{i+n} \\
&= \int_{[0,1]^2} \log c_{i,i+n;S_{i,n}}^{\mathrm{SVC}}(t_i, t_{i+n}) \\
&\quad \times \Bigg(\int_{[0,1]^{n-1}} c_{i,i+n;S_{i,n}}\left(F_{i|S_{i,n}}\left(G_{i|S_{i,n}}(t_i|u_{S_{i,n}})\big|u_{S_{i,n}}\right), F_{i|S_{i,n}}\left(G_{i+n|S_{i,n}}(t_{i+n}|u_{S_{i,n}})\big|u_{S_{i,n}}\right)\Big|u_{S_{i,n}}\right) \\
&\quad \times \frac{\prod_{k=i,i+n} f_{k|S_{i,n}}\left(G_{k|S_{i,n}}(t_k|u_{S_{i,n}})\big|u_{S_{i,n}}\right)}{\prod_{k=i,i+n} f_{k|S_{i,n}}^{\mathrm{PVC}}\left(G_{k|S_{i,n}}(t_k|u_{S_{i,n}})\big|u_{S_{i,n}}\right)} c_{S_{i,n}}(u_{S_{i,n}}) \mathrm{d}u_{S_{i,n}}\Bigg) \mathrm{d}t_i \mathrm{d}t_{i+n} \\
&= \int_{[0,1]^2} \log c_{i,i+n;S_{i,n}}^{\mathrm{SVC}}(t_i, t_{i+n}) c_{i,i+n;S_{i,n}}^{\mathrm{PVC}}(t_i, t_{i+n}) \mathrm{d}t_i \mathrm{d}t_{i+n},
\end{aligned}
$$

which is maximized for $c_{i,i+n;S_{i,n}}^{\mathrm{SVC}} = c_{i,i+n;S_{i,n}}^{\mathrm{PVC}} = c_{i,i+(j+1);S_{i,j+1}}^{\mathrm{PVC}}$ by Gibbs' inequality.

A.6   Proof of Theorem 5.2

Equation (5.3) is obvious, since $C_{1:d}^{\mathrm{PVC}}$ is the data generating process. Equation (5.5) immediately follows from the equations (5.1) and (5.4). Using the same arguments as in Appendix A.2, the validity of (5.4) for $d = 3$ implies the validity of (5.4) for $d \geq 3$. However, even for $d = 3$, the KLD is a triple integral and does not exhibit an analytical expression if the data generating process is a non-simplified vine copula. Thus, the hard part is to show that there exists a data generating copula which does not satisfy the simplifying assumption and for which the PVC does not minimize the KLD. We prove equation (5.4) for $d = 3$ by means of the following example.

Example A.1

Let $g\colon [0,1] \to [-1,1]$ be a measurable function. Consider the data generating process

$$C_{1:3}(u_{1:3}) = \int_0^{u_2} C^{FGM_2}\big(u_1, u_3; g(z)\big)\, dz,$$

i.e., the two unconditional bivariate margins (C_{12}, C_{23}) are independence copulas and the conditional copula is a FGM copula with varying parameter $g(u_2)$. The first-order partial copula is also a FGM copula given by

$$C_{13;2}^{SVC}(u_1, u_3; \theta_{13;2}^{PVC}) = u_1 u_3[1 + \theta_{13;2}^{PVC}(1 - u_1)(1 - u_3)], \quad \theta_{13;2}^{PVC} := \int_0^1 g(u_2)\, du_2.$$

We set $C_{23}^{SVC} = C_{23}$, $C_{13;2}^{SVC} = C_{13;2}^{PVC}$, and specify a parametric copula $C_{12}^{SVC}(\theta_{12})$, $\theta_{12} \in \Theta_{12} \subset \mathbb{R}$, with conditional cdf $F_{1|2}^{SVC}(u_1|u_2; \theta_{12})$ and such that $C_{12}^{SVC}(0)$ corresponds to the independence copula. Thus, $(C_{12}^{SVC}(0), C_{23}, C_{13;2}^{PVC}) = (C_{12}, C_{23}, C_{13;2}^{PVC})$. We also assume that $c_{12}^{SVC}(u_1, u_2; \theta_{12})$ and $\partial_{\theta_{12}} c_{12}^{SVC}(u_1, u_2; \theta_{12})$ are both continuous on $(u_1, u_2, \theta_{12}) \in (0,1)^2 \times \Theta_{12}$.

We now derive necessary and sufficient conditions such that

$$D_{KL}(C_{1:3} \| C_{12}^{\mathrm{SVC}}(\theta_{12}), C_{23}, C_{13;2}^{\mathrm{PVC}}) := D_{KL}(C_{1:3} \| ((C_{12}^{\mathrm{SVC}}(\theta_{12}), C_{23}), (C_{13;2}^{\mathrm{PVC}})))$$

attains an extremum at $\theta_{12} = 0$.

Lemma A.1 (Extremum of the KLD in Example A.1)

Let $C_{1:3}$ be given as in Example A.1. For $u_1 \in (0,1)$, we define

$$h(u_1; g) := \int_0^1 \partial_{\theta_{12}} F_{1|2}^{SVC}(u_1|u_2; \theta_{12})\Big|_{\theta_{12}=0} g(u_2)\, du_2,$$

$$K(u_1; \theta_{13;2}^{PVC}) := \frac{1}{h(u_1; g)} \int_0^1 \int_0^1 \partial_{\theta_{12}} \log c_{13;2}^{SVC}\big(F_{1|2}^{SVC}(u_1|u_2; \theta_{12}), u_3; \theta_{13;2}^{PVC}\big)\Big|_{\theta_{12}=0} c_{1:3}(u_{1:3})\, du_2\, du_3.$$

Then, $\forall u_1 \in (0, 0.5)$: $K(0.5+u_1; \theta_{13;2}^{PVC}) > 0 \Leftrightarrow \theta_{13;2}^{PVC} > 0$, and $D_{KL}(C_{1:3} \| C_{12}^{SVC}(\theta_{12}), C_{23}, C_{13;2}^{PVC})$ has an extremum at $\theta_{12} = 0$ if and only if

$$\partial_{\theta_{12}} D_{KL}(C_{1:3} \| C_{12}^{SVC}(\theta_{12}), C_{23}, C_{13;2}^{PVC})\Big|_{\theta_{12}=0} = \int_0^{0.5} K(0.5 + u_1; \theta_{13;2}^{PVC})$$
$$\times\, [h(0.5 + u_1; g) - h(0.5 - u_1; g)]\, du_1 = 0. \tag{A.16}$$

Proof. See Appendix A.7. ∎

It depends on the data generating process whether the condition in Lemma A.1 is satisfied and $D_{KL}(C_{1:3}||C_{12}^{\text{SVC}}(0), C_{23}, C_{13;2}^{\text{PVC}})$ is an extremum or not as we illustrate in the following. If $\theta_{13;2}^{\text{PVC}} = 0$, then $K(u_1; \theta_{13;2}^{\text{PVC}}) = 0$ for all $u_1 \in (0,1)$, or if g does not depend on u_2, then $h(u_1; g) = 0$ for all $u_1 \in (0,1)$. Thus, the integrand in (A.16) is zero and we have an extremum if one of these conditions is true. Assuming $\theta_{13;2}^{\text{PVC}} \neq 0$ and that g depends on u_2, we see from (A.16) that g and C_{12}^{SVC} determine whether we have an extremum at $\theta_{12} = 0$. Depending on the copula family that is chosen for C_{12}^{SVC}, it may be possible that the copula family alone determines whether $D_{KL}(C_{1:3}||C_{12}^{\text{SVC}}(0), C_{23}, C_{13;2}^{\text{PVC}})$ is an extremum. For instance, if C_{12}^{SVC} is a FGM copula we obtain

$$h(u_1; g) = u_1(1 - u_1) \int_0^1 (1 - 2u_2)g(u_2)\mathrm{d}u_2$$

so that

$$h(0.5 + u_1; g) = h(0.5 - u_1; g), \quad \forall u_1 \in (0, 0.5).$$

This symmetry of h across 0.5 implies that (A.16) is satisfied for all functions g.

If we do not impose any constraints on the bivariate copulas in the first tree of the simplified vine copula approximation, then $D_{KL}(C_{1:3}||C_{12}^{\text{SVC}}(0), C_{23}, C_{13;2}^{\text{PVC}})$ may not even be a local minimizer of the KLD. For instance, if C_{12}^{SVC} is the asymmetric FGM copula given in (4.1), we find that

$$h(u_1; g) = u_1^2(1 - u_1) \int_0^1 (1 - 2u_2)g(u_2)\mathrm{d}u_2.$$

If $\Lambda := \int_0^1 (1 - 2u_2)g(u_2)\mathrm{d}u_2 \neq 0$, e.g., g is a non-negative function which is increasing, say $g(u_2) = u_2$, then, depending on the sign of Λ, either

$$h(0.5 + u_1; g) > h(0.5 - u_1; g), \quad \forall u_1 \in (0, 0.5),$$

or

$$h(0.5 + u_1; g) < h(0.5 - u_1; g), \quad \forall u_1 \in (0, 0.5),$$

so that the integrand in (A.16) is either strictly positive or negative and thus $D_{KL}(C_{1:3}||C_{12}, C_{23}, C_{13;2}^{\text{PVC}})$ can not be an extremum. Since $\theta_{12} \in [-1, 1]$, it follows that $D_{KL}(C_{1:3}||C_{12}^{\text{SVC}}(0), C_{23}, C_{13;2}^{\text{PVC}})$ is not a local minimum. As a result, we can, relating to the PVC, further decrease the KLD from the true copula if we adequately specify "wrong" copulas in the first tree and choose the first-order partial copula in the second tree of the simplified vine copula approximation.

A.7 Proof of Lemma A.1

The KLD attains an extremum if and only if the negative cross entropy attains an extremum. The negative cross entropy is given by

$$
\begin{aligned}
H_{1:3}(C_{1:3}||C_{1:3}^{\mathrm{SVC}}(\theta_{12}; \theta_{13;2}^{\mathrm{PVC}})) &= \mathbb{E}[\log c_{1:3}^{\mathrm{SVC}}(U_{1:3}; \theta_{12}, \theta_{13;2}^{\mathrm{PVC}})] \\
&= \mathbb{E}[\log[c_{12}^{\mathrm{SVC}}(U_{1:2}; \theta_{12})c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(U_1|U_2; \theta_{12}), U_3; \theta_{13;2}^{\mathrm{PVC}})]] \\
&= \mathbb{E}[\log c_{12}^{\mathrm{SVC}}(U_{1:2}; \theta_{12})] + \mathbb{E}[\log c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(U_1|U_2; \theta_{12}), U_3; \theta_{13;2}^{\mathrm{PVC}})].
\end{aligned}
$$

If the negative cross entropy attains an extremum then the derivative of $\mathbb{E}[\log c_{1:3}^{\mathrm{SVC}}(U_{1:3}; \theta_{12})]$ w.r.t. θ_{12} is zero. Since $c_{12}^{\mathrm{SVC}}(u_1, u_2; \theta_{12})$ and $\partial_{\theta_{12}} c_{12}^{\mathrm{SVC}}(u_1, u_2; \theta_{12})$ are both continuous on $(u_1, u_2, \theta_{12}) \in (0, 1)^2 \times \Theta_{12}$, we can apply Leibniz's rule for differentiation under the integral sign to conclude that $\partial_{\theta_{12}} \mathbb{E}[\log c_{12}^{\mathrm{SVC}}(U_{1:2}; \theta_{12})]\big|_{\theta_{12}=0} = \mathbb{E}[\partial_{\theta_{12}} \log c_{12}^{\mathrm{SVC}}(U_{1:2}; \theta_{12})\big|_{\theta_{12}=0}] = 0$ because $C_{12}^{\mathrm{SVC}}(0)$ is the true copula of $U_{1:2}$. Thus, the derivative evaluated at $\theta_{12} = 0$ becomes

$$
\begin{aligned}
&\partial_{\theta_{12}} \mathbb{E}[\log c_{1:3}^{\mathrm{SVC}}(U_{1:3}; \theta_{12}, \theta_{13;2}^{\mathrm{PVC}})]\Big|_{\theta_{12}=0} \\
&= \partial_{\theta_{12}} \mathbb{E}[\log c_{12}^{\mathrm{SVC}}(U_{1:2}; \theta_{12})]\Big|_{\theta_{12}=0} + \partial_{\theta_{12}} \mathbb{E}[\log c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(U_1|U_2; \theta_{12}), U_3; \theta_{13;2}^{\mathrm{PVC}})]\Big|_{\theta_{12}=0} \\
&= \partial_{\theta_{12}} \mathbb{E}[\log c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(U_1|U_2; \theta_{12}), U_3; \theta_{13;2}^{\mathrm{PVC}})]\Big|_{\theta_{12}=0} \\
&= \partial_{\theta_{12}} \int_{[0,1]^3} \log c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12}), u_3; \theta_{13;2}^{\mathrm{PVC}}) c_{1:3}(u_{1:3}) \mathrm{d}u_{1:3}\Big|_{\theta_{12}=0} \\
&= \int_{[0,1]^3} \frac{\partial_1 c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12}), u_3; \theta_{13;2}^{\mathrm{PVC}})}{c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12}), u_3; \theta_{13;2}^{\mathrm{PVC}})} \partial_{\theta_{12}} F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12})\Big|_{\theta_{12}=0} c_{1:3}(u_{1:3}) \mathrm{d}u_{1:3},
\end{aligned}
$$

where $\partial_1 c_{13;2}^{\mathrm{SVC}}(u, v; \theta_{13;2}^{\mathrm{PVC}})$ is the partial derivative w.r.t. u and we have used Leibniz's integral rule to perform the differentiation under the integral sign for the second last equality which is valid since the integrand and its partial derivative w.r.t. θ_{12} are both continuous in $u_{1:3}$ and θ_{12} on $(0, 1)^3 \times (-1, 1)$.

To compute the integral we observe that

$$
\begin{aligned}
\partial_1 c_{13;2}^{\mathrm{SVC}}(u, v; \theta_{13;2}^{\mathrm{PVC}}) &= -2\theta_{13;2}^{\mathrm{PVC}}(1 - 2v), \\
\partial_1 c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(u_1|u_2; 0), u_3; \theta_{13;2}^{\mathrm{PVC}}) &= -2\theta_{13;2}^{\mathrm{PVC}}(1 - 2u_3), \\
\frac{\partial_1 c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12}), u_3; \theta_{13;2}^{\mathrm{PVC}})}{c_{13;2}^{\mathrm{SVC}}(F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12}), u_3; \theta_{13;2}^{\mathrm{PVC}})}\Big|_{\theta_{12}=0} &= \frac{-2\theta_{13;2}^{\mathrm{PVC}}(1 - 2u_3)}{1 + \theta_{13;2}^{\mathrm{PVC}}(1 - 2u_1)(1 - 2u_3)} =: m(u_1, u_3; \theta_{13;2}^{\mathrm{PVC}}).
\end{aligned}
$$

Note that $m(u_1, u_3; \theta_{13;2}^{\mathrm{PVC}})$ does not depend on u_2. Moreover, with $c_{1:3}(u_{1:3}) =$

$1 + g(u_2)(1 - 2u_1)(1 - 2u_3),$

$$\int_0^1 \partial_{\theta_{12}} F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12}, \theta_{13;2}^{\mathrm{PVC}})\Big|_{\theta_{12}=0} c_{1:3}(u_{1:3})\mathrm{d}u_2$$

$$= \int_0^1 \partial_{\theta_{12}} F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12})\Big|_{\theta_{12}=0} \mathrm{d}u_2$$

$$+ (1 - 2u_1)(1 - 2u_3) \int_0^1 \partial_{\theta_{12}} F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12})\Big|_{\theta_{12}=0} g(u_2)\mathrm{d}u_2$$

$$= (1 - 2u_1)(1 - 2u_3)h(u_1; g),$$

where the second equality follows because

$$\int_0^1 \partial_{\theta_{12}} F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12})\mathrm{d}u_2 = \partial_{\theta_{12}} \int_0^1 F_{1|2}^{\mathrm{SVC}}(u_1|u_2; \theta_{12})\mathrm{d}u_2 = \partial_{\theta_{12}} u_1 = 0.$$

Thus, integrating out u_2, we obtain

$$\partial_{\theta_{12}} \mathbb{E}[\log c_{1:3}^{\mathrm{SVC}}(U_{1:3}; \theta_{12})]\Big|_{\theta_{12}=0} = \int_{[0,1]^2} m(u_1, u_3; \theta_{13;2}^{\mathrm{PVC}})(1 - 2u_1)(1 - 2u_3)h(u_1; g)\mathrm{d}u_1\mathrm{d}u_3$$

$$= \int_{[0,1]^2} f(u_1, u_3; \theta_{13;2}^{\mathrm{PVC}})h(u_1; g)\mathrm{d}u_1\mathrm{d}u_3, \tag{A.17}$$

where $f(u_1, u_3; \theta_{13;2}^{\mathrm{PVC}}) := m(u_1, u_3; \theta_{13;2}^{\mathrm{PVC}})(1 - 2u_1)(1 - 2u_3)$. We note that $\forall u_1 \in (0, 0.5)$, $u_3 \in (0, 1)$:

$$f(0.5 + u_1, u_3; \theta_{13;2}^{\mathrm{PVC}}) > 0 \Leftrightarrow \theta_{13;2}^{\mathrm{PVC}} > 0,$$
$$f(0.5 - u_1, u_3; \theta_{13;2}^{\mathrm{PVC}}) = -f(0.5 + u_1, 1 - u_3; \theta_{13;2}^{\mathrm{PVC}}).$$

So, if $u_1 \in (0, 0.5)$ then

$$\int_0^1 f(0.5 - u_1, u_3; \theta_{13;2}^{\mathrm{PVC}})\mathrm{d}u_3 = \int_0^1 f(0.5 - u_1, 1 - u_3; \theta_{13;2}^{\mathrm{PVC}})\mathrm{d}u_3$$

$$= -\int_0^1 f(0.5 + u_1, u_3; \theta_{13;2}^{\mathrm{PVC}})\mathrm{d}u_3.$$

Thus, if we define $K(u_1; \theta_{13;2}^{\mathrm{PVC}}) := \int_0^1 f(u_1, u_3; \theta_{13;2}^{\mathrm{PVC}})\mathrm{d}u_3$ we have that $\forall u_1 \in (0, 0.5)$:

$$K(0.5 + u_1; \theta_{13;2}^{\mathrm{PVC}}) > 0 \Leftrightarrow \theta_{13;2}^{\mathrm{PVC}} > 0,$$
$$K(0.5 - u_1; \theta_{13;2}^{\mathrm{PVC}}) = -K(0.5 + u_1; \theta_{13;2}^{\mathrm{PVC}}). \tag{A.18}$$

Plugging this into our integral (A.17) yields

$$
\partial_{\theta_{12}} \mathbb{E}[\log c_{1:3}^{\mathrm{SVC}}(U_{1:3}; \theta_{12}, \theta_{13;2}^{\mathrm{PVC}})]\Big|_{\theta_{12}=0}
$$

$$
= \int_0^1 \int_0^1 f(u_1, u_3; \theta_{13;2}^{\mathrm{PVC}}) h(u_1; g) \mathrm{d}u_1 \mathrm{d}u_3
$$

$$
= \int_0^{0.5} h(0.5 - u_1; g) \left(\int_0^1 f(0.5 - u_1, u_3; \theta_{13;2}^{\mathrm{PVC}}) \mathrm{d}u_3 \right) \mathrm{d}u_1
$$

$$
+ \int_0^{0.5} h(0.5 + u_1; g) \left(\int_0^1 f(0.5 + u_1, u_3; \theta_{13;2}^{\mathrm{PVC}}) \mathrm{d}u_3 \right) \mathrm{d}u_1
$$

$$
= \int_0^{0.5} h(0.5 - u_1; g) K(0.5 - u_1; \theta_{13;2}^{\mathrm{PVC}}) \mathrm{d}u_1
$$

$$
+ \int_0^{0.5} h(0.5 + u_1; g) K(0.5 + u_1; \theta_{13;2}^{\mathrm{PVC}}) \mathrm{d}u_1
$$

$$
\overset{(\mathrm{A.18})}{=} \int_0^{0.5} K(0.5 + u_1; \theta_{13;2}^{\mathrm{PVC}})[h(0.5 + u_1; g) - h(0.5 - u_1; g)] \mathrm{d}u_1.
$$

Note that if $\theta_{13;2}^{\mathrm{PVC}} = 0$, then $K(u_1; \theta_{13;2}^{\mathrm{PVC}}) = 0$ for all $u_1 \in (0, 1)$, or if g does not depend on u_2, then $h(u_1; g) = 0$ for all $u_1 \in (0, 1)$, so in both cases the integrand is zero and we have an extremum.

A.8   Proof of Corollary 6.1

Corollary 6.1 (i) and (ii) follow directly from Theorem 1 in Spanhel and Kurz (2016a), which states the asymptotic distribution of approximate rank Z-estimators if the data generating process is not nested in the parametric model family. Corollary 6.1 (iii) follows then from Theorem 5.2 and Theorem 5.1.

A.9   An example where the difference between $\hat{\theta}^S$ and $\hat{\theta}^J$ is more pronounced

Example A.2

Let $C^{BB1}(\theta, \delta)$ denote the BB1 copula with dependence parameter (θ, δ) and $C^{Sar}(\alpha)$ be the Sarmanov copula with cdf $C(u, v; \alpha) = uv\big(1 + (3\alpha + 5\alpha^2 \prod_{i=u,v}(1 - 2i)) \prod_{i=u,v}(1 - i)\big)$ for $|\alpha| \leq \sqrt{7}/5$. The partial Sarmanov copula is given by $C^{P\text{-}Sar}(u, v; a, b) = uv\big(1 + (3a + 5b \prod_{i=u,v}(1 - 2i)) \prod_{i=u,v}(1 - i)\big)$, where $|a| \leq \sqrt{7}/5$ and $a^2 \leq b \leq (\sqrt{1 - 3a^2} + 1)/5$. Define $S(u_2) = (1 + \exp(u_2))^{-1}$ and $f(u_2) = 1 - 2S(10u_2 - 0.5)) + 2(1 - 2u_2)S(-5)$ so that $g(u_2) = 0.1\big(\sqrt{7} + 1\big)(1 - f(u_2)) - 0.2$. Let $C_{1:3}$ be the true copula with $(C_{12}, C_{23}, C_{13;2}) =$

$(C^{BB1}(2,2), C^{BB1}(2,2), C^{Sar}(g(u_2))$ *and* $C^{SVC}_{1:3} = (C^{BB1}(2,2), C^{BB1}(2,2), C^{P\text{-}Sar}(a,b))$ *be the parametric SVC that is fitted to data generated from* $C_{1:3}$.

Note that g is a sigmoid function, with $(g(0), g(1)) = (-0.2, \sqrt{7}/5)$, so that Spearman's rho of the conditional copula $C^{Sar}(g(u_2))$ varies in the interval $(g(0), g(1)) = (-0.2, \sqrt{7}/5)$ because $\rho_{C^{Sar}} = \alpha$. Figure 5 shows that the difference between step-by-step and joint ML estimates for the two parameters of the first copula in the first tree is already (individually) significant at the 5% level if the sample size is 500 observations. Thus, the difference between step-by-step and joint ML estimates can be relevant for moderate sample sizes if the variation in the conditional copula is strong enough. Once again, the difference between step-by-step and joint ML estimates is less pronounced for the parameters of $C^{SVC}_{13;2}$ but it also becomes highly significant with sufficient sample size.
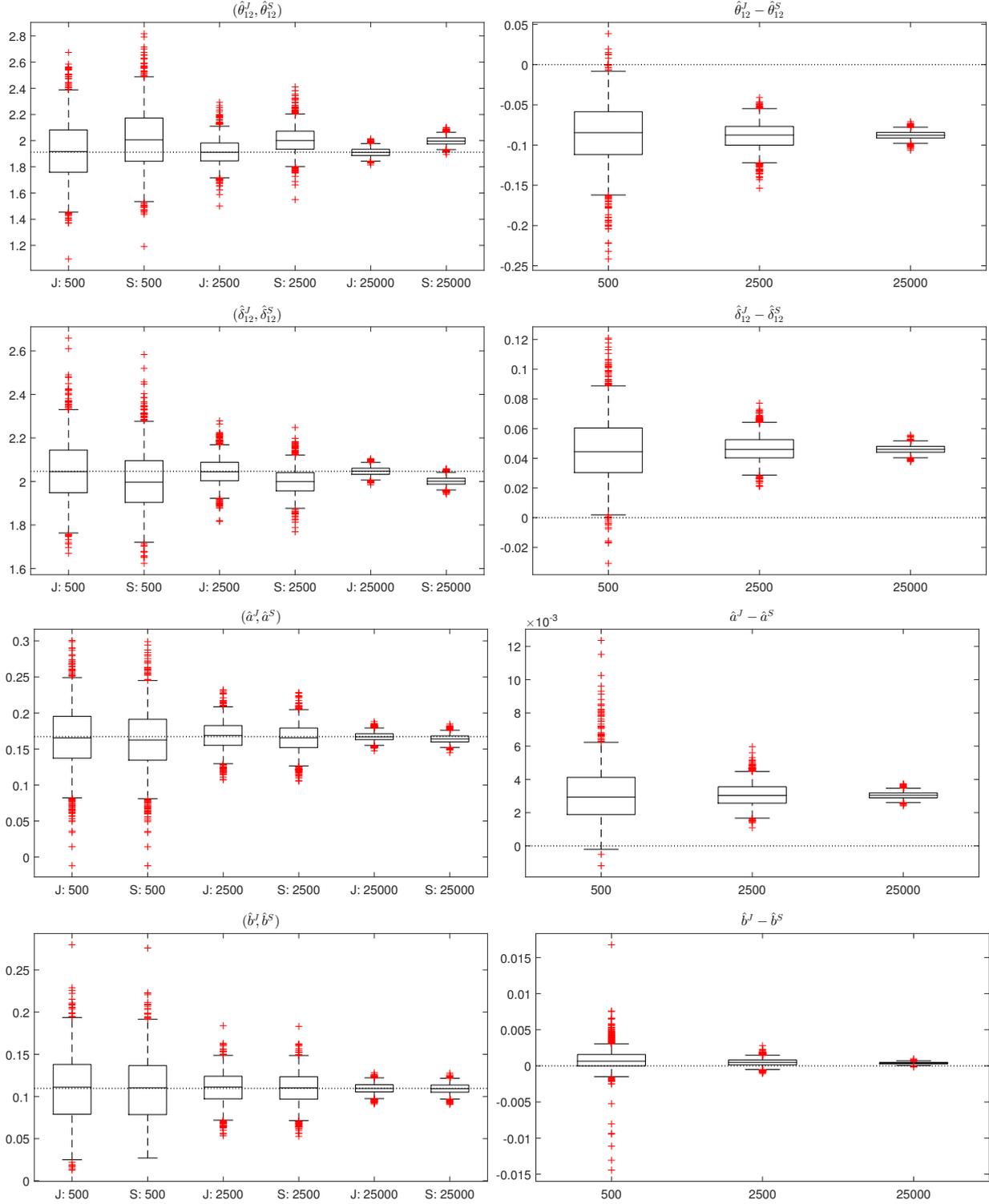
Figure 5: Box plots of joint (J) and sequential (S) ML estimates and their difference for sample sizes $N = 500, 2500, 25000$, if the data is generated from $C_{1:3}$ in Example A.2 and the pair-copula families of the SVC are given by the corresponding PVC. The dotted line indicates the pseudo-true parameter and zero, respectively. The end of the whiskers is 0.953 times the inter-quartile range, corresponding to approximately 95% coverage if the data is generated by a normal distribution.

Chapter IV

Testing the simplifying assumption in high-dimensional vine copulas

This chapter is a reprint* of:

M. S. Kurz and F. Spanhel. 2018. "Testing the simplifying assumption in high-dimensional vine copulas". ArXiv e-prints `arXiv:1706.02338`.

Revision note:

This chapter is a reprint of the revised manuscript available on arXiv since April 18, 2018. On June 07, 2017 the first version was made publicly available on arXiv.

Author contributions:

The project was initiated jointly by Malte Kurz and Fabian Spanhel. Most parts of the manuscript have been drafted by Malte Kurz and both authors contributed to the manuscript. The test idea and algorithms have been developed in close collaboration. All proofs have been developed in close collaboration and both authors contributed equally to the proofs. A revision has been done jointly. The graphical illustrations of the decision tree approach have been developed by Malte Kurz. All implementations in the accompanying software package and the implementations of the simulations and applications in `R` and `C++` have been done by Malte Kurz.

Accompanying software:

M. S. Kurz. 2018c. *pacotest: Testing for Partial Copulas and the Simplifying Assumption in Vine Copulas.* URL: `https://CRAN.R-project.org/package=pacotest`, R-Package version 0.3.

---

*All reprints in this cumulative dissertation have been slightly adapted in order to harmonize the overall appearance, like the format, page margins, caption setup, citation style and bibliography style.

Testing the simplifying assumption in high-dimensional vine copulas

Malte S. Kurz [*][†][1] and Fabian Spanhel [‡][1]

[1]Department of Statistics,
Ludwig-Maximilians-Universität München,
Akademiestr. 1, 80799 Munich, Germany.

Abstract

Testing the simplifying assumption in high-dimensional vine copulas is a difficult task because tests must be based on estimated observations and amount to checking constraints on high-dimensional distributions. So far, corresponding tests have been limited to single conditional copulas with a low-dimensional set of conditioning variables. We propose a novel testing procedure that is computationally feasible for high-dimensional data sets and that exhibits a power that decreases only slightly with the dimension. By discretizing the support of the conditioning variables and incorporating a penalty in the test statistic, we mitigate the curse of dimensions by looking for the possibly strongest deviation from the simplifying assumption. The use of a decision tree renders the test computationally feasible for large dimensions. We derive the asymptotic distribution of the test and analyze its finite sample performance in an extensive simulation study. The utility of the test is demonstrated by its application to 10 data sets with up to 49 dimensions.

Keywords: Conditional copula, Pair-copula construction, Partial vine copula, Simplifying assumption, Test for constant conditional correlation, Vine copula.

---

[*]Corresponding author
[†]malte.kurz@stat.uni-muenchen.de
[‡]spanhel@stat.uni-muenchen.de

1 Introduction

Vine copulas (Joe 1997, Bedford and Cooke 2002, Aas et al. 2009) are one of the most popular tools for multivariate dependence modeling. The central model assumption for statistical inference is the so-called simplifying assumption (Hobæk Haff et al. 2010). It is the assumption that every conditional copula (Patton 2006) within the vine copula can be represented by an unconditional copula. An extensive literature containing methodological advances and applied research for vine copulas has been based on the simplifying assumption (Dißmann et al. 2013, Grothe and Nicklas 2013, Joe et al. 2010, Kauermann and Schellhase 2014, Kurowicka and Joe 2011, Nagler and Czado 2016, Nikoloulopoulos et al. 2012).

In the early papers discussing the simplifying assumption, multivariate distributions which can be represented as simplified vine copulas have been identified (Hobæk Haff et al. 2010, Stöber et al. 2013). More recently the simplifying assumption has again attracted a lot of attention (Acar et al. 2012, Derumigny and Fermanian 2017, Gijbels et al. 2017a, 2017b, Killiches et al. 2016, 2017, Kraus and Czado 2017, Nagler and Czado 2016, Portier and Segers 2018, Spanhel and Kurz 2017). In the context of (bivariate) conditional copulas non- and semiparametric tests for the simplifying assumption have been developed (Acar et al. 2013, Gijbels et al. 2017a, 2017b). See also Derumigny and Fermanian (2017) for a recent survey. In the simulations and application of these studies the conditioning variables are either assumed to be one- or rather low-dimensional and the simplifying assumption is only tested for one single conditional copula but not for a vine copula where several conditional copulas need to be tested for constancy.

We contribute to the literature by proposing a framework for testing the simplifying assumption in high-dimensional vine copulas. To this end, we use the partial vine copula to introduce a stochastic interpretation of the simplifying assumption which is particularly useful for testing it in high dimensions. We apply a semi-parametric framework to test the simplifying assumption. First, we use ranks to obtain pseudo-observations from the partial vine copula. On the basis of these pseudo-observations we apply a stepwise maximum likelihood estimator to obtain pseudo-observations from the partial probability integral transforms which can be used to test the simplifying assumption. We consider the null hypothesis that the correlation of the conditional copula associated to an edge of a vine is constant w.r.t. the conditioning variables if the simplifying assumption is true. To obtain a test whose power does not collapse dramatically with the dimension of the conditioning variables, we discretize the support of the conditioning variables into a finite number of partitions and incorporate a penalty in the test statistic. To render the test computationally feasible in high-dimensions, we apply a decision tree to find the possibly largest difference

in the set of conditional correlations using a greedy approach.

The proposed test is computationally feasible even in high dimensions which is demonstrated by its application to a 49-dimensional data set. An accompanying `R`-package `pacotest` (Kurz 2018) with a `C++`-kernel is publicly available and has already been applied to even higher-dimensional data sets (Kraus and Czado 2017). Moreover, the proposed test can be used to detect building blocks of a vine copula where the modeling of a conditional copula might be more reasonable than the use of an unconditional copula (Schellhase and Spanhel 2018) and it can also be applied to construct new methods for the structure selection of vine copulas (Kraus and Czado 2017).

The organization of the paper is as follows. The partial vine copula (PVC) and stochastic interpretations of the simplifying assumption are discussed in Section 2. A hierarchical procedure to test the simplifying assumption in vine copulas is introduced in Section 3. In Section 4 we present the test for the constancy of the conditional correlation by discretizing the support of the conditioning variables into a finite number of subsets and derive its asymptotic distribution. A decision tree algorithm for finding the largest deviation from the simplifying assumption with high-dimensional conditioning variables is proposed in Section 5. An extensive analysis of the finite sample performance of the proposed test is provided in Section 6. Applications to real data sets are presented in Section 7 and Section 8 concludes.

Throughout the paper we rely on the following notation and assumptions.[1] The cdf of a d-dimensional random vector $X_{1:d}$ is denoted by $F_{X_{1:d}} := \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$. The distribution function or copula of a random vector $U_{1:d}$ with uniform margins is denoted by $F_{1:d} = C_{1:d}$. For simplicity, we assume that all random variables are real-valued and continuous with positive density. If X and Y are stochastically independent we write $X \perp Y$. For the indicator function we use $\mathbb{1}_{\{A\}} = 1$ if A is true, and $\mathbb{1}_{\{A\}} = 0$ otherwise. $\partial_\theta g(\theta)$ denotes the gradient w.r.t. θ and if $h(\gamma)$ is a d-dimensional function then $\partial_i h(\gamma)$ is the partial derivative w.r.t. the i-th element. To shorten the notation for D-vine copulas we use the sets $\mathcal{I}_l^d := \{(i,j) : j = l, \ldots, d-1, i = 1, \ldots, d-j\}$ and $S_{ij} := i+1 : i+j-1 := i+1, \ldots, i+j-1$.

---

[1] We use the same notational conventions as in Spanhel and Kurz (2017) and refer to Table 1 therein for an overview of the used notation.

2 The partial vine copula (PVC) and stochastic interpretations of the simplifying assumption

In this section, we discuss vine copulas and the simplifying assumption. Thereafter, we establish a probabilistic interpretation of the simplifying assumption in terms of vectorial independencies, which can be used to check the validity of the simplifying assumption.

Definition 1 (D-vine copula – Kurowicka and Cooke (2006))
Let $d \geq 3$ and $U_{1:d}$ be a uniform random vector with cdf $F_{1:d} = C_{1:d}$. Define $u_{k|S_{ij}} := F_{k|S_{ij}}(u_k|u_{S_{ij}})$ for $(i,j) \in \mathcal{I}_2^d$, $k = i, i+j$ and denote the conditional copula of $F_{i,i+j|S_{ij}}$ by $C_{i,i+j;S_{ij}}$ (Definition 3). For $j = 1$ and $k = i, i+j$ we set $u_{k|S_{ij}} = u_k$ and $C_{i,i+j;S_{ij}} = C_{i,i+1}$. The density of a D-vine copula is given by

$$c_{1:d}(u_{1:d}) = \prod_{(i,j)\in\mathcal{I}_1^d} c_{i,i+j;S_{ij}}\big(u_{i|S_{ij}}, u_{i+j|S_{ij}}|u_{S_{ij}}\big),$$

where $c_{i,i+j;S_{ij}}$ are bivariate conditional copula densities.

D-vine copulas, or regular vine copulas in general, can be considered as an ordered sequence of trees, where j refers to the number of the tree and a bivariate conditional copula $C_{i,i+j;S_{ij}}$ is assigned to each of the $d - j$ edges of tree j (Bedford and Cooke 2002). For notational simplicity, we will discuss D-vine copulas but all results can easily be generalized to regular vine copulas.[2] Using conditional copulas as building blocks there exists a D-vine copula representation for every multivariate copula. However, in order to simplify the modeling process and to overcome the curse of dimensions, it is commonly assumed that the simplifying assumption holds for the data generating vine copula.

Definition 2 (The simplifying assumption – Hobæk Haff et al. (2010))
The D-vine copula in Definition 1 satisfies the simplifying assumption if $c_{i,i+j;S_{ij}}(\cdot, \cdot|u_{S_{ij}})$ does not depend on $u_{S_{ij}}$ for all $(i,j) \in \mathcal{I}_2^d$.

If the simplifying assumption holds for the data generating vine copula its density collapses to a product of $d(d-1)/2$ bivariate unconditional copula densities. Similarly, if a vine copula model consists of bivariate unconditional copula densities we call it a simplified vine copula (SVC) model. Definition 2 characterizes the simplifying assumption in terms of restrictions that are placed on the functional form of conditional copulas. That is, the simplifying assumption holds if each $(j + 1)$-dimensional function $c_{i,i+j;S_{ij}}(\cdot, \cdot|u_{S_{ij}})$

---

[2]In the accompanying `R`-package `pacotest` (Kurz 2018) all functions are implemented for regular vine copulas and we also make use of regular vine copulas for real data applications.

only depends on its first two arguments, but the other $(j-1)$ arguments $u_{S_{ij}}$ have no effect. In the remainder of this section we derive equivalent statements to the simplifying assumption which are especially useful for testing it. The new characterizations of the simplifying assumption are related to the partial copula and the partial vine copula.

Definition 3 (Bivariate conditional and partial copula)
Let $U_{1:d} \sim F_{1:d} = C_{1:d}$, $(i,j) \in \mathcal{I}_2^d$ and $k = i, i+j$.

(i) $U_{k|S_{ij}} := F_{k|S_{ij}}(U_k|U_{S_{ij}})$ *is the conditional probability integral transform (CPIT) of U_k w.r.t. $U_{S_{ij}}$.*

(ii) *The bivariate conditional copula $C_{i,i+j;S_{ij}}$ of $F_{i,i+j|S_{ij}}$ (Patton (2006)) is defined as*

$$C_{i,i+j;S_{ij}}(a,b|u_{S_{ij}}) := \mathbb{P}(U_{i|S_{ij}} \le a, U_{i+j|S_{ij}} \le b | U_{S_{ij}} = u_{S_{ij}}).$$

(iii) *The bivariate partial copula $C_{i,i+j;S_{ij}}^P$ of $F_{i,i+j|S_{ij}}$ (Bergsma 2004, Gijbels et al. 2015, Spanhel and Kurz 2016b) is defined as*

$$C_{i,i+j;S_{ij}}^P(a,b) := \mathbb{P}(U_{i|S_{ij}} \le a, U_{i+j|S_{ij}} \le b).$$

Whereas the bivariate conditional copula is the conditional distribution of a pair of CPITs, the partial copula is the bivariate unconditional distribution of a pair of CPITs. Therefore, assuming that a conditional copula $C_{i,i+j;S_{ij}}(\cdot,\cdot|u_{S_{ij}})$ does not depend on $u_{S_{ij}}$ is equivalent to assuming that the conditional copula equals the partial copula. From Definition 3 it immediately follows that the conditional copula is equal to the partial copula if and only if the vectorial independence $(U_{i|S_{ij}}, U_{i+j|S_{ij}}) \perp U_{S_{ij}}$ holds.

A special simplified vine copula is the so-called partial vine copula which is of great importance for testing the simplifying assumption. By means of the partial vine copula one can obtain a useful probabilistic interpretation of the simplifying assumption and develop tests being feasible even in high dimensions.

Definition 4 (Partial vine copula (PVC) – Spanhel and Kurz (2017))
For $i = 1, \ldots, d-1$ set $C_{i,i+1}^{PVC} = C_{i,i+1}$ and for $i = 1, \ldots, d-2$: $C_{i,i+2;i+1}^{PVC} = C_{i,i+2;i+1}^P$. For $(i,j) \in \mathcal{I}_2^d$ define the partial probability integral transforms (PPITs) for $j = 2$ as $U_{i|S_{ij}}^{PVC} := U_{i|S_{ij}}$ and $U_{i+j|S_{ij}}^{PVC} := U_{i+j|S_{ij}}$ and for $j \ge 3$ as

$$U_{i|S_{ij}}^{PVC} := \partial_2 C_{i,i+j-1;S_{i,j-1}}^{PVC}(U_{i|S_{i,j-1}}^{PVC}, U_{i+j-1|S_{i,j-1}}^{PVC}), \tag{2.1}$$

$$U_{i+j|S_{ij}}^{PVC} := \partial_1 C_{i+1,i+j;S_{i+1,j-1}}^{PVC}(U_{i+1|S_{i+1,j-1}}^{PVC}, U_{i+j|S_{i+1,j-1}}^{PVC}), \tag{2.2}$$

and the $(j-1)$-th order partial copula as

$$C_{i,i+j;S_{ij}}^{PVC}(a,b) := \mathbb{P}(U_{i|S_{ij}}^{PVC} \leq a, U_{i+j|S_{ij}}^{PVC} \leq b).$$

The resulting simplified vine copula $C_{1:d}^{PVC}$ is called the partial vine copula (PVC) of $C_{1:d}$ and its density is given by

$$c_{1:d}^{PVC}(u_{1:d}) := \prod_{(i,j)\in\mathcal{I}_1^d} c_{i,i+j;S_{ij}}^{PVC}(u_{i|S_{ij}}^{PVC}, u_{i+j|S_{ij}}^{PVC}).$$

The partial copula as well as the PVC give rise to the following stochastic interpretations of the simplifying assumption.

Proposition 1 (Stochastic interpretations of the simplifying assumption)
For $d \geq 3$, the following statements are equivalent:

(i) The D-vine copula given in Definition 1 satisfies the simplifying assumption (Definition 2).

(ii) $\forall (i,j) \in \mathcal{I}_2^d: \qquad (U_{i|S_{ij}}, U_{i+j|S_{ij}}) \perp U_{S_{ij}}$

(iii) $\forall (i,j) \in \mathcal{I}_2^d: \qquad (U_{i|S_{ij}}^{PVC}, U_{i+j|S_{ij}}^{PVC}) \perp U_{S_{ij}}$

Proof. That (i) and (ii) are equivalent follows directly from the definition of the simplifying assumption in Definition 2 and the definition of the conditional and partial copula in Definition 3. By means of Lemma 3.1 in Spanhel and Kurz (2017) and the definition of the PVC in Definition 4 it can be readily verified that (ii) and (iii) are equivalent. ∎

Proposition 1 can be formulated in the same manner for any regular vine copula and highlights that the simplifying assumption is equivalent to $(d-1)(d-2)/2$ vectorial independence assumptions. Note that $(U_{i|S_{ij}}, U_{i+j|S_{ij}}) \perp U_{S_{ij}}$ in (ii) can be replaced by $C_{i,i+j;S_{ij}} = C_{i,i+j;S_{ij}}^{\mathrm{P}}$ and that $(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}) \perp U_{S_{ij}}$ in (iii) can be replaced by $C_{i,i+j;S_{ij}} = C_{i,i+j;S_{ij}}^{\mathrm{PVC}}$. That is, the simplifying assumption holds if and only if all conditional, partial and $(j-1)$-th order partial copulas coincide. While the different stochastic interpretations (ii) and (iii) in Proposition 1 are equivalent in theory, the stochastic interpretation (iii) is much more useful for testing the simplifying assumption. In practice, observations from the pair of CPITs $(U_{i|S_{ij}}, U_{i+j|S_{ij}})$ or the pair of PPITs $(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}})$ are not observable and have to be estimated from data so that a test for the simplifying assumption can be established. Observations from the CPIT $U_{k|S_{ij}}$ can be obtained by estimating a j-dimensional function, namely, the conditional distribution function $F_{k|S_{ij}}$ of U_k given $U_{S_{ij}}$. In contrast to the CPIT, observations from the PPIT $U_{k|S_{ij}}^{\mathrm{PVC}}$ can be obtained as a composition of $j(j-1)/2$

bivariate functions, namely, the bivariate unconditional copulas being the building blocks of the corresponding PVC given in (2.1) and (2.2) in Definition 4. Thus, one possible cause for a curse of dimensions is eliminated when one uses the stochastic interpretation (iii) in Proposition 1 instead of (ii). Therefore, we use the stochastic interpretation given in (iii) to construct a test of the simplifying assumption which is based on pseudo-observations from the PPITs.

3 A hierarchical procedure for testing the simplifying assumption in vine copulas

To obtain pseudo-observations from the PPITs, we use in the following a semi-parametric approach. To this end, let $X_{1:d} \sim F_{X_{1:d}}$ be the data generating process and $C_{1:d}$ be the copula of $X_{1:d}$, i.e., $U_{1:d} \sim C_{1:d}$, where $U_i = F_{X_i}(X_i)$ for $i = 1, \ldots, d$. Let $\{C_{1:d}^{\mathrm{SVC}}(\cdot; \theta_{1:d-1}) \colon \theta_{1:d-1} \in \Upsilon\}$ be a parametric simplified vine copula model for the PVC $C_{1:d}^{\mathrm{PVC}}$ of $C_{1:d}$ such that $\exists \theta_{1:d-1;0} \in \Upsilon$ so that $C_{1:d}^{\mathrm{SVC}}(\cdot; \theta_{1:d-1;0}) = C_{1:d}^{\mathrm{PVC}}(\cdot)$. The density of the parametric simplified vine copula model for the PVC is given by

$$c_{1:d}^{\mathrm{SVC}}(u_{1:d}; \theta_{1:d-1}) = \prod_{(i,j) \in \mathcal{I}_1^d} c_{i,i+j; S_{ij}}^{\mathrm{SVC}}(u_{i|S_{ij}}^{\mathrm{SVC}}(\theta_{1:j-1}), u_{i+j|S_{ij}}^{\mathrm{SVC}}(\theta_{1:j-1}); \theta_{j,i}),$$

where $c_{i,i+j; S_{ij}}^{\mathrm{SVC}}$ is the density of a bivariate (unconditional) copula for each $(i,j) \in \mathcal{I}_1^d$. For $(i,j) \in \mathcal{I}_1^d$ the parameter of the copula $C_{i,i+j; S_{ij}}^{\mathrm{SVC}}$ is denoted by $\theta_{j,i}$. For $1 \le j \le d-1$, the vector $\theta_j := (\theta_{j,i})_{i=1,\ldots,d-j}^T$ collects all parameter of the copulas in the j-th tree and the vector $\theta_{1:j} := (\theta_1, \ldots, \theta_j)^T$ collects all parameters up to and including the j-th tree. A sequential algorithm to test the set of hypotheses

$$\forall (i,j) \in \mathcal{I}_2^d : H_0 : (U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}) \perp U_{S_{ij}},$$

while controlling the family-wise error rate is stated in Definition 5.

Definition 5 (Hierarchical procedure for testing the simplifying assumption in vine copulas)
Let $(X_{1:d}^k)_{k=1,\ldots,n}$ be n independent copies from $X_{1:d}$ and $C_{1:d}^{SVC}(\cdot; \theta_{1:d-1})$ be a parametric simplified vine copula model for the PVC $C_{1:d}^{PVC}$ of $X_{1:d}$. Let α be the family-wise error rate, i.e., the probability of making at least one type 1 error among the considered hypotheses and set $M = (d-1)(d-2)/2$.

1. *Use the rescaled ecdf to compute the pseudo-observations from $C_{1:d}^{PVC}$ via*

$$V_i^k := \frac{1}{n+1} \sum_{m=1}^n \mathbb{1}_{\{X_i^m \leq X_i^k\}}, \quad \text{for } i = 1, \ldots, d \text{ and } k = 1, \ldots, n.$$

2. *Loop over the trees $j = 2, \ldots, d-1$:*

   (a) *Apply the stepwise ML estimator (Hobæk Haff 2013) to estimate θ_{j-1}, i.e., the parameters of the pair-copulas in tree $j-1$.*

   (b) *For $i = 1, \ldots, d-j$, obtain the pseudo-observations*

   $$\left(V_{i|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1}), V_{i+j|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1}) \right)_{k=1,\ldots,n}$$

   from the pair of PPITs $(U_{i|S_{ij}}^{PVC}, U_{i+j|S_{ij}}^{PVC})$ using

   $$V_{i|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1}) := \partial_2 C_{i,i+j-1;S_{i,j-1}}^{SVC}(V_{i|S_{i,j-1}}^{SVC,k}(\hat{\theta}_{1:j-2}), V_{i+j-1|S_{i,j-1}}^{SVC,k}(\hat{\theta}_{1:j-2}); \hat{\theta}_{j-1,i}),$$
   $$V_{i+j|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1}) := \partial_1 C_{i+1,i+j;S_{i+1,j-1}}^{SVC}(V_{i+1|S_{i+1,j-1}}^{SVC,k}(\hat{\theta}_{1:j-2}), V_{i+j|S_{i+1,j-1}}^{SVC,k}(\hat{\theta}_{1:j-2}); \hat{\theta}_{j-1,i+1}).$$

   where for $j = 1$ and $i = 1, \ldots, d-1$ we set $V_{i|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1}) := V_i^k$ and $V_{i+j|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1}) := V_{i+j}^k$.

   (c) *For all $i = 1, \ldots, d-j$, test the hypothesis $H_0 \colon (U_{i|S_{ij}}^{PVC}, U_{i+j|S_{ij}}^{PVC}) \perp U_{S_{ij}}$ using a significance level of α/M for each individual test.*

   (d) *If at least one H_0 is rejected, stop the testing procedure and reject the null hypothesis that the simplifying assumption holds. Otherwise, increment the tree index j by one and start over with step 2a.*

The hierarchical procedure tests $M = (d-1)(d-2)/2$ hypotheses for a d-dimensional vine copula. In order to control the family-wise error rate, we apply the Bonferroni correction and test each hypothesis at a level of α/M. The rejection of an individual hypothesis $H_0 \colon (U_{i|S_{ij}}^{\text{PVC}}, U_{i+j|S_{ij}}^{\text{PVC}}) \perp U_{S_{ij}}$ means that the $(j-1)$-th order partial copula $C_{i,i+j;S_{ij}}^{\text{PVC}}$ does not equal the conditional copula $C_{i,i+j;S_{ij}}$. Thus, the hierarchical procedure detects critical building blocks of a vine copula model where an unconditional copula does not seem to be adequate and the modeling of a conditional copula may be required. See Schellhase and Spanhel (2018) who use this procedure to estimate (non-)simplified vine copulas. The testing procedure given in Definition 5 is also in line with the common sequential specification and estimation of simplified vine copulas and can be integrated in model selection algorithms as demonstrated in Kraus and Czado (2017).

4 Tests for higher-order partial copulas

In this and the following section, we develop a test that can be used in step 2c in Definition 5 to test the null hypothesis $H_0 \colon (U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}) \perp U_{S_{ij}}$. The main challenge of testing this null hypothesis is that the dimension $j-1$ of $U_{S_{ij}}$ can be rather large so that the power of consistent tests is not satisfying in practice if we do not only consider the second or third tree of a vine copula. For instance, a consistent test for the H_0 could be obtained using a Cramér-von Mises type test for vectorial independence testing (Kojadinovic and Holmes (2009) and Quessy (2010)). However, as it is pointed out by Gijbels et al. (2017a) and shown in our simulation results, such a consistent test suffers dramatically from the curse of dimensions, i.e., its power rapidly approaches the significance level if the dimension increases. Therefore, we do not aim to develop a consistent test which is only useful for a very small dimension $j-1$ but intend to develop a test that is quite robust to the dimension of the data set and exhibits good power properties for alternatives that one encounters in practical applications. In the following, we introduce a test that considers the null hypothesis that the conditional correlation of the PPITs associated to one edge of a vine is constant w.r.t. the conditioning variables $U_{S_{ij}}$ if the simplifying assumption is true. To obtain a test whose power does not collapse substantially with the dimension of the conditioning variables, we now discretize the support of the conditioning variables into a finite number of subsets and later introduce a penalty in the test statistic.

4.1 Constant conditional correlation (CCC) test for higher-order partial copulas

We first introduce the idea of discretizing the support of the conditioning variables into a finite number of partitions.[3] For the ease of exposition, assume for the moment that a sample from the PPITs is directly observable and there is no need to estimate their pseudo-observations. By $(U_{i|S_{ij}}^{\mathrm{PVC},k}, U_{i+j|S_{ij}}^{\mathrm{PVC},k}, U_{S_{ij}}^{k})_{k=1:n}$ we denote n independent copies of the vector $(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}, U_{S_{ij}})$. Let $\Lambda_0 := \mathrm{supp}(U_{S_{ij}}) \subset [0,1]^{j-1}$, $\Lambda_1, \Lambda_2 \subset \Lambda_0$ with $\Lambda_1 \cap \Lambda_2 = \emptyset$, and $\mathbb{P}(U_{S_{ij}} \in \Lambda_1), \mathbb{P}(U_{S_{ij}} \in \Lambda_2) > 0$. We call $\Gamma := \{\Lambda_1, \Lambda_2\}$ a partition of the support Λ_0 into two disjoint subsets. We are interested in the correlation between $U_{i|S_{ij}}^{\mathrm{PVC}}$ and $U_{i+j|S_{ij}}^{\mathrm{PVC}}$ in the

---

[3]The idea of discretization has some similarity to the boxes approach of Derumigny and Fermanian (2017) but differs substantially. We only discretize the conditioning variables and the rejection of our null hypothesis is still a rejection of the simplifying assumption which is not always true for the approach of Derumigny and Fermanian (2017). Moreover, we present a data-driven approach to select the partition so that the idea of discretization can also be applied in high-dimensional settings without the need to impose strong a priori assumptions on the form of the partition.

two subgroups determined by Γ, i.e.,

$$r_l := \mathrm{Corr}(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}|U_{S_{ij}} \in \Lambda_l)$$

$$= \frac{\mathrm{Cov}(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}|U_{S_{ij}} \in \Lambda_l)}{\sqrt{\mathrm{Var}(U_{i|S_{ij}}^{\mathrm{PVC}}|U_{S_{ij}} \in \Lambda_l)\mathrm{Var}(U_{i+j|S_{ij}}^{\mathrm{PVC}}|U_{S_{ij}} \in \Lambda_l)}},$$

for $l = 1, 2$. Note that if the simplifying assumption holds up to the j-th tree these conditional correlations r_l coincide with conditional Spearman's ρ. Under the H_0 : $(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}) \perp U_{S_{ij}}$, it follows that

$$\mathrm{Corr}(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}}) = r_1 = r_2,$$

i.e., the conditional correlations are constant w.r.t. the conditioning event.

To estimate the correlation in the l-th group we use the sample version

$$\hat{r}_l := \frac{\hat{\sigma}_l}{\sqrt{\hat{\sigma}_{1,l}^2 \hat{\sigma}_{2,l}^2}},$$

with

$$\hat{\sigma}_l := \frac{1}{n\hat{\pi}_l} \sum_{k=1}^n \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_l\}}(U_{i|S_{ij}}^{\mathrm{PVC},k} - \hat{\mu}_{1,l})(U_{i+j|S_{ij}}^{\mathrm{PVC},k} - \hat{\mu}_{2,l}),$$

$$\hat{\sigma}_{1,l}^2 := \frac{1}{n\hat{\pi}_l} \sum_{k=1}^n \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_l\}}(U_{i|S_{ij}}^{\mathrm{PVC},k} - \hat{\mu}_{1,l})^2$$

and

$$\hat{\mu}_{1,l} := \frac{1}{n\hat{\pi}_l} \sum_{k=1}^n \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_l\}} U_{i|S_{ij}}^{\mathrm{PVC},k}.$$

For the second PPIT $U_{i+j|S_{ij}}^{\mathrm{PVC}}$, the estimates $\hat{\sigma}_{2,l}$ and $\hat{\mu}_{2,l}$ are defined analogously. The random scaling factor

$$\hat{\pi}_l := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_l\}}$$

is the fraction of data corresponding to the subset Λ_l.

A statistic for testing the equality of the correlations in the two samples is given by

$$T_n^\star(\Gamma) = n\frac{(\hat{r}_1 - \hat{r}_2)^2}{\hat{\sigma}^2(\hat{r}_1) + \hat{\sigma}^2(\hat{r}_2)},$$

where $\hat{\sigma}^2(\hat{r}_l), l = 1, 2$, is a consistent estimator (see Appendix A.2) for the asymptotic variance of $\sqrt{n}(\hat{r}_l - r_l)$. By construction of the estimators and because a sample from the

PPITs is observable by assumption, the asymptotic covariance between $\hat{r}_1$ and $\hat{r}_2$ is zero. Thus, under regularity conditions and the H_0 it can be readily verified that $T_n^\star(\Gamma) \xrightarrow{d} \chi^2(1)$.

In a more general setting, one can also use a partition of the support Λ_0 into $L \in \mathbb{N}$ pairwise disjoint subsets $\Gamma := \{\Lambda_1, \ldots, \Lambda_L\}$ and test whether

$$H_0 : r_1 = \ldots = r_L \quad \text{vs.} \quad H_A : \exists l_1, l_2 \in \{1, \ldots, L\}, l_1 \neq l_2 : r_{l_1} \neq r_{l_2}.$$

For this purpose, denote the vector of sample correlations in the groups by $\hat{R}_\Gamma^\star = (\hat{r}_1, \ldots, \hat{r}_L)^T$. Further, define the diagonal $L \times L$ matrix $\hat{\Sigma}_{R_\Gamma}^\star$, with diagonal elements $\hat{\Sigma}_{R_\Gamma, l, l}^\star = \hat{\sigma}^2(\hat{r}_l)$ and the $(L-1) \times L$ first-order difference matrix

$$A = \begin{bmatrix} 1 & -1 & 0 & \cdots & & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -1 & 0 \\ 0 & \cdots & & 0 & 1 & -1 \end{bmatrix},$$

so that $(A\hat{R}_\Gamma^\star)^T A\hat{R}_\Gamma^\star = \sum_{l=1}^{L-1} (\hat{r}_l - \hat{r}_{l+1})^2$. A statistic to test the equality of correlations in L groups is then defined by the quadratic form[4]

$$T_n^\star(\Gamma) = n(A\hat{R}_\Gamma^\star)^T (A\hat{\Sigma}_{R_\Gamma}^\star A^T)^{-1} A\hat{R}_\Gamma^\star.$$

The asymptotic distribution of the resulting test statistic when pseudo-observations from the PPITs are estimated is stated in Proposition 2.

Proposition 2

Let $(X_{1:d}^k)_{k=1,\ldots,n}$ be n independent copies from $X_{1:d}$, $C_{1:d}^{SVC}(\cdot; \theta_{1:d-1})$ be a parametric simplified vine copula model for the PVC $C_{1:d}^{PVC}$ of $X_{1:d}$ and $(i,j) \in \mathcal{I}_2^d$ be fixed. Assume that the regularity conditions stated in Theorem 1 in Hobæk Haff (2013) hold and that the partition $\Gamma := \{\Lambda_1, \ldots, \Lambda_L\}$, where $\Lambda_1, \ldots, \Lambda_L \subset \Lambda_0 \subset [0,1]^{j-1}$, satisfies

(i) $\Lambda_{l_1} \cap \Lambda_{l_2} = \emptyset$, for $l_1 \neq l_2$ with $1 \leq l_1, l_2 \leq L$,

(ii) $\mathbb{P}(U_{S_{ij}} \in \Lambda_l) > 0$ for all $1 \leq l \leq L$.

Let $\hat{\Sigma}_{R_\Gamma} := \hat{\Sigma}_{R_\Gamma}^\star + \hat{\Sigma}_{PVC} + \hat{\Sigma}_r$, where $\hat{\Sigma}_{R_\Gamma}$ is defined in Appendix A.2, and $\hat{R}_\Gamma$ denote the vector of sample correlations that are computed using the pseudo-observations from the

---

[4] The statistic $T_n^\star(\Gamma)$ also follows from $(B\hat{R}_\Gamma^\star)^T B\hat{R}_\Gamma^\star = \sum_{l=1}^{L-1} \hat{\pi}_l(\hat{r}_l - \bar{r})^2$, where $\bar{r}$ is the average correlation (see Appendix A.1). That is, the statistic can be written in terms of weighted differences of the correlations to the average correlation.

PPITs. Construct the test statistic

$$T_n(\Gamma) = n(A\hat{R}_\Gamma)^T(A\hat{\Sigma}_{R_\Gamma}A^T)^{-1}A\hat{R}_\Gamma.$$

Under the $H_0 : (U^{PVC}_{i|S_{ij}}, U^{PVC}_{i+j|S_{ij}}) \perp U_{S_{ij}}$ *it holds that*

$$T_n(\Gamma) \xrightarrow{d} \chi^2(L-1).$$

Proof. The proof is given in Appendix A.2. ∎

The matrices $\Sigma_{\text{PVC}} = \text{plim}_{n\to\infty}\hat{\Sigma}_{\text{PVC}}$ and $\Sigma_r = \text{plim}_{n\to\infty}\hat{\Sigma}_r$ quantify the change in the asymptotic covariance matrix due to the estimation of pseudo-observations from the PPITs. If the marginal distributions are known and we don't have to estimate ranks to obtain pseudo-observations from $C^{\text{PVC}}_{1:d}$ it follows that $\Sigma_r = 0$. If the PVC $C^{\text{PVC}}_{1:d}$ is known it follows that $\Sigma_{PVC} = 0$. Note that the off-diagonal elements of $\Sigma_{R_\Gamma} = \text{plim}_{n\to\infty}\hat{\Sigma}_{R_\Gamma}$, i.e., the asymptotic covariances between estimated correlations in different groups, are, in general, no longer zero if observations from the PPITs are estimated.

4.2   CCC test for higher-order partial copulas: Combining partitions

Whether the test proposed in Proposition 2 rejects the null hypothesis if the conditional correlation $\text{Corr}(U^{\text{PVC}}_{i|S_{ij}}, U^{\text{PVC}}_{i+j|S_{ij}}|U_{S_{ij}})$ is not constant depends on the chosen partition Γ. To illustrate that the choice of the partition matters and to motivate the construction of a test based on the combination of several partitions, we use the following Example 1. It consists of the D-vine copula representation of the four-dimensional Clayton copula where the conditional copula in the last tree is replaced by a Frank copula having a parameter which depends on the two conditioning variables U_2 and U_3.

Example 1
Let $C^{Cl}(\theta)$ *and* $C^{Fr}(\theta)$ *be the cdfs of the Clayton and Frank copula with parameter* θ, *respectively. The building-blocks of the four-dimensional D-vine copula are chosen to be*

$$C_{12} = C_{23} = C_{34} = C^{Cl}(\theta_1),$$
$$C_{13;2} = C^{PVC}_{13;2} = C_{24;3} = C^{PVC}_{24;3} = C^{Cl}(\theta_2),$$
$$C_{14;23} = C^{Fr}(\alpha(u_{2:3};\lambda)),$$
$$\alpha(u_{2:3};\lambda) = 1 + 2.5\lambda(1 - 1.5(u_2 + u_3))^2,$$
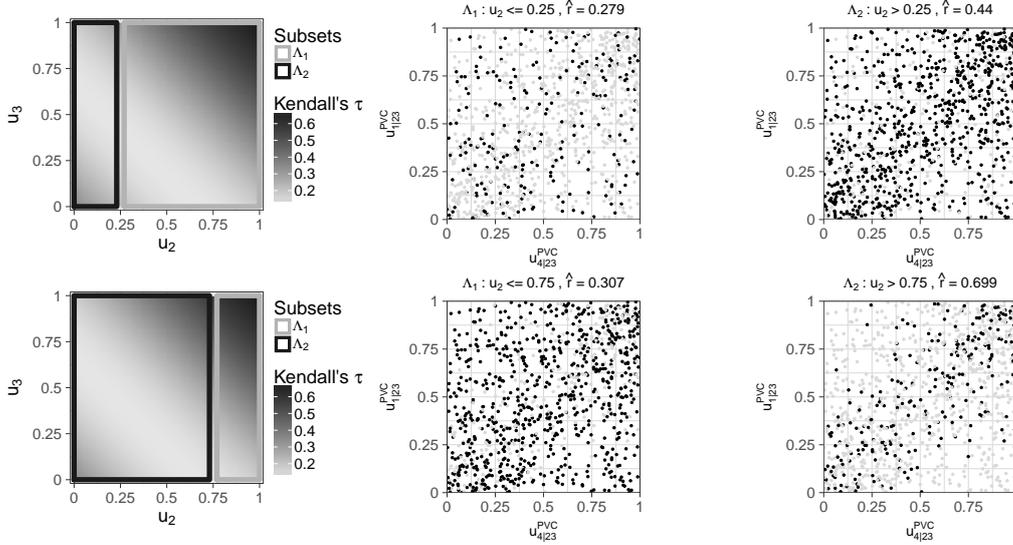
Figure 1: The plots correspond to a random sample of size 1000 from the data generating process defined in Example 1. On the left hand side the shaded background shows the value of Kendall's τ of $C_{14;23}$ as a function of U_2 and U_3. Areas with darker gray correspond to higher values of Kendall's τ. In the middle and on the right hand side, realizations from the PPITs of $C_{14;23}^{\text{PVC}}$ grouped according to Γ_1 in the upper row and Γ_2 in the lower row are shown. The black points show observations being assigned to the corresponding subset of the support Λ_0 and the light-gray points correspond to the observations which have been assigned to the other subset.

with $\theta_1 := \frac{2\tau}{1-\tau}$ and $\theta_2 := \frac{\theta_1}{1+\theta_1}$, where τ is the value of Kendall's τ.[5]

For the illustration we set in Example 1 $\tau = 0.4$ and $\lambda = 1$ and simulate a sample of size $n = 1000$. For instance, if we choose $\Lambda_1 = \{(u_2, u_3) \in [0,1]^2 : u_3 \leq u_2\}$ and $\Lambda_2 = \Lambda_0 \setminus \Lambda_1$, then $r_1 = r_2$ and the power of the test is asymptotically equal to the level of the test. Instead, we could use partitions such as $\Gamma_1 := \{\Lambda_1 = [0, 0.25] \times [0,1], \Lambda_2 = (0.25, 1] \times [0,1]\}$ or $\Gamma_2 := \{\Lambda_1 = [0, 0.75] \times [0,1], \Lambda_2 = (0.75, 1] \times [0,1]\}$. In Figure 1, we illustrate the resulting tests $T_n(\Gamma_1)$ and $T_n(\Gamma_2)$. The upper row corresponds to the first partition Γ_1 where the difference of the correlations in the two groups is $\hat{r}_2 - \hat{r}_1 = 0.161$, yielding a test statistic value of $T_n(\Gamma_1) = 5.41$. In contrast, if we consider the second partition Γ_2 shown in the lower row of Figure 1, we get $\hat{r}_2 - \hat{r}_1 = 0.392$ and $T_n(\Gamma_2) = 65.72$.

In order to increase the probability that the test will reject the null hypothesis if the conditional correlation is not constant, it seems naturally to consider not only one but a finite number of partitions $\Gamma_0, \ldots, \Gamma_M$, $M \geq 1$, where each partition Γ_m is a collection of $L_m \in \mathbb{N}$ subsets of the support Λ_0. A test statistic using a combination of several partitions

---

[5] The parameter function is a generalization of the function $\theta(X) = 1 + 2.5(3 - X)^2$ used by Acar et al. (2013) for a Frank copula with a one-dimensional conditioning set, where the conditioning variable X is assumed to be uniformly distributed on the interval $[2, 5]$. The four-dimensional Clayton copula, and not the Frank copula, is used in the lower trees as it can be represented as a simplified vine copula.

is given by

$$\Theta_n = \max\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_1), \ldots, T_n(\Gamma_M)\} - n\lambda_n, \qquad (4.1)$$

where λ_n is a penalty function. The construction of a test in such a manner has some similarity to the approach of Lavergne and Patilea (2008). The idea is that by choosing an appropriate penalty function the asymptotic distribution of Θ_n under the H_0 should be equivalent to the asymptotic distribution of $T_n(\Gamma_0)$. Precise conditions are given in the following proposition.

Proposition 3

Assume that the conditions stated in Proposition 2 hold and that the partitions $\Gamma_0, \ldots, \Gamma_M$ fulfill the conditions stated for Γ in Proposition 2. Additionally, let $\lambda_n : \mathbb{N} \to \mathbb{R}^+$ be a penalty function such that

(i) $n\lambda_n \to \infty$ for $n \to \infty$,

(ii) $\lambda_n \to 0$ for $n \to \infty$.

Set $\Theta_n = \max\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_1), \ldots, T_n(\Gamma_M)\} - n\lambda_n$, where $T_n(\Gamma_m), 0 \leq m \leq M$, is given as in Proposition 2. Under the $H_0 : (U^{PVC}_{i|S_{ij}}, U^{PVC}_{i+j|S_{ij}}) \perp U_{S_{ij}}$ it holds that

$$\Theta_n \xrightarrow{d} \chi^2(L_0 - 1).$$

If there is a partition $\Gamma_{m^\star}, m^\star \in \{0, \ldots, M\}$, such that $\mathrm{plim}_{n\to\infty} \frac{1}{n} T_n(\Gamma_{m^\star}) =: c > 0$ it follows that

$$\Theta_n \xrightarrow{p} \infty.$$

Proof. See Appendix A.3   ∎

Thus, the critical value of Θ_n under the H_0 only depends on Γ_0 but not on $\Gamma_1, \ldots, \Gamma_M$. Moreover, if there is a partition $\Gamma_{m^\star}$ such that the correlations conditional on the subsets of the partition are not identical, i.e., $c > 0$, the power of the test approaches 1 if the sample size goes to infinity.

5 A data-driven algorithm for testing with high-dimensional conditioning variables

In the previous chapter we introduced statistical tests for the hypothesis that the conditional correlation $\mathrm{Corr}(U^{PVC}_{i|S_{ij}}, U^{PVC}_{i+j|S_{ij}} | U_{S_{ij}})$ is constant. Both constant conditional correla-

tion (CCC) tests, $T_n(\Gamma)$ and Θ_n, are based on partitions of $\Lambda_0 := supp(U_{S_{ij}})$. We will now explain how such partitions can be defined and efficiently selected in a data-driven fashion.

5.1 Naive approaches

For a conditional copula $C_{i,i+2;i+1}$, $i = 1, \ldots, d-2$, in the second tree of a D-vine copula, we obtain a one-dimensional conditioning variable U_{i+1}. The support $[0,1]$ of U_{i+1} can be partitioned into two subsets according to the sample median $Q_{0.5}[(U_{i+1}^k)_{k=1:n}]$ of U_{i+1}. The subsets Λ_1^{i+1} and Λ_2^{i+1} are then defined as

$$\Lambda_1^{i+1} := \left\{ u_{i+1} \in [0,1] : u_{i+1} \leq Q_{0.5}[(U_{i+1}^k)_{k=1:n}] \right\} \text{ and } \Lambda_2^{i+1} := [0,1] \setminus \Lambda_1^{i+1},$$

resulting in the partition $\Gamma_{med}^{i+1} := \{\Lambda_1^{i+1}, \Lambda_2^{i+1}\}$. The corresponding groups of observations are given by $\left\{ (U_{i|i+1}^{\mathrm{PVC},k}, U_{i+2|i+1}^{\mathrm{PVC},k}) : 1 \leq k \leq n, U_{i+1}^k \in \Lambda_l^{i+1} \right\}$ with $l = 1, 2$.

For $j = 3$ one could consider the sample median of each conditioning variable to split the observations into groups and then consider the partitions that result from the combinations of these groups. That is, for the conditional copula $C_{i,i+3;i+1:i+2}$ we would obtain

$$\Gamma_{(1,1)} = \Lambda_1^{i+1} \cap \Lambda_1^{i+2}, \quad \Gamma_{(1,2)} = \Lambda_1^{i+1} \cap \Lambda_2^{i+2}, \quad \Gamma_{(2,1)} = \Lambda_2^{i+1} \cap \Lambda_1^{i+2}, \quad \Gamma_{(2,2)} = \Lambda_2^{i+1} \cap \Lambda_2^{i+2}.$$

Generalizing this strategy for $j \geq 3$ a resulting partition Γ_m has the form

$$\Gamma_m = \bigcap_{k=1}^{j-1} \Lambda_{m_k}^{i+k},$$

where m_k is the k-th entry of a vector $m \in \{1,2\}^{j-1}$ and $\{1,2\}^{j-1}$ is the cartesian power of the set $\{1,2\}$. However, this approach is computationally only feasible for small j, since the number of partitions 2^{j-1} increases exponentially with the dimension of the conditioning variables. Moreover, the number of observations that are contained in a partition Γ_m might get too small.

Alternatively, one could use maps from $[0,1]^{j-1}$ to $[0,1]$ to aggregate the information. For example, one can use the mean

$$g : [0,1]^{j-1} \to [0,1], \qquad u_{S_{ij}} \mapsto g\left(u_{S_{ij}}\right) = \frac{1}{j-1} \sum_{k=1}^{j-1} u_{i+k}. \tag{5.1}$$

The resulting partition $\Gamma_{med} := \{\Lambda_1, \Lambda_2\}$ is then given by

$$\Lambda_1 := \left\{ u_{S_{ij}} \in \Lambda_0 : g(u_{S_{ij}}) \leq Q_{0.5}[(g(U_{S_{ij}}^k))_{k=1:n}] \right\} \text{ and } \Lambda_2 := \Lambda_0 \setminus \Lambda_1,$$

with $\Lambda_0 = \text{supp}(U_{S_{ij}}) \subset [0,1]^{j-1}$. This mean aggregation with the median as fixed split point provides a useful partition for the CCC test to detect a possible difference in the conditional correlations if there is a moderate positive pairwise dependence between all pairs in the conditioning set and the influence of the conditioning variables on the conditional copula is similar, e.g. high (low) values of the conditioning variable result in a copula with high (low) monotone dependence. However, in practice, we typically have no information on the functional form of the conditional copula, so that a justification for this and any other a priori determination of the partition is difficult. Therefore, we introduce in the following a decision tree algorithm which selects the partitions in a data-driven way and is computationally feasible also for large j.

5.2 A decision tree algorithm

The test statistic Θ_n can be rewritten in the following way

$$\Theta_n = \max\left\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_1), \ldots, T_n(\Gamma_M)\right\} - n\lambda_n$$
$$= \max\left\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_{max})\right\} - n\lambda_n,$$

with $\Gamma_{max} := \text{argmax}_{\Gamma_m \in \{\Gamma_1, \ldots, \Gamma_M\}} T_n(\Gamma_m)$. The set Γ_{max} denotes the partition for which a possible violation of the H_0 is most pronounced. To find Γ_{max} in a data-driven and computationally efficient way we use the decision tree shown in Figure 2. The decision tree recursively uses binary splits to partition the support $\Lambda_0 = \text{supp}(U_{S_{ij}})$ into disjoint subsets to obtain $\Gamma_{\max} := \{\Lambda_{(0,\gamma_1,\ldots,\gamma_{J_{\max}})} \subset \Lambda_0 : (0, \gamma_1, \ldots, \gamma_{J_{\max}}) \in \{0\} \times \{l,r\}^{J_{\max}}\}$, where $J_{\max}$ is the maximum depth of the tree. E.g., if $J_{\max} = 2$, then $\Gamma_{\max} := \{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$. The possible split points for each leaf $\Lambda_{(\gamma_0,\gamma_1,\ldots,\gamma_J)}, 0 \leq J \leq J_{\max}$, are given by the empirical quartiles of each conditioning variable and by the empirical quartiles of the mean aggregation of the conditioning vector given in (5.1). Among these possible splits the split is chosen that maximize the statistic of the CCC test. For algorithmic details and a formal definition of the decision tree algorithm we refer to Appendix A.4. In all simulations in Section 6 and the real data applications in Section 7, we choose $J_{\max} = 2$, and $\lambda_n = \frac{1}{\sqrt{n}}$.[6] We further set $\Gamma_0 = \Gamma_{med}$, because we have no a priori information about the relative importance of each conditioning variable and because the median as split point guarantees well-balanced sample sizes in the groups.

We illustrate the decision tree-based algorithm using the four-dimensional D-vine copula defined in Example 1. In Figure 3, the decision tree approach to test whether $C_{14;23}$ is a

---

[6] A detailed finite sample analysis and explanations of how to choose the penalty function are provided in Appendix A.6.
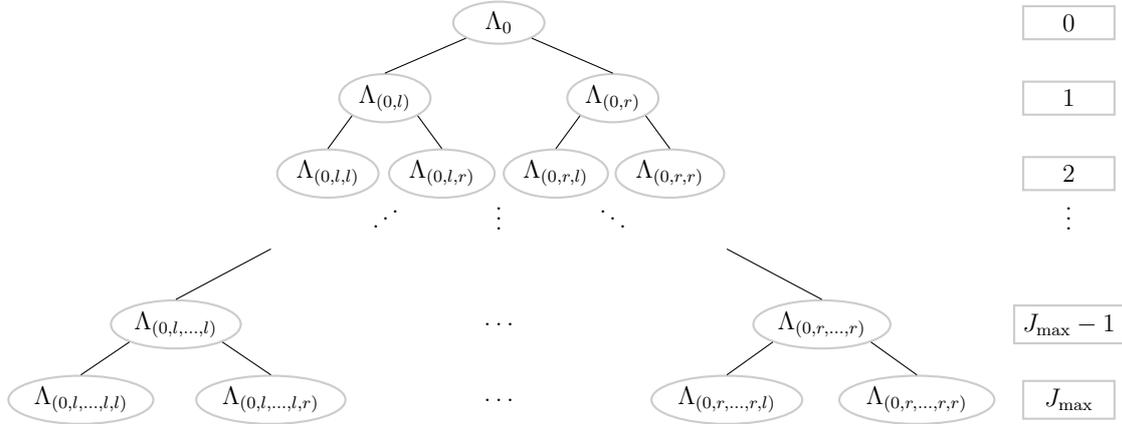
Figure 2: Partitioning of the support $\Lambda_0 = \mathrm{supp}(U_{S_{ij}}) \subset [0,1]^{j-1}$ of the random vector $U_{S_{ij}}$ into disjoint subsets $\Lambda_{\gamma_{0:J}}$, where $\gamma_{0:J} := (\gamma_0, \gamma_1, \ldots, \gamma_J) \in \{0\} \times \{l, r\}^J$, $0 \leq J \leq J_{\max}$, using a decision tree algorithm with maximum depth $J_{\max}$.

second-order partial copula is visualized via colored frames showing the subsets building the selected partition. On the left hand side of Figure 3, the shaded area in the background corresponds to the variation in Kendall's τ of the conditional copula $C_{14;23}$ as a function of U_2 and U_3. Areas with darker gray correspond to higher values of Kendall's τ. The decision tree partitions Λ_0 into subsets where the variation of the conditional correlation is rather constant. The chosen partition by the decision tree $\Gamma_{\max} := \{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$ results in $T_n(\Gamma_{\max}) = 130.54$ and $\Theta_n = 98.91$. Compared with the tests $T_n(\Gamma_1) = 5.41$ and $T_n(\Gamma_2) = 65.72$ for naively constructed partitions (see Figure 1), the decision tree identifies a more suitable partition $\Gamma_{\max}$ when applied to the same data set. On the right hand side of Figure 3 it is illustrated how the dependence within the conditioning set, determined by the Clayton copula $C_{23}(\theta_1)$, influences the variation of $C_{14;23}$. Shown are the realized values of (U_2, U_3) and their grouping into the subsets forming the partition chosen by the decision tree algorithm. The two plots on top correspond to the first binary split, which is done according to the 75% quantile of U_2. The two plots at the bottom show the splits in the second level of the decision tree.

The partitioning of Λ_0 into $\Gamma_{\max}$ is visualized in a different way in Figure 4 which shows the grouping of the observations from the PPITs $(U_{1|23}^{\mathrm{PVC}}, U_{4|23}^{\mathrm{PVC}})$ according to $\Gamma_{\max}$. In each scatter plot the black observations have been assigned to this leaf while the observations in light gray have been assigned to the other leaf due to the binary split of the observations from the mother leaf. Furthermore, the estimated correlations in each group, which are used for the CCC test, are shown. We see that the decision tree chooses a partition that results in estimated correlations that are quite different with a maximal difference of $\widehat{\mathrm{Corr}}(U_{1|23}^{\mathrm{PVC}}, U_{4|23}^{\mathrm{PVC}} | U_{2:3} \in \Lambda_{(0,l,l)}) - \widehat{\mathrm{Corr}}(U_{1|23}^{\mathrm{PVC}}, U_{4|23}^{\mathrm{PVC}} | U_{2:3} \in \Lambda_{(0,r,r)}) = 0.596$.
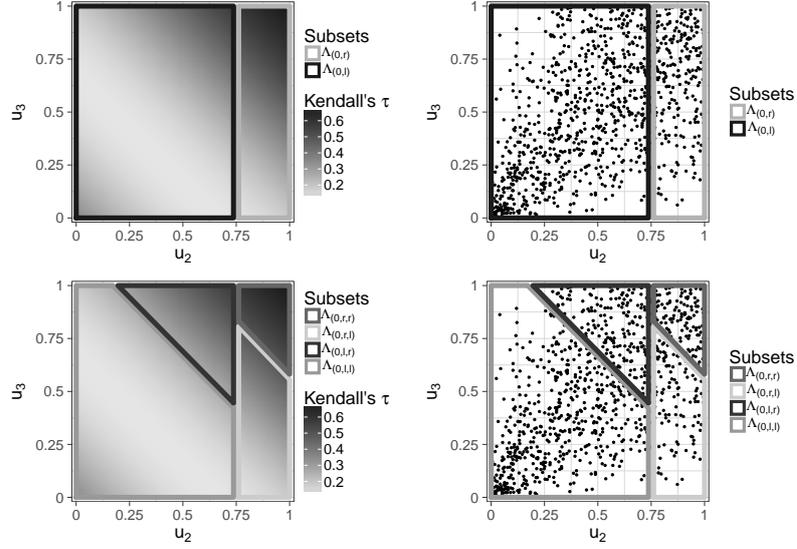
Figure 3: The plot correspond to the same 1000 observations from the D-vine copula (Example 1) as in Figure 1. On the left hand side the shaded background shows the value of Kendall's τ of $C_{14;23}$ as a function of U_2 and U_3. Areas with darker gray correspond to higher values of Kendall's τ. On the right hand side the realizations of the conditioning variables U_2 and U_3 are shown. The upper plots correspond to the first binary split of the decision tree and the lower plots to the second and third binary split. The colored frames show the different subsets $\{\Lambda_{(0,l)}, \Lambda_{(0,r)}, \Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$ of the support $\Lambda_0 := [0,1]^2$.
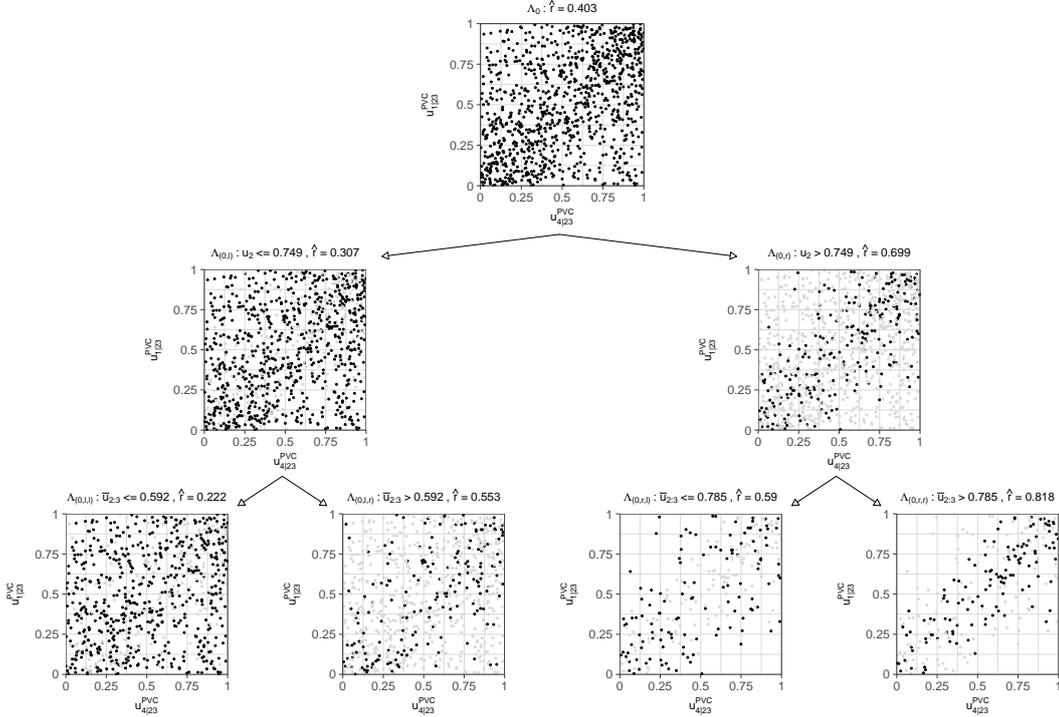
Figure 4: The plot correspond to the same 1000 observations from the D-vine copula (Example 1) as in Figure 1. Shown is the decision tree-based grouping of PPITs from $C_{14;23}^{\text{PVC}}$ into four different groups according to the partition $\Gamma_{\max} := \{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$, which is also shown in the lower plots of Figure 3. The black points show observations being assigned to the corresponding subset of the support Λ_0 and the light-gray points correspond to the observations which have been assigned to the other subset due to the binary split. The mean of the conditioning variables u_2 and u_3 is denoted by $\bar{u}_{2:3}$.

6 Simulation study

In the following, the finite-sample performance of the CCC test is analyzed and compared to the performance of the vectorial independence (VI) test of Kojadinovic and Holmes (2009). We will analyze the power of both tests w.r.t. the variation in the conditional copula, illustrate the power gain of the CCC test due to the decision tree algorithm, and investigate the performance of both tests w.r.t. the dimensionality of the testing problem. Additionally, we discuss two practical issues of testing the simplifying assumption in high-dimensional vine copulas, namely, the impact of misspecified copula families and vine structure selection.

All results for the CCC test are computed with estimated pseudo-observations using the steps described in Definition 5. Since the asymptotic distribution of the VI test with estimated pseudo-observations is unknown, we use the true observations from the PPITs for the VI test and compute p-values on the basis of 1000 bootstrap samples (Quessy 2010).

6.1 Power study: The functional form of the conditional copula

To alter the variation of the conditional copula in Example 1, we vary the parameter λ in the function

$$\alpha(u_{2:3}; \lambda) = 1 + 2.5\lambda(1 - 1.5(u_2 + u_3))^2 \tag{6.1}$$

between zero and one. For $\lambda = 0$, the copula $C_{14;23}$ does not vary at all and for $\lambda = 1$ the variation is most pronounced. In Figure 5, the variation in Kendall's τ of the Frank copula $C_{14;23}$ as a function of the mean $\bar{u}_{2:3} = \frac{1}{2}(u_2 + u_3)$ is shown on the left hand side.[7] For the sample sizes $n = 500, 1000$, and $\lambda = 0, 0.2, 0.4, 0.6, 0.8, 1$, we apply the CCC test and the VI test (Kojadinovic and Holmes 2009) for the hypothesis $H_0 : C_{14;23} = C_{14;23}^{\mathrm{PVC}}$.

On the right hand side of Figure 5, empirical power values are plotted for different values of the parameter λ. The numbers are based on 1000 samples for each combination of λ and n. The level of the tests is chosen to be 5%. For both tests and sample sizes the empirical size (i.e. the case $\lambda = 0$) is close to the theoretical level of the test. The empirical power of both tests is clearly increasing for all values of λ if one doubles the sample size from 500 to 1000 observations. Furthermore, both tests are more powerful the more the variation in the conditional copula is pronounced, i.e., the larger the parameter λ. Overall the empirical power is much better for the CCC test than for the VI test. In terms of empirical power,

---

[7] Note that we have already seen $\tau_{14|23}$ for $\lambda = 1$ as a function of u_2 and u_3 as shaded background in the plots on the left hand side of Figure 1 and Figure 3.

116
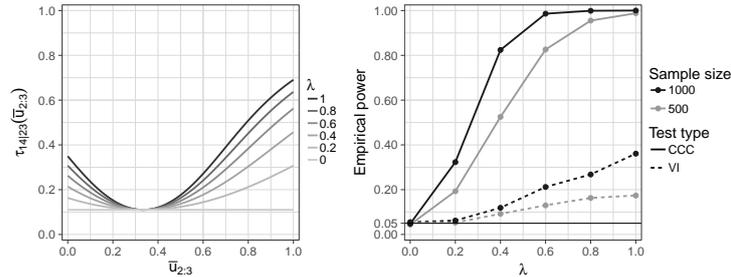
Figure 5: On the left hand side the value of Kendall's τ of the conditional Frank copula $C_{14;23}$ as a function of the mean $\bar{u}_{2:3}$ of the conditioning variables u_2 and u_3 is shown for different values of the parameter λ. The parameter function is stated in Equation 6.1 and the data generating process is defined in Example 1. The plot on the right hand side shows the empirical power for different values of λ and the theoretical 5% level of the tests. Empirical power values are based on 1000 samples. The solid lines correspond to the CCC test and the dashed lines to the VI test. Different sample sizes are shown color coded.

the CCC test outperforms the VI test in all settings with a relative improvement that often exceeds 300%.

6.2 Power study: Gain of power by using the decision tree algorithm

We now compare the CCC test based on the decision tree approach Θ_n with the CCC test $T_n(\Gamma_0)$ where the partition Γ_0 is pre-selected. By construction, $\Theta_n \geq T_n(\Gamma_0)$ always holds, meaning that if we reject based on $T_n(\Gamma_0)$, we also reject based on Θ_n. As a consequence, the empirical power of Θ_n is never smaller than the empirical power of $T_n(\Gamma_0)$. The improvement in power due to the use of Θ_n instead of $T_n(\Gamma_0)$ depends on the data generating process and will be investigated in the following

As in Section 5, we choose $\Gamma_0 = \Gamma_{med}$, i.e., we use mean aggregation and the median as fixed split point. As data generating processes we consider the vine copula in Example 1 and the resulting vine copulas that arise if the parameter of the conditional Frank copula $C_{14;23}$ in the last tree of Example 1 is given by

$$\alpha_I(u_{2:3}; \lambda) = 1 + 2.5\lambda(1 - 2u_2(u_2 + u_3))^2, \quad \text{or} \quad \alpha_D(u_{2:3}; \lambda) = 1 + 2.5\lambda(1 - 2(u_2 - u_3))^2.$$

Instead of summing up the two conditioning variables as in $\alpha(\cdot)$ (Equation 6.1), $\alpha_I(\cdot)$ exhibits an interaction effect between the conditioning variables and $\alpha_D(\cdot)$ takes the difference of the conditioning variables. Figure 6 shows the empirical power of the CCC tests Θ_n and $T_n(\Gamma_0)$ for the hypothesis $H_0 : C_{14;23} = C_{14;23}^{\text{PVC}}$. For the case of Example 1 (left panel in Figure 6), the test with the fixed partition $\Gamma_0 = \Gamma_{med}$ delivers a test $T_n(\Gamma_0)$ which per-

forms almost as good as Θ_n. That is because the parameter $\alpha(u_{2:3}; \lambda)$ of the conditional copula in Example 1 can be written as a function of the mean of the conditioning variables $\bar{u}_{2:3} = \frac{1}{2}(u_2 + u_3)$. Furthermore, the conditioning variables are positively associated due to the Clayton copula with $\tau_{23} = 0.4$. As a result, the decision tree rarely finds a better partition than the fixed partition $\Gamma_0 = \Gamma_{med}$.

For the other two cases, the partition Γ_{med} is not a good choice and the decision tree algorithm finds substantially better partitions in a data-driven way. The varying parameter $\alpha_I(\cdot)$ (middle panel in Figure 6) introduces an interaction effect between the two conditioning variables. Although the test with the fixed Γ_{med} partition can detect some of the variation in the conditional copula, the decision tree finds better partitions which can increase the empirical power by more than 20 percentage points. The gain of power is even more pronounced if the parameter $\alpha_D(\cdot)$ (right panel in Figure 6) of the conditional copula is a function of the difference of the conditioning variables. Even if $\lambda = 1$ and the conditional copula is strongly varying, the test with the fixed partition Γ_{med} can not recognize the variation in the conditional copula because it uses a partition that is based on the mean of the conditioning variables. As a result, the empirical power is identical to the level of the test. On the contrary, the data-driven selection of the partition results in a substantial power increase even though the algorithm can not directly split the support of the conditioning variables w.r.t. the difference of the conditioning variables. For $\lambda = 1$ and $n = 1000$, the data-driven selection of the partition increases the power from 5% to 99%.
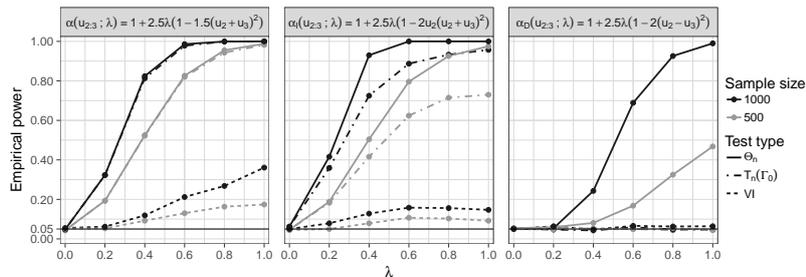
Figure 6: Empirical power of the CCC tests Θ_n and $T_n(\Gamma_0)$ and the VI test for the hypothesis that the copula $C_{14;23}$ in Example 1 is a second-order partial copula. The three panels correspond the parameter functionals $\alpha(\cdot)$, $\alpha_I(\cdot)$ and $\alpha_D(\cdot)$, respectively. Empirical power values are based on 1000 samples and plotted against the parameter λ. The solid lines correspond to the CCC test Θ_n, the dashdotted lines to the CCC test $T_n(\Gamma_0)$ and the dashed lines to the VI test. Different sample sizes are shown color coded.

In summary, the choice of Γ_0 determines a lower bound for the empirical power of the test Θ_n and by applying Θ_n with the decision tree one obtains a more powerful test. The magnitude of the power improvement depends on the data generating process and ranges from negligible (e.g., $\alpha(\cdot)$) to huge (e.g., $\alpha_D(\cdot)$). For all data generating processes, the power of the data-driven test Θ_n is much better than the power of the VI test. The difference is

most pronounced for $\alpha_D(\cdot)$ where the empirical power of the VI test is always approximately 5% while the empirical power of the CCC test Θ_n can be 99%.

6.3 Power study: The dimension of the conditioning set

For high-dimensional vine copulas, the dimension of the conditioning set of a conditional copula increases rapidly. Therefore, it is substantial that a test for the constancy of a conditional copula still has power if the dimension of the conditioning set is not small. To investigate the performance of the CCC test w.r.t. the dimension of the conditioning set, we start with a up to twelve-dimensional Clayton copula which can be represented as a D-vine copula consisting of Clayton copulas. We then replace the Clayton copula in the edge of the last tree by a Frank copula with varying parameter and investigate the performance of the CCC test. The data generating process being analyzed is defined in Example 2.

Example 2
For $d \geq 4$, the building-blocks of the d-dimensional D-vine copula are chosen to be

$$
\begin{aligned}
C_{i,i+j;S_{i,j}} &= C^{Cl}(\theta_j), \qquad 1 \leq j \leq d-2, 1 \leq i \leq d-j, \\
C_{1,d;2:(d-1)} &= C^{Fr}(\alpha(u_{2:3}; \lambda)), \\
\alpha(u_{2:3}; \lambda) &= 1 + 2.5\lambda(1 - 1.5(u_2 + u_3))^2, \\
\theta_j &= \frac{\theta_1}{1 + (j-1)\theta_1}, \qquad 2 \leq j \leq d-2, \\
\theta_1 &= \frac{2\tau}{1 - \tau}.
\end{aligned}
$$

For $d = 4$ Example 2 coincides with Example 1 and as before we set $\tau = 0.4$ and consider different values for λ. For $d = 4, 6, 8, 10, 12$, we always use the same varying Frank copula in the last tree where the functional form of the parameter only depends on the conditioning variables U_2 and U_3. Therefore, the variation in the conditional copula is always the same and not influenced by d but the testing problem is getting higher-dimensional with d. Grouped by the dimension d, the empirical power and size of the VI and the CCC test are shown in Figure 7.

While the VI test suffers a lot from the curse of dimensions if the dimension of the conditioning set is increased, the empirical power of the CCC test is only slightly decreasing for higher values of d. Indeed, it is remarkable how the CCC test performs in comparison to the VI test. In particular, for the setup $\lambda = 1$ and $n = 1000$, the power of the VI test drops from 36% to 5% if the dimension is increased from $d = 4$ to $d = 12$. On the contrary, the power of the CCC test for this setup is always 100%. Moreover, even when the power
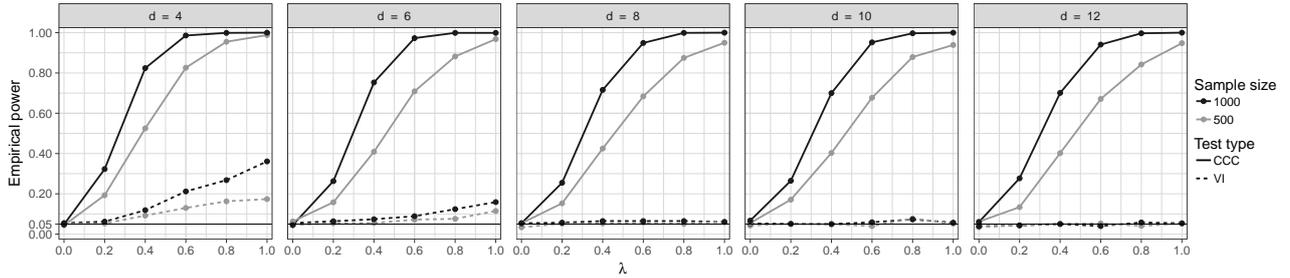
Figure 7: Empirical power of the CCC and VI test for the hypothesis that the copula $C_{1,d;2:(d-1)}$ in Example 2 is a $(d-2)$-th order partial copula. Empirical power values are based on 1000 samples. The empirical power is plotted against the parameter λ. Each plot corresponds to a specific dimension d of the D-vine copula (Example 2). For $\lambda = 0$ the plot shows the empirical size of the tests. The solid lines correspond to the CCC test and the dashed lines to the VI test. Different sample sizes are shown color coded.

of the CCC test is not 100% for $d = 4$, the decrease in its power is still marginal. For instance, for $\lambda = 0.6$ and $n = 500$, the power of the CCC test only decreases from 83% to 67% while the power of the VI test quickly drops to 5% if the dimension is increased from $d = 4$ to $d = 12$. Thus, the introduction of a penalty in the CCC test statistic and the data-driven selection of the partition $\Gamma_{\max}$ by means of a decision tree yields a test whose power decreases only slightly with the dimension of the conditioning variables.

6.4 Practical issues: Misspecification of the copulas in the lower trees

The true family of the five copulas in the first and second tree $(C_{12}, C_{23}, C_{34}, C_{13;2}, C_{24;3})$ in Example 1 is the Clayton copula. To analyze the effect of misspecified copula families, we now vary the pairwise value of Kendall's tau $\tau_{23} = \tau_{12} = \tau_{34} = \tau_{13} = \tau_{24}$ between 0 and 0.8 and estimate either survival Gumbel or Gumbel copulas for all five copulas in the lower trees. The top row of Figure 8 shows the results for correctly specified Clayton copulas in the lower trees as a benchmark. Since the strength of the variation of $C_{14;23}$ is more pronounced for higher values of τ_{23}, the empirical power of the tests is also increasing in τ_{23}. The empirical size of the tests ($\lambda = 0$) is not influenced by τ_{23} and always close to the theoretical level of 5%.

The second row of Figure 8 corresponds to a rather mild misspecification where we estimate survival Gumbel copulas in the first and second tree. We see that the empirical size is still very close to the theoretical level of 5% ($\lambda = 0$). Moreover, the power of the test with misspecified survival Gumbel copulas is almost indistinguishable from the power of the test with correctly specified Clayton copulas. If the degree of misspecification is severe
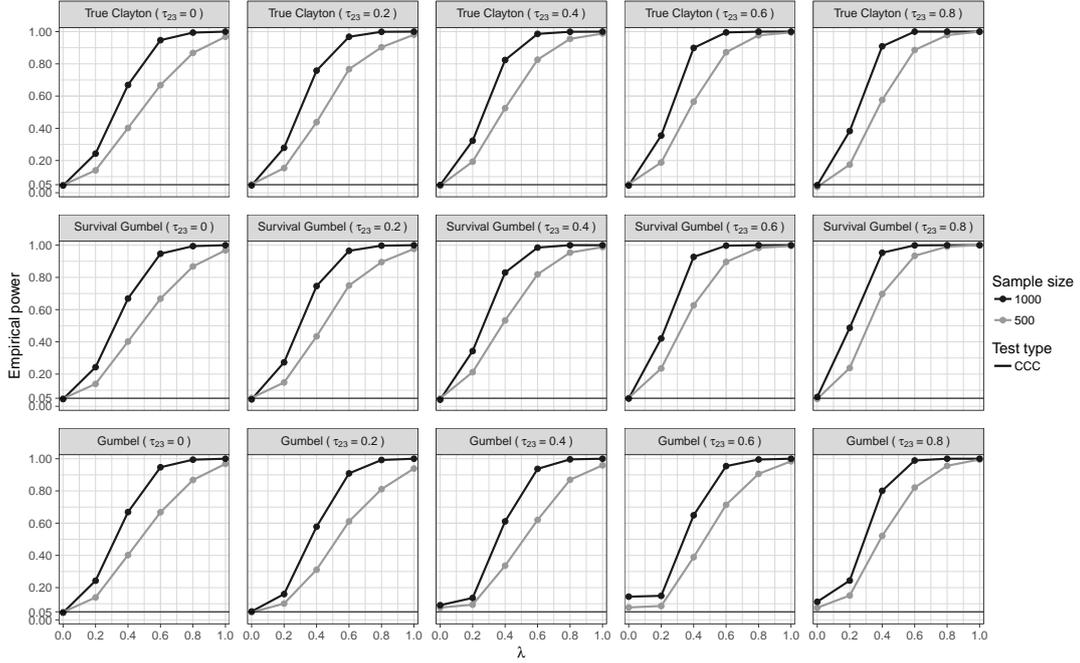
Figure 8: Empirical power of the CCC test for the hypothesis that the copula $C_{14;23}$ in Example 1 is a second-order partial copula. Empirical power values are based on 1000 samples and are plotted against the parameter λ. For $\lambda = 0$ the plot shows the empirical size of the test. Each plot corresponds to a specific value of Kendall's τ_{23} for the true copulas in the first and second tree. In the top row the PPITs are estimated with the true Clayton copula family in the first and second tree ($C_{12}, C_{23}, C_{34}, C_{13;2}, C_{24;3}$). The second and third row show the results if the PPITs are estimated with misspecified copula families (survival Gumbel and Gumbel, respectively). Different sample sizes are shown color coded.

and we fit Gumbel copulas (with upper tail dependence) to data generated from Clayton copulas (with lower tail dependence), differences in the empirical power of the CCC test become visible in the third row of Figure 8. In the majority of the considered scenarios the empirical power is now smaller. In cases with high dependence, i.e., $\tau_{23} = 0.4, 0.6, 0.8$, the empirical size is increased and no longer as close to the theoretical level than without misspecification. This shows that the test might not control the size if the copula families in the lower trees are severely misspecified. Note that we misspecify five copula families and that the misspecification in the second tree might be even worse because the data in the edges of the second tree is no longer generated by Clayton copulas if the copulas in the first tree are misspecified. Thus, the performance of the CCC test is relatively robust w.r.t. such a severe misspecification.

6.5 Practical issues: Model selection and the hierarchical test procedure

Throughout the paper we consider the scenario where the hypothesis about the simplifying assumption is formulated for a given vine copula structure. In practice, the vine structure and copula families are not known and have to be selected by some model selection algorithms. Note that, in general, depending on the vine structure, a copula can satisfy or violate the simplifying assumption (see Hobæk Haff et al. (2010) for examples). Thus, in order to investigate the effect of vine copula model selection on the empirical power of the CCC test, it is reasonable to consider exchangeable data generating processes for which the simplifying assumption is either satisfied or violated for any structure. In this case, the null hypothesis is independent of the vine structure and it is sound to analyze the resulting size and power of the CCC test due to model selection.

For this reason, we generate data from the Clayton, Gaussian, Gumbel, and Frank copula in four dimensions with pair-wise values of Kendall's τ of $0.2, 0.4, 0.6, 0.8$. Then, we use the algorithm of Dißmann et al. (2013) to select a regular vine copula structure and to specify the copula families via the AIC. We apply the hierarchical procedure (Definition 5) to test the simplifying assumption at a theoretical level of 5%. Thus, in the second tree two conditional copulas are tested with an individual level of 1.67%. If we do not reject the H_0 for both copulas in the second tree, we continue in the third tree and test $C_{14;23}$ at an individual level of 1.67%. Therefore, each test of the simplifying assumption for the considered four-dimensional copulas consists of up to three individual tests.

The first two panels in Figure 9 report the results for the four-dimensional Clayton and Gaussian copula for which any structure satisfies the simplifying assumption (Stöber et al. 2013). The empirical size of the CCC test is still very close to the theoretical level of the
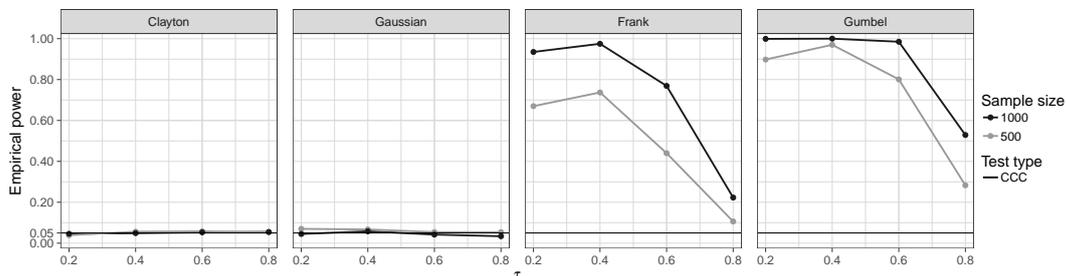
Figure 9: Empirical power for the sequential procedure (Definition 5) with the CCC test applied to test the simplifying assumption for the four-dimensional Clayton, Gaussian, Frank and Gumbel copula. Regular vine copula models are selected with the algorithm of Dißmann et al. (2013). Empirical power values are based on 1000 samples and plotted against the pairwise value of Kendall's τ. Different sample sizes are shown color coded.

test even under the consideration of vine copula model selection and possibly misspecified

copula families. On the right of Figure 9, the empirical power of testing the simplifying assumption for the four-dimensional Frank and Gumbel copula is plotted. The Frank and Gumbel copula slightly violate the simplifying assumption for every vine structure (Stöber et al. 2013) as long as $\tau \notin \{0, 1\}$. Although the variation in the conditional copulas induced by the four-dimensional Frank and Gumbel copulas is rather mild,[8] the CCC test often rejects the simplifying assumption. That the power has a minimum at $\tau = 0.8$ can be explained by the fact that both copulas satisfy the simplifying assumption for $\tau \to 1$.

7 Real data applications

We now analyze the performance of the proposed test procedure (Definition 5 and Definition 6) for regular vine copulas fitted to ten different data sets. Two different kinds of real data are considered: On the one hand, prominent data sets from the vine copula literature, and on the other hand, filtered financial returns which have been the subject of study in many applied vine copula research papers. The dimensionality of the data varies between 3 and 49 and the number of observations between 312 and 23,909. The two non-financial data sets are `uranium` (Cook and Johnson 1986) and `concrete` (Yeh 1998). For both data sets, normalized ranks as pseudo-observations from the copula are obtained by means of the rescaled ecdf. All eight financial data sets are from the *Kenneth R. French – Data Library* and we apply ARMA(1,1)-GARCH(1,1)-filtering (Engle 1982, Bollerslev 1986) and the rescaled ecdf to the residuals to obtain pseudo-observations from the d-dimensional copulas. Note that by applying the rescaled ecdf to the residuals, our test with pseudo-observations is still an asymptotically correct test (Chen and Fan (2006)).

To obtain parametric models for the PVC we apply the standard regular vine model selection algorithm proposed by Dißmann et al. (2013) which is implemented in the R-package `VineCopula` (Schepsmeier et al. 2017). The pair-copula families are selected according to the AIC and we use the option to test for the independence copula in each node of the vine copula by applying the test of Genest and Favre (2007) to obtain a sparse parametric simplified vine copula model.

In Table 1 and Table 2 we provide detailed information for all ten data sets being studied and also report the test results with the hierarchical procedure (Definition 5 and Definition 6) to check the validity of the simplifying assumption. For all cases where we reject the H_0 that the simplifying assumption is satisfied, we also report the first tree in which

---

[8] For example, the induced conditional copula $C_{13;2}$ of a three dimensional Frank copula $C_{1:3}$ is a Ali-Mikhail-Haq copula which only slightly varies (see Fig. 1. in Spanhel and Kurz (2016b)). In general, under regularity conditions, which hold for the Clayton and Frank copula, the value of Kendall's $\tau_{12|3}$ is restricted to the interval $[0, \frac{1}{3}]$ if $C_{1:3}$ is an Archimedean copula (Mesfioui and Quessy 2008).

Table 1: Test results for non-financial data sets. Rescaled ecdfs are used as univariate models and the regular vine copula models are selected with the algorithm of Dißmann et al. (2013). To test the simplifying assumption we apply the hierarchical procedure (Definition 5 and Definition 6) with the CCC test.

Name	concrete	uranium
Description	Concrete Compressive Strength Data Set	Uranium Exploration Data Set
Source	Yeh (1998), R-package `randomUniformForest`	Cook and Johnson (1986), R-package `copula`
Variables	`Cement, Coarse Aggregate, Fine Aggregate, Concrete compressive strength`	log concentration of `Uranium, Lithium, Colbalt, Potassium, Cesium, Scandum, Titanium`
Dimension	4 (6 pair-copulas)	7 (21 pair-copulas)
n	$1,030$	655
CCC test results (at a 5% confidence level)	The simplifying assumption can be **rejected** in the **second** tree.	The simplifying assumption can be **rejected** in the **second** tree.

we reject at least one null hypothesis of the form $H_0 : (U^{\mathrm{PVC}}_{i|S_{ij}}, U^{\mathrm{PVC}}_{i+j|S_{ij}}) \perp U_{S_{ij}}$ and therefore stop the hierarchical test procedure. For both non-financial data examples `uranium` and `concrete` we reject the simplifying assumption already in the second tree (Table 1). This is in line with the results reported by Gijbels et al. (2017a) for the data set `uranium` and by Kraus and Czado (2017) for both data sets. The results for the filtered financial returns are mixed (Table 2). For three out of eight data sets we reject the simplifying assumption. In the cases where the simplifying assumption is rejected, the simplifying assumption is not already rejected in the second tree. This indicates that the violation of the simplifying assumption for the vine copula selected by Dißmann's algorithm might be less severe for this kind of data as compared to `uranium` and `concrete`. This is consistent with the findings of Kraus and Czado (2017) who also use the CCC test and report that the simplifying assumption seems to be rather appropriate for filtered financial returns. Kraus and Czado (2017) argue that a possible explanation is that multivariate t-copulas are often well suited for modeling filtered financial returns and Stöber et al. (2013) has proven that t-copulas can be represented as simplified vine copulas.

Summing up, the analysis of the real data sets demonstrates that the CCC test can successfully be applied to investigate the simplifying assumption in high-dimensional simplified vine copula models. For further applications of the CCC test we refer the reader to Kraus and Czado (2017) and Schellhase and Spanhel (2018).

8 Conclusion

We introduce a test for the simplifying assumption in high-dimensional vine copulas. In practical applications, a test for the simplifying assumption in high-dimensional vine copulas

Table 2: Test results for financial data sets. The data source for all financial data sets is the *Kenneth R. French – Data Library* (available under: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`). ARMA(1,1)-GARCH(1,1) with t-distributed innovations are used as univariate models. The rescaled ecdf of these innovations provides pseudo-observations which are used for the regular vine copula which is selected with the algorithm of Dißmann et al. (2013). To test the simplifying assumption we apply the sequential procedure (Definition 5 and Definition 6) with the CCC test.

Name	FF3F	FF5F	Ind5	Ind10
Description	Fama/French 3 Factors	Fama/French 5 Factors	5 Industry Portfolios	10 Industry Portfolios
Variables	SMB, HML, $R_m - R_f$	SMB, HML, RMW, CMA, $R_m - R_f$	5 industry portfolios formed according to four-digit SIC codes.	10 industry portfolios formed according to four-digit SIC codes.
Period	01-Jul-1926 to 31-Jan-2017 (daily)	01-Jul-1963 to 31-Jan-2017 (daily)	02-Jan-1997 to 31-Jan-2017 (daily)	02-Jan-1997 to 31-Jan-2017 (daily)
Dimension	3 (3 pair-copulas)	5 (10 pair-copulas)	5 (10 pair-copulas)	10 (45 pair-copulas)
n	23,909	13,489	5,054	5,054
CCC test results (at a 5% confidence level)	The simplifying assumption can **not** be rejected.	The simplifying assumption can **not** be rejected.	The simplifying assumption can **not** be rejected.	The simplifying assumption can **not** be rejected.

Name	Countries	Pfs25	Ind30	Ind49
Description	20 Country Portfolios	25 Portfolios Formed on Size and Book-to-Market	30 Industry Portfolios	49 Industry Portfolios
Variables	AUT, AUS, BEL, CAN, DNK, FIN, FRA, GER, HKG, IRL, ITA, JPN, NLD, NZL, NOR, SGP, ESP, SWE, CHE, GBR	Intersections of five portfolios formed on size and five portfolios formed on the ratio of book equity to market equity	30 Industry Portfolios formed according to four-digit SIC codes.	49 Industry Portfolios formed according to four-digit SIC codes.
Period	Jan-1975 to Dec-2016 (monthly)	03-Jan-2007 to 31-Jan-2017 (daily)	Jul-1926 to Jan-2017 (monthly)	Jul-1969 to Jan-2017 (monthly)
Dimension	20 (190 pair-copulas)	25 (300 pair-copulas)	30 (435 pair-copulas)	49 (1,176 pair-copulas)
n	312	2,538	1,087	571
CCC test results (at a 5% confidence level)	The simplifying assumption can **not** be rejected.	The simplifying assumption can be **rejected** in the **ninth** tree.	The simplifying assumption can be **rejected** in the **ninth** tree.	The simplifying assumption can be **rejected** in the **fourth** tree.

must be computationally feasible and tackle the curse of dimensions. The introduced CCC test addresses these two issues.

The asymptotic distribution of the CCC test statistic is derived under the assumption of semi-parametrically estimated pseudo-observations from the partial probability integral transforms. Since the test has a known asymptotic distribution and is based on the stepwise maximum likelihood estimator it is computationally feasible also in high dimensions. To prevent suffering from the curse of dimensions if the number of conditioning variables increases, the CCC test utilizes a novel stochastic interpretation of the simplifying assumption based on the partial vine copula. Moreover, we propose a discretization of the support

of the conditioning variables into a finite number of subsets and incorporate a penalty in the test statistic. A decision tree algorithm detects the possibly largest deviation from the simplifying assumption measured in terms of conditional correlations and also contributes to a computationally feasible test. In a simulation study we provide a thorough analysis of the finite sample performance of the CCC test for various kinds of data generating processes. The CCC test outperforms the vectorial independence test by a large margin if the conditional correlation is varying. Even more important for high-dimensional applications, the simulation study demonstrates that the power of the test decreases only slightly with the dimension of the conditioning variables. Moreover, vine structure model selection and a mild misspecification of the parametric copula families do not affect the power properties of the CCC test. An application to 10 data sets with up to 49 dimensions demonstrates the usefulness of the test and indicates that the validity of the simplifying assumption should be checked individually for each data set.

Beside its application as a specification test of simplified vine copula models, the CCC test can also be utilized to improve the modeling of data with vine copulas. Schellhase and Spanhel (2018) make use of the CCC test to identify building blocks of vine copulas where the modeling of a conditional copula is more appropriate than the specification of an unconditional copula. Additionally, Kraus and Czado (2017) introduce model selection algorithms that use the CCC test to find appropriate structures for vine copulas which outperform the popular Dißmann algorithm (Dißmann et al. 2013).

References for Chapter IV

Aas, K., C. Czado, A. Frigessi, and H. Bakken. 2009. "Pair-copula constructions of multiple dependence". *Insurance: Mathematics and Economics* 44 (2): 182–198.

Acar, E. F., C. Genest, and J. Nešlehová. 2012. "Beyond simplified pair-copula constructions". *Journal of Multivariate Analysis* 110: 74–90.

Acar, E. F., R. V. Craiu, and F. Yao. 2013. "Statistical testing of covariate effects in conditional copula models". *Electronic Journal of Statistics* 7: 2822–2850.

Bedford, T., and R. M. Cooke. 2002. "Vines – A new graphical model for dependent random variables". *The Annals of Statistics* 30 (4): 1031–1068.

Bergsma, W. P. 2004. "Testing conditional independence for continuous random variables". URL: https://www.eurandom.tue.nl/reports/2004/048-report.pdf.

Bollerslev, T. 1986. "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics* 31 (3): 307–327.

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

Cook, R. D., and M. E. Johnson. 1986. "Generalized Burr-Pareto-Logistic Distributions With Applications to a Uranium Exploration Data Set". *Technometrics* 28 (2): 123–131.

Derumigny, A., and J.-D. Fermanian. 2017. "About tests of the "simplifying" assumption for conditional copulas". *Dependence Modeling* 5 (1): 154–197.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka. 2013. "Selecting and estimating regular vine copulae and application to financial returns". *Computational Statistics & Data Analysis* 59: 52–69.

Engle, R. F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". *Econometrica* 50 (4): 987–1007.

Fréchet, M. 1951. "Sur les tableaux de corrélation dont les marges sont données". *Annales de l'Université de Lyon, Section A, Series 3* 14: 53–77.

Genest, C., and A. Favre. 2007. "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask". *Journal of Hydrologic Engineering* 12 (4): 347–368.

Genest, C., K. Ghoudi, and L.-P. Rivest. 1995. "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions". *Biometrika* 82 (3): 543–552.

Gijbels, I., M. Omelka, and N. Veraverbeke. 2017a. "Nonparametric testing for no covariate effects in conditional copulas". *Statistics* 51 (3): 475–509.

— . 2015. "Partial and average copulas and association measures". *Electronic Journal of Statistics* 9 (2): 2420–2474.

Gijbels, I., M. Omelka, M. Pešta, and N. Veraverbeke. 2017b. "Score tests for covariate effects in conditional copulas". *Journal of Multivariate Analysis* 159: 111–133.

Grothe, O., and S. Nicklas. 2013. "Vine constructions of Lévy copulas". *Journal of Multivariate Analysis* 119: 1–15.

Hobæk Haff, I. 2013. "Parameter estimation for pair-copula constructions". *Bernoulli* 19 (2): 462–491.

Hobæk Haff, I., K. Aas, and A. Frigessi. 2010. "On the simplified pair-copula construction – Simply useful or too simplistic?" *Journal of Multivariate Analysis* 101 (5): 1296–1310.

Hoeffding, W. 1940. *Maßtabinvariante Korrelationstheorie*. Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik an der Universität Berlin, 5.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Joe, H., H. Li, and A. K. Nikoloulopoulos. 2010. "Tail dependence functions and vine copulas". *Journal of Multivariate Analysis* 101 (1): 252–270.

Kauermann, G., and C. Schellhase. 2014. "Flexible pair-copula estimation in D-vines using bivariate penalized splines". *Statistics and Computing* 24 (6): 1081–1100.

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

— . 2016. "Using model distances to investigate the simplifying assumption, goodness-of-fit and truncation levels for vine copulas". ArXiv e-prints `arXiv:1610.08795`.

Kojadinovic, I., and M. Holmes. 2009. "Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process". *Journal of Multivariate Analysis* 100 (6): 1137–1154.

Kraus, D., and C. Czado. 2017. "Growing simplified vine copula trees: improving Dißmann's algorithm". ArXiv e-prints `arXiv:1703.05203`.

Kurowicka, D., and R. Cooke. 2006. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Chichester: Wiley.

Kurowicka, D., and H. Joe, eds. 2011. *Dependence Modeling*. Singapore: World Scientific.

Kurz, M. S. 2018. *pacotest: Testing for Partial Copulas and the Simplifying Assumption in Vine Copulas.* URL: `https://CRAN.R-project.org/package=pacotest`, R-Package version 0.3.

Lavergne, P., and V. Patilea. 2008. "Breaking the curse of dimensionality in nonparametric testing". *Journal of Econometrics* 143 (1): 103–122.

Mesfioui, M., and J.-F. Quessy. 2008. "Dependence structure of conditional Archimedean copulas". *Journal of Multivariate Analysis* 99 (3): 372–385.

Nagler, T., and C. Czado. 2016. "Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas". *Journal of Multivariate Analysis* 151: 69–89.

Newey, W. K., and D. McFadden. 1994. "Large sample estimation and hypothesis testing". *Handbook of Econometrics* 4: 2111–2245.

Nikoloulopoulos, A. K., H. Joe, and H. Li. 2012. "Vine copulas with asymmetric tail dependence and applications to financial return data". *Computational Statistics & Data Analysis* 56 (11): 3659–3673.

Patton, A. J. 2006. "Modelling asymmetric exchange rate dependence". *International Economic Review* 47 (2): 527–556.

Portier, F., and J. Segers. 2018. "On the weak convergence of the empirical conditional copula under a simplifying assumption". *Journal of Multivariate Analysis* 166: 160–181.

Quessy, J.-F. 2010. "Applications and asymptotic power of marginal-free tests of stochastic vectorial independence". *Journal of Statistical Planning and Inference* 140 (11): 3058–3075.

Schellhase, C., and F. Spanhel. 2018. "Estimating non-simplified vine copulas using penalized splines". *Statistics and Computing* 28: 387–409.

Schepsmeier, U., J. Stoeber, E. C. Brechmann, B. Graeler, T. Nagler, and T. Erhardt. 2017. *VineCopula: Statistical Inference of Vine Copulas.* URL: `https://CRAN.R-project.org/package=VineCopula`, R-Package version 2.1.1.

Spanhel, F., and M. S. Kurz. 2016a. "Estimating parametric simplified vine copulas under general conditions". Unpublished working paper.

— . 2016b. "The partial copula: Properties and associated dependence measures". *Statistics & Probability Letters* 119: 76–83.

— . 2017. "The partial vine copula: A dependence measure and approximation based on the simplifying assumption". ArXiv e-prints `arXiv:1510.06971`.

Stöber, J., H. Joe, and C. Czado. 2013. "Simplified pair copula constructions – Limitations and extensions". *Journal of Multivariate Analysis* 119: 101–118.

Tsukahara, H. 2005. "Semiparametric estimation in copula models". *Canadian Journal of Statistics* 33 (3): 357–375.

Yeh, I.-C. 1998. "Modeling of strength of high-performance concrete using artificial neural networks". *Cement and Concrete Research* 28 (12): 1797–1808.

Appendix

A.1 Representation of the CCC test statistic in terms of weighted differences to the average correlation

The test statistic $T_n(\Gamma)$ (and also $T_n^\star(\Gamma)$) are based on the first-order difference matrix A defined in Section 4.1. In the following, we show that an equivalent test statistic can be obtained using a matrix B which results in a test statistic based on weighted squared differences to the average correlation. To see this, consider the $(L-1) \times L$ matrix

$$B = \begin{bmatrix} (1-\hat{\pi}_1)\sqrt{\hat{\pi}_1} & -\hat{\pi}_2\sqrt{\hat{\pi}_1} & -\hat{\pi}_3\sqrt{\hat{\pi}_1} & \cdots & -\hat{\pi}_L\sqrt{\hat{\pi}_1} \\ -\hat{\pi}_1\sqrt{\hat{\pi}_2} & (1-\hat{\pi}_2)\sqrt{\hat{\pi}_2} & -\hat{\pi}_3\sqrt{\hat{\pi}_2} & \cdots & -\hat{\pi}_L\sqrt{\hat{\pi}_2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -\hat{\pi}_1\sqrt{\hat{\pi}_{L-1}} & \cdots & & (1-\hat{\pi}_{L-1})\sqrt{\hat{\pi}_{L-1}} & -\hat{\pi}_L\sqrt{\hat{\pi}_{L-1}} \end{bmatrix}.$$

If we multiply B by the vector of estimated correlations $\hat{R}_\Gamma$, we get the vector of weighted differences to the average correlation, i.e.,

$$B\hat{R}_\Gamma = \left(\sqrt{\hat{\pi}_1}(\hat{r}_1 - \bar{r}), \ldots, \sqrt{\hat{\pi}_{L-1}}(\hat{r}_{L-1} - \bar{r}) \right)^T,$$

with $\bar{r} = \sum_{l=1}^L \hat{\pi}_l \hat{r}_l$, so that $(B\hat{R}_\Gamma)^T B\hat{R}_\Gamma = \sum_{l=1}^{L-1} \hat{\pi}_l(\hat{r}_l - \bar{r})^2$.

W.l.o.g., let $\sum_{l=1}^L \hat{\pi}_l = 1$. We can rewrite the matrix B as

$$B := DA = \text{Diag}(\sqrt{\hat{\pi}_1}, \ldots, \sqrt{\hat{\pi}_{L-1}}) \left(I_{L-1} - 1_{L-1} \hat{\pi}_{1:L-1}^T \right) C A,$$

where C is the $(L-1) \times (L-1)$ upper triangular matrix where all non-zero entries are one and A is the first-order difference matrix used to define the test statistic $T_n(\Gamma)$. Note that by the matrix determinant lemma D is invertible. It follows that the χ^2-statistics are equal, i.e.,

$$T_n(\Gamma) = n(A\hat{R}_\Gamma)^T (A\hat{\Sigma}_{R_\Gamma} A^T)^{-1} A\hat{R}_\Gamma = n(B\hat{R}_\Gamma)^T (B\hat{\Sigma}_{R_\Gamma} B^T)^{-1} B\hat{R}_\Gamma.$$

A.2 Proof of Proposition 2

We first prove the following lemma stating the asymptotic distribution of the test statistic $T_n^\star(\Gamma) = n(A\hat{R}_\Gamma^\star)^T (A\hat{\Sigma}_{R_\Gamma}^\star A^T)^{-1} A\hat{R}_\Gamma^\star$ under the $H_0 : (U_{i|S_{ij}}^{\text{PVC}}, U_{i+j|S_{ij}}^{\text{PVC}}) \perp U_{S_{ij}}$ and the assumption that observations from the PPITs are observable.

Lemma 1

Let $(U_{1:d}^k)_{k=1:n}$ be n independent copies of $U_{1:d} \sim C_{1:d}$, $\Lambda_0 := supp(U_{S_{ij}})$ and $(i,j) \in \mathcal{I}_2^d$ be fixed. Assume that the partition $\Gamma := \{\Lambda_1, \ldots, \Lambda_L\}$ satisfies the conditions stated in Proposition 2. Under the $H_0 : (U_{i|S_{ij}}^{PVC}, U_{i+j|S_{ij}}^{PVC}) \perp U_{S_{ij}}$ it holds that

$$T_n^\star(\Gamma) \xrightarrow{d} \chi^2(L-1).$$

Proof. We first derive the asymptotic distribution of $\hat{R}_\Gamma^\star = (\hat{r}_1, \ldots, \hat{r}_L)$ under the H_0 before showing that $T_n^\star(\Gamma)$ has an asymptotic Chi-square distribution under the H_0. For this purpose, let $e_5 := (0,0,0,0,1)^T$, $\otimes$ denote the Kronecker product, 1_L be a $L \times 1$ column vector of ones and I_L be the $L \times L$ identity matrix, so that $(I_L \otimes e_5)^T$ is a $L \times 5L$ matrix that can be used to extract every fifth element from a $5L$-dimensional column vector. The correlations are then given by $\hat{R}_\Gamma^\star = (I_L \otimes e_5)^T \hat{\alpha}$, with $\hat{\alpha}$ being the unique solution of the estimating equation

$$\frac{1}{n} \sum_{k=1}^n g_\Gamma(U_{1:d}^k, \hat{\pi}, \alpha) \stackrel{!}{=} 0, \tag{8.1}$$

where the estimating function g_Γ will be stated in the following.

Define

$$g_\pi(U_{1:d}^k, \pi) = \left(\pi_1 - \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_1\}}, \ldots, \pi_L - \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_L\}}\right)^T,$$

where $\pi := \pi_{1:L} \in (0,1)^L$. The solution $\hat{\pi}$ of $\frac{1}{n}\sum_{k=1}^n g_\pi(U_{1:d}^k, \pi) \stackrel{!}{=} 0$ denotes the random fraction of data corresponding to Λ_l, i.e.,

$$\hat{\pi}_l := \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_l\}}, \quad l = 1, \ldots, L.$$

Define

$$h(U_{1:d}^k, \phi) = \begin{pmatrix} \phi_1 - U_{i|S_{ij}}^{\text{PVC},k} \\ \phi_2 - U_{i+j|S_{ij}}^{\text{PVC},k} \\ \phi_3 - (U_{i|S_{ij}}^{\text{PVC},k} - \phi_1)^2 \\ \phi_4 - (U_{i+j|S_{ij}}^{\text{PVC},k} - \phi_2)^2 \\ \phi_5 - (U_{i|S_{ij}}^{\text{PVC},k} - \phi_1)(U_{i+j|S_{ij}}^{\text{PVC},k} - \phi_2)(\phi_3\phi_4)^{-\frac{1}{2}} \end{pmatrix},$$

where $\phi := \phi_{1:5} \in \mathbb{R}^5$. For $\Lambda_l \in \Gamma = \{\Lambda_1, \dots, \Lambda_L\}$ we set

$$f(U_{1:d}^k, \pi_l, \phi) = \pi_l^{-1} \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_l\}} h(U_{1:d}^k, \phi).$$

The estimating function g_Γ in (8.1) is given by

$$g_\Gamma(U_{1:d}^k, \pi, \alpha) := \left(f(U_{1:d}^k, \pi_1, \alpha_1)^T, \dots, f(U_{1:d}^k, \pi_L, \alpha_L)^T \right)^T.$$

Let π_0 be the unique solution of $\mathbb{E}[g_\pi(U_{1:d}^k, \pi)] = 0$, ϕ_0 be the unique solution of $\mathbb{E}[h(U_{1:d}^k, \phi)] = 0$, and $\alpha_0 = (\phi_0^T, \dots, \phi_0^T)^T$ so that $\mathbb{E}[g_\Gamma(U_{1:d}^k, \pi, \alpha_0)] = 0$ for all $\pi \in (0,1)^L$ under the H_0 because for each l-th block element of $\mathbb{E}[g_\Gamma(U_{1:d}^k, \pi, \alpha_0)]$ it holds that

$$\left(\mathbb{E}[g_\Gamma(U_{1:d}^k, \pi, \alpha_0)] \right)_l := \mathbb{E}[f(U_{1:d}^k, \pi_l, \phi_0)] = \mathbb{E}\left[\pi_l^{-1} \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_1\}} h(U_{1:d}^k, \phi_0) \right]$$

$$\stackrel{H_0}{=} \mathbb{E}\left[\pi_l^{-1} \mathbb{1}_{\{U_{S_{ij}}^k \in \Lambda_l\}} \right] \underbrace{\mathbb{E}[h(U_{1:d}^k, \phi_0)]}_{=0} = 0.$$

Using the same steps it can be readily verified that $\mathbb{E}[\partial_{\pi^T} g_\Gamma(U_{1:d}^k, \pi, \alpha_0)] = 0$ for all $\pi \in (0,1)^L$ under the H_0. Thus, under the H_0, the standard theory of estimating equations for two-step estimators, e.g., Theorem 6.1 in Newey and McFadden (1994), yields that

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \stackrel{d}{\to} N_{5L}(0, G_\Gamma^{-1} \Omega_\Gamma (G_\Gamma^{-1})^T), \tag{8.2}$$

where $G_\Gamma := \mathbb{E}\left[\partial_{\alpha^T} g_\Gamma(U_{1:d}^k, \pi_0, \alpha_0) \right]$, $\Omega_\Gamma := \mathrm{Cov}\left[g_\Gamma(U_{1:d}^k, \pi_0, \alpha_0) \right]$ and $N_d(\mu, \Sigma)$ denotes a d-dimensional normal distribution with mean vector μ and covariance matrix Σ.

If we now extract every fifth element from $\hat{\alpha}$ using $\hat{R}_\Gamma^\star = (I_L \otimes e_5)^T \hat{\alpha}$, we obtain the joint asymptotic distribution of the estimated correlations under the H_0 as

$$\sqrt{n}((\hat{r}_1, \dots, \hat{r}_L)^T - (r_1, \dots, r_L)^T) = \sqrt{n}(\hat{R}_\Gamma^\star - R_\Gamma) \stackrel{d}{\to} N_L(0, \Sigma_{R_\Gamma}^\star),$$

so that

$$\Sigma_{R_\Gamma}^\star = (I_L \otimes e_5)^T G_\Gamma^{-1} \Omega_\Gamma (G_\Gamma^{-1})^T (I_L \otimes e_5).$$

Under the H_0 it holds that $r_1 = \dots = r_L = r = \mathrm{Corr}(U_{i|S_{ij}}^{\mathrm{PVC}}, U_{i+j|S_{ij}}^{\mathrm{PVC}})$ and therefore it follows with the $(L-1) \times L$ first-order difference matrix A and the continuous mapping

theorem, that

$$\sqrt{n}A\hat{R}^{\star}_{\Gamma} \xrightarrow{d} N_{L-1}(0, A\Sigma^{\star}_{R_{\Gamma}}A^{T}).$$

To obtain the statistic of the CCC test when a sample from the PPITs is observable, the covariance matrix

$$\Sigma^{\star}_{R_{\Gamma}} = \text{Cov}\left[(I_L \otimes e_5)^T G_{\Gamma}^{-1} g_{\Gamma}(U_{1:d}^k, \pi_0, \alpha_0)\right]$$

has to be consistently estimated, e.g., by $\hat{\Sigma}^{\star}_{R_{\Gamma}} = \widehat{\text{Cov}}[(I_L \otimes e_5)^T \hat{G}_{\Gamma}^{-1} g_{\Gamma}(U_{1:d}^k, \hat{\pi}, \hat{\alpha})]$, where $\widehat{\text{Cov}}[X]$ denotes the sample covariance of the random vector X. By applying once more the continuous mapping theorem and Slutsky's theorem, we get

$$T^{\star}_n(\Gamma) = n(A\hat{R}^{\star}_{\Gamma})^T(A\hat{\Sigma}^{\star}_{R_{\Gamma}}A^T)^{-1}A\hat{R}^{\star}_{\Gamma} \xrightarrow{d} \chi^2(L-1)$$

and Lemma 1 is proven. ∎

The remaining part of the proof of Proposition 2 requires the definition of the pseudo stepwise maximum likelihood estimator of the vine copula parameters. This estimator can be obtained as the solution of estimating equations (Hobæk Haff (2013), Spanhel and Kurz (2016a), Tsukahara (2005)).[9] By extending these estimating equations by the ones for the correlations defined in the proof of Lemma 1 we derive the asymptotic distribution of the CCC test when pseudo-observations from the PPITs are estimated. Let $\{C^{\text{SVC}}_{1:d}(\cdot; \theta_{1:d-1}): \theta_{1:d-1} \in \Upsilon\}$ be a parametric simplified vine copula such that $\exists \theta_{1:d-1;0} \in \Upsilon$ so that $C^{\text{SVC}}_{1:d}(\cdot; \theta_{1:d-1;0}) = C^{\text{PVC}}_{1:d}(\cdot)$ with density given by

$$c^{\text{SVC}}_{1:d}(u_{1:d}; \theta_{1:d-1}) = \prod_{(i,j)\in\mathcal{I}^d_1} c^{\text{SVC}}_{i,i+j;S_{ij}}(u^{\text{SVC}}_{i|S_{ij}}(\theta_{1:j-1}), u^{\text{SVC}}_{i+j|S_{ij}}(\theta_{1:j-1}); \theta_{j,i}),$$

where $C^{\text{SVC}}_{i,i+j;S_{ij}}$ is a bivariate unconditional copula for each $(i,j) \in \mathcal{I}^d_1$. The individual stepwise pseudo score functions for the copulas in the j-th tree are given by

$$\partial_{\theta_j}\ell_j(V_{1:d}; \theta_{1:j}) := \partial_{\theta_j}\sum_{i=1}^{d-j}\ln\left(c^{\text{SVC}}_{i,i+j;S_{ij}}(V^{\text{SVC},k}_{i|S_{ij}}(\theta_{1:j-1}), V^{\text{SVC},k}_{i+j|S_{ij}}(\theta_{1:j-1}); \theta_{j,i})\right), \quad j = 1,\ldots,d-1.$$

Here, the pseudo-observations of the PPITs for $j = 1$, $i = 1,\ldots,d-1$ are defined by

$$V^{\text{SVC},k}_{i|S_{ij}}(\theta_{1:j-1}) := V^k_i := \frac{1}{n+1}\sum_{m=1}^n \mathbb{1}_{\{X^m_i \le X^k_i\}},$$

---

[9]For details on the pseudo stepwise maximum likelihood estimator of the vine copula parameters we refer to the stated references (Hobæk Haff 2013, Spanhel and Kurz 2016a).

$$V_{i+j|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) := V_{i+j}^{k} := \frac{1}{n+1}\sum_{m=1}^{n}\mathbb{1}_{\{X_{i+j}^{m}\leq X_{i+j}^{k}\}},$$

and for $(i,j)\in\mathcal{I}_2^d$ as

$$V_{i|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) := \partial_2 C_{i,i+j-1;S_{i,j-1}}^{\text{SVC}}(V_{i|S_{i,j-1}}^{\text{SVC},k}(\theta_{1:j-2}), V_{i+j-1|S_{i,j-1}}^{\text{SVC},k}(\theta_{1:j-2}); \theta_{j-1,i}),$$

$$V_{i+j|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) := \partial_1 C_{i+1,i+j;S_{i+1,j-1}}^{\text{SVC}}(V_{i+1|S_{i+1,j-1}}^{\text{SVC},k}(\theta_{1:j-2}), V_{i+j|S_{i+1,j-1}}^{\text{SVC},k}(\theta_{1:j-2}); \theta_{j-1,i+1}).$$

Set $\theta := \theta_{1:d-1}$ and define the estimating function

$$g_S(V_{1:d}^k, \theta) = (\partial_{\theta_1}\ell_1(V_{1:d}^k; \theta_1)^T, \ldots, \partial_{\theta_{d-1}}\ell_{d-1}(V_{1:d}^k; \theta_{1:d-1})^T)^T,$$

so that the solution $\hat{\theta}$ of $\frac{1}{n}\sum_{k=1}^n g_S(V_{1:d}^k, \theta) \overset{!}{=} 0$ is the pseudo stepwise maximum likelihood estimator.

Moreover, $g_\Gamma(V_{1:d}^k, \pi, \theta, \alpha)$ denotes the estimating function of the correlations when pseudo-observations from the PPITs are used, i.e.,

$$g_\Gamma(V_{1:d}^k, \pi, \theta, \alpha) := \left(f(V_{1:d}^k, \pi_1, \theta, \alpha_1)^T, \ldots, f(V_{1:d}^k, \pi_L, \theta, \alpha_L)^T \right)^T,$$

where

$$f(V_{1:d}^k, \pi_l, \theta, \phi) = \pi_l^{-1}\mathbb{1}_{\{U_{S_{ij}}^k\in\Lambda_l\}}\begin{pmatrix} \phi_1 - V_{i|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) \\ \phi_2 - V_{i+j|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) \\ \phi_3 - (V_{i|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) - \phi_1)^2 \\ \phi_4 - (V_{i+j|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) - \phi_2)^2 \\ \phi_5 - (V_{i|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) - \phi_1)(V_{i+j|S_{ij}}^{\text{SVC},k}(\theta_{1:j-1}) - \phi_2)(\phi_3\phi_4)^{-\frac{1}{2}} \end{pmatrix}.$$

Let $\beta := (\theta^T, \alpha^T)^T$ so that the estimating function of the vine copula parameters and the correlations is given by

$$g(V_{1:d}^k, \pi, \beta) = \begin{pmatrix} g_S(V_{1:d}^k, \theta) \\ g_\Gamma(V_{1:d}^k, \pi, \theta, \alpha) \end{pmatrix}.$$

The rank approximate estimator $\hat{\beta}$ is then given as the solution of $\frac{1}{n}\sum_{k=1}^n g(V_{1:d}^k, \hat{\pi}, \beta) \overset{!}{=} 0$ where $\hat{\pi}$ is given as in the proof of Lemma 1. To derive the asymptotic distribution of $\hat{\beta}$, introduce

$$W_i := \int \partial_{u_i} g(u_{1:d}, \pi_0, \beta_0)\mathbb{1}_{\{U_i\leq u_i\}}\mathrm{d}C_{1:d}(u_{1:d}), \qquad i = 1, \ldots, d,$$

$$\tilde{\Omega} := \text{Cov}\left(g(U_{1:d}^k, \pi_0, \beta_0) + \sum_{i=1}^d W_i \right),$$

where $\beta_0 := (\theta_{1:d-1;0}^T, \alpha_0^T)^T$ is the unique solution of $\mathbb{E}[g(U_{1:d}^k, \pi, \beta_0)] = 0$ for all $\pi \in (0,1)^L$ under the H_0. By the same reasoning as in the proof of Lemma 1 and because $g_S(U_{1:d}^k, \theta)$ does not depend on π it follows that $\mathbb{E}[\partial_{\pi^T} g(U_{1:d}^k, \pi, \beta_0)] = 0$ for all $\pi \in (0,1)^L$ under the H_0. Moreover, under the H_0 the simplifying assumption is true for the D-vine copula of $C_{i:i+j}$ (Lemma 3.1 in Spanhel and Kurz (2017)). Thus, provided the regularity conditions in Theorem 1 in Hobæk Haff (2013) are satisfied, it follows that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N_{n_\theta + 5L}(0, \tilde{\Sigma}),$$

where $\tilde{\Sigma} = G^{-1}\tilde{\Omega}(G^{-1})^T$ with $G := -\mathbb{E}[\partial_{\beta^T} g(U_{1:d}^n, \pi_0, \beta_0)]$ and let n_θ be the number of vine copula parameters, i.e., the length of the vector θ.

To extract the estimated correlations $\hat{R}_\Gamma$ from $\hat{\beta}$ and to obtain the corresponding asymptotic covariance matrix, we can exploit the block-structure of G as follows

$$G = \begin{pmatrix} \mathbb{E}[\partial_{\theta^T} g_S(U_{1:d}^k, \theta_0)] & 0 \\ \mathbb{E}[\partial_{\theta^T} g_\Gamma(U_{1:d}^k, \pi_0, \theta_0, \alpha_0)] & \mathbb{E}[\partial_{\alpha^T} g_\Gamma(U_{1:d}^k, \pi_0, \theta_0, \alpha_0)] \end{pmatrix}$$
$$= \begin{pmatrix} \mathbb{E}[\partial_{\theta^T} g_S(U_{1:d}^k, \theta_0)] & 0 \\ \mathbb{E}[\partial_{\theta^T} g_\Gamma(U_{1:d}^k, \pi_0, \theta_0, \alpha_0)] & G_\Gamma \end{pmatrix}.$$

Denote the $n_\theta \times L$ matrix consisting of zeros by $0_{n_\theta \times L}$ and define $\delta := (0_{n_\theta \times L}^T, (I_L \otimes e_5)^T)^T$ so that $\hat{R}_\Gamma = \delta^T \hat{\beta}$. The asymptotic covariance matrix of $\sqrt{n}(\hat{R}_\Gamma - R_\Gamma)$ is then

$$\begin{aligned} \Sigma_{R_\Gamma} &= \mathrm{Cov}\Big[\delta^T G^{-1}\big(g(U_{1:d}, \pi_0, \beta_0) + \sum_{i=1}^d W_i\big)\Big] \\ &= \Sigma_{R_\Gamma}^\star + \big(\mathrm{Cov}\big[\delta^T G^{-1} g(U_{1:d}, \pi_0, \beta_0)\big] - \Sigma_{R_\Gamma}^\star\big) \\ &\quad + \Big(\mathrm{Cov}\big[\delta^T G^{-1}\big(g(U_{1:d}, \pi_0, \beta_0) + \sum_{i=1}^d W_i\big)\big] - \mathrm{Cov}\big[\delta^T G^{-1} g(U_{1:d}, \pi_0, \beta_0)\big]\Big) \\ &= \Sigma_{R_\Gamma}^\star + \Sigma_{\mathrm{PVC}} + \Sigma_r. \end{aligned}$$

Thus, under the H_0 it follows that

$$\sqrt{n}(\hat{R}_\Gamma - R_\Gamma) \xrightarrow{d} N_L(0, \Sigma_{R_\Gamma}).$$

With the same arguments as in the proof of Lemma 1 this implies under the H_0

$$T_n(\Gamma) = n(A\hat{R}_\Gamma)^T(A\hat{\Sigma}_{R_\Gamma}A^T)^{-1}A\hat{R}_\Gamma \xrightarrow{d} \chi^2(L-1),$$

where $\hat{\Sigma}_{R_\Gamma} = \widehat{\mathrm{Cov}}\left(\delta^T \hat{G}^{-1}(g(U_{1:d}^k, \hat{\pi}, \hat{\beta}) + \sum_{i=1}^d \hat{W}_i)\right)$ is a consistent estimator of Σ_{R_Γ}.[10]

A.3 Proof of Proposition 3

To obtain the asymptotic distribution of the test statistic Θ_n, we need the following lemmas.

Lemma 2

Let $Y \sim F_Y$, where F_Y is the cdf of a continuous probability distribution. Additionally, let λ_n be a penalty function satisfying the conditions stated in Proposition 3. If $Y_n \overset{d}{\to} Y$ it holds that $Y_n - n\lambda_n \overset{p}{\to} -\infty$, i.e.,

$$\forall \alpha \in \mathbb{R} : \lim_{n \to \infty} \mathbb{P}(Y_n - n\lambda_n \leq \alpha) = 1.$$

Proof. Let $\alpha \in \mathbb{R}$. Since $n\lambda_n \to \infty$ it holds that

$$\forall \epsilon > 0 \ \exists n_1 \in \mathbb{N} \ \forall n \geq n_1 : F_Y(\alpha + n\lambda_n) > 1 - \frac{\epsilon}{2}. \tag{8.3}$$

By assumption Y_n converges in distribution to $Y \sim F_Y$, therefore

$$\forall \epsilon > 0 \ \forall n_1 \in \mathbb{N} \ \exists n_2 \in \mathbb{N} \ \forall n \geq n_2 : |F_{Y_n}(\alpha + n_1\lambda_{n_1}) - F_Y(\alpha + n_1\lambda_{n_1})| < \frac{\epsilon}{2}. \tag{8.4}$$

Moreover, $\exists n_3 \in \mathbb{N} \ \forall n \geq n_3 : n\lambda_n \geq n_1\lambda_{n_1}$. Thus, $\forall \epsilon > 0 \ \forall n \geq \max(n_1, n_2, n_3)$ it holds that

$$\begin{aligned}
\mathbb{P}(Y_n \leq \alpha + n\lambda_n) &\geq \mathbb{P}(Y_n \leq \alpha + n_1\lambda_{n_1}) = F_{Y_n}(\alpha + n_1\lambda_{n_1}) \\
&= F_{Y_n}(\alpha + n_1\lambda_{n_1}) - F_Y(\alpha + n_1\lambda_{n_1}) + F_Y(\alpha + n_1\lambda_{n_1}) \\
&\overset{(8.4)}{>} F_Y(\alpha + n_1\lambda_{n_1}) - \frac{\epsilon}{2} \overset{(8.3)}{>} 1 - \epsilon.
\end{aligned}$$

Thus,

$$\lim_{n \to \infty} \mathbb{P}(Y_n - n\lambda_n \leq \alpha) = 1.$$

$\blacksquare$

In the following Lemma 3 the asymptotic behavior of

$$\delta_n := \max\left\{Y_0^n + n\lambda_n, Y_1^n, Y_2^n, \ldots, Y_M^n\right\} - n\lambda_n$$

is analyzed.

---

[10] See Genest et al. (1995) for a consistent estimator of $W_i = \int \partial_{u_i} g(u_{1:d}, \pi_0, \beta_0) \mathbb{1}_{\{U_i \leq u_i\}} dC_{1:d}(u_{1:d})$.

Lemma 3

Let $(Y_0^k)_{k=1:n}, (Y_1^k)_{k=1:n}, \ldots, (Y_M^k)_{k=1:n}$ be $(M + 1)$ sequences of random variables and $Y_m \sim F_{Y_m}$, $0 \leq m \leq M$, random variables with continuous cumulative distribution functions. Further let $\lambda_n : \mathbb{N} \to \mathbb{R}^+$ be a penalty function satisfying the conditions stated in Proposition 3.

Define $\delta_n := \max\{Y_0^n + n\lambda_n, Y_1^n, Y_2^n, \ldots, Y_M^n\} - n\lambda_n$.

  (i) If $Y_m^n \xrightarrow{d} Y_m$ for each $0 \leq m \leq M$, it holds that $\delta_n \xrightarrow{d} Y_0$.

  (ii) If there is an $m^\star \in \{0, \ldots, M\}$ such that $\operatorname{plim}_{n\to\infty} \frac{1}{n} Y_{m^\star}^n = y_{m^\star} > 0$ then $\delta_n \xrightarrow{p} \infty$.

Proof. Proof of (i). Let $\alpha \in \mathbb{R}$, then

$$
\begin{aligned}
\mathbb{P}(\delta_n \leq \alpha) &= \mathbb{P}\left(\max\{Y_0^n + n\lambda_n, Y_1^n, \ldots, Y_M^n\} - n\lambda_n \leq \alpha\right) \\
&= \mathbb{P}\left(\max\{Y_0^n, Y_1^n - n\lambda_n, \ldots, Y_M^n - n\lambda_n\} \leq \alpha\right) \\
&= \mathbb{P}\left(Y_0^n \leq \alpha, Y_1^n - n\lambda_n \leq \alpha, \ldots, Y_M^n - n\lambda_n \leq \alpha\right).
\end{aligned}
$$

Using the Fréchet-Hoeffding inequalities (Fréchet 1951, Hoeffding 1940) we have

$$
\mathbb{P}(\delta_n \leq \alpha) \geq \max\left\{0, \mathbb{P}(Y_0^n \leq \alpha) + \sum_{m=1}^{M} \mathbb{P}(Y_m^n - n\lambda_n \leq \alpha) - M\right\}
$$

and

$$
\mathbb{P}(\delta_n \leq \alpha) \leq \min\left\{\mathbb{P}(Y_0^n \leq \alpha), \min_{1\leq m\leq M}\left\{\mathbb{P}(Y_m^n - n\lambda_n \leq \alpha)\right\}\right\}.
$$

Due to the continuity of the minimum and maximum as well as Lemma 2 it follows that

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(\delta_n \leq \alpha) &\geq \lim_{n\to\infty} \max\left\{0, \mathbb{P}(Y_0^n \leq \alpha) + \sum_{m=1}^{M} \mathbb{P}(Y_m^n - n\lambda_n \leq \alpha) - M\right\} \\
&= \max\left\{0, \lim_{n\to\infty} \mathbb{P}(Y_0^n \leq \alpha) + \sum_{m=1}^{M} \lim_{n\to\infty} \mathbb{P}(Y_m^n - n\lambda_n \leq \alpha) - M\right\} \\
&= \max\left\{0, F_{Y_0}(\alpha) + \sum_{m=1}^{M} 1 - M\right\} = F_{Y_0}(\alpha)
\end{aligned}
$$

and

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(\delta_n \leq \alpha) &\leq \lim_{n\to\infty} \min\left\{\mathbb{P}(Y_0^n \leq \alpha), \min_{1\leq m\leq M}\left\{\mathbb{P}(Y_m^n - n\lambda_n \leq \alpha)\right\}\right\} \\
&= \min\left\{\lim_{n\to\infty} \mathbb{P}(Y_0^n \leq \alpha), \min_{1\leq m\leq M}\left\{\lim_{n\to\infty} \mathbb{P}(Y_m^n - n\lambda_n \leq \alpha)\right\}\right\} \qquad (8.5) \\
&= \min\left\{F_{Y_0}(\alpha), \min_{1\leq m\leq M}\{1\}\right\} = F_{Y_0}(\alpha).
\end{aligned}
$$

138

Thus,

$$\delta_n \xrightarrow{d} Y_0.$$

Proof of (ii). For $m = 0, \ldots, M$, define $Z_m^n := Y_m^n - n\lambda_n \mathbb{1}_{\{m \neq 0\}}$. Because $\lambda_n \to 0$ it follows that $\mathrm{plim}_{n \to \infty} \frac{1}{n} Z_{m^\star}^n = y_{m^\star}$. Note that $\mathbb{P}(Z_{m^\star}^n \leq \alpha) \xrightarrow{n \to \infty} 0$ for all $\alpha \in \mathbb{R}$ because $\frac{1}{n} Z_{m^\star}^n \xrightarrow{p} y_{m^\star} > 0$.

Thus, the Fréchet-Hoeffding upper bound implies that for any $\alpha \in \mathbb{R}$,

$$\lim_{n \to \infty} \mathbb{P}(\delta_n \leq \alpha) \leq \lim_{n \to \infty} \min\Big\{ \mathbb{P}(Z_{m^\star}^n \leq \alpha), \min_{m \in \{0, \ldots, M\} \setminus m^\star} \big\{ \mathbb{P}\left(Z_m^n \leq \alpha \right) \big\} \Big\}$$

$$= \min\Big\{ \lim_{n \to \infty} \mathbb{P}(Z_{m^\star}^n \leq \alpha), \min_{m \in \{0, \ldots, M\} \setminus m^\star} \big\{ \lim_{n \to \infty} \mathbb{P}\left(Z_m^n \leq \alpha \right) \big\} \Big\}$$

$$= \min\Big\{ 0, \underbrace{\min_{m \in \{0, \ldots, M\} \setminus m^\star} \big\{ \lim_{n \to \infty} \mathbb{P}\left(Z_m^n \leq \alpha \right) \big\}}_{\in [0,1]} \Big\} = 0,$$

and the proof is complete. ∎

Using Proposition 2 and setting $Y_m^n = T_n(\Gamma_m)$ in Lemma 3 (i) it follows that the statistic Θ_n converges under the H_0 to a $\chi^2(L_0 - 1)$ distribution.

Now assume that the correlations conditional on the subsets of the partition $\Gamma_{m^\star}$ are not identical, i.e., $\mathrm{plim}_{n \to \infty} \frac{1}{n} T_n(\Gamma_{m^\star}) = c > 0$ for some $m^\star \in \{0, \ldots, M\}$. Setting $Y_m^n = T_n(\Gamma_m)$ in Lemma 3 (ii) shows that the test statistic Θ_n converges in probability to infinity.

A.4 The decision tree: Algorithmic details

Every leaf in the tree represents a subset of the support Λ_0 of the random vector $U_{S_{ij}}$. The maximum depth of the decision tree is denoted by $J_{\max}$ and every leaf is assigned to a level J in the decision tree ($0 \leq J \leq J_{\max}$). The level of a leaf refers to the number of splits which have already been used to arrive at the leaf, starting from the root leaf Λ_0 (see Figure 2).

A leaf is denoted by $\Lambda_{\gamma_{0:J}}$, where the $(J+1)$-dimensional vector $\gamma_{0:J} := (\gamma_0, \gamma_1, \ldots, \gamma_J) \in \{0\} \times \{l, r\}^J$ is the unique identifier for a leaf in the J-th level of the decision tree. That is, the two leaves in the $(J+1)$-th level of the decision tree being connected via edges to the leaf $\gamma_{0:J} = (\gamma_0, \gamma_1, \ldots, \gamma_J)$ in the J-th level are identified by $\gamma_{0:J+1} := (\gamma_{0:J}, k) := (\gamma_0, \gamma_1, \ldots, \gamma_J, k)$ with $k \in \{l, r\}$. The subsets assigned to the leafs in the $(J+1)$-th level by a binary split are given by

$$\Lambda_{(\gamma_{0:J}, k)} \subset \Lambda_{\gamma_{0:J}}, \qquad J \geq 0, \qquad k \in \{l, r\}, \qquad \Lambda_{(\gamma_{0:J}, l)} \cap \Lambda_{(\gamma_{0:J}, r)} = \emptyset.$$

Every split is chosen out of a finite number $\mathcal{M}$ of possible splits. A possible split $S_{\gamma_{0:J}}^m$ in the leaf $\gamma_{0:J}$ is defined as a pair of disjoint subsets of $\Lambda_{\gamma_{0:J}}$, i.e., $S_{\gamma_{0:J}}^m := (\Lambda_{(\gamma_{0:J},l)}^m, \Lambda_{(\gamma_{0:J},r)}^m) \subset \Lambda_{\gamma_{0:J}} \times \Lambda_{\gamma_{0:J}}$ with $\Lambda_{(\gamma_{0:J},l)}^m \cap \Lambda_{(\gamma_{0:J},r)}^m = \emptyset$. From these possible splits, the split is selected that maximizes the statistic of the CCC test. Meaning that every split is defined as

$$S_{\gamma_{0:J}} = (\Lambda_{(\gamma_{0:J},l)}, \Lambda_{(\gamma_{0:J},r)}) = \underset{S_{\gamma_{0:J}}^m \in \{S_{\gamma_{0:J}}^1, \ldots, S_{\gamma_{0:J}}^{\mathcal{M}}\}}{\operatorname{argmax}} T_n\left(S_{\gamma_{0:J}}^m\right).$$

Thus, the subsets that are transferred to leaf $(\gamma_{0:J}, k)$, $k \in \{l, r\}$, after using the optimal split $S_{\gamma_{0:J}}$, are given by $\Lambda_{(\gamma_{0:J},l)}$ and $\Lambda_{(\gamma_{0:J},r)}$. In the last level $J_{\max}$ we obtain a final partition of the support Λ_0 into $2^{J_{\max}}$ disjoint subsets given by

$$\Gamma_{max} := \{\Lambda_{(0,l,\ldots,l,l)}, \Lambda_{(0,l,\ldots,l,r)}, \ldots, \Lambda_{(0,r,\ldots,r,l)}, \Lambda_{(0,r,\ldots,r,r)}\}.$$

For the final partition we compute the value of the test statistic

$$\Theta_n = \max\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_{max})\} - n\lambda_n.$$

The following definition summarizes the decision tree-based CCC test.

Definition 6 (CCC test with a decision tree)
To test the hypothesis $H_0 \colon (U_{i|S_{ij}}^{PVC}, U_{i+j|S_{ij}}^{PVC}) \perp U_{S_{ij}}$ for a fixed $(i,j) \in \mathcal{I}_2^d$ using the CCC test with a decision tree and a significance level of α do the following:

1. *Obtain the pseudo-observations $\left(V_{i|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1}), V_{i+j|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1})\right)_{k=1,\ldots,n}$ using steps 1 - 2b in Definition 5.*

2. *Choose a null partition Γ_0 consisting of L_0 disjoint subsets of the support of $U_{S_{ij}}$.*

3. *Select a maximal depth $J_{\max}$ for the decision tree and a finite number of possible splitting points $\{S_{\gamma_{0:J}}^1, \ldots, S_{\gamma_{0:J}}^{\mathcal{M}}\}$ in every leaf $\gamma_{0:J}$ for all $0 \leq J \leq J_{\max} - 1$.*

4. *Use $\left(V_{i|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1}), V_{i+j|S_{ij}}^{SVC,k}(\hat{\theta}_{1:j-1})\right)_{k=1,\ldots,n}$, the conditioning variables $(V_{S_{ij}}^k)_{k=1,\ldots,n}$, and the decision tree explained in Section 5.2 to get $\Gamma_{\max}$.*

5. *Compute $\Theta_n = \max\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_{\max})\} - n\lambda_n$.*

6. *Reject the H_0 if $\Theta_n \geq F_{\chi^2(L_0-1)}^{-1}(1-\alpha)$, where $F_{\chi^2(L_0-1)}^{-1}$ is the quantile function of the χ^2-distribution with $L_0 - 1$ degrees of freedom.*

In all simulations in Section 6 and the real data applications in Section 7, we choose $\lambda_n = \frac{1}{\sqrt{n}}$ and $\Gamma_0 = \Gamma_{med}$. Further tuning parameters of the decision tree are the maximum

depth $J_{\max}$ of the tree and the set of possible splits $\{S^1_{\gamma_{0:J}}, \ldots, S^{\mathcal{M}}_{\gamma_{0:J}}\}$. To keep the test computationally feasible and because it performs well in simulations, we consider a maximum depth of $J_{\max} = 2$ and the number of possible splits in each leaf $\gamma_{0:J}$ is restricted to be at most $\mathcal{M} = 3(j - 1 + \mathbb{1}_{\{j \geq 3\}})$. The formal definition of the set of possible splits is given in Appendix A.5 and we provide here a short explanation. To obtain the sets $\Lambda_{(0,l)}$ and $\Lambda_{(0,r)}$ for the two leaves in level 1, we consider the empirical 0.25, 0.5, and 0.75 quantiles for each conditioning variable U_k, $k \in S_{ij}$. If $j \geq 3$, we additionally take the empirical 0.25, 0.5 and 0.75 quantiles of the mean aggregation given in (5.1) into account, resulting in $3 \cdot j$ possible splits. The sets $\{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}\}$ and $\{\Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$ for the four leaves in level 2 are obtained in the same fashion except that we now condition on $U_{S_{ij}} \in \Lambda_{(0,l)}$ or $U_{S_{ij}} \in \Lambda_{(0,r)}$, respectively. Furthermore, we use several restrictions in the decision tree algorithm to guarantee that the final data sets do not become too small, so that we can still rely on approximations of finite-sample distributions using asymptotic distributions.[11]

A.5 Formal definition of the set of possible splits for the decision tree

If $Q_q[(X_i)_{i \in \mathcal{I}}]$ denotes the empirical q-quantile of the vector $(X_i)_{i \in \mathcal{I}}$, the set of possible splits in the leaf $\gamma_{0:J}$, for $J = 0, 1$ and $j \geq 3$ is given by

$$\begin{aligned}
\mathcal{S}^{1:3 \cdot j}_{\gamma_{0:J}} &:= \left\{ S^1_{\gamma_{0:J}}, \ldots, S^{3 \cdot j}_{\gamma_{0:J}} \right\} \\
&= \left\{ (\Lambda^1_{(\gamma_{0:J},l)}, \Lambda^1_{(\gamma_{0:J},r)}), \ldots, (\Lambda^{3 \cdot j}_{(\gamma_{0:J},l)}, \Lambda^{3 \cdot j}_{(\gamma_{0:J},r)}) \right\},
\end{aligned}$$

with

$$\Lambda^m_{(\gamma_{0:J},l)} := \begin{cases}
\left\{ u_{S_{ij}} \in \Lambda_{\gamma_{0:J}} : u_{i+m} \leq Q_{0.25}[(V^k_{i+m})_{k \in \mathcal{I}_{\gamma_{0:J}}}] \right\} & , 1 \leq m \leq j-1 \\
\left\{ u_{S_{ij}} \in \Lambda_{\gamma_{0:J}} : u_{i+m-(j-1)} \leq Q_{0.5}[(V^k_{i+m-(j-1)})_{k \in \mathcal{I}_{\gamma_{0:J}}}] \right\} & , j \leq m \leq 2j-2 \\
\left\{ u_{S_{ij}} \in \Lambda_{\gamma_{0:J}} : u_{i+m-2(j-1)} \leq Q_{0.75}[(V^k_{i+m-2(j-1)})_{k \in \mathcal{I}_{\gamma_{0:J}}}] \right\} & , 2j-1 \leq m \leq 3j-3 \\
\left\{ u_{S_{ij}} \in \Lambda_{\gamma_{0:J}} : g(u_{S_{ij}}) \leq Q_{(m-3(j-1)) \cdot 0.25}[(g(V^k_{S_{ij}}))_{k \in \mathcal{I}_{\gamma_{0:J}}}] \right\} & , 3j-2 \leq m \leq 3j
\end{cases}$$

and $\Lambda^m_{(\gamma_{0:J},r)} = \Lambda_{\gamma_{0:J}} \setminus \Lambda^m_{(\gamma_{0:J},l)}$, where the index set $\mathcal{I}_{\gamma_{0:J}}$ is defined as $\mathcal{I}_{\gamma_{0:J}} := \{k \in \{1, \ldots, n\} : V^k_{S_{ij}} \in \Lambda_{\gamma_{0:J}}\}$.

---

[11] A decision tree with two or three splits is only applied if we have a certain amount of data. This is implemented by introducing a tuning parameter which controls the minimum sample size per leaf in the decision tree (the default value is 100 observations). As a result we do not always use the 0.25, 0.5 and 0.75 quantiles as thresholds but depending on the available sample size we may only use the 0.5 quantile or even don't apply any additional split at all.

A.6 Choosing the penalty function: A finite sample analysis

To apply the test based on the statistic Θ_n, a penalty function λ_n has to be specified and any choice satisfying the conditions stated in Lemma 3 results in an asymptotically valid test. However, the size and power for finite sample sizes depends on the chosen penalty function λ_n. The choice of the penalty function in finite samples will be analyzed in a simulation study under the H_0, i.e., with a focus on the empirical size.

In all simulations in Section 6 and the real data applications in Section 7, we choose $\lambda_n = \frac{1}{\sqrt{n}}$ and $\Gamma_0 = \Gamma_{med}$.[12] We will now show how testing based on Θ_n is related to testing based on $T_n(\Gamma_0)$, i.e., the CCC test with fixed partition Γ_0.

For $\Theta_n = \max\{T_n(\Gamma_0), T_n(\Gamma_{max}) - n\lambda_n\}$, with $\Gamma_{max} := \operatorname{argmax}_{\Gamma_m \in \{\Gamma_1,\dots,\Gamma_M\}} T_n(\Gamma_m)$, it holds

$$T_n(\Gamma_0) \leq \Theta_n,$$

meaning that if we reject based on $T_n(\Gamma_0)$, we also reject based on Θ_n. It follows that the empirical size of Θ_n is bounded from below by the empirical size of $T_n(\Gamma_0)$ when both test are applied to the same collection of data sets in a monte carlo simulation to compute the empirical size.

We now derive a condition on λ_n such that Θ_n and $T_n(\Gamma_0)$ result in equivalent test decisions. This means that the test statistic Θ_n is analyzed relative to $T_n(\Gamma_0)$.[13] Let $\tau := F_{\chi^2(L_0-1)}^{-1}(1-\alpha)$. If the penalty function λ_n satisfies

$$\lambda_n > \frac{1}{n}\Big(T_n(\Gamma_{max}) - \tau\Big) =: b_n, \tag{8.6}$$

it follows that $\Theta_n = \max\{T_n(\Gamma_0), T_n(\Gamma_{max}) - n\lambda_n\} \leq \max\{T_n(\Gamma_0), \tau\}$. Therefore, if we can not reject at a α-level based on $T_n(\Gamma_0)$ and if λ_n satisfies (8.6), we also can not reject based on Θ_n, i.e., if $\lambda_n > b_n$ it holds

$$T_n(\Gamma_0) < \tau \quad \Rightarrow \quad \Theta_n < \tau.$$

As a result, if $\lambda_n > b_n$, both tests result in the same α-level test decisions, i.e.,

$$T_n(\Gamma_0) \geq \tau \quad \Leftrightarrow \quad \Theta_n \geq \tau.$$

Note that $T_n(\Gamma)$ converges in distribution to a χ^2-distribution under the H_0 and by

---

[12]The partition Γ_{med} is defined in Section 5.1.

[13] An extensive simulation study of the finite sample performance of the proposed test Θ_n is presented in Section 6 where the empirical size relative to the theoretical level of the test is studied.

Slutsky's theorem it follows that $b_n \overset{p}{\to} 0$. Therefore, the lower bound b_n is bounded in probability, i.e.,

$$\forall \epsilon > 0 \; \exists B > 0 \; \forall n \geq 1 : \; \mathbb{P}(|b_n| \geq B) < \epsilon.$$

Meaning that for any $\epsilon > 0$, we can choose λ_n such that $\mathbb{P}(|b_n| \geq \lambda_n) < \epsilon$, which restricts the probability of different test decisions (i.e., rejecting the H_0 with Θ_n but not rejecting the H_0 with $T_n(\Gamma_0)$) at a α-level to ϵ because

$$
\begin{aligned}
\mathbb{P}(\Theta_n \geq \tau, T_n(\Gamma_0) < \tau) &= \mathbb{P}(\max\{T_n(\Gamma_0), T_n(\Gamma_{max}) - n\lambda_n\} \geq \tau, T_n(\Gamma_0) < \tau) \\
&= \mathbb{P}(T_n(\Gamma_{max}) - n\lambda_n \geq \tau, T_n(\Gamma_0) < \tau) \\
&\leq \mathbb{P}(T_n(\Gamma_{max}) - n\lambda_n \geq \tau) = \mathbb{P}(b_n \geq \lambda_n) < \epsilon.
\end{aligned}
$$

This implies that for any $\epsilon > 0$, we can choose λ_n such that $\mathbb{P}(|b_n| \geq \lambda_n) < \epsilon$ and therefore

$$
\begin{aligned}
\mathbb{P}\Big(T_n(\Gamma_0) \geq \tau \quad \Leftrightarrow \quad \Theta_n \geq \tau \Big) &= \mathbb{P}(T_n(\Gamma_0) \geq \tau, \Theta_n \geq \tau) + \mathbb{P}(T_n(\Gamma_0) < \tau, \Theta_n < \tau) \\
&= 1 - \underbrace{\mathbb{P}(T_n(\Gamma_0) \geq \tau, \Theta_n < \tau)}_{=0} - \mathbb{P}(T_n(\Gamma_0) < \tau, \Theta_n \geq \tau) \\
&\geq 1 - \epsilon.
\end{aligned}
$$

In practical applications, we are interested in the finite sample distribution of the lower bound b_n of the penalty function λ_n. Using resampling techniques, we can determine this lower bound for λ_n under the H_0.

To illustrate how one can use resampling techniques to determine the parameters c and β of the penalty function $\lambda_n = cn^{-\beta}$, we again consider the data generating process given in Example 1. For $\lambda = 0$, the Frank copula $C_{14;23}$ in the third tree of the D-vine copula defined in Example 1 is not varying with the conditioning variables and therefore fulfills the H_0 of being a second-order partial copula. For each considered sample size n we generate 1000 random samples of size n from the four-dimensional D-vine copula and compute for each sample the lower bound b_n of the penalty function λ_n. In Figure 10, the maximum of all 1000 lower bounds in the different samples is plotted for different sample sizes n as dots.[14] By taking the maximum over all resampled lower bounds we identify a lower bound for the penalty which would guarantee that in every of the 1000 samples the asymptotic α-level tests are equivalent, i.e., $T_n(\Gamma_0) \geq \tau \quad \Leftrightarrow \quad \Theta_n \geq \tau$. The lines correspond to different

---

[14] The upward-jump at $n = 500$ is caused by restrictions on the decision tree for small samples which are less restrictive for 500 than for 250 observations.
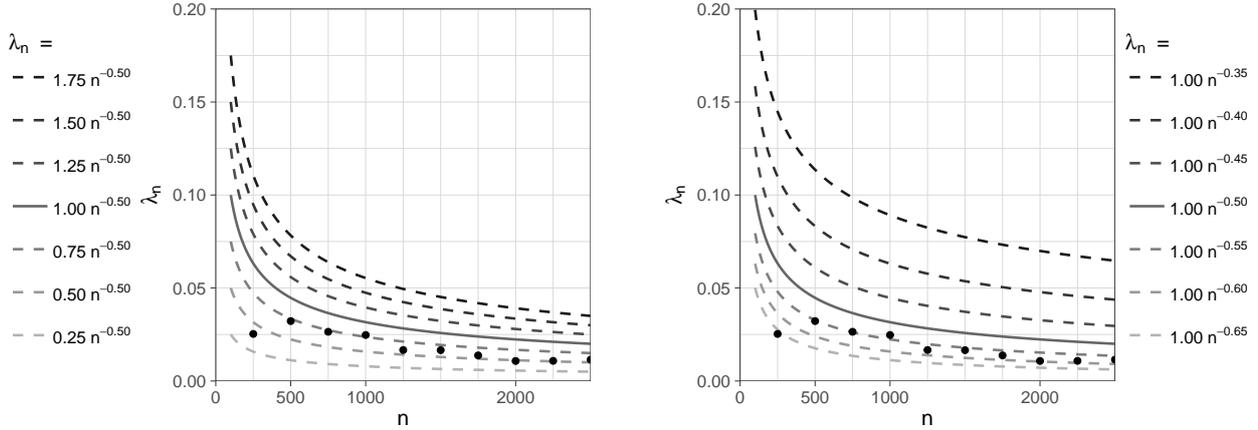
Figure 10: The dots correspond to the maximum of the lower bound for the penalty function in 1000 simulated samples each consisting of n observations. The data generating process is the four-dimensional D-vine copula defined in Example 1. The lines correspond to different choices of the penalty function $\lambda_n = cn^{-\beta}$. The solid line corresponds to the default penalty function $\lambda_n = \frac{1}{\sqrt{n}}$ used in all simulations and applications. On the left hand side the dashed lines correspond to different choices of the level c of the penalty function λ_n and on the right hand side to different choices for the power β of the penalty function.

choices of the penalty function $\lambda_n = cn^{-\beta}$. The level c of the penalty function is varied for a fixed power of $\beta = 0.5$ in the plot on the left hand side of Figure 10 and the power β of the penalty function is varied for a fixed level of $c = 1$ on the right hand side. The solid line corresponds to the penalty function $\lambda_n = \frac{1}{\sqrt{n}}$ which we use in all the simulations and applications. One can see that the choice of the penalty function is reasonable in comparison to the lower bounds obtained via resampling techniques for all sample sizes between 250 and 2500 observations, as the penalty is for all sample sizes considerably larger than the lower bounds.

Chapter V

Risk Assessment and Spurious Seasonality

This chapter is a reprint[*] of:

M. S. Kurz and S. Mittnik. 2018. "Risk Assessment and Spurious Seasonality". Available at SSRN: `https://dx.doi.org/10.2139/ssrn.2990772`, *Center for Quantitative Risk Analysis (CEQURA), Working Paper Number 19.*

Revision note:

This chapter is a reprint of the revised manuscript available at SSRN since April 25, 2018. The first version was made publicly available at SSRN on June 22, 2017 under the title "Risk Estimation and Spurious Seasonality".

Author contributions:

The origin of the research project was an estimation algorithm for recursive risk assessment designed by Stefan Mittnik. Analyses by Malte Kurz revealed that the resulting estimates displayed certain peculiar properties. Malte Kurz identified the causes of this behavior and came up with solution strategies, including mathematical derivations, proofs, `MATLAB` implementations, and simulation studies. Various issues were jointly discussed throughout research. The manuscript was drafted by Malte Kurz, iteratively revised by Stefan Mittnik and jointly finalized.

---

[*]All reprints in this cumulative dissertation have been slightly adapted in order to harmonize the overall appearance, like the format, page margins, caption setup, citation style and bibliography style.

Risk Assessment and Spurious Seasonality*

Malte S. Kurz[†1] and Stefan Mittnik[‡1]

[1]Department of Statistics,
Ludwig-Maximilians-Universität München,
Akademiestr. 1, 80799 Munich, Germany

Abstract

To determine the appropriate level of risk capital financial institutions are required to empirically estimate and predict specific risk measures. Although regulation commonly prescribes the forecasting horizon and the frequency with which risk assessments have to be reported, the scheme with which the underlying data are sampled typically remains unspecified. We show that, given assessment frequency and forecasting horizon, the choice of the sampling scheme can greatly affect the results of risk assessment procedures. Specifically, sequences of variance estimates are prone to exhibit spurious seasonality when the assessment frequency is higher than the sampling frequency of non-overlapping return data. We derive the autocorrelation function of such sequences for a general class of weak white noise processes and for a general class of variance estimators. To overcome the problem of spurious seasonality, we present a boundary-corrected exponentially-weighted moving-average version of the two-scales variance estimator introduced in the realized-volatility literature.

JEL Classification: C18, C58, G17, G28.

Keywords: Basel III, GARCH, overlapping data, temporal aggregation, two-scales variance estimator.

---

1 Introduction

Reliable estimation and prediction of the volatility of financial instruments is key to sound financial risk management. In practice, the return interval, the forecasting horizon and the assessment frequency are specified by regulation or management policies. Typically, however, the sampling frequency of the data underlying the empirical analysis remains unspecified. If the sampling frequency of the return data is more granular than the horizon for risk assessment, three strategies for estimating and forecasting risk measures are commonly adopted: (a) derive a risk estimate that matches the return interval specified (e.g., one-day volatility from daily return data) and then either use (square-root) scaling or derive model-based multi-step forecasts to obtain estimates for longer (e.g., monthly, annual) horizons; (b) temporally aggregate the underlying data so that they match the horizon for risk assessment, leading to analyses with overlapping samples; or (c) temporally aggregate the data so that samples do not overlap.[1]

In this paper, we address consequences of assessing risk for horizons that exceed the assessment frequency of the risk estimates. This is, for example, the case when asset managers rebalance weekly or monthly but assess and report risk at a daily frequency. Similar situations arise in banking (Basel III) and insurance (Solvency II) regulation. According to the Basel Committee on Banking Supervision (BCBS 2016), in Basel III, banks have to estimate the ten-day-ahead expected shortfall (ES) on a daily basis; and BCBS (2016, §181 c) states that the ten-day ES estimates need to be derived without scaling from shorter horizons and allows using overlapping return data. Various studies have investigated possible consequences of using overlapping returns for risk estimation or, more general, for statistical inference.[2] What has not been studied are the consequences of the implicit overlap that arises when assessing risk measures at a higher frequency than the sampling frequency of the data.

In the following, we restrict our analysis to the return *variance*, since other risk measures, such as volatility, value-at-risk or expected shortfall, are directly or indirectly related to variance. Moreover, for the sake of simplicity, we assume that returns are recorded at a daily frequency – implying that the most granular sampling and assessment frequency is daily.[3] To illustrate the estimation strategies (a)–(c) outlined above, Figure 1 depicts possible specifications for return interval and data sampling schemes in h-day-ahead assessments.

---

[1] See Andersen and Bollerslev (1998) and Andersen et al. (1999) for a detailed analysis and discussion of the tradeoff between sampling frequency and forecast horizon. More recently Kole et al. (2017) studied the impact of temporal and portfolio aggregation on the quality of ten-day ahead VaR forecasts.

[2] See, for example, Bod et al. (2002), Daníelsson et al. (2016), Daníelsson and Zhou (2016), Hansen and Hodrick (1980), Hedegaard and Hodrick (2016), Kluitman and Franses (2002) and Mittnik (2011).

[3] Note, however, our results also apply to intraday analyses.
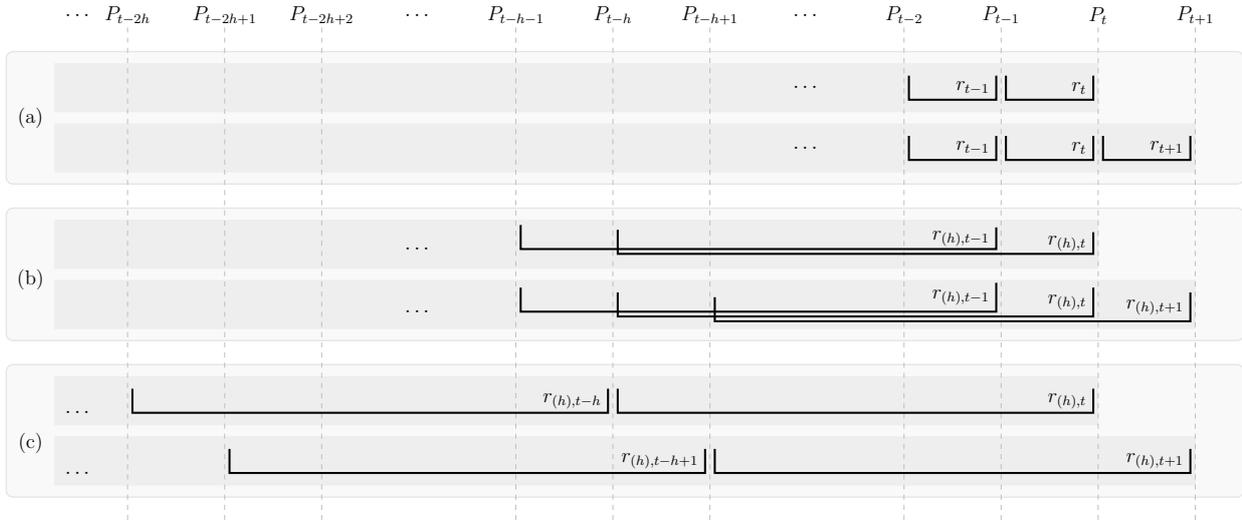
Figure 1: Illustration of different combinations of return intervals and sampling schemes for deriving h-day-ahead risk measures. Each panel consists of two rows: The first row sketches the data used for estimation at time t and the second row those at $t+1$. Panel (a) shows a scheme with daily sampling of daily returns. Here, risk estimates have to be scaled up to derive h-day-ahead risk estimates. Panel (b) illustrates the sampling scheme when using overlapping h-day returns. Panel (c) indicates the scheme when using non-overlapping h-day returns.

The two rows in each panel indicate the return data used for estimation on day t and $t+1$, respectively. Panel (a) reflects the sampling scheme for risk estimation based on daily return data. In this case, to derive h-day-horizon estimates, one needs to either rely on a scaling rule that approximates h-day risk from one-day estimates or on some multi-step forecasting procedure. Panel (b) illustrate the sampling when estimating with overlapping h-day returns at times t and $t+1$. Finally, Panel (c) shows the sampling scheme for returns when estimates are based on non-overlapping return intervals, revealing the implicit overlap when the assessment frequency is higher than the data sampling frequency. It is the latter scheme that is the main focus of this study.

In a recent study, Daníelsson and Zhou (2016) consider sampling strategies (a)–(c) for obtaining h-day-ahead risk estimates. They concentrate on a comparison of strategy (a) (square-root-of-time scaling) with strategy (b) (overlapping returns) with regards to bias and variance of risk estimates. Our work differs in two regards: First, we focus directly on strategies based on longer, namely, h-day return intervals. Given that BCBS (2016) explicitly rules out any risk assessment based on scaling, but also to avoid excessive clutter, we do not consider scaling strategies.[4] Second, we are not only interested in the accuracy of risk estimates (i.e., bias, variance, mean squared error etc.) at a given period, but also

---

[4]For a discussion of the square-root-of-time scaling see Christoffersen et al. (1998), Daníelsson and Zigrand (2006), Diebold et al. (1997) and McNeil and Frey (2000). Scaling rules under other than multivariate normal processes and, especially, serially dependent observations are derived in Embrechts et al. (2005).

in the dynamic properties of risk estimates.

If data availability is not an issue, estimates based on non-overlapping returns are, from an econometric point of view, the preferred choice (Daníelsson and Zhou 2016). We need to be concerned, however, when assessing variances at a higher frequency (e.g., daily) from return data that have longer return interval (e.g., weekly or monthly returns). We show that standard variance estimators, such as the moving-window sample variance or the exponentially-weighted moving average (EWMA) variance estimator (Riskmetrics 1996), tend to exhibit strong but spurious saw-tooth patterns. Clearly, risk managers who are obliged to assess risk more often (e.g., daily) than the horizon for risk-assessment implies (e.g., ten days in the Basel III or 259 days in the Solvency II framework) need to be aware of the fact that strong seasonal patterns may be induced. We demonstrate this phenomenon both empirically for real data and theoretically for well-behaved data-generating processes (DGPs), such as Gaussian white noise or GARCH(p,q) processes. We derive the theoretical autocorrelation function (ACF) for sequences of successive variance estimates for a broad class of DGPs and variance estimators. Moreover, we present variance estimators, based on overlapping h-day return intervals, that overcome the problem of spurious seasonality. Specifically, we introduce a boundary-corrected exponentially-weighted moving-average (EWMA) version of the two-scales estimator developed by Zhang et al. (2005). Our estimator does not suffer from spurious seasonality and performs best when compared to a range of alternative estimators.

The paper is organized as follows. In Section 2, using real data, we empirically illustrate and explain the presence of spurious seasonality in sequentially estimated variances. Section 3 defines the DGPs considered in this study, summarizes relevant results pertaining to stochastic processes and temporal aggregation, and derives quadratic-form representations for variance estimators. The theoretical ACF for sequences of daily estimated variances is derived in Section 4. Moreover, the phenomenon of spurious seasonality is illustrated and explained on theoretical grounds. Alternative variance estimators based on overlapping return intervals, but not suffering from spurious seasonality, are discussed in Section 5. Section 6 compares all variance estimators considered with respect to bias, variance, mean squared error (MSE) as well as their responsiveness to shocks in the data. Finally, Section 7 summarizes and concludes.

2 Spurious Seasonality in Variance Estimates from Temporally Aggregated, Non-Overlapping Returns

We are especially interested in the dynamic properties of sequential variance estimates. To illustrate our concern, we consider bi-weekly returns (i.e., returns over ten trading days) of the Dow Jones Industrial Average and look at two ways of displaying sequential variance estimates. First, we compute ten different bi-weekly return series, one for each of the ten trading days in the two-week window. For each of the ten return series we derive series of bi-weekly variance estimates, using an EWMA variance estimator (Riskmetrics 1996)

$$\sigma^2_{(h),t,\lambda} = \frac{h}{\text{tr}(\boldsymbol{\mathcal{Q}}_{(h),\Delta,\lambda})} \frac{1-\lambda}{1-\lambda^\Delta} \sum_{\delta=0}^{\Delta-1} \lambda^\delta \big(r_{(h),t-h\delta} - \mu_{(h),t,\lambda}\big)^2, \tag{1}$$

where $r_{(h),t}$ is the h-day return at time t and $\mu_{(h),t,\lambda} = \frac{1-\lambda}{1-\lambda^\Delta} \sum_{\delta=0}^{\Delta-1} \lambda^\delta r_{(h),t-h\delta}$ is the EWMA estimator for the first moment.[5] We set $\lambda = 0.96$. The left graph in Figure 2 shows the ten different series of variance estimates, $(\sigma^2_{(10),10t+\tau,\lambda})_{t\in\mathbb{Z}}$, for $1 \leq \tau \leq 10$, each corresponding to a specific starting day. The right graph in Figure 2 is obtained by combining the ten bi-weekly variance estimates to a single, daily sequence. In other words, we appropriately connect the bi-weekly estimates, $(\sigma^2_{(10),t,\lambda})_{t\in\mathbb{Z}}$, obtained at a daily frequency and shown in the plot on the left. This means that the distance between two adjacent points of the sequence of variance estimates is always one day rather than ten days, as is the case with the plots on the left.

Each variance estimate shown in Figure 2 is based on non-overlapping ten-day returns. However, the assessment frequency of the estimates is higher than the sampling frequency of the underlying data set. As a consequence, there is a substantial overlap in data used for successive estimates.

In the following, we study the pronounced sawtooth pattern of the series of daily variance estimates shown on the right of Figure 2. Furthermore, we explain the reason for the slowly changing patterns in the ten variance series plotted on the left of Figure 2. To characterize the properties of estimated variance sequences we examine their autocorrelation function (ACF). The sample ACF of the daily series of estimates, $(\sigma^2_{(10),t,\lambda})_{t\in\mathbb{Z}}$, based on bi-weekly data, shown in Figure 3, displays a systematic periodic pattern, a feature we refer to as *spurious seasonality*. As will be shown below, this seasonal pattern is due to the sampling scheme for the data used for variance estimation.

---

[5] The multiplicative constant $\frac{h}{\text{tr}(\boldsymbol{\mathcal{Q}}_{(h),\Delta,\lambda})} = \left(1 - \frac{(1-\lambda)^2(1-\lambda^{2\Delta})}{(1-\lambda^\Delta)^2(1-\lambda^2)}\right)^{-1}$ is the bias-correction factor, see (3).
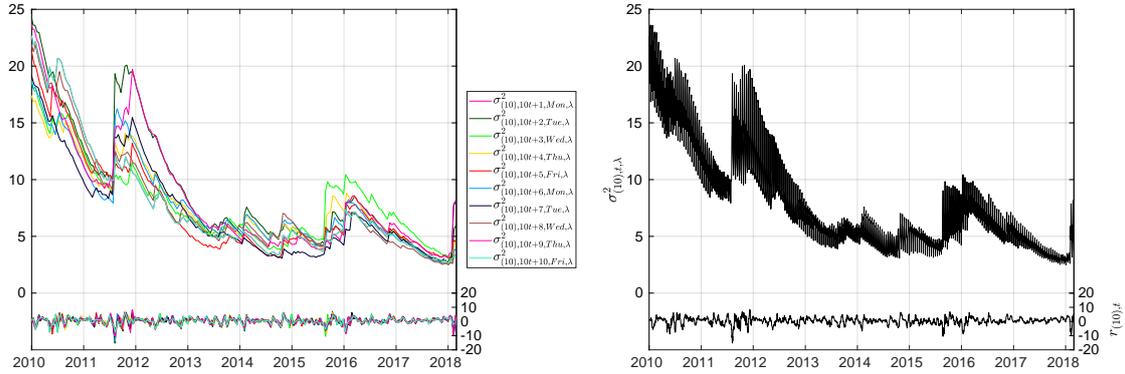
Figure 2: Estimated EWMA variances of the Dow Jones Industrial Average (DJIA) based on ten-day log-returns with a window length of 100 bi-weekly returns and an EWMA parameter of $\lambda = 0.96$. The first (last) estimates in both graphs are for 01-Jan-2010 (for 28-Feb-2018). The graph on the left shows at the top ten series of bi-weekly variance estimates, each corresponding to a specific weekday and start date, and the one on the right the daily series of bi-weekly variance estimates. The corresponding ten-day log-returns are plotted at the bottom of both graphs.
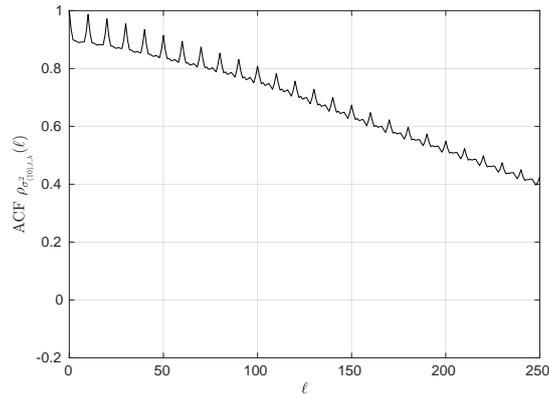
Figure 3: Sample autocorrelation function (ACF) for the daily series of bi-weekly EWMA variance estimates based on non-overlapping ten-day log-returns. The plot shows the sample ACF for the series of EWMA variance estimates for the Dow Jones Industrial Average (DJIA) from 01-Jan-2010 to 28-Feb-2018.

3 Some Prerequisites and Notation

In this section we introduce the two stochastic processes used in the analysis below, establish necessary notation, and briefly summarize relevant results on the temporal aggregation of stochastic processes. Finally, we introduce the (conditional) variance estimators that are the focus of this study.

3.1 Data Generating Stochastic Processes

We consider two data generating processes, the Gaussian white noise process and the GARCH(p,q) process. Both processes are so-called weak white noise processes.

Definition 1. *A stochastic process,* $(x_t)_{t\in\mathbb{Z}}$, *is called weak white noise process, if* $\forall t, t_1, t_2 \in \mathbb{Z},\ t_1 \neq t_2$:

(i) $\mathbb{E}(x_t) = \mu,$ *with* $|\mu| < \infty$,

(ii) $\mathrm{Var}(x_t) = \sigma^2,$ *with* $0 < \sigma^2 < \infty$,

(iii) $\mathrm{Cov}(x_{t_1}, x_{t_2}) = 0$.

The Gaussian white noise process (Example 1) is the special case of independent and identically distributed (i.i.d.) random variables with normal distribution.

Example 1. *A stochastic process,* $(x_t)_{t\in\mathbb{Z}}$, *is called Gaussian white noise process, if* $(x_t)_{t\in\mathbb{Z}}$ *is a weak white noise process and* $x_t \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$.

As a second case we consider the generalized autoregressive conditional heteroskedasticity (GARCH) process (Example 2) introduced by Engle (1982) and Bollerslev (1986), a model class that is widely used in academic research and in practice in order to model the volatility of financial returns.

Example 2. *Let* $(\epsilon_t)_{t\in\mathbb{Z}}$ *be a sequence of i.i.d. random variables and let* $p \in \mathbb{N}$ *and* $q \in \mathbb{N}_0$. *Further, let* $\alpha_0 > 0$, $\alpha_1, \ldots, \alpha_q \geq 0$ *and* $\beta_1, \ldots, \beta_p \geq 0$ *and assume* $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$, *such that the process is weakly stationary.*[6] *Then, a GARCH(p,q) process,* $(x_t)_{t\in\mathbb{Z}}$, *with strictly positive volatility process,* $(\sigma_t)_{t\in\mathbb{Z}}$, *is defined by*

$$x_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2.$$

3.2 Temporal Aggregation of Returns and Stochastic Processes

Let $(P_t)_{t\in\mathbb{Z}}$ denote the process of daily prices of an asset, $(r_t)_{t\in\mathbb{Z}}$ with $r_t = \ln(P_t) - \ln(P_{t-1})$ the process of daily log-returns, and let vector $\boldsymbol{r}_{t,\delta} := [r_{t-\delta+1}, r_{t-\delta+2}, \ldots, r_{t-1}, r_t]'$ collect the

---

[6]These restrictions on the parameter space guarantee a positive conditional variance σ_t^2 in the case of normally distributed innovations (Bollerslev 1986). Weaker necessary and sufficient conditions for a positive conditional variance are given in Nelson and Cao (1992).

δ daily returns from day $t - \delta + 1$ up to and including day t. h-day returns, $h > 1$, are then given by

$$r_{(h),t} = \ln(P_t) - \ln(P_{t-h}) = \sum_{j=0}^{h-1} r_{t-j} = \mathbf{1}_h' \boldsymbol{r}_{t,h},$$

where $\mathbf{1}_h$ is an $h \times 1$ column vector of ones. We call h the aggregation horizon.

In the following, we will always assume that the process of daily log-returns, $(r_t)_{t \in \mathbb{Z}}$, is generated by a weak white noise process (Definition 1) and, in some instances, consider the Gaussian white noise (Example 1) and the GARCH(p,q) process (Example 2) as special cases.

If we assume that the daily log-return series, $(r_t)_{t \in \mathbb{Z}}$, is a Gaussian white noise process with $\mathbb{E}(r_t) = \mu = 0$ and variance $\mathbb{E}(r_t^2) = \sigma^2 < \infty$, the temporally aggregated series, $(r_{(h),th})_{t \in \mathbb{Z}}$, where the sampling frequency coincides with the aggregation horizon, is again a Gaussian white noise process but with variance $\mathbb{E}(r_{(h),th}^2) = h\sigma^2 < \infty$. The situation changes, however, when the sampling frequency is lower than the aggregation horizon. This would, for example, be the case when sampling h-day returns, $h > 1$, on a, say, daily basis. Then, $(r_{(h),t})_{t \in \mathbb{Z}}$ turns out to be a non-invertible moving average process of order $h - 1$ (in short: MA($h - 1$) process) (Hansen and Hodrick 1980), with parameters $\theta_j = 1$ for $1 \leq j \leq h - 1$, i.e., $r_{(h),t} = \sum_{j=0}^{h-1} r_{t-j} = \sum_{j=1}^{h-1} \theta_j r_{t-j} + r_t$, where $(r_t)_{t \in \mathbb{Z}}$ is the weak white noise series of daily log-returns. The autocorrelation function (ACF) $\rho_{r_{(h),t}}(\ell)$ for the process $(r_{(h),t})_{t \in \mathbb{Z}}$ is given by (cf. Mittnik (1988))

$$\rho_{r_{(h),t}}(\ell) = \text{Corr}(r_{(h),t}, r_{(h),t-\ell}) = \begin{cases} \frac{\sum_{j=0}^{h-1-|\ell|} \theta_j \theta_{j+|\ell|}}{\sum_{j=0}^{h-1} \theta_j^2} = \frac{h-|\ell|}{h} & , |\ell| \leq h - 1, \\ 0 & , |\ell| \geq h, \end{cases}$$

where we set $\theta_0 = 1$ for notational simplicity. Similar results can also be obtained under some regularity conditions for the GARCH(p,q) process.[7]

3.3 Estimating Variances

Analogous to the vector of daily returns, let $\boldsymbol{r}_{(h),t,\Delta} = \big[r_{(h),t-h(\Delta-1)}, r_{(h),t-h(\Delta-2)}, \ldots, r_{(h),t-h},$ $r_{(h),t} \big]'$ be the Δ-period vector of non-overlapping h-day returns up to and including time t. Denoting the $\Delta \times \Delta$ identity matrix by $\boldsymbol{I}_\Delta$, we define the $h\Delta \times \Delta$ matrix $\boldsymbol{H} = \boldsymbol{I}_\Delta \otimes \mathbf{1}_h$, where $\otimes$ is the Kronecker product, so that $\boldsymbol{r}_{(h),t,\Delta} = \boldsymbol{H}' \boldsymbol{r}_{t,h\Delta}$.

---

[7]Temporal aggregation of GARCH processes has been investigated by Drost and Nijman (1993), and a survey of studies on temporal aggregation of various types of univariate and multivariate time series processes is provided by Silvestrini and Veredas (2008).

The most common estimator for the dispersion of returns is the sample variance. Defining the idempotent matrix $\boldsymbol{D} \in \mathbb{R}^{\Delta \times \Delta}$, $\boldsymbol{D} = \boldsymbol{I}_\Delta - \frac{1}{\Delta} \boldsymbol{1}_\Delta \boldsymbol{1}'_\Delta$, the moving-window sample variance for non-overlapping h-day returns is given by

$$\sigma^2_{(h),t} = \frac{1}{\Delta - 1} \sum_{\delta=0}^{\Delta-1} (r_{(h),t-h\delta} - \mu_{(h),t})^2 = \frac{1}{\Delta - 1} \boldsymbol{r}'_{(h),t,\Delta} \boldsymbol{D}' \boldsymbol{D} \boldsymbol{r}_{(h),t,\Delta} = \frac{1}{\Delta - 1} \boldsymbol{r}'_{t,h\Delta} \boldsymbol{H} \boldsymbol{D} \boldsymbol{H}' \boldsymbol{r}_{t,h\Delta},$$

with $\mu_{(h),t} = \frac{1}{\Delta} \boldsymbol{1}'_\Delta \boldsymbol{r}_{(h),t,\Delta} = \frac{1}{\Delta} \boldsymbol{1}'_\Delta \boldsymbol{H}' \boldsymbol{r}_{t,h\Delta}$ being the sample mean.

Below, we only discuss moving-window-type estimators. We restrict ourselves to this kind of estimators because in practice estimation is always based on a finite amount of data, so that finite-sample properties are of a relevance. The generalization of the results to the increasing window case is straightforward.[8]

Many variance estimators can be written as quadratic forms of the daily return vector, $\boldsymbol{r}_{t,h\Delta}$, i.e., $\sigma^2_{(h),t} = \boldsymbol{r}'_{t,h\Delta} \boldsymbol{\mathcal{Q}} \boldsymbol{r}_{t,h\Delta}$, where $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{h\Delta \times h\Delta}$ is a positive definite, symmetric matrix. Examples are the sample variance given above, but also the exponentially-weighted moving average (EWMA) variance estimator (Riskmetrics 1996).

If we assume a weak white noise process (Definition 1) for $(r_t)_{t \in \mathbb{Z}}$ with $\mathrm{Var}(r_t) = \sigma^2$, we have $\mathbb{E}(\boldsymbol{r}'_{t,h\Delta} \boldsymbol{\mathcal{Q}} \boldsymbol{r}_{t,h\Delta}) = \sigma^2 \mathrm{tr}(\boldsymbol{\mathcal{Q}})$, and the bias of the variance estimator, $\boldsymbol{r}'_{t,h\Delta} \boldsymbol{\mathcal{Q}} \boldsymbol{r}_{t,h\Delta}$, is

$$\mathrm{Bias}(\boldsymbol{r}'_{t,h\Delta} \boldsymbol{\mathcal{Q}} \boldsymbol{r}_{t,h\Delta}) = \mathbb{E}(\boldsymbol{r}'_{t,h\Delta} \boldsymbol{\mathcal{Q}} \boldsymbol{r}_{t,h\Delta}) - \mathrm{Var}(r_{(h),t}) = \sigma^2(\mathrm{tr}(\boldsymbol{\mathcal{Q}}) - h). \tag{2}$$

Therefore, variance estimates of the form $\boldsymbol{r}'_{t,h\Delta} \boldsymbol{\mathcal{Q}} \boldsymbol{r}_{t,h\Delta}$, can be bias-corrected by multiplying with factor $\frac{h}{\mathrm{tr}(\boldsymbol{\mathcal{Q}})}$, i.e., by using

$$\sigma^2_{(h),t} = \boldsymbol{r}'_{t,h\Delta} \boldsymbol{Q} \boldsymbol{r}_{t,h\Delta}, \tag{3}$$

as variance estimate with $\boldsymbol{Q} = \frac{h}{\mathrm{tr}(\boldsymbol{\mathcal{Q}})} \boldsymbol{\mathcal{Q}}$. Throughout the paper, we will use the bias-corrected versions of the variance estimators but will, in general, only define $\boldsymbol{\mathcal{Q}}$. Quantities $\boldsymbol{Q}$ and $\boldsymbol{\mathcal{Q}}$ are always related by $\boldsymbol{Q} = \frac{h}{\mathrm{tr}(\boldsymbol{\mathcal{Q}})} \boldsymbol{\mathcal{Q}}$.

Specifically, the sample variance for non-overlapping h-day returns is given by

$$\sigma^2_{(h),t} = \boldsymbol{r}'_{t,h\Delta} \boldsymbol{Q}_{(h),\Delta} \boldsymbol{r}_{t,h\Delta}, \tag{4}$$

with $\boldsymbol{\mathcal{Q}}_{(h),\Delta} = \frac{1}{\Delta} \boldsymbol{H} \boldsymbol{D} \boldsymbol{H}' = \frac{1}{\Delta} (\boldsymbol{I}_\Delta \otimes \boldsymbol{1}_h)(\boldsymbol{I}_\Delta - \frac{1}{\Delta} \boldsymbol{1}_\Delta \boldsymbol{1}'_\Delta)(\boldsymbol{I}_\Delta \otimes \boldsymbol{1}'_h)$, and the EWMA variance

---

[8]Asymptotic properties of sample variances when data are generated by a GARCH process are derived in Horváth et al. (2006).

for non-overlapping h-day returns (1) by

$$\sigma^2_{(h),t,\lambda} = \frac{h}{\text{tr}(\boldsymbol{Q}_{(h),\Delta,\lambda})} \frac{1-\lambda}{1-\lambda^\Delta} \sum_{\delta=0}^{\Delta-1} \lambda^\delta (r_{(h),t-h\delta} - \mu_{(h),t,\lambda})^2 = \boldsymbol{r}'_{t,h\Delta} \boldsymbol{Q}_{(h),\Delta,\lambda} \boldsymbol{r}_{t,h\Delta}, \quad (5)$$

with $\boldsymbol{Q}_{(h),\Delta,\lambda} = \boldsymbol{H}\boldsymbol{E}'\boldsymbol{\Lambda}\boldsymbol{E}\boldsymbol{H}' = (\boldsymbol{I}_\Delta \otimes \mathbf{1}_h)(\boldsymbol{I}_\Delta - \boldsymbol{w}\mathbf{1}'_\Delta)\boldsymbol{\Lambda}(\boldsymbol{I}_\Delta - \mathbf{1}_\Delta \boldsymbol{w}')(\boldsymbol{I}_\Delta \otimes \mathbf{1}'_h)$ and $\lambda \in (0,1)$. Vector $\boldsymbol{w} \in \mathbb{R}^{\Delta \times 1}$ and matrices $\boldsymbol{\Lambda}, \boldsymbol{E} \in \mathbb{R}^{\Delta \times \Delta}$ are defined by $\boldsymbol{w} = \frac{1-\lambda}{1-\lambda^\Delta} \cdot [\lambda^{\Delta-1}, \lambda^{\Delta-2}, \dots, \lambda^1, 1,]'$, $\boldsymbol{\Lambda} = \text{Diag}(\boldsymbol{w}) = (\boldsymbol{w}\mathbf{1}'_\Delta) \odot \boldsymbol{I}_\Delta$ and $\boldsymbol{E} = \boldsymbol{I}_\Delta - \mathbf{1}_\Delta \boldsymbol{w}'$, respectively, with $\odot$ denoting the Hadamard product.

4 Autocorrelation of Estimated Variances

4.1 Theoretical Derivation

Let matrices $\boldsymbol{K}, \boldsymbol{L} \in \mathbb{R}^{h\Delta+\ell \times h\Delta}$ be defined by $\boldsymbol{K} = [\mathbf{0}_{(h\Delta \times \ell)}, \boldsymbol{I}_{h\Delta}]'$ and $\boldsymbol{L} = [\boldsymbol{I}_{h\Delta}, \mathbf{0}_{(h\Delta \times \ell)}]'$, $\ell \geq 0$, where $\mathbf{0}_{(h\Delta \times \ell)}$ denotes an $h\Delta \times \ell$ matrix of zeros, so that $\boldsymbol{r}_{t,h\Delta} = \boldsymbol{K}'\boldsymbol{r}_{t,h\Delta+\ell}$ and $\boldsymbol{r}_{t-\ell,h\Delta} = \boldsymbol{L}'\boldsymbol{r}_{t,h\Delta+\ell}$. Variance estimators are then given by the quadratic-form

$$\sigma^2_{(h),t} = \boldsymbol{r}'_{t,h\Delta} \boldsymbol{Q} \boldsymbol{r}_{t,h\Delta} = \boldsymbol{r}'_{t,h\Delta+\ell} \boldsymbol{K}\boldsymbol{Q}\boldsymbol{K}' \boldsymbol{r}_{t,h\Delta+\ell}. \quad (6)$$

We obtain the sample variance specified in (4) for $\boldsymbol{Q} = \boldsymbol{Q}_{(h),\Delta}$ and the EWMA variance specified in (5) for $\boldsymbol{Q} = \boldsymbol{Q}_{(h),\Delta,\lambda}$. Similarly, the ($\ell$ days) lagged variance estimator is given by

$$\sigma^2_{(h),t-\ell} = \boldsymbol{r}'_{t-\ell,h\Delta} \boldsymbol{Q} \boldsymbol{r}_{t-\ell,h\Delta} = \boldsymbol{r}'_{t,h\Delta+\ell} \boldsymbol{L}\boldsymbol{Q}\boldsymbol{L}' \boldsymbol{r}_{t,h\Delta+\ell}. \quad (7)$$

Expressions (6) and (7) allow us to write the variance estimator, $\sigma^2_{(h),t}$, and its lagged version, $\sigma^2_{(h),t-\ell}$, as quadratic forms of the very same vector of daily returns, $\boldsymbol{r}_{t,h\Delta+\ell}$. The quadratic forms $\boldsymbol{K}\boldsymbol{Q}\boldsymbol{K}'$ and $\boldsymbol{L}\boldsymbol{Q}\boldsymbol{L}'$ turn out to be block-diagonal matrices, with $\boldsymbol{K}\boldsymbol{Q}\boldsymbol{K}' = \text{blkDiag}(\mathbf{0}_{(\ell \times \ell)}, \boldsymbol{Q})$ and $\boldsymbol{L}\boldsymbol{Q}\boldsymbol{L}' = \text{blkDiag}(\boldsymbol{Q}, \mathbf{0}_{(\ell \times \ell)})$.

Next, to further analyze the properties of estimated variances based on non-overlapping h-day returns assessed at a frequency higher than the aggregation horizon, we derive the ACF of the series of estimated variances, $(\sigma^2_{(h),t})_{t \in \mathbb{Z}}$. Theorem 1 states a well-known result about the covariance of two quadratic forms of the same multivariate normally distributed random vector. It follows directly from results in Magnus (1978) on moments of products of quadratic forms for multivariate normally distributed random variables.

Theorem 1. *Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ be symmetric matrices and $\boldsymbol{X}$ be multivariate normally distributed $n \times 1$ vector with $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{X})$ and $\boldsymbol{\Sigma} = \mathbb{E}((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})') = \mathbb{E}(\boldsymbol{X}\boldsymbol{X}') - \boldsymbol{\mu}\boldsymbol{\mu}'.$*

For the quadratic forms $\boldsymbol{X'AX}$ and $\boldsymbol{X'BX}$ we have

$$\mathrm{Cov}(\boldsymbol{X'AX}, \boldsymbol{X'BX}) = 2\mathrm{tr}(\boldsymbol{A\Sigma B\Sigma}) + 4\boldsymbol{\mu'A\Sigma B\mu}.$$

Proof. This follows directly from Lemma 6.2 in Magnus (1978). $\qquad\square$

The following corollary to Theorem 1 establishes the autocovariance function of the variances given by the quadratic forms (3) when the daily log-returns, $(r_t)_{t\in\mathbb{Z}}$, follow a Gaussian white noise process.[9]

Corollary 1. *Let $(r_t)_{t\in\mathbb{Z}}$ be a Gaussian white noise process (Example 1) with $\mathbb{E}(r_t) = 0$ and variance $Var(r_t) = \sigma^2$ and consider variance estimates of the form $\sigma^2_{(h),t} = \boldsymbol{r'}_{t,h\Delta}\boldsymbol{Q}\boldsymbol{r}_{t,h\Delta}$ (Eq. (3)). Then, the autocovariance of the series $(\sigma^2_{(h),t})_{t\in\mathbb{Z}}$, for $\ell \geq 0$, is given by*

$$\gamma_{\sigma^2_{(h),t}}(\ell) = \mathrm{Cov}(\sigma^2_{(h),t}, \sigma^2_{(h),t-\ell}) = 2\sigma^4\,\mathrm{tr}(\boldsymbol{KQK'LQL'}).$$

Proof. See Appendix A.1. $\qquad\square$

Note that for $\ell > h\Delta$

$$\gamma_{\sigma^2_{(h),t}}(\ell) = 2\sigma^4\,\mathrm{tr}(\boldsymbol{KQK'LQL'}) = 2\sigma^4\,\mathrm{tr}(\boldsymbol{0}_{(h\Delta\times h\Delta)}\,\boldsymbol{Q}\,\boldsymbol{0}_{(h\Delta\times h\Delta)}\,\boldsymbol{Q}) = 0,$$

and, by definition, the ACF for $\ell \geq 0$ is given by $\rho_{\sigma^2_{(h),t}}(\ell) = \gamma_{\sigma^2_{(h),t}}(\ell)/\gamma_{\sigma^2_{(h),t}}(0)$.

In the following, we extend Theorem 1 to a more general class of weak white noise processes which contains many zero-mean weak white noise processes – especially, the Gaussian white noise process with $\mu = 0$ and GARCH(p,q) processes.

Theorem 2. *Let $(x_t)_{t\in\mathbb{Z}}$ be a stochastic process with $\mathbb{E}(|x_t|^i) < \infty$, for $t \in \mathbb{Z}$ and $i \leq 4$. For $t_1, t_2, t_3, t_4 \in \mathbb{Z}$ with $\forall i, j \in \{t_1, t_2, t_3, t_4\}$, $i \neq j$, we assume*

$$\mathbb{E}(x_{t_1}) = 0, \tag{8}$$

$$\mathbb{E}(x_{t_1}x_{t_2}x_{t_3}x_{t_4}) = 0, \tag{9}$$

$$\mathbb{E}(x^2_{t_1}x_{t_2}x_{t_3}) = 0, \tag{10}$$

$$\mathbb{E}(x^3_{t_1}x_{t_2}) = 0. \tag{11}$$

---

[9]For the sake of simplicity, we assume a Gaussian white noise process with zero mean. In case of $\mathbb{E}(r_t) = \mu \neq 0$, we have $\gamma_{\sigma^2_{(h),t}}(\ell) = 2\sigma^4\,\mathrm{tr}(\boldsymbol{KQK'LQL'}) + 4\mu^2\sigma^2\boldsymbol{1'}_{h\Delta+\ell}\boldsymbol{KQK'LQL'}\boldsymbol{1}_{h\Delta+\ell}$.

Let $\boldsymbol{X} = [x_1, \ldots, x_n]'$ and define $\boldsymbol{X}^{2\odot} = \boldsymbol{X} \odot \boldsymbol{X} = [x_1^2, \ldots, x_n^2]'$. Furthermore, define vector $\boldsymbol{\mu}_{\boldsymbol{X}^{2\odot}} \in \mathbb{R}^{n \times 1}$ and matrices $\boldsymbol{\Sigma}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{X}^{2\odot}} \in \mathbb{R}^{n \times n}$ by

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathbb{E}(\boldsymbol{X}\boldsymbol{X}'), \quad \boldsymbol{\mu}_{\boldsymbol{X}^{2\odot}} = \mathbb{E}(\boldsymbol{X}^{2\odot}) \quad and \quad \boldsymbol{\Sigma}_{\boldsymbol{X}^{2\odot}} = \mathbb{E}(\boldsymbol{X}^{2\odot}\boldsymbol{X}^{2\odot'}) - \boldsymbol{\mu}_{\boldsymbol{X}^{2\odot}}\,\boldsymbol{\mu}'_{\boldsymbol{X}^{2\odot}},$$

respectively. Then, for symmetric matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$, we have

$$\text{Cov}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}, \boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}) = \text{tr}(\boldsymbol{C}(\boldsymbol{\Sigma}_{\boldsymbol{X}^{2\odot}} + \boldsymbol{\mu}_{\boldsymbol{X}^{2\odot}}\,\boldsymbol{\mu}'_{\boldsymbol{X}^{2\odot}})) - \text{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{X}})\text{tr}(\boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{X}}),$$

where $\boldsymbol{C} = \boldsymbol{a}\boldsymbol{b}' + 2\boldsymbol{A} \odot \boldsymbol{B} \odot (\mathbf{1}_n\mathbf{1}'_n - \boldsymbol{I}_n)$, with $\boldsymbol{a} = \text{diag}(\boldsymbol{A}) = (\boldsymbol{A} \odot \boldsymbol{I}_n)\mathbf{1}_n$ and $\boldsymbol{b} = \text{diag}(\boldsymbol{B}) = (\boldsymbol{B} \odot \boldsymbol{I}_n)\mathbf{1}_n$.

Proof. See Appendix A.2. □

Again, a corollary to Theorem 2 establishes the autocovariance function of the quadratic-form variance estimator when the daily log-return process, $(r_t)_{t \in \mathbb{Z}}$, is a weak white noise process (Definition 1) satisfying the moment conditions (8)-(11).

Corollary 2. *Let $(r_t)_{t \in \mathbb{Z}}$ be a weak white noise process fulfilling the moment conditions (8)-(11). Moreover, let $\sigma^2 = \text{Var}(r_t) = \mathbb{E}(r_t^2)$ and $\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell} = \boldsymbol{r}_{t,h\Delta+\ell} \odot \boldsymbol{r}_{t,h\Delta+\ell}$, and define vector $\boldsymbol{\mu}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}} \in \mathbb{R}^{h\Delta+\ell \times 1}$ and matrix $\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}} \in \mathbb{R}^{h\Delta+\ell \times h\Delta+\ell}$ by*

$$\boldsymbol{\mu}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}} = \mathbb{E}(\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}) \qquad and \qquad \boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}} = \mathbb{E}(\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}\boldsymbol{r}^{2\odot'}_{t,h\Delta+\ell}) - \boldsymbol{\mu}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}\boldsymbol{\mu}'_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}},$$

respectively. Then, considering variance estimates of the form $\sigma^2_{(h),t} = \boldsymbol{r}'_{t,h\Delta}\boldsymbol{Q}\boldsymbol{r}_{t,h\Delta}$, the autocovariance of the series $(\sigma^2_{(h),t})_{t \in \mathbb{Z}}$, for $\ell \geq 0$, is given by

$$\gamma_{\sigma^2_{(h),t}}(\ell) = \text{tr}\big(\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}\big) + 2\sigma^4\big(\text{tr}(\boldsymbol{K}\boldsymbol{Q}\boldsymbol{K}'\boldsymbol{L}\boldsymbol{Q}\boldsymbol{L}') - \boldsymbol{a}'\boldsymbol{b}\big),$$

with $\boldsymbol{C} = \boldsymbol{a}\boldsymbol{b}' + 2(\boldsymbol{K}\boldsymbol{Q}\boldsymbol{K}') \odot (\boldsymbol{L}\boldsymbol{Q}\boldsymbol{L}') \odot (\mathbf{1}_{h\Delta+\ell}\mathbf{1}'_{h\Delta+\ell} - \boldsymbol{I}_{h\Delta+\ell})$, where $\boldsymbol{a} = \text{diag}(\boldsymbol{K}\boldsymbol{Q}\boldsymbol{K}')$ and $\boldsymbol{b} = \text{diag}(\boldsymbol{L}\boldsymbol{Q}\boldsymbol{L}')$.

Proof. See Appendix A.3. □

Remark 1. *The following processes satisfy the conditions of Corollary 2 and especially the moment conditions (8)-(11):*

(i) For $\mu = 0$, the Gaussian white noise process (Example 1) clearly fulfills all conditions.

(ii) *Let $(x_t)_{t \in \mathbb{Z}}$ be a GARCH(p,q) process as defined in Example 2. For the innovations, $(\epsilon_t)_{t \in \mathbb{Z}}$, we assume a sequence of i.i.d. random variables being symmetrically distributed such that odd moments are zero. We further assume that the first four moments of $(x_t)_{t \in \mathbb{Z}}$ exist.[10] In Appendix A.4 we show that under these conditions the GARCH(p,q) satisfies all conditions such that Corollary 2 holds.*

(iii) *Under some regularity conditions, even more general classes of GARCH processes satisfy the conditions of Corollary 2. For discussions on families of GARCH processes and conditions on stationarity and the existence of moments see He and Teräsvirta (1999b) and Ling and McAleer (2002b).*

(iv) *Due to the moment conditions (8)-(11), the autocovariance in Corollary 2 only depends on the variance $\sigma^2 = \mathrm{Var}(r_t)$, $\boldsymbol{\mu}_{\boldsymbol{r}_{t,h\Delta+\ell}^{2\odot}}$, and the variance-covariance matrix of the vector of squares, $\boldsymbol{\Sigma}_{\boldsymbol{r}_{t,h\Delta+\ell}^{2\odot}}$. If the daily returns are, for example, asymmetrically distributed or follow a GARCH process with leverage, the moment conditions have to be weakened and additional terms, like unconditional skewness, are necessary to compute autocovariances. Note that these moments are often not available in closed form (He et al. 2008).*

The functional form of the relevant unconditional moments for different GARCH processes have been derived in He and Teräsvirta (1999a) and Karanasos (1999).

4.2 Illustration

We illustrate the theoretical results of the previous section by presenting plots of the theoretical ACF and those obtained from a simulation study. All illustrations in this section are for a GARCH(1,1) data generating process and the EWMA variance estimator (5).[11]

Let $(r_t)_{t \in \mathbb{Z}}$ be generated by the GARCH(1,1) process

$$r_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{12}$$

with $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and parameter vector, $[\alpha_0, \alpha_1, \beta_1]' := [0.01, 0.05, 0.94]'$, parameter values that are typical for daily stock returns. As estimator for the h-day variance at time t we use the EWMA estimator (5) with $\lambda = 0.96$. In view of the Basel III rules (BCBS 2016),

---

[10] Conditions for the existence of moments can be found in He and Teräsvirta (1999b) and Bollerslev (1986) for the GARCH(1,1) case and in Ling and McAleer (2002a) for GARCH(p,q).

[11] Appendix B.1 presents plots for a Gaussian white noise data generating process and variance being estimated by the sample variance (4).
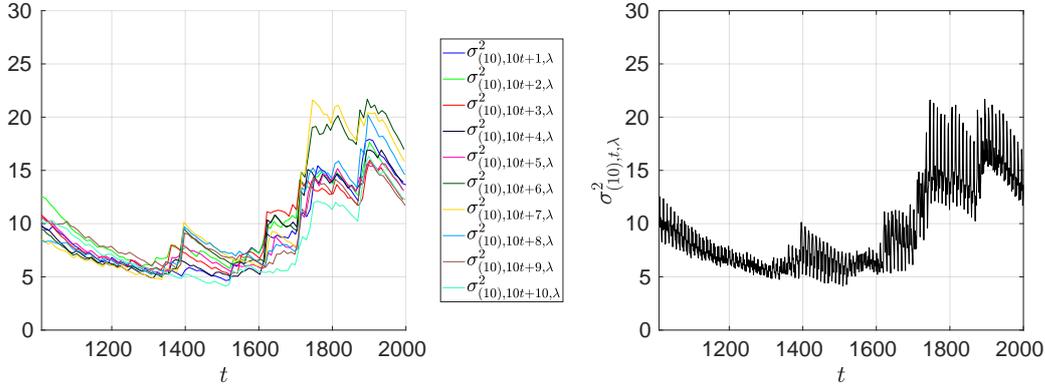
Figure 4: Time series of EWMA variance estimates (5), $\sigma^2_{(h),t,\lambda}$, for simulated daily return series from GARCH(1,1) process (12). The plot on the left shows the estimates $(\sigma^2_{(10),10t+\tau,\lambda})_{t\in\mathbb{Z}}$, for $1 \leq \tau \leq 10$. The right plot shows the series $(\sigma^2_{(10),t,\lambda})_{t\in\mathbb{Z}}$. Both plots are based on bi-weekly ($h = 10$) returns and estimation window $\Delta = 100$.

we chose $h = 10$ as aggregation horizon, i.e., we consider a bi-weekly target horizon as, for example, in Kole et al. (2017).

As for the ten-day return series itself, we can obtain ten different series of estimates for the variance if we synchronize assessment and sampling frequencies to be equal to the aggregation horizon of $h = 10$ days, namely, $(\sigma^2_{(h),ht+\tau,\lambda})_{t\in\mathbb{Z}}$, for $1 \leq \tau \leq h$.[12] As window length for the rolling-window estimates we choose $\Delta = 100$, giving rise to $h\Delta = 1000$ daily observations which corresponds to roughly four years of return data. At any point in time, each estimate is based on Δ non-overlapping h-day returns or $h\Delta$ daily returns. Two consecutive estimates, $\sigma^2_{(h),t,\lambda}$ and $\sigma^2_{(h),t+1,\lambda}$ (or $\sigma^2_{(h),t-1,\lambda}$), have $h\Delta - 2$ daily return observations in common. Our simulations are based on the GARCH(1,1) process defined in (12) with a sample size of $8 \times 250 = 2000$ trading days or about eight calendar years.

The left graph in Figure 4 shows the ten different series of variance estimates, $(\sigma^2_{(h),th+\tau,\lambda})_{t\in\mathbb{Z}}$, $1 \leq \tau \leq h$, obtained when assessment and sampling frequencies are synchronized. In each of the ten plots, two consecutive points, $(\sigma^2_{(h),th+\tau,\lambda})_{t\in\mathbb{Z}}$, $1 \leq \tau \leq h$, have distance $h = 10$. The graph on the right shows the sequence of daily EWMA variance estimates, $(\sigma^2_{(h),t,\lambda})_{t\in\mathbb{Z}}$, based on non-overlapping h-day returns.

The plots for the simulations in Figure 4 are constructed as those for the DJIA returns in Figure 2. The daily series of EWMA variance estimates (right graph in Figure 4) fluctuates in a highly regular fashion, mimicking a strong seasonal pattern. From a risk management

---

[12]By synchronization of the assessment and sampling frequency we mean that the value of $\tau \in \{1,\ldots,h\}$ is the same for the series of h-day returns, $(r_{(h),ht+\tau,\lambda})_{t\in\mathbb{Z}}$, and the series of (assessed) variance estimates, $(\sigma^2_{(h),ht+\tau,\lambda})_{t\in\mathbb{Z}}$. That is, both series are sampled on the same equidistant grid where we observe an h-day return and estimate the variance on every h-th day.
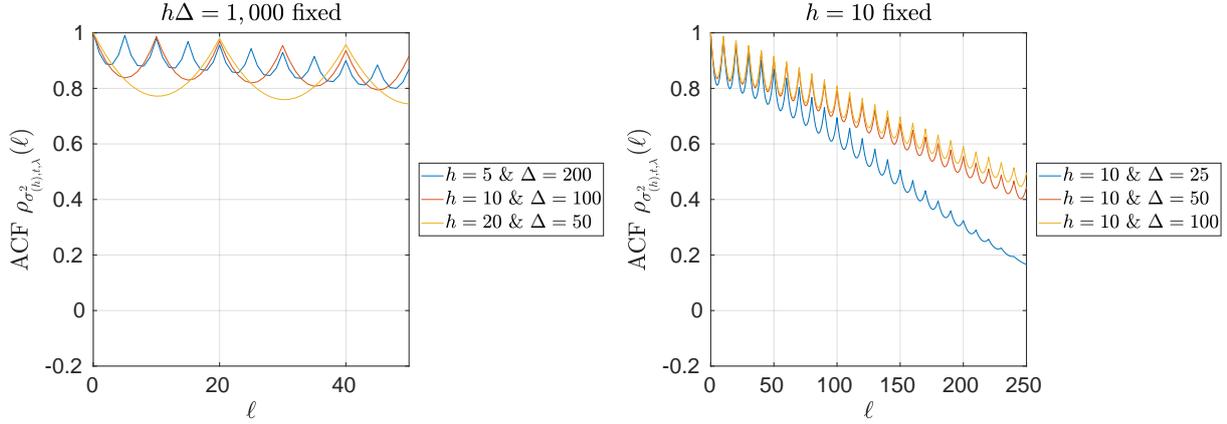
Figure 5: The ACF of EWMA variances (5), $\sigma^2_{(h),t,\lambda}$, for daily returns from GARCH(1,1) process (12). For the left plot we use a fixed number of daily returns to derive the EWMA variances. The right plot depicts the ACF of EWMA variances based on bi-weekly ($h = 10$) returns and estimation windows $\Delta = 25, 50, 100$.

perspective, such strong fluctuations are bound to have detrimental implications as they induce volatile risk capital charges and risk mitigation activities.

To derive the autocovariances of the series of variance estimates, $(\sigma^2_{(h),t,\lambda})_{t\in\mathbb{Z}}$, assuming a GARCH(1,1) process for the daily returns, $(r_t)_{t\in\mathbb{Z}}$, we use the fact that the variance-covariance matrix of the vector of squared returns, $\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell} = [r^2_{t-h\Delta-\ell+1}, \ldots, r^2_t]'$, is given by the symmetric Toeplitz matrix (cf. He and Teräsvirta (1999a) or Karanasos (1999))

$$\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell},[i,i+j]} = \begin{cases} \gamma_{r^2_t}(0) & , 1 \leq i \leq h\Delta + \ell, j = 0, \\ \frac{\gamma_{r^2_t}(1)}{(\alpha_1+\beta_1)^{1-j}} & , 1 \leq i \leq h\Delta + \ell - 1, 1 \leq j \leq h\Delta + \ell - i, \end{cases}$$

where $\gamma_{r^2_t}(0)$ and $\gamma_{r^2_t}(1)$ denote the variance and first-order autocovariance, respectively, of the squared returns from a GARCH(1,1) process. Thus, if daily returns follow a GARCH(1,1) process, the autocorrelation of the process of daily exponentially-weighted moving-average variances (5), $(\sigma^2_{(h),t,\lambda})_{t\in\mathbb{Z}}$, based on non-overlapping h-day returns, can be computed via Corollary 2 by plugging in $\boldsymbol{Q}_{(h),\Delta,\lambda}$, $\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}$ and $\mathrm{Var}(r_t) = \alpha_0/(1 - \alpha_1 - \beta_1)$ into the formula for the autocovariance and scaling via $\rho_{\sigma^2_{(h),t,\lambda}}(\ell) = \gamma_{\sigma^2_{(h),t,\lambda}}(\ell)/\gamma_{\sigma^2_{(h),t,\lambda}}(0)$.

The ACF with estimates based on daily return samples of size $h\Delta = 1000$ is presented on the left in Figure 5. The graph shows the effect for the aggregation horizons $h = 5, 10, 20$, amounting to quasi-weekly, quasi-bi-weekly and quasi-monthly return periods. It demonstrates that the ACF of EWMA variances, $\rho_{\sigma^2_{(h),t,\lambda}}(\ell)$, based on non-overlapping h-day returns, is highly cyclical and slowly decaying. The (spurious) seasonality that is present in the sample ACF of estimated variances for the DJIA data (Figure 3) is compatible with the (spurious) seasonality in the theoretical ACF in Figure 5. The right graph in Figure 5 fur-

ther illustrates the interaction between aggregation horizon, h, and the window length, Δ. The aggregation horizon is bi-weekly ($h = 10$) and the window length, Δ, assumes values 25, 50 and 100, i.e., roughly one, two and four calendar years of daily returns, respectively.

The formula for the autocovariance $\gamma_{\sigma^2_{(h),t,\lambda}}(\ell)$ given in Corollary 2 and for the ACF are rather handy. But it offers little insight into where the spurious seasonality exactly comes from or how the amplitude of the periodic spurious seasonality in the ACF depends on the variance estimator and the data generating process. Appendix B.2 expresses the ACF as the sum of three components, which provide more insight and show that the term $(\boldsymbol{KQK'}) \odot (\boldsymbol{LQL'})$ is crucially responsible for the spurious seasonality in the ACF.

The left graph in Figure 4 shows a pronounced periodicity, though it is not always the same observation within the h-day periods that assumes the highest or lowest value. In other words, the order statistics of the different estimates within an h-day period fluctuate, but do so rather slowly. Therefore, if the focus is on bi-weekly risk estimation but assessment occurs at a daily frequency, then, by construction, the ordering of the ten different variance estimates in a two-week period gradually changes over time. The color-coded series of EWMA variances on the left in Figure 4 make clear that, at some point in time, any particular color may be on top (or bottom) and that there is a high probability that this will also hold for the following h-day period.

To get further insights into why the order statistics gradually change over time, we take a look at the ACF of the first difference of the estimated variances,

$$\gamma_{\sigma^2_{(h),t}-\sigma^2_{(h),t-1}}(\ell) = 2\gamma_{\sigma^2_{(h),t}}(\ell) - \gamma_{\sigma^2_{(h),t}}(\ell+1) - \gamma_{\sigma^2_{(h),t}}(|\ell-1|),$$

$\ell \geq 0$. The ACF of the first differences of the estimated variances, $\rho_{\sigma^2_{(h),t,\lambda}-\sigma^2_{(h),t-1,\lambda}}(\ell)$, is plotted in Figure 6, where we used the same settings as for the ACF of the EWMA variances shown in Figure 5. The plots in Figure 6 demonstrate that the series of first differences is highly autocorrelated for lags being multiples of the aggregation horizon, h. This means, the change of the estimate from one day to another is highly autocorrelated with that of h days ago. At lags that are not multiples of h, the autocorrelations of the first-order differences are quite small and slightly negative. This behavior explains the slowly changing ordering of the h different variance series in the left graph of Figure 4.

5 The Case of Overlapping Aggregated Returns

The previous section showed that daily variance estimates based on non-overlapping h-day returns suffer from spurious seasonality. In cases where the aggregation horizon is fixed,
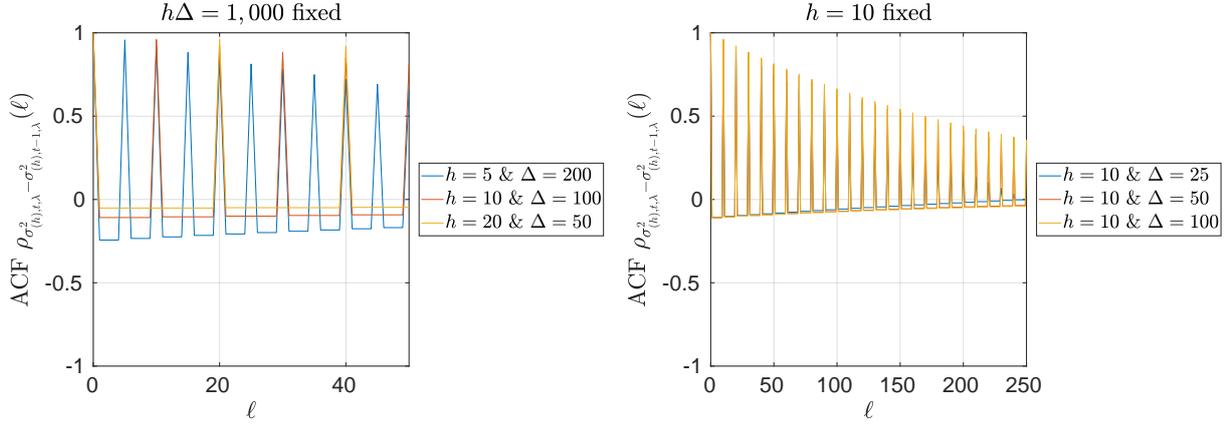
Figure 6: The ACF of the first difference of EWMA variances (5), $\sigma^2_{(h),t,\lambda}$, for daily returns from GARCH(1,1) process (12). For the left plot we use a fixed number of daily returns to derive the EWMA variances. The right plot depicts the ACF of the first difference of EWMA variances based on bi-weekly ($h = 10$) returns and estimation windows $\Delta = 25, 50, 100$.

the only alternative is to synchronize both assessment and sampling at a daily frequency, i.e., to use overlapping h-day returns for daily risk estimations. In the following we consider several variance estimators for overlapping returns that avoid spurious seasonality.

The simplest variance estimator based on overlapping h-day returns is to apply standard formulas for variance estimators to sets of overlapping return observations. The quadratic-form representations for the sample variance for overlapping h-day returns is given by

$$\check{\sigma}^2_{(h),t} = \frac{h}{\text{tr}(\check{\mathcal{Q}}_{(h),\Delta})} \frac{1}{h(\Delta-1)+1} \sum_{\tau=0}^{h(\Delta-1)} \left(r_{(h),t-\tau} - \check{\mu}_{(h),t}\right)^2 = \boldsymbol{r}'_{t,h\Delta} \check{\boldsymbol{Q}}_{(h),\Delta} \boldsymbol{r}_{t,h\Delta}, \qquad (13)$$

with $\check{\mathcal{Q}}_{(h),\Delta} = \frac{1}{h(\Delta-1)+1} \sum_{j=0}^{h-1} \boldsymbol{S}'_j \boldsymbol{S}_j$, $\boldsymbol{S}_0 = \boldsymbol{H}' - \boldsymbol{1}_\Delta \boldsymbol{1}'_\Delta \boldsymbol{J}'$ and $\boldsymbol{S}_j = \boldsymbol{H}'_j - \boldsymbol{a}\boldsymbol{1}'_\Delta \boldsymbol{J}'$, for $1 \le j \le h-1$, where $\boldsymbol{J} = \boldsymbol{H} + \sum_{j=1}^{h-1} \boldsymbol{H}_j$, $\boldsymbol{a} = \begin{bmatrix} \boldsymbol{1}'_{\Delta-1} & 0 \end{bmatrix}'$ and

$$\boldsymbol{H}_j = \begin{bmatrix} \boldsymbol{0}_{(h-j\times 1)} & \boldsymbol{0}_{(h-j\times\Delta-1)} \\ \boldsymbol{0}_{(h(\Delta-1)\times 1)} & \boldsymbol{I}_{\Delta-1} \otimes \boldsymbol{1}_h \\ \boldsymbol{0}_{(j\times 1)} & \boldsymbol{0}_{(j\times\Delta-1)} \end{bmatrix},$$

and the EWMA variance for overlapping h-day returns by

$$\check{\sigma}^2_{(h),t,\lambda} = \frac{h}{\text{tr}(\check{\mathcal{Q}}_{(h),\Delta,\lambda})} \frac{1-\lambda^{\frac{1}{h}}}{1-\lambda^{\frac{h(\Delta-1)+1}{h}}} \sum_{\tau=0}^{h(\Delta-1)} \lambda^{\frac{\tau}{h}} \left(r_{(h),t-\tau} - \check{\mu}_{(h),t,\lambda}\right)^2 = \boldsymbol{r}'_{t,h\Delta} \check{\boldsymbol{Q}}_{(h),\Delta,\lambda} \boldsymbol{r}_{t,h\Delta}, \quad (14)$$

with $\check{\mathcal{Q}}_{(h),\Delta,\lambda} = \sum_{j=0}^{h-1} \lambda^{\frac{j}{h}} \tilde{\boldsymbol{S}}'_j \boldsymbol{\Gamma} \tilde{\boldsymbol{S}}_j$, $\tilde{\boldsymbol{S}}_0 = \boldsymbol{H}' - \boldsymbol{1}_\Delta \boldsymbol{\gamma}' \boldsymbol{G}'$ and $\tilde{\boldsymbol{S}}_j = \boldsymbol{H}'_j - \boldsymbol{a}\boldsymbol{\gamma}' \boldsymbol{G}'$, for $1 \le j \le h-1$,

with $\boldsymbol{G} = \boldsymbol{H} + \sum_{j=1}^{h-1} \lambda^{\frac{j}{h}} \boldsymbol{H}_j$ and, for $\lambda \in (0,1)$, $\boldsymbol{\Gamma} = \mathrm{Diag}(\boldsymbol{\gamma}) = (\boldsymbol{\gamma}\boldsymbol{1}'_\Delta) \odot \boldsymbol{I}_\Delta$, where $\boldsymbol{\gamma} = \frac{1-\lambda^{1/h}}{1-\lambda^{(h(\Delta-1)+1)/h}} \left[\lambda^{\Delta-1}, \lambda^{\Delta-2}, \ldots, \lambda^1, 1,\right]'$. Graphical evidence (not shown here) indicates that these variance estimators do not suffer from spurious seasonality.[13]

Another approach to avoid spurious seasonality is to simply take the average of the last h sample variances based on non-overlapping h-day returns. In the (ultra-)high-frequency context, this type of post-averaging of subsampling-based variance estimates has been proposed in Zhang et al. (2005) to overcome problems arising from microstructure noise. It is referred to as two-scales realized volatility and provides a consistent estimator of integrated volatility under the assumption of additive white noise. In the following, we show that the two-scales estimator has the potential to solve the spurious seasonality problem in variance estimation.[14]

For our setting, we obtain the two-scales sample variance[15]

$$\bar{\sigma}^2_{(h),t} = \frac{1}{h} \boldsymbol{r}'_{t,h\Delta} \boldsymbol{Q}_{(h),\Delta} \boldsymbol{r}_{t,h\Delta} + \frac{1}{h} \sum_{j=1}^{h-1} \boldsymbol{r}'_{t-j,h(\Delta-1)} \boldsymbol{Q}_{(h),\Delta-1} \boldsymbol{r}_{t-j,h(\Delta-1)} = \boldsymbol{r}'_{t,h\Delta} \bar{\boldsymbol{Q}}_{(h),\Delta} \boldsymbol{r}_{t,h\Delta}, \quad (15)$$

with $\bar{\boldsymbol{Q}}_{(h),\Delta} = \frac{1}{h} \boldsymbol{Q}_{(h),\Delta} + \frac{1}{h} \sum_{j=1}^{h-1} \boldsymbol{T}_j(\boldsymbol{Q}_{(h),\Delta-1})$, where for symmetric matrices $\boldsymbol{Q} \in \mathbb{R}^{h(\Delta-1) \times h(\Delta-1)}$

$$\boldsymbol{T}_j(\boldsymbol{Q}) = \begin{bmatrix} \boldsymbol{0}_{(h-j \times h-j)} & \boldsymbol{0}_{(h-j \times h(\Delta-1))} & \boldsymbol{0}_{(h-j \times j)} \\ \boldsymbol{0}_{(h(\Delta-1) \times h-j)} & \boldsymbol{Q} & \boldsymbol{0}_{(h(\Delta-1) \times j)} \\ \boldsymbol{0}_{(j \times h-j)} & \boldsymbol{0}_{(j \times h(\Delta-1))} & \boldsymbol{0}_{(j \times j)} \end{bmatrix}.$$

For the two-scales EWMA variance we have

$$\bar{\sigma}^2_{(h),t,\lambda} = \frac{1-\lambda^{\frac{1}{h}}}{1-\lambda} \left(\boldsymbol{r}'_{t,h\Delta} \boldsymbol{Q}_{(h),\Delta,\lambda} \boldsymbol{r}_{t,h\Delta} + \sum_{j=1}^{h-1} \lambda^{\frac{j}{h}} \boldsymbol{r}'_{t-j,h(\Delta-1)} \boldsymbol{Q}_{(h),\Delta-1,\lambda} \boldsymbol{r}_{t-j,h(\Delta-1)} \right)$$

$$= \boldsymbol{r}'_{t,h\Delta} \bar{\boldsymbol{Q}}_{(h),\Delta,\lambda} \boldsymbol{r}_{t,h\Delta}, \quad (16)$$

---

[13]It turns out that there are two additional variance estimators based on overlapping returns that overcome spurious seasonality. In this section we report results for only one of the three variance estimators and refer to Appendix B.3 for graphical results and a comparison of all four EWMA variance estimators considered.

[14]In the high-frequency literature, several other variance estimators, such as the the multi-scales realized volatility (Zhang 2006) and the pre-averaging approach (Jacod et al. 2009), have been proposed. As the two-scales estimator, they can overcome spurious seasonality. We restrict ourselves, however, to the two-scales estimator of Zhang et al. (2005), since it is the simplest variance estimator handling the problem of spurious seasonality.

[15]Use of $\boldsymbol{\mathcal{Q}}_{(h),\Delta}$ instead of $\boldsymbol{Q}_{(h),\Delta}$ in (15) yields the biased "second-best" two-scales variance estimator of Zhang et al. (2005). The biased-corrected version (15), $\bar{\sigma}^2_{(h),t}$, is obtained by using $\boldsymbol{Q}_{(h),\Delta}$.
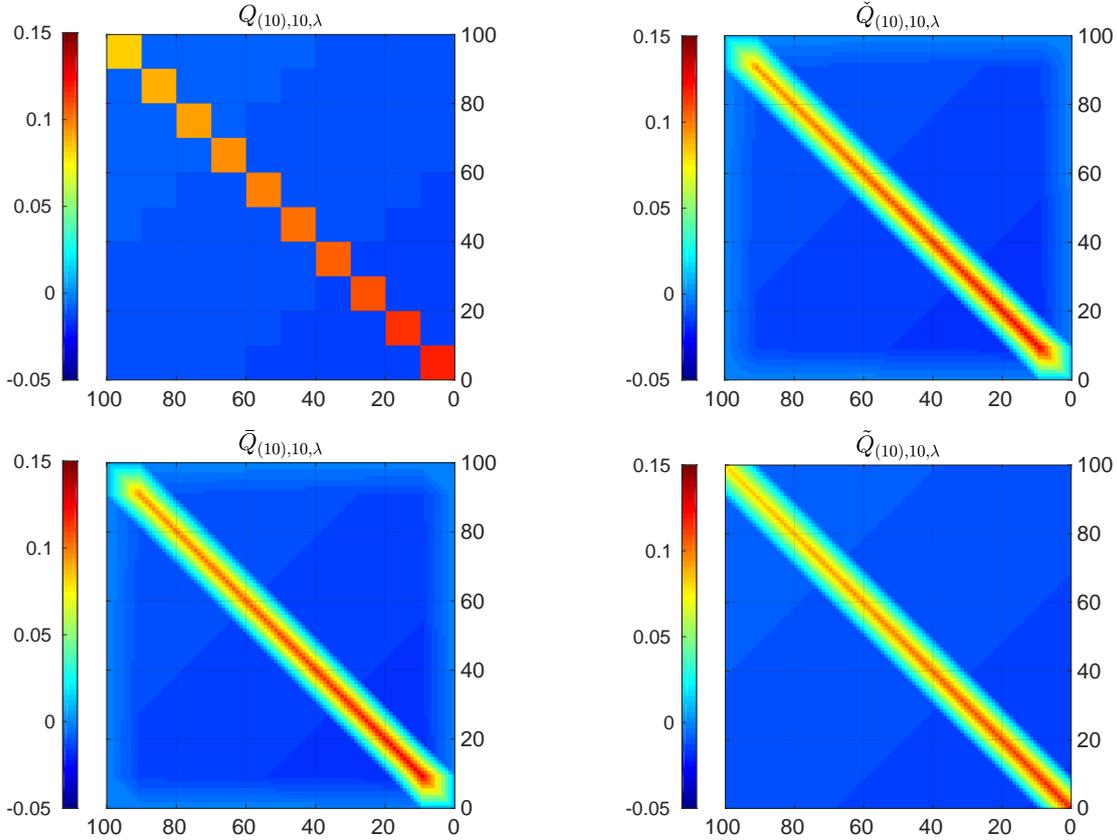
Figure 7: Heatmaps of the quadratic-form matrices $\boldsymbol{Q}$ for different EWMA variance estimators of the form (3). For this illustration, we set $h = 10$, $\Delta = 10$ and $\lambda = 0.96$. The pixels correspond to the entries of $\boldsymbol{Q}$ and reflect the weights of cross-products $r_{t-x}r_{t-y}$. The magnitude of a weight is indicated by the color scale. The upper plots show the standard estimator based on non-overlapping h-day returns (left; Eq. (5)) and that based on overlapping h-day returns (right; Eq. (14)). The lower plots show subsampling-based post-average EWMA variance estimators. The left plot corresponds to the two-scales variance (16) and the right to the boundary corrected version (18).

with $\bar{\boldsymbol{Q}}_{(h),\Delta,\lambda} = \frac{1-\lambda^{\frac{1}{h}}}{1-\lambda}\boldsymbol{Q}_{(h),\Delta,\lambda} + \frac{1-\lambda^{\frac{1}{h}}}{1-\lambda}\sum_{j=1}^{h-1}\lambda^{\frac{j}{h}}\boldsymbol{T}_j(\boldsymbol{Q}_{(h),\Delta-1,\lambda})$.

The variance estimators studied so far can be written as quadratic forms and, thus, straightforwardly visualized in form of heatmaps as shown in Figure 7. Each pixel in a heatmap corresponds to an entry in matrix $\boldsymbol{Q}$ in quadratic form $\boldsymbol{r}'_{t,h\Delta}\boldsymbol{Q}\boldsymbol{r}_{t,h\Delta}$ in (3). The entries in $\boldsymbol{Q}$ can be interpreted as weights of cross-products $r_{t-x}r_{t-y}$, whose magnitude is indicated by the color scale in Figure 7. It is evident that the subsampling-based post-average EWMA variance estimator (16), shown in the lower left plot in Figure 7 as well as the EWMA variance based on overlapping h-day returns (top right) given by (14) have boundary problems. The weights assigned to the most recent squared return observations, r_{t-j}^2, $0 \leq j \leq h-1$, are much lower than the weight of the lagged squared return, r_{t-h}^2.

To eliminate such undesirable boundary effects, we propose modifications to the two-

scales estimator of Zhang et al. (2005) and the EWMA variant. The two-scales sample variance with boundary-correction is given by

$$\tilde{\sigma}^2_{(h),t} = \boldsymbol{r}'_{t,h\Delta}\tilde{\boldsymbol{Q}}_{(h),\Delta}\boldsymbol{r}_{t,h\Delta}, \tag{17}$$

with $\tilde{\boldsymbol{Q}}_{(h),\Delta} = \frac{h}{\mathrm{tr}(\tilde{\boldsymbol{\mathcal{Q}}}_{(h),\Delta})}\tilde{\boldsymbol{\mathcal{Q}}}_{(h),\Delta}$ and $\tilde{\boldsymbol{\mathcal{Q}}}_{(h),\Delta} \in \mathbb{R}^{h\Delta\times h\Delta}$ being a symmetric Toeplitz matrix with the j-th (off-)diagonal element given by

$$\tilde{\boldsymbol{\mathcal{Q}}}_{(h),\Delta,[i,i+j]} = \begin{cases} \frac{1}{\Delta}\left(1 - \frac{1}{\Delta}\right)\left(1 - \frac{j\Delta}{h\Delta - j}\right) & ,1 \le i \le h\Delta,\, 0 \le j \le \min\{h\Delta - i, h - 1\}, \\ -\frac{1}{\Delta^2} & ,1 \le i \le h(\Delta - 1),\, h \le j \le h\Delta - i. \end{cases}$$

The two-scales EWMA variance with boundary-correction becomes

$$\tilde{\sigma}^2_{(h),t,\lambda} = \boldsymbol{r}'_{t,h\Delta}\tilde{\boldsymbol{Q}}_{(h),\Delta,\lambda}\boldsymbol{r}_{t,h\Delta}, \tag{18}$$

where $\tilde{\boldsymbol{Q}}_{(h),\Delta,\lambda} = \frac{h}{\mathrm{tr}(\tilde{\boldsymbol{\mathcal{Q}}}_{(h),\Delta,\lambda})}\tilde{\boldsymbol{\mathcal{Q}}}_{(h),\Delta,\lambda}$, with symmetric matrix $\tilde{\boldsymbol{\mathcal{Q}}}_{(h),\Delta,\lambda} \in \mathbb{R}^{h\Delta\times h\Delta}$ being defined by $\tilde{\boldsymbol{\mathcal{Q}}}_{(h),\Delta,\lambda} = \boldsymbol{\Psi} - \boldsymbol{\Xi}$, where $\boldsymbol{\Psi}, \boldsymbol{\Xi} \in \mathbb{R}^{h\Delta\times h\Delta}$ are symmetric matrices with entries

$$\boldsymbol{\Psi}_{[i,i+j]} = \begin{cases} \frac{(h-j)(1-\lambda^{\frac{1}{h}})}{1-\lambda^{\frac{h\Delta-j}{h}}}\lambda^{\frac{h\Delta-i}{h}} & ,1 \le i \le h\Delta,\, 0 \le j \le \min\{h\Delta - i, h - 1\}, \\ 0 & ,1 \le i \le h(\Delta - 1),\, h \le j \le h\Delta - i, \end{cases}$$

and, for $1 \le i \le h\Delta,\, 0 \le j \le h\Delta - i$,

$$\boldsymbol{\Xi}_{[i,i+j]} = \lambda^{\frac{2(h\Delta-i-j)}{h}+\delta}\frac{(1-\lambda)^2(1-\lambda^{\frac{2}{h}})\big[(h-k)(1-\lambda^{2(\Delta-\delta)}) + k\lambda(1-\lambda^{2(\Delta-\delta-1)})\big]}{(1-\lambda^\Delta)^2(1-\lambda^2)(1-\lambda^{\frac{2(h\Delta-j)}{h}})},$$

with $\delta = \lfloor\frac{j}{h}\rfloor$, $k = j - h\lfloor\frac{j}{h}\rfloor = j - h\delta$ and $\lambda \in (0,1)$. The heatmap of the weighting scheme for the two-scales EWMA variance estimator with boundary-correction (18), associated with the quadratic-form matrix $\tilde{\boldsymbol{Q}}_{(h),\Delta,\lambda}$, is depicted in the lower right of Figure 7.

If we apply the subsampling-based post-average EWMA variance estimator with boundary correction (18) to the same simulated GARCH(1,1) series used in Figure 4, we obtain the series of variance estimators plotted in Figure 8. The right panel in Figure 8 clearly shows the absence of spurious seasonality as compared to the right panel in Figure 4. The ten different series of variance estimates (left in Figure 8), where assessment and sampling frequencies are in sync and equal to the aggregation horizon, h, turn out to be much more stable; and long-memory effects in the order statistics, associated with h-day periods, are no longer present.
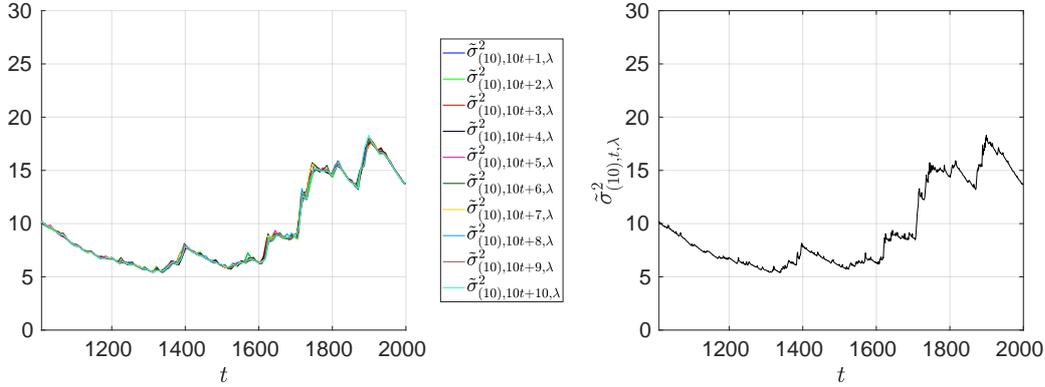
Figure 8: Time series of subsampling-based post-average EWMA variance estimates with boundary correction (18), $\tilde{\sigma}^2_{(h),t,\lambda}$, for simulated daily return series from GARCH(1,1) process (12). The plot on the left shows the estimates $(\tilde{\sigma}^2_{(10),10t+\tau,\lambda})_{t\in\mathbb{Z}}$, for $1 \leq \tau \leq 10$. The right plot shows the series $(\tilde{\sigma}^2_{(10),t,\lambda})_{t\in\mathbb{Z}}$. Both plots are based on bi-weekly ($h = 10$) returns and estimation window $\Delta = 100$.
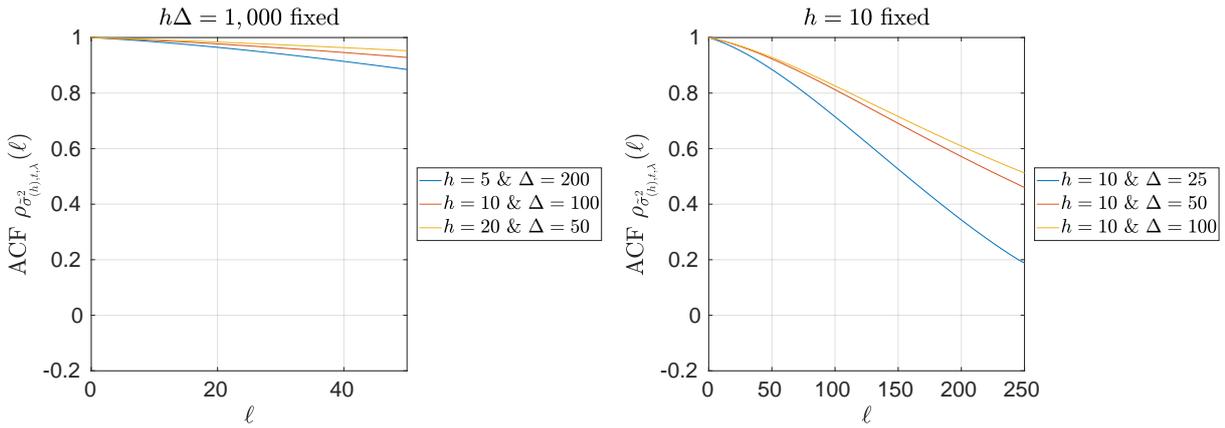
Figure 9: The ACF of subsampling-based post-average EWMA variances with boundary correction (18), for daily returns from GARCH(1,1) process (12). For the left plot we use a fixed number of daily returns to derive the EWMA variances. The right plot depicts the ACF of EWMA variances based on bi-weekly ($h = 10$) returns and estimation windows $\Delta = 25, 50, 100$.

Note that the autocovariances for the daily subsampling-based post-average EWMA variance estimates (with boundary correction) can again be obtained from Corollary 2. The corresponding ACF of the estimated variances is shown in Figure 9. The data generating process and the combinations of aggregation horizons, h, and estimation window length, Δ, are the same, as in Figure 5. The periodicity in the ACF has been eliminated, and the functional form of the ACF seems reasonable for an EWMA type estimator of the variance.

In Figure 10 we repeat the DJIA analysis (Section 2) using the boundary-corrected two-scales variance estimates (18) instead. The construction of the plot is exactly as in Figure 2. We see that the problem of spurious seasonality is no longer present and that the ten different variance series, shown on the left in Figure 10, do no longer slowly change
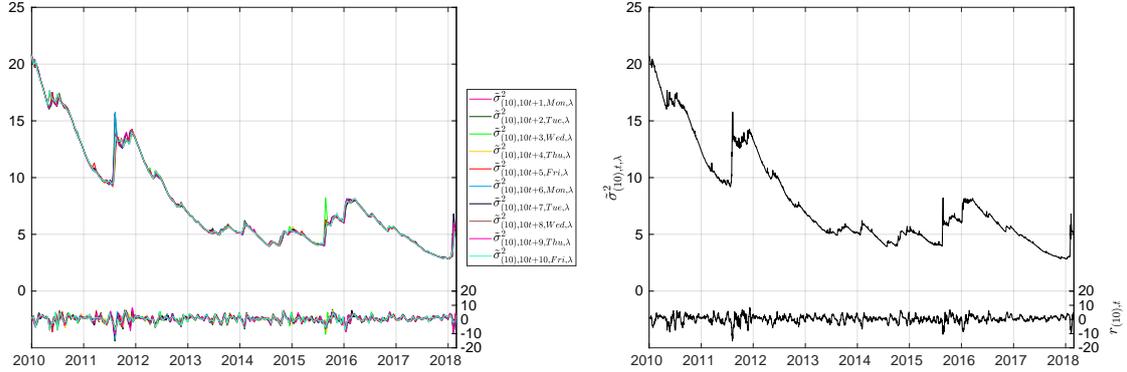
Figure 10: Estimated boundary-corrected two-scales EWMA variances of the Dow Jones Industrial Average (DJIA) based on ten-day log-returns with a window length of 100 bi-weekly returns and an EWMA parameter of $\lambda = 0.96$. The first (last) estimates in both graphs are for 01-Jan-2010 (for 28-Feb-2018). The graph on the left shows at the top ten series of bi-weekly variance estimates, each corresponding to a specific weekday and start date, and the one on the right the daily series of bi-weekly variance estimates. The corresponding ten-day log-returns are plotted at the bottom of both graphs.

their position.

6 Properties of the Variance Estimators

Apart from the presence or absence of spurious seasonality, properties such as bias, variance and mean-squared error (MSE) of variance estimators are typically of interest. Another concern is the question whether or not the volatility dynamics are captured in an adequate fashion (Engle and Patton 2001). This is reflected by the responsiveness of variance estimators with respect to shocks. Both issues are addressed next.

6.1 Bias, Variance and MSE

We use the quadratic-form representation (3) for estimating the unconditional variance of a weak white noise process that satisfies the moment conditions (8)-(11) to derive the bias, variance and mean squared error (MSE) of the variance estimators. Denoting the unconditional variance of process $(r_t)_{t\in\mathbb{Z}}$ by $\sigma^2 = \mathrm{Var}(r_t)$, the bias is given by (2) as $\mathrm{Bias}(\boldsymbol{r}'_{t,h\Delta}\boldsymbol{\mathcal{Q}}\boldsymbol{r}_{t,h\Delta}) = \sigma^2(\mathrm{tr}(\boldsymbol{\mathcal{Q}}) - h)$. From Corollary 2 we obtain

$$\mathrm{Var}(\boldsymbol{r}'_{t,h\Delta}\boldsymbol{\mathcal{Q}}\boldsymbol{r}_{t,h\Delta}) = \mathrm{tr}\big(\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}\big) + \sigma^4\big(\mathrm{tr}(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{Q}}) - \boldsymbol{q}'\boldsymbol{q}\big),$$

with $\boldsymbol{C} = \boldsymbol{q}\boldsymbol{q}' + 2\boldsymbol{\mathcal{Q}}^{2\odot} \odot (\boldsymbol{1}_{h\Delta}\boldsymbol{1}'_{h\Delta} - \boldsymbol{I}_{h\Delta})$ and $\boldsymbol{q} = \mathrm{diag}(\boldsymbol{\mathcal{Q}}) = (\boldsymbol{\mathcal{Q}} \odot \boldsymbol{I}_{h\Delta})\boldsymbol{1}_{h\Delta}$.
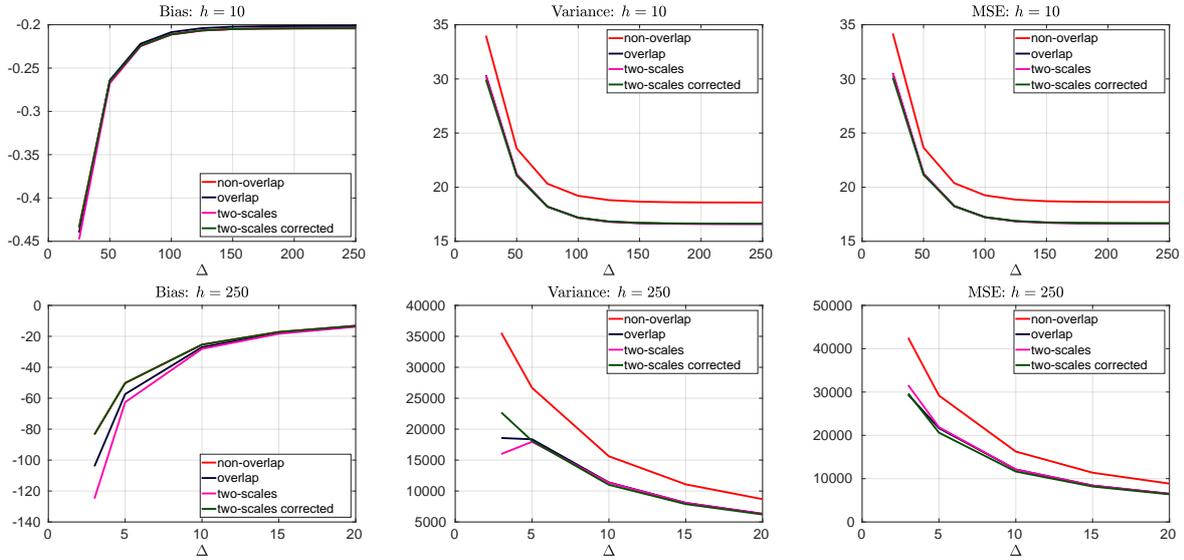
Figure 11: Bias, variance and mean squared error (MSE) of EWMA variance estimators (5), (14), (16), and (18). For the series of daily returns, $(r_t)_{t \in \mathbb{Z}}$, GARCH(1,1) process (12) is assumed. The upper (lower) panel shows results for aggregation horizon $h = 10$ ($h = 250$) for different window sizes, Δ. The results for bias, variance and MSE are shown in left, center and right panels, respectively.

In the following, we assume $(r_t)_{t \in \mathbb{Z}}$ follows the GARCH(1,1) process (12). As estimators for the variance we analyze the different EWMA variance estimators $\sigma^2_{(h),t,\lambda}$ (non-overlapping h-day returns, (5)), $\check{\sigma}^2_{(h),t,\lambda}$ (overlapping h-day returns, (14)), $\bar{\sigma}^2_{(h),t,\lambda}$ (two-scales, (16)), and $\tilde{\sigma}^2_{(h),t,\lambda}$ (corrected two-scales, (18)). The top panel in Figure 11 shows the bias, variance and MSE for $h = 10$ and window sizes, Δ, ranging from 25 to 250. For a low aggregation horizon ($h = 10$) all estimators have a similar bias. This is in line with the findings of Bod et al. (2002). With respect to variance and MSE, the three estimators based on overlapping returns produce smaller values. The results differ, however, when the aggregation level increases to $h = 250$ (bottom panel in Figure 11). The standard overlapping estimator, $\check{\sigma}^2_{(h),t}$, and the two-scales estimator, $\bar{\sigma}^2_{(h),t}$, produce the highest absolute bias. In terms of the MSE, the corrected two-scales estimator performs best.

6.2 Responsiveness to Shocks

Especially the EWMA variance estimator, which is not only used for estimating the unconditional but also the conditional variance, turns out to be more responsive to recent shocks. To illustrate and compare the responsiveness of the estimators we focus on the last 100 diagonal elements of the quadratic-form matrices $\boldsymbol{Q}$. These elements correspond to the weights the respective variance estimators assign to the squared daily returns $r_t^2, r_{t-1}^2, \ldots, r_{t-99}^2$. The left (right) graph in Figure 12 plots the weights of the different sample (EWMA) variance
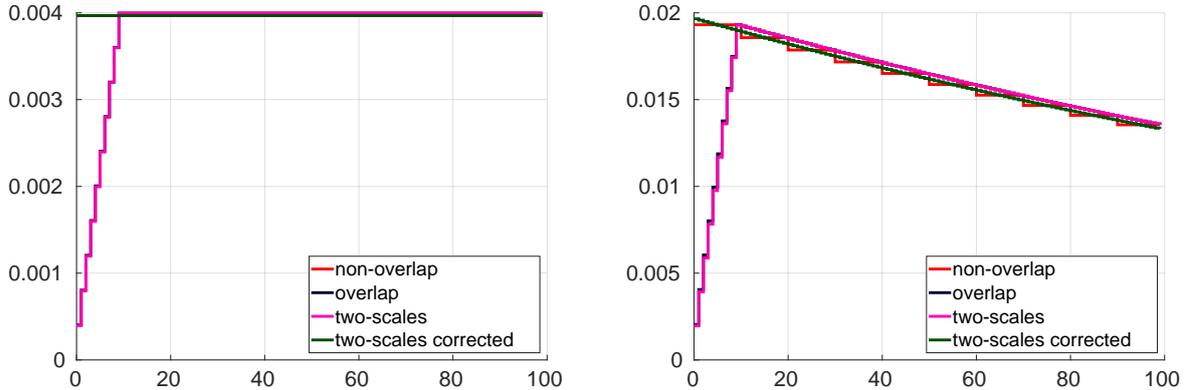
Figure 12: Values of the diagonal elements of the quadratic-form matrices for variance estimators of the form (3), $r'_{t,h\Delta}Qr_{t,h\Delta}$. The horizontal axis indicates the diagonal position in matrix Q. The diagonal entries reflect the weights assigned to the squared daily returns, r^2_{t-x}. The left plot shows the entries for variance estimators (4), (13), (15), and (17); and the one on the right for EWMA variance estimators (5), (14), (16), and (18).

estimators. As becomes evident, the corrected two-scales estimators do not suffer from the boundary issue as do the standard two-scales estimators and the standard variance estimators based on overlapping returns. Furthermore, in the EWMA case, the corrected two-scales estimator allocates the weights more smoothly to past squared shocks than the estimators based on non-overlapping returns.

In summary, the corrected two-scales estimator does not suffer from spurious seasonality and dominates other overlapping-return estimators in terms of bias, variance and MSE as well as the responsiveness with respect to recent shocks. A shortcoming of the corrected two-scales estimator is the fact that, in contrast to the other estimators discussed, it cannot be directly expressed as an estimator based on (non-)overlapping h-day returns. But, as the other estimators, it has a quadratic-form representation in terms of the daily return vector $r_{t,h\Delta}$, i.e., $r'_{t,h\Delta}Qr_{t,h\Delta}$ (3).

7 Concluding Remarks

We have investigated the phenomenon of spurious seasonality in sequentially estimated variances. It arises when the assessment frequency is higher than the sampling frequency of the (non-overlapping) return data used for estimation. The phenomenon, which, to our knowledge, has not yet been addressed in the literature, is attributable to an implicit overlap in the return data used for estimation. To provide a better understanding of this phenomenon, we have analyzed the properties of series of variance estimates in terms of their theoretical autocorrelation functions, considering a large class of data generating processes

and various alternative variance estimators. We have shown ways how to overcome the problem of spurious seasonality, introducing an EWMA-based estimator and a boundary correction for the two-scales estimator of Zhang et al. (2005).

In our analysis, we have focused exclusively on variance estimation. However, the phenomenon of spurious seasonality also translates directly to other risk measures, such as value-at-risk or expected shortfall, which are widely used in order to determine the capital requirements of financial institutions. As a consequence, capital charges based on such risk estimates will be subject to spurious seasonality. Risk managers and regulators need to be aware of that phenomenon and, more importantly, understand it in order to establish sound risk management practices. Our findings also provide an explanation for the variation in daily GARCH-parameter estimates derived from different non-overlapping monthly samples reported in Hedegaard and Hodrick (2016). Finally, although we have simplified our discussion by focussing on a daily data frequency, it should be understood that spurious seasonality also arises with other frequencies, such as in (ultra-)high-frequency realized-volatility analysis.

References for Chapter V

Andersen, T. G., and T. Bollerslev. 1998. "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts". *International Economic Review* 39 (4): 885–905.

Andersen, T. G., T. Bollerslev, and S. Lange. 1999. "Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon". *Journal of Empirical Finance* 6 (5): 457–477.

BCBS. 2016. "Minimum capital requirements for market risk". `http://www.bis.org/bcbs/publ/d352.htm`.

Bod, P., D. Blitz, P. H. Franses, and R. Kluitman. 2002. "An unbiased variance estimator for overlapping returns." *Applied Financial Economics* 12 (3): 155–158.

Bollerslev, T. 1986. "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics* 31 (3): 307–327.

Christoffersen, P., F. X. Diebold, and T. Schuermann. 1998. "Horizon problems and extreme events in financial risk management". *Economic Policy Review* 4 (3): 98–16.

Daníelsson, J., and C. Zhou. 2016. "Why risk is so hard to measure". DNB Working Paper, No. 494 / January 2016, URL: `https://www.dnb.nl/en/news/dnb-publications/dnb-working-papers-series/dnb-working-papers/working-papers-2016/dnb335907.jsp`.

Daníelsson, J., and J.-P. Zigrand. 2006. "On time-scaling of risk and the square-root-of-time rule". *Journal of Banking & Finance* 30 (10): 2701–2713.

Daníelsson, J., K. R. James, M. Valenzuela, and I. Zer. 2016. "Model risk of risk models". *Journal of Financial Stability* 23: 79–91.

Diebold, F. X., A. Hickman, A. Inoue, and T. Schuermann. 1997. "Converting 1-day Volatility to h-day Volatility: Scaling by h is Worse than You Think". *Wharton Financial Institutions Center, Working Paper 97-34*.

Drost, F. C., and T. E. Nijman. 1993. "Temporal Aggregation of GARCH Processes". *Econometrica* 61 (4): 909–927.

Embrechts, P., R. Kaufmann, and P. Patie. 2005. "Strategic Long-Term Financial Risks: Single Risk Factors". *Computational Optimization and Applications* 32 (1): 61–90.

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 A. J. Patton. 2001. "What good is a volatility model?" *Quantitative Finance* 1 (2): 237–245.

Hansen, L. P., and R. J. Hodrick. 1980. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis". *Journal of Political Economy* 88 (5): 829–853.

He, C., and T. Teräsvirta. 1999a. "Fourth Moment Structure of the GARCH(p,q) Process". *Econometric Theory* 15 (6): 824–846.

— . 1999b. "Properties of moments of a family of GARCH processes". *Journal of Econometrics* 92 (1): 173–192.

He, C., A. Silvennoinen, and T. Teräsvirta. 2008. "Parameterizing Unconditional Skewness in Models for Financial Time Series". *Journal of Financial Econometrics* 6 (2): 208–230.

Hedegaard, E., and R. J. Hodrick. 2016. "Estimating the risk-return trade-off with overlapping data inference". *Journal of Banking & Finance* 67: 135–145.

Horváth, L., P. Kokoszka, and R. Zitikis. 2006. "Sample and Implied Volatility in GARCH Models". *Journal of Financial Econometrics* 4 (4): 617–635.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. 2009. "Microstructure noise in the continuous case: The pre-averaging approach". *Stochastic Processes and their Applications* 119 (7): 2249–2276.

Karanasos, M. 1999. "The second moment and the autocovariance function of the squared errors of the GARCH model". *Journal of Econometrics* 90 (1): 63–76.

Kluitman, R., and P. H. Franses. 2002. "Estimating volatility on overlapping returns when returns are autocorrelated." *Applied Mathematical Finance* 9 (3): 179–188.

Kole, E., T. Markwat, A. Opschoor, and D. van Dijk. 2017. "Forecasting Value-at-Risk under Temporal and Portfolio Aggregation". *Journal of Financial Econometrics* 15 (4): 649–677.

Ling, S., and M. McAleer. 2002a. "Necessary and Sufficient Moment Conditions for the GARCH(r,s) and Asymmetric Power GARCH(r,s) Models". *Econometric Theory* 18 (3): 722–729.

— . 2002b. "Stationarity and the existence of moments of a family of GARCH processes". *Journal of Econometrics* 106 (1): 109–117.

Magnus, J. R. 1978. "The moments of products of quadratic forms in normal variables". *Statistica Neerlandica* 32 (4): 201–210.

McNeil, A. J., and R. Frey. 2000. "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach". Special issue on Risk Management, *Journal of Empirical Finance* 7 (3–4): 271–300.

Mittnik, S. 1988. "Derivation of the theoretical autocovariance and autocorrelation function of autoregessive moving average processes". *Communications in Statistics - Theory and Methods* 17 (11): 3825–3831.

— . 2011. "Solvency II Calibrations: Where Curiosity Meets Spuriosity". Center for Quantitative Risk Analysis (CEQURA), Working Paper Number 04, URL: `http://www.cequra.uni-muenchen.de/download/cequra_wp_041.pdf`.

Nelson, D. B., and C. Q. Cao. 1992. "Inequality Constraints in the Univariate GARCH Model". *Journal of Business & Economic Statistics* 10 (2): 229–235.

Riskmetrics, T. 1996. *JP Morgan Technical Document*.

Silvestrini, A., and D. Veredas. 2008. "Temporal aggregation of univariate and multivariate time series models: A survey". *Journal of Economic Surveys* 22 (3): 458–497.

Zhang, L. 2006. "Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach". *Bernoulli* 12 (6): 1019–1043.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia. 2005. "A Tale of Two Time Scales". *Journal of the American Statistical Association* 100 (472): 1394–1411.

Appendix

A Proofs

A.1 Proof of Corollary 1

Assume for the daily returns, $(r_t)_t$, a Gaussian white noise process (Example 1) with $\mathbb{E}(r_t) = 0$ and variances $\mathrm{Var}(r_t) = \sigma^2$. Then $\mathbb{E}(\boldsymbol{r}_{t,h\Delta+\ell}) = \boldsymbol{0}_{(h\Delta+\ell\times 1)}$ and $\mathbb{E}(\boldsymbol{r}_{t,h\Delta+\ell}\boldsymbol{r}'_{t,h\Delta+\ell}) = \sigma^2\boldsymbol{I}_{h\Delta+\ell}$, so that, due to the independence, the joint distribution of the vector $\boldsymbol{r}_{t,h\Delta+\ell}$ is a multivariate normal distribution with zero mean vector and variance-covariance matrix $\sigma^2\boldsymbol{I}_{h\Delta+\ell}$. Using Theorem 1 it follows immediately

$$\mathrm{Cov}(\sigma^2_{(h),t}, \sigma^2_{(h),t-\ell}) = 2\,\mathrm{tr}(\boldsymbol{KQK}'\sigma^2\boldsymbol{I}_{h\Delta+\ell}\boldsymbol{LQL}'\sigma^2 I_{h\Delta+\ell}) = 2\sigma^4\,\mathrm{tr}(\boldsymbol{KQK}'\boldsymbol{LQL}'). \quad \square$$

A.2 Proof of Theorem 2

$$\mathrm{Cov}(\boldsymbol{X}'\boldsymbol{AX}, \boldsymbol{X}'\boldsymbol{BX}) = \mathbb{E}(\boldsymbol{X}'\boldsymbol{AXX}'\boldsymbol{BX}) - \mathbb{E}(\boldsymbol{X}'\boldsymbol{AX})\mathbb{E}(\boldsymbol{X}'\boldsymbol{BX})$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} a_{ij}b_{kl}\mathbb{E}(x_i x_j x_k x_l) - \mathrm{tr}(\mathbb{E}(\boldsymbol{X}'\boldsymbol{AX}))\mathrm{tr}(\mathbb{E}(\boldsymbol{X}'\boldsymbol{BX}))$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ii}b_{jj}\mathbb{E}(x_i^2 x_j^2) + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(a_{ij}b_{ij} + a_{ij}b_{ji})\mathbb{E}(x_i^2 x_j^2)$$

$$\quad - \mathbb{E}(\mathrm{tr}(\boldsymbol{X}'\boldsymbol{AX}))\mathbb{E}(\mathrm{tr}(\boldsymbol{X}'\boldsymbol{BX}))$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ii}b_{jj} + \mathbb{1}_{\{i\neq j\}}2a_{ij}b_{ij})\mathbb{E}(x_i^2 x_j^2) - \mathbb{E}(\mathrm{tr}(\boldsymbol{X}'\boldsymbol{AX}))\mathbb{E}(\mathrm{tr}(\boldsymbol{X}'\boldsymbol{BX}))$$

$$= \mathbb{E}(\boldsymbol{X}^{2\odot'}\boldsymbol{C}\boldsymbol{X}^{2\odot}) - \mathbb{E}(\mathrm{tr}(\boldsymbol{AXX}'))\mathbb{E}(\mathrm{tr}(\boldsymbol{BXX}'))$$

$$= \mathrm{tr}(\boldsymbol{C}\mathbb{E}(\boldsymbol{X}^{2\odot}\boldsymbol{X}^{2\odot'})) - \mathrm{tr}(\boldsymbol{A}\mathbb{E}(\boldsymbol{XX}'))\mathrm{tr}(\boldsymbol{B}\mathbb{E}(\boldsymbol{XX}'))$$

$$= \mathrm{tr}(\boldsymbol{C}(\boldsymbol{\Sigma}_{\boldsymbol{X}^{2\odot}} + \boldsymbol{\mu}_{\boldsymbol{X}^{2\odot}}\boldsymbol{\mu}'_{\boldsymbol{X}^{2\odot}})) - \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{X}})\mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{X}})$$

$$= \mathrm{tr}(\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{X}^{2\odot}}) + \boldsymbol{\mu}'_{\boldsymbol{X}^{2\odot}}\boldsymbol{C}\boldsymbol{\mu}_{\boldsymbol{X}^{2\odot}} - \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{X}})\mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{X}}) \quad \square$$

A.3 Proof of Corollary 2

From Theorem 2 we get

$$\gamma_{\sigma^2_{(h),t}}(\ell) = \mathrm{tr}(\boldsymbol{C}(\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}} + \boldsymbol{\mu}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}\boldsymbol{\mu}'_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}})) - \mathrm{tr}(\boldsymbol{KQK}'\boldsymbol{\Sigma}_{\boldsymbol{r}_{t,h\Delta+\ell}})\mathrm{tr}(\boldsymbol{LQL}'\boldsymbol{\Sigma}_{\boldsymbol{r}_{t,h\Delta+\ell}}),$$

with $\boldsymbol{C} = \boldsymbol{a}\boldsymbol{b}' + 2(\boldsymbol{KQK}') \odot (\boldsymbol{LQL}') \odot (\boldsymbol{1}_{h\Delta+\ell}\boldsymbol{1}'_{h\Delta+\ell} - \boldsymbol{I}_{h\Delta+\ell})$, where $\boldsymbol{a} = \text{diag}(\boldsymbol{KQK}')$ and $\boldsymbol{b} = \text{diag}(\boldsymbol{LQL}')$. By assumption $(r_t)_{t\in\mathbb{Z}}$ follows a weak white noise process (Definition 1) with zero mean, which directly implies

$$\boldsymbol{\mu}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}} = \mathbb{E}(\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}) = \text{Var}(r_t)\boldsymbol{1}_{h\Delta+\ell} = \sigma^2\boldsymbol{1}_{h\Delta+\ell}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{r}_{t,h\Delta+\ell}} = \mathbb{E}(\boldsymbol{r}_{t,h\Delta+\ell}\boldsymbol{r}'_{t,h\Delta+\ell}) = \text{Var}(r_t)\boldsymbol{I}_{h\Delta+\ell} = \sigma^2\boldsymbol{I}_{h\Delta+\ell}.$$

Plugging in gives

$$\begin{aligned}
\gamma_{\sigma^2_{(h),t}}(\ell) &= \text{tr}(\boldsymbol{C}(\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}} + \sigma^4\boldsymbol{1}_{h\Delta+\ell}\boldsymbol{1}'_{h\Delta+\ell})) - \sigma^4\text{tr}(\boldsymbol{KQK}')\text{tr}(\boldsymbol{LQL}') \\
&= \text{tr}(\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}) + \sigma^4(\boldsymbol{1}'_{h\Delta+\ell}\boldsymbol{C}\boldsymbol{1}_{h\Delta+\ell} - \text{tr}(\boldsymbol{Q})^2).
\end{aligned}$$

Furthermore, it holds with $\boldsymbol{A} := \boldsymbol{KQK}'$ and $\boldsymbol{B} := \boldsymbol{LQL}'$

$$\begin{aligned}
\boldsymbol{1}'_{h\Delta+\ell}\boldsymbol{C}\boldsymbol{1}_{h\Delta+\ell} &= \boldsymbol{1}'_{h\Delta+\ell}(\boldsymbol{a}\boldsymbol{b}' + 2\boldsymbol{A}\odot\boldsymbol{B}\odot(\boldsymbol{1}_{h\Delta+\ell}\boldsymbol{1}'_{h\Delta+\ell} - \boldsymbol{I}_{h\Delta+\ell}))\boldsymbol{1}_{h\Delta+\ell} \\
&= \underbrace{\boldsymbol{1}'_{h\Delta+\ell}\boldsymbol{a}}_{=\text{tr}(\boldsymbol{A})}\underbrace{\boldsymbol{b}'\boldsymbol{1}_{h\Delta+\ell}}_{=\text{tr}(\boldsymbol{B})} + 2\boldsymbol{1}'_{h\Delta+\ell}(\boldsymbol{A}\odot\boldsymbol{B}\odot\boldsymbol{1}_{h\Delta+\ell}\boldsymbol{1}'_{h\Delta+\ell})\boldsymbol{1}_{h\Delta+\ell} \\
&\quad - 2\boldsymbol{1}'_{h\Delta+\ell}(\boldsymbol{A}\odot\boldsymbol{B}\odot\boldsymbol{I}_{h\Delta+\ell})\boldsymbol{1}_{h\Delta+\ell} \\
&= \text{tr}(\boldsymbol{Q})^2 + 2\boldsymbol{1}'_{h\Delta+\ell}(\boldsymbol{A}\odot\boldsymbol{B})\boldsymbol{1}_{h\Delta+\ell} - 2\boldsymbol{a}'\boldsymbol{b} \\
&= \text{tr}(\boldsymbol{Q})^2 + 2\text{tr}(\boldsymbol{AB}) - 2\boldsymbol{a}'\boldsymbol{b} \\
&= \text{tr}(\boldsymbol{Q})^2 + 2\text{tr}(\boldsymbol{KQK}'\boldsymbol{LQL}') - 2\boldsymbol{a}'\boldsymbol{b}.
\end{aligned}$$

Implying

$$\begin{aligned}
\gamma_{\sigma^2_{(h),t}}(\ell) &= \text{tr}(\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}) + \sigma^4(\text{tr}(\boldsymbol{Q})^2 + 2\text{tr}(\boldsymbol{KQK}'\boldsymbol{LQL}') - 2\boldsymbol{a}'\boldsymbol{b} - \text{tr}(\boldsymbol{Q})^2) \\
&= \text{tr}(\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}) + 2\sigma^4(\text{tr}(\boldsymbol{KQK}'\boldsymbol{LQL}') - \boldsymbol{a}'\boldsymbol{b}). \qquad \square
\end{aligned}$$

A.4 GARCH(p,q) Fulfills the Conditions of Corollary 2

Let $(r_t)_{t\in\mathbb{Z}}$ be a GARCH(p,q) as defined in Example 2. We further assume that the first four moments of r_t exist and are finite.[16]

It is well known that GARCH processes are weak white noise process (Definition 1), so

---

[16] Conditions for the existence of moments can be found in He and Teräsvirta (1999b) and Bollerslev (1986) for the GARCH(1,1) model and for the GARCH(p,q) model in Ling and McAleer (2002a). The functional form of the moments are given in He and Teräsvirta (1999a) and Karanasos (1999).

it remains to show that the moment conditions (8)-(11) of Theorem 2 are satisfied. The first moment condition (8) is obviously fulfilled for the zero-mean process $(r_t)_{t\in\mathbb{Z}}$.

Let $\mathcal{I}_\tau := \{r_s : s \le \tau\}$. W.l.o.g. assume $t_1 < t_2 < t_3 < t_4$, then it holds

$$\mathbb{E}(r_{t_1}r_{t_2}r_{t_3}r_{t_4}) = \mathbb{E}(\mathbb{E}(r_{t_1}r_{t_2}r_{t_3}r_{t_4}|\mathcal{I}_{t_4-1})) = \mathbb{E}(r_{t_1}r_{t_2}r_{t_3}\sigma_{t_4}\mathbb{E}(\epsilon_{t_4}|\mathcal{I}_{t_4-1})) = 0,$$

which shows that moment condition (9) holds for GARCH(p,q) processes with symmetric innovation distributions and existing and finite fourth moments. If $t_1 < t_2$, we get

$$\mathbb{E}(r_{t_1}^3 r_{t_2}) = \mathbb{E}(\mathbb{E}(r_{t_1}^3 r_{t_2}|\mathcal{I}_{t_2-1})) = \mathbb{E}(r_{t_1}^3\sigma_{t_2}\mathbb{E}(\epsilon_{t_2}|\mathcal{I}_{t_2-1})) = 0$$

and if $t_1 > t_2$ it follows

$$\mathbb{E}(r_{t_1}^3 r_{t_2}) = \mathbb{E}(\mathbb{E}(r_{t_1}^3 r_{t_2}|\mathcal{I}_{t_1-1})) = \mathbb{E}(r_{t_2}\sigma_{t_1}^3\mathbb{E}(\epsilon_{t_1}^3|\mathcal{I}_{t_1-1})) = 0,$$

which shows that moment condition (11) holds for GARCH(p,q) processes with symmetric innovation distributions and existing and finite fourth moments. Let $t_1 < \max\{t_2, t_3\}$ and w.l.o.g. $t_3 > t_2$ then it follows

$$\mathbb{E}(r_{t_1}^2 r_{t_2} r_{t_3}) = \mathbb{E}(\mathbb{E}(r_{t_1}^2 r_{t_2} r_{t_3}|\mathcal{I}_{t_3-1})) = \mathbb{E}(r_{t_1}^2 r_{t_2}\sigma_{t_3}\mathbb{E}(\epsilon_{t_3}|\mathcal{I}_{t_3-1})) = 0.$$

If $t_1 > \max\{t_2, t_3\}$, $\mathbb{E}(\epsilon_t^2) = \sigma^2$ and w.l.o.g. $t_1 = t$, $t_2 = t-1$ and $t_3 = t-2$ it holds

$$\mathbb{E}(r_{t_1}^2 r_{t_2} r_{t_3}) = \mathbb{E}(r_t^2 r_{t-1} r_{t-2}) = \mathbb{E}(\mathbb{E}(r_t^2 r_{t-1} r_{t-2}|\mathcal{I}_{t-1})) = \mathbb{E}(r_{t-1} r_{t-2}\mathbb{E}(\sigma_t^2\epsilon_t^2|\mathcal{I}_{t-1}))$$

$$= \mathbb{E}(r_{t-1} r_{t-2}\sigma_t^2\mathbb{E}(\epsilon_t^2)) = \sigma^2\mathbb{E}(r_{t-1} r_{t-2}(\alpha_0 + \sum_{i=1}^{q}\alpha_i r_{t-i}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2))$$

$$= \sigma^2\alpha_0\mathbb{E}(r_{t-1} r_{t-2}) + \sigma^2\sum_{j=1}^{p}\beta_j\mathbb{E}(r_{t-1} r_{t-2}\sigma_{t-j}^2) + \sigma^2\alpha_1\mathbb{E}(r_{t-1}^3 r_{t-2}) + \sigma^2\sum_{i=2}^{q}\alpha_i\mathbb{E}(r_{t-1} r_{t-2} r_{t-i}^2)$$

$$= 0,$$

which shows that moment condition (10) holds for GARCH(p,q) processes with symmetric innovation distributions and existing and finite fourth moments.

B Additional Results and Figures

B.1 Gaussian White Noise Process with the Sample Variance

Figure 13 and Figure 14 in this section of the appendix are analogously to Figure 4 and Figure 5 but with GARCH(1,1) (12) being replaced by the Gaussian white noise as data generating process and the EWMA variance (5) being being substituted by the sample variance (4).
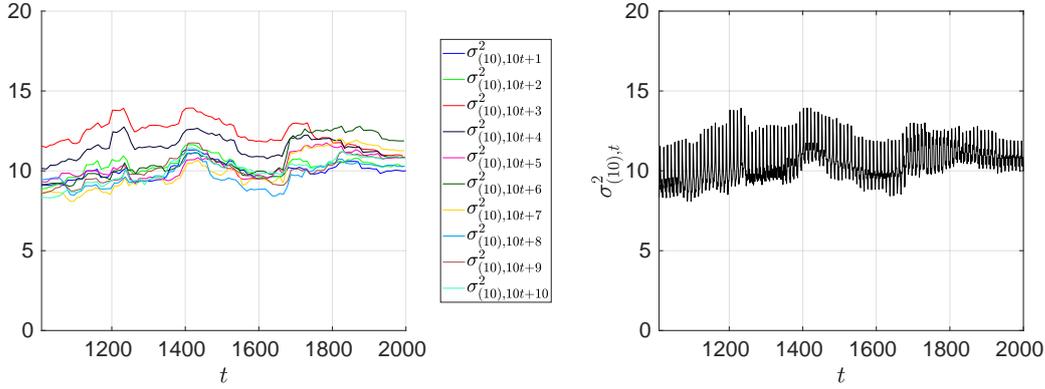

Figure 13: Time series of sample variance estimates (4), $\sigma^2_{(h),t}$, for simulated daily return series from the Gaussian white noise process with variance $\sigma^2 = 1$. The plot on the left shows the estimates $(\sigma^2_{(10),10t+\tau})_{t\in\mathbb{Z}}$, for $1 \leq \tau \leq 10$. The right plot shows the series $(\sigma^2_{(10),t})_{t\in\mathbb{Z}}$. Both plots are based on bi-weekly ($h = 10$) returns and estimation window $\Delta = 100$.
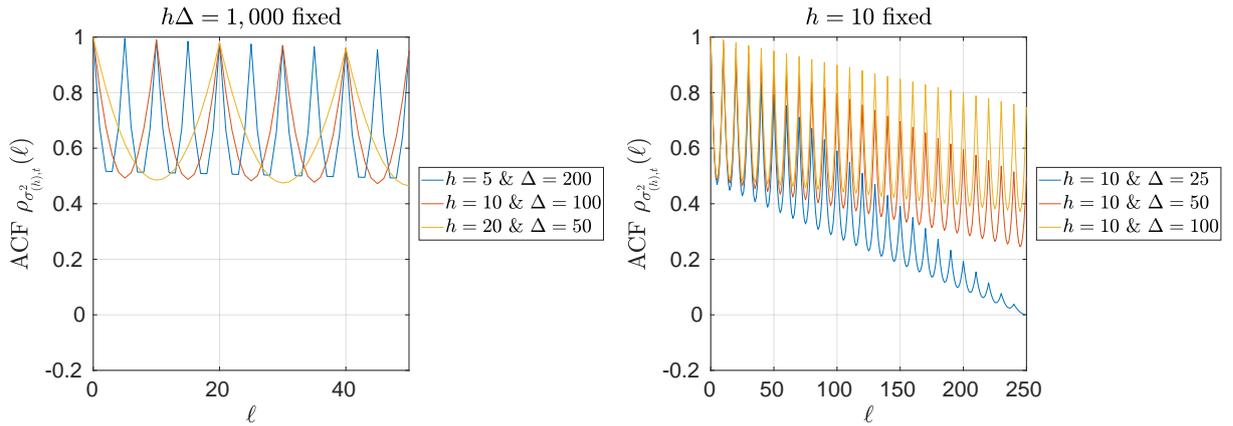

Figure 14: The ACF of sample variances (4), $\sigma^2_{(h),t}$, for daily returns from the Gaussian white noise process with $\sigma^2 = 1$. For the left plot we use a fixed number of daily returns to derive the sample variances. The right plot depicts the ACF of sample variances based on bi-weekly ($h = 10$) returns and estimation windows $\Delta = 25, 50, 100$.

B.2 The Functional Form and Amplitude of the Periodic Spurious Seasonality in the ACF

The autocovariance function of the quadratic-form variance estimator when the daily log-return process, $(r_t)_{t \in \mathbb{Z}}$, is a weak white noise process satisfying the moment conditions (8)-(11) is given in Corollary 2 for $\ell \geq 0$ by

$$\gamma_{\sigma^2_{(h),t}}(\ell) = \text{tr}\big(\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}\big) + 2\sigma^4\big(\text{tr}(\boldsymbol{KQK'LQL'}) - \boldsymbol{a'b}\big),$$

with $\boldsymbol{C} = \boldsymbol{ab'} + 2(\boldsymbol{KQK'}) \odot (\boldsymbol{LQL'}) \odot (\boldsymbol{1}_{h\Delta+\ell}\boldsymbol{1}'_{h\Delta+\ell} - \boldsymbol{I}_{h\Delta+\ell})$, where $\boldsymbol{a} = \text{diag}(\boldsymbol{KQK'})$ and $\boldsymbol{b} = \text{diag}(\boldsymbol{LQL'})$. The expression is nicely compact, but lacks intuition. It is not obvious, where the spurious seasonality is exactly coming from and how the amplitude of the periodic spurious seasonality in the ACF depends on the variance estimator and the data generating process. To provide more insight, we re-write the autocovariance as a sum of three components

$$\gamma_{\sigma^2_{(h),t}}(\ell) = s_1(\boldsymbol{Q}) + s_2(\boldsymbol{Q}) + s_3(\boldsymbol{Q}),$$

with

$$s_1(\boldsymbol{Q}) := \boldsymbol{b'}\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}\boldsymbol{a} - 2(\sigma^4 + \gamma_{r^2_t}(0))\boldsymbol{a'b},$$

$$s_2(\boldsymbol{Q}) := 2\text{tr}\big(((\boldsymbol{KQK'}) \odot (\boldsymbol{LQL'}))\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}\big),$$

$$s_3(\boldsymbol{Q}) := 2\sigma^4 \boldsymbol{1}'_{h\Delta+\ell}((\boldsymbol{KQK'}) \odot (\boldsymbol{LQL'}))\boldsymbol{1}_{h\Delta+\ell}.$$

The three terms depend on the variance estimator, defining $\boldsymbol{Q}$, and the data generating process, which impacts σ^4, $\gamma_{r^2_t}(0)$ and $\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}$. In the following we consider again the Gaussian white noise process with $\sigma^2 = 1$ and GARCH(1,1) process (12). As variance estimators, we study the sample variance and the EWMA variance based on non-overlapping h-day returns with $h = 10$. The window length is set to $\Delta = 100$.

Figure 15 shows the ACF of the estimated variances, $\rho_{\sigma^2_{(h),t}}(\ell)$, and the three components $s_i(\rho_{\sigma^2_{(h),t}}(\ell)) = s_i(\boldsymbol{Q}_{(h),\Delta})/\gamma_{\sigma^2_{(h),t}}(0)$. Note that the components add up to the ACF, i.e., $\rho_{\sigma^2_{(h),t}}(\ell) = \sum_{i=1}^3 s_i(\rho_{\sigma^2_{(h),t}}(\ell))$. The two top rows correspond to the Gaussian white noise process with the sample variance in the first row and the EWMA variance in the second row. Accordingly, the third and fourth row show the results for the GARCH(1,1) process and the respective variance estimators. One can see that the first component $s_1(\rho_{\sigma^2_{(h),t}}(\ell))$ is not contributing to the periodic spurious seasonality effect and the functional form depends on the variance estimator and the data generating process. The second component, $s_2(\rho_{\sigma^2_{(h),t}}(\ell))$, is not periodic for the Gaussian white noise process and periodic for the GARCH(1,1) pro-
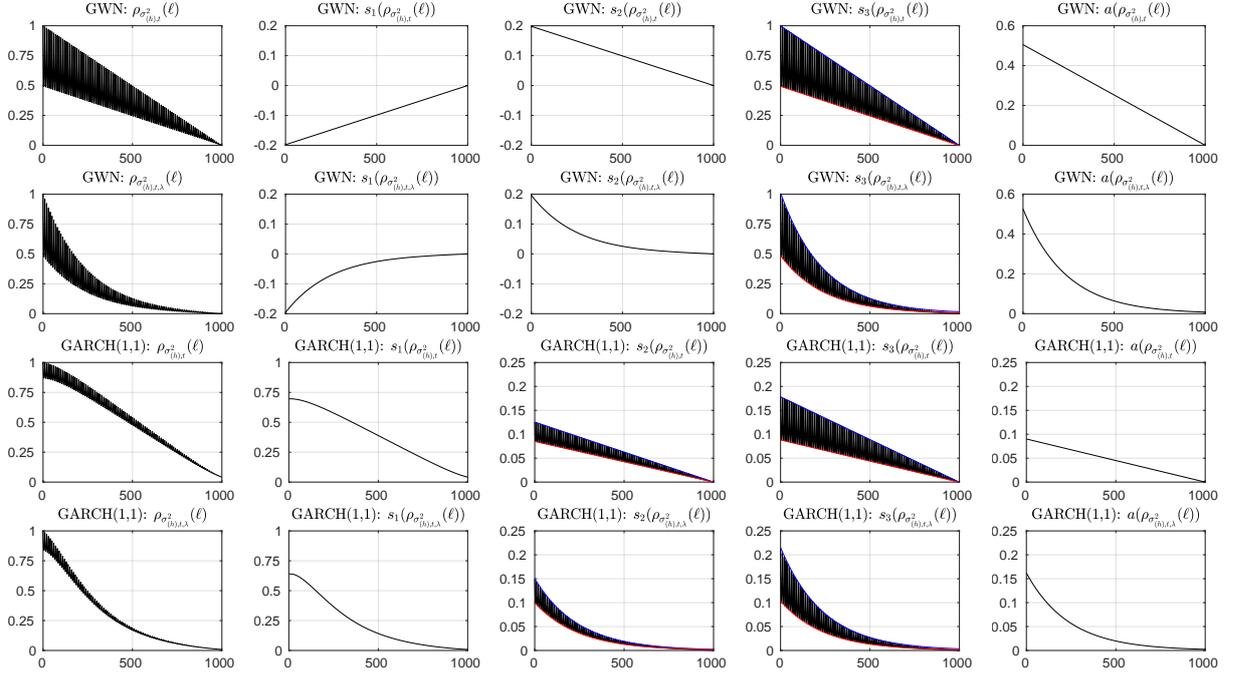
Figure 15: The ACF, its components $s_1(\cdot)$, $s_2(\cdot)$, $s_3(\cdot)$, and peak-to-peak amplitudes, $a(\cdot)$, of sample variances (first and third row), $\sigma^2_{(h),t}$, and EWMA variances (second and fourth row), $\sigma^2_{(h),t,\lambda}$, for different lags ℓ. The time-scale on the horizontal axis represents lags of ℓ days. In the first two rows, for the daily returns the Gaussian white noise process with $\sigma^2 = 1$ is assumed and in the third and fourth row GARCH(1,1) process (12). All plots are based on bi-weekly ($h = 10$) returns and estimation window $\Delta = 100$.

cess. This is due to the fact that $\boldsymbol{\Sigma}_{\boldsymbol{r}^{2\odot}_{t,h\Delta+\ell}}$ is diagonal for Gaussian white noise but not for GARCH(1,1) processes since squared observations are autocorrelated. In case of the sample variance, the peak-to-peak amplitude in $s_2(\rho_{\sigma^2_{(h),t}}(\ell))$ is decreasing more slowly than for the EWMA variance estimator. The third term, $s_3(\rho_{\sigma^2_{(h),t}}(\ell))$, is periodic in all four cases and the functional form of the amplitude is comparable to those of the second component. The fifth column shows the contribution of the second and third component to the peak-to-peak amplitudes, denoted by $a(\rho_{\sigma^2_{(h),t}}(\ell))$.[17] One can see that the peak-to-peak amplitude of the sample variance is decreasing in ℓ, and it is larger for the Gaussian white noise process than for the GARCH(1,1) process. The peak-to-peak amplitudes for the EWMA variances are comparable to those of the sample variance for small values of ℓ, but the amplitudes are decreasing much faster in ℓ.

The crucial term in both periodic components, $s_2(\boldsymbol{Q})$ and $s_3(\boldsymbol{Q})$, is $(\boldsymbol{KQK}') \odot (\boldsymbol{LQL}')$. The block-structure of $\boldsymbol{Q}_{(h),\Delta,\lambda}$ and $\boldsymbol{Q}_{(h),\Delta}$ (see, for example, the top-left plot in Figure 7)

---

[17] The peak-to-peak amplitudes have been approximated by fitting linear functions (for the sample variance) and exponential functions (for the EWMA variance) through the peaks and taking the pointwise differences between the fitted functions. The fitted curves are shown in blue and red in the plots of $s_2(\cdot)$ and $s_3(\cdot)$.

and the fact that $\boldsymbol{KQK'} = \mathrm{blkDiag}(\boldsymbol{0}_{(\ell \times \ell)}, \boldsymbol{Q})$ and $\boldsymbol{LQL'} = \mathrm{blkDiag}(\boldsymbol{Q}, \boldsymbol{0}_{(\ell \times \ell)})$ have a block-diagonal structure reveal how the periodicity of length h is generated, when different lags, ℓ, are considered and the Hadamard product of the matrices $\boldsymbol{KQK'}$ and $\boldsymbol{LQL'}$ is formed.

B.3   Comparison of EWMA Variance Estimators

In this section of the appendix we present plots for the four different EWMA variance estimators: $\sigma^2_{(h),t,\lambda}$ (non-overlapping h-day returns, (5)), $\breve{\sigma}^2_{(h),t,\lambda}$ (overlapping h-day returns, (14)), $\bar{\sigma}^2_{(h),t,\lambda}$ (two-scales, (16)), and $\tilde{\sigma}^2_{(h),t,\lambda}$ (corrected two-scales, (18)). In Figure 16, time series of variance estimates for simulated data from GARCH(1,1) process (12) are shown. The most left and right plots, for $\sigma^2_{(h),t,\lambda}$ and $\tilde{\sigma}^2_{(h),t,\lambda}$, have already been shown on the right
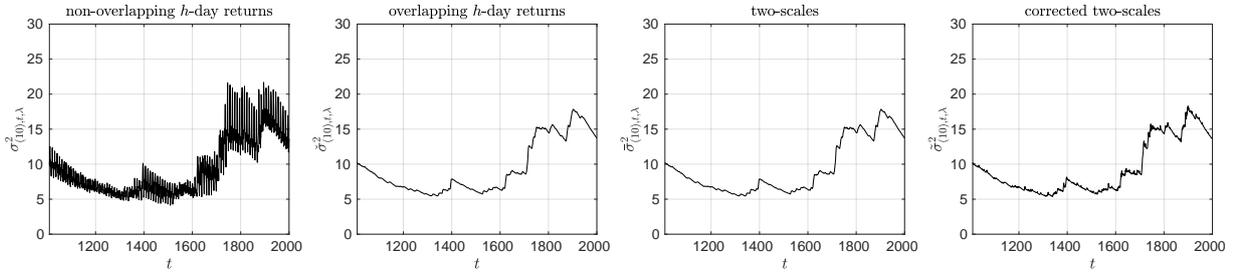

Figure 16: Time series of EWMA variance estimates (5), (14), (16), (18) for simulated daily return series from GARCH(1,1) process (12). The plots are based on bi-weekly ($h = 10$) returns and estimation window $\Delta = 100$.

of Figure 4 and Figure 8, respectively. In Figure 17, the ACF for all four EWMA variances is plotted. Again, the most left and right plot, for $\sigma^2_{(h),t,\lambda}$ and $\tilde{\sigma}^2_{(h),t,\lambda}$, have already been shown on the right of Figure 5 and Figure 9, respectively.
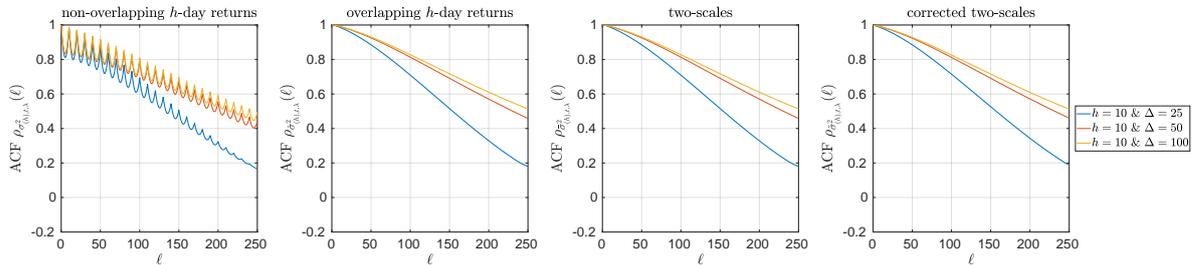

Figure 17: The ACF of EWMA variances (5), (14), (16), (18), for daily returns from GARCH(1,1) process (12). The ACF of EWMA variances is based on bi-weekly ($h = 10$) returns and estimation windows $\Delta = 25, 50, 100$.

Chapter VI

A note on low-dimensional Kalman smoothers for systems with lagged states in the measurement equation

This chapter is a reprint[*] of:

M. S. Kurz. 2018a. "A note on low-dimensional Kalman smoothers for systems with lagged states in the measurement equation". `https://doi.org/10.1016/j.econlet.2018.03.037`, *Economics Letters* 168: 42–45.

Copyright:

© 2018 Elsevier B.V. All rights reserved.

Accompanying software:

M. S. Kurz. 2018b. *State Space Models with Lagged State (SSMwLS) in the measurement equation.* URL: `https://github.com/MalteKurz/SSMwLS`, MATLAB-code package.

---

[*]All reprints in this cumulative dissertation have been slightly adapted in order to harmonize the overall appearance, like the format, page margins, caption setup, citation style and bibliography style.

A note on low-dimensional Kalman smoothers for systems with lagged states in the measurement equation

Malte S. Kurz[*][1]

[1]Department of Statistics,
Ludwig-Maximilians-Universität München,
Akademiestr. 1, 80799 Munich, Germany.

Abstract

In this paper we derive a modified Kalman smoother for state space systems with lagged states in the measurement equation. This modified Kalman smoother minimizes the mean squared error (MSE). Computationally efficient algorithms that can be used to implement it in practice are discussed. We also show that the conjecture in Nimark (2015) that the output of his modified Kalman filter for this type of systems can be plugged into the standard Kalman smoother is in general not correct. The competing smoothers are compared with regards to the MSE.

JEL Classification: C18, C22, C32.

Keywords: Kalman filter, Kalman smoother, Lagged states.

---

[*]malte.kurz@stat.uni-muenchen.de

In this note, we consider state space systems with a lagged state in the measurement equation for which Nimark (2015) derives a modified low-dimensional Kalman filter. Nimark (2015) also states, without a formal derivation, that the filtered state estimates from the modified filter can be plugged into the standard, i.e., unmodified, Kalman smoother of Hamilton (1994). In this paper we show that to use the filtered state estimates from the modified filter, we also need to modify the Kalman smoother to obtain the MSE-minimizing smoothed state estimates. That is, the claim that the filtered estimates from Nimark's (2015) modified filter can be plugged into the standard Kalman smoother is in general not correct. In what follows, we derive three modified Kalman smoothers that all can be used in combination with the modified Kalman filter of Nimark (2015). The first is based on the same principles as the one in Hamilton (1994). The second and third, and computationally more efficient, smoothers are a modified version of the smoother of de Jong (1988, 1989) and Kohn and Ansley (1989), and a modified version of the disturbance-smoother-based state smoother of Koopman (1993). Finally, the minimum variance estimator for the smoothed states will be compared to the Nimark (2015) smoother.

1 The state space model

In this note we stick as close as possible to the notation of Nimark (2015) and consider the state space model

$$X_t = AX_{t-1} + Cu_t, \qquad\qquad Z_t = D_1 X_t + D_2 X_{t-1} + Ru_t, \qquad\qquad (1.1)$$

where u_t is a m-dimensional vector of disturbances being multivariate normally distributed with zero mean and the identity as variance-covariance matrix. The observable at time t, Z_t, is a $p \times 1$ vector and the state vector X_t is of dimension $n \times 1$. Similar to Nimark (2015), we use for the conditional expectation and variance the notations

$$X_{t|t-s} = \mathbb{E}(X_t | Z_{1:t-s}, X_{0|0}), \qquad\qquad P_{t|t} = \mathbb{E}((X_t - X_{t|t})(X_t - X_{t|t})'),$$

with $Z_{1:t} = (Z_1', \ldots, Z_t')'$ and we initialize the system by $X_0 \sim N(X_{0|0}, P_{0|0})$.

2 The modified Kalman filter

The standard solution to apply the Kalman filter to the state space system (1.1) is obtained by augmenting the state vector with lagged states. A modified Kalman filter, which oper-

ates with an n-dimensional state vector, was derived by Nimark (2015). Nimark's (2015) modified Kalman filter can be summarized by the following recursion

$$\tilde{Z}_t = Z_t - \tilde{D}X_{t-1|t-1}, \qquad\qquad P_{t|t-1} = AP_{t-1|t-1}A' + CC', \qquad (2.1)$$

$$X_{t|t} = AX_{t-1|t-1} + K_t\tilde{Z}_t, \qquad\qquad P_{t|t} = P_{t|t-1} - K_tF_tK_t', \qquad (2.2)$$

with $\tilde{D} = (D_1A + D_2)$ and where the Kalman gain is given by $K_t = U_tF_t^{-1}$ with

$$U_t = \mathbb{E}(X_t\tilde{Z}_t') = AP_{t-1|t-1}\tilde{D}' + CC'D_1' + CR', \qquad (2.3)$$

$$F_t = \mathbb{E}(\tilde{Z}_t\tilde{Z}_t') = \tilde{D}P_{t-1|t-1}\tilde{D}' + (D_1C + R)(D_1C + R)'. \qquad (2.4)$$

3 On the Kalman smoother for systems with a lagged state in the measurement equation

To derive the updating equations which are purely based on filtered states and not on the observables, Hamilton (1994) uses the following approach.[1] By the formula for updating linear projections (Eq. [4.5.30] in Hamilton (1994)) one gets

$$\mathbb{E}(X_t|X_{t+1}, Z_{1:t}, X_{0|0}) = X_{t|t} + \hat{J}_t(X_{t+1} - X_{t+1|t}),$$

with $\hat{J}_t = P_{t|t}A'P_{t+1|t}^{-1}$. In a next step, Hamilton (1994) argues that $\mathbb{E}(X_t|X_{t+1}, Z_{1:t}, X_{0|0})$ is equal to $\mathbb{E}(X_t|X_{t+1}, Z_{1:T}, X_{0|0})$, as the error

$$X_t - \mathbb{E}(X_t|X_{t+1}, Z_{1:t}, X_{0|0})$$

is uncorrelated with Z_{t+j}, for $0 < j \leq T - t$. While this is true for a standard Kalman filter, as shown in Hamilton (1994), this is (in general) not the case for state space systems with a lagged state in the measurement equation, i.e., in general for state space systems of the form (1.1)

$$\text{Corr}(X_t - \mathbb{E}(X_t|X_{t+1}, Z_{1:t}, X_{0|0}), Z_{t+1}) \neq 0$$

and therefore

$$\mathbb{E}(X_t|X_{t+1}, Z_{1:t}, X_{0|0}) \neq \mathbb{E}(X_t|X_{t+1}, Z_{1:T}, X_{0|0}). \qquad (3.1)$$

---

[1] This state smoothing algorithm goes back to Anderson and Moore (1979) and Rauch et al. (1965).

As a consequence, the smoother stated in Eq. (4.2) in Nimark (2015)[2]

$$\hat{X}_{t|T} = X_{t|t} + \hat{J}_t(X_{t+1|T} - X_{t+1|t}), \qquad \hat{J}_t = P_{t|t}A'P_{t+1|t}^{-1}, \qquad (3.2)$$

is in general not equal to $\mathbb{E}(X_t|Z_{1:T}, X_{0|0})$ as claimed by Nimark (2015). Note that in general the smoothed estimate, $\hat{X}_{t|T}$, (Eq. (3.2)) is also not minimizing the MSE to X_t conditional on the complete history of the observables $Z_{1:T}$.

This can be easily verified, e.g., by considering the special case $A = 0_{n \times n}$. Then, by (3.2), we get

$$\hat{X}_{T-1|T} = X_{T-1|T-1} \quad \Rightarrow \quad \text{Var}(X_{T-1} - \hat{X}_{T-1|T}) = P_{T-1|T-1} \qquad (3.3)$$

and in contrast for

$$X_{T-1|T} = X_{T-1|T-1} + P_{T-1|T-1}D_2'F_T^{-1}\tilde{Z}_T \qquad (3.4)$$

we obtain

$$\text{Var}(X_{T-1} - X_{T-1|T}) = P_{T-1|T-1} - P_{T-1|T-1}D_2'F_T^{-1}D_2P_{T-1|T-1}. \qquad (3.5)$$

Both smoothers, (3.2) and (3.4), are obviously unbiased and as $P_{T-1|T-1}D_2'F_T^{-1}D_2P_{T-1|T-1}$ is positive semidefinite if F_T is positive semidefinite it follows with (3.3) and (3.5)

$$\begin{aligned} \text{MSE}(X_{T-1|T}) &= \text{tr}(P_{T-1|T-1}) - \text{tr}(P_{T-1|T-1}D_2'F_T^{-1}D_2P_{T-1|T-1}) \\ &\leq \text{tr}(P_{T-1|T-1}) = \text{MSE}(\hat{X}_{T-1|T}), \end{aligned}$$

i.e., the smoother, $\hat{X}_{T-1|T}$, is not the MSE-minimizing estimator of X_{T-1} given the complete history of the observables $Z_{1:T}$.

4   Kalman smoothing algorithms for the modified system

Similar to Hamilton (1994), the MSE-minimizing smoother for the modified system can be obtained using the updating equation for linear projections but with an adaption for systems with a lagged state in the measurement equation. Start by considering the conditional expectation $\mathbb{E}(X_t|X_{t+1}, Z_{1:t+1}, X_{0|0})$ and by applying the formula for updating a

---

[2] Note that there is a typo in Eq. (4.2) in Nimark (2015), where the index of $\hat{J}$ was $t-1$ instead of t, as in Hamilton (1994).

linear projection (Eq. [4.5.30] in Hamilton (1994))

$$\mathbb{E}(X_t|X_{t+1}, Z_{1:t+1}, X_{0|0}) = X_{t|t+1} + \mathbb{E}((X_t - X_{t|t+1})(X_{t+1} - X_{t+1|t+1})')$$
$$\cdot \mathbb{E}((X_{t+1} - X_{t+1|t+1})(X_{t+1} - X_{t+1|t+1})')^{-1}(X_{t+1} - X_{t+1|t+1})$$
$$= X_{t|t+1} + P'_{t+1,t|t+1}P_{t+1|t+1}^{-1}(X_{t+1} - X_{t+1|t+1}),$$

where $P_{t+1,t|t+1} = \mathbb{E}((X_{t+1} - X_{t+1|t+1})(X_t - X_{t|t+1})') = AP_{t|t} - K_{t+1}\tilde{D}P_{t|t}$. From the standard theory on state smoothing (see, e.g., Durbin and Koopman (2012)), we get the one-step ahead smoothed state as

$$X_{t|t+1} = X_{t|t} + P_{t|t}\tilde{D}'F_{t+1}^{-1}\tilde{Z}_{t+1}.$$

Future observables, Z_{t+j}, for $1 < j \leq T - t$, can be written as

$$Z_{t+j} = \tilde{D}X_{t+j-1} + (D_1C + R)u_{t+j} = \tilde{D}\Big(A^{j-2}X_{t+1} + \sum_{i=2}^{j-1} A^{j-1-i}Cu_{t+i}\Big) + (D_1C + R)u_{t+j},$$

where we use the notational convention that A^0 is the identity and A^n denotes the n-th power of the square matrix A. Therefore, using the same reasoning as in Hamilton (1994), we see that the prediction error

$$X_t - \mathbb{E}(X_t|X_{t+1}, Z_{1:t+1}, X_{0|0}) = X_t - X_{t|t+1} - P'_{t+1,t|t+1}P_{t+1|t+1}^{-1}(X_{t+1} - X_{t+1|t+1}) \quad (4.1)$$

is uncorrelated with Z_{t+j} for $1 < j \leq T - t$. This follows because the prediction error (4.1) is by construction uncorrelated with X_{t+1}, and by assumption uncorrelated with $u_{t+j}, u_{t+j-1}, \ldots, u_{t+2}$. As a consequence, we get

$$\mathbb{E}(X_t|X_{t+1}, Z_{1:T}, X_{0|0}) = \mathbb{E}(X_t|X_{t+1}, Z_{1:t+1}, X_{0|0}) \quad (4.2)$$

and by applying the law of iterated projections, as Hamilton (1994), we obtain the smoothed estimate, $\mathbb{E}(X_t|Z_{1:T}, X_{0|0})$, by projecting (4.2) on $Z_{1:T}$. The smoothed estimate is given by

$$X_{t|T} = \mathbb{E}(X_t|Z_{1:T}, X_{0|0}) = X_{t|t+1} + J_t(X_{t+1|T} - X_{t+1|t+1}), \quad (4.3)$$

with $J_t = P'_{t+1,t|t+1}P_{t+1|t+1}^{-1}$.

4.1 MSE of the smoothed state

Analogously to Hamilton (1994), by subtracting X_t from Eq. (4.3) and rearranging, we obtain

$$X_t - X_{t|T} + J_t X_{t+1|T} = X_t - X_{t|t+1} + J_t X_{t+1|t+1}. \tag{4.4}$$

Multiplying (4.4) with its transpose and applying the expectation implies

$$P_{t|T} + J_t \mathbb{E}(X_{t+1|T} X'_{t+1|T}) J'_t = P_{t|t+1} + J_t \mathbb{E}(X_{t+1|t+1} X'_{t+1|t+1}) J'_t, \tag{4.5}$$

where we used $\mathbb{E}((X_t - X_{t|T}) X'_{t+1|T}) = 0$ and $\mathbb{E}((X_t - X_{t|t+1}) X'_{t+1|t+1}) = 0$. Rearranging (4.5) results in the backward-recursion

$$P_{t|T} = P_{t|t+1} + J_t (P_{t+1|T} - P_{t+1|t+1}) J'_t, \tag{4.6}$$

where $P_{t|t+1} = P_{t|t} - P_{t|t} \tilde{D}' F_{t+1}^{-1} \tilde{D} P_{t|t}$.

4.2 Computationally more efficient smoothers for the modified system

We first reformulate the state space problem (1.1) in a comparable way as the so-called innovation analogue stated in Durbin and Koopman (2012). Using the updating equations of the modified filter (2.1)–(2.2), we get with $L_t = A - K_t \tilde{D}$ and $M_t = C - K_t(D_1 C + R)$

$$\tilde{Z}_t = Z_t - Z_{t|t-1} = \tilde{D} \tilde{X}_{t-1} + (D_1 C + R) u_t, \tag{4.7}$$

$$\tilde{X}_t = X_t - X_{t|t} = (A - K_t \tilde{D}) \tilde{X}_{t-1} + (C - K_t(D_1 C + R)) u_t = L_t \tilde{X}_{t-1} + M_t u_t. \tag{4.8}$$

As shown by Durbin and Koopman (2012), Z_T is fixed if Z_t and $\tilde{Z}_{t+1}, \ldots, \tilde{Z}_T$ are fixed. Note that the errors $\tilde{Z}_{t+1}, \ldots, \tilde{Z}_T$ are uncorrelated and $\mathbb{E}(\tilde{Z}_j | Z_{1:t}, X_{0|0}) = 0$ for $j = t+1, \ldots, T$. By the formula for updating a linear projection (Eq. [4.5.30] in Hamilton (1994)), it follows for the smoothed state

$$X_{t|T} = X_{t|t} + \sum_{j=t+1}^{T} \mathbb{E}(\tilde{X}_t \tilde{Z}'_j) \mathbb{E}(\tilde{Z}_j \tilde{Z}'_j)^{-1} \tilde{Z}_j, \tag{4.9}$$

where $E(\tilde{Z}_j \tilde{Z}'_j)^{-1} = F_j^{-1}$ is the second term of the Kalman gain (2.4). Using Eq. (4.7)–(4.8),

we get for $j = t+1, \ldots, T$

$$\mathbb{E}(\tilde{X}_t \tilde{Z}'_j) = \mathbb{E}(\tilde{X}_t \tilde{X}'_{j-1})\tilde{D}' + \mathbb{E}(\tilde{X}_t u'_j)(D_1 C + R)' = P_{t|t} L'_{t+1} \cdots L'_{j-1}\tilde{D}', \qquad (4.10)$$

where we apply the notational convention that $L'_{t+1} \cdots L'_{j-1}$ is the identity I_n for $j = t+1$ and L'_{t+1} for $j = t+2$. Inserting Eq. (4.10) into Eq. (4.9) results in the backward recursion

$$X_{t|T} = X_{t|t} + P_{t|t} r_t, \qquad\qquad r_t = \tilde{D}' F^{-1}_{t+1} \tilde{Z}_{t+1} + L'_{t+1} r_{t+1}, \qquad (4.11)$$

with initial conditions $r_T = 0_{n \times 1}$ and $L_T = 0_{n \times n}$. By the theory for updating linear projections (Eq. [4.5.31] in Hamilton (1994)), we obtain for the variance of the smoothed state vector

$$P_{t|T} = P_{t|t} - \sum_{j=t+1}^{T} \mathbb{E}(\tilde{X}_t \tilde{Z}'_j)\mathbb{E}(\tilde{Z}_j \tilde{Z}'_j)^{-1}\mathbb{E}(\tilde{Z}_j \tilde{X}'_t), \qquad (4.12)$$

and by inserting Eq. (4.10) into (4.12), we obtain the backward recursion

$$P_{t|T} = P_{t|t} - P_{t|t} N_t P_{t|t}, \qquad\qquad N_t = \tilde{D}' F^{-1}_{t+1} \tilde{D} + L'_{t+1} N_{t+1} L_{t+1}, \qquad (4.13)$$

with initial condition $N_T = 0_{n \times n}$. Note that the backward recursions for the smoothed state (4.11) and its variance (4.13) are very similar to the smoother proposed in de Jong (1988, 1989) and Kohn and Ansley (1989) but with a modification to be applicable in the context of Nimark's (2015) modified Kalman filter.

An even more efficient fast state smoothing recursion, similar to Koopman (1993), can be obtained by computing the smoothed disturbances via the backward recursion

$$u_{t|T} = (D_1 C + R)' F^{-1}_t \tilde{Z}_t + M'_t r_t, \qquad (4.14)$$

with the recursively defined r_t from (4.11). Then, like Koopman (1993), we obtain via a forward recursion the smoothed states as

$$X_{t|T} = A X_{t-1|T} + C u_{t|T}, \qquad (4.15)$$

with initial condition $X_{0|T} = X_{0|0} + P_{0|0} r_0$.

The gains in computational costs of the modified de Jong (1988, 1989) and Kohn and Ansley (1989) smoother (4.11) and the modified Koopman (1993) smoother (4.14)–(4.15)

are comparable to the classical case, i.e., without lagged state in the measurement equation.[3] Further, note that the three different recursions for the smoothed states (4.3), (4.11) and (4.14)–(4.15) are equivalent, i.e., return the same smoothed states $X_{t|T} = \mathbb{E}(X_t|Z_{1:T}, X_{0|0})$. The same applies to the two different recursions for obtaining the variance of the smoothers (4.6) and (4.13).

4.3 The MSE of the Nimark (2015) smoother

Nimark (2015) claims that by Hamilton (1994) the variance of the smoothed state, $\hat{X}_{t|T}$, is given by $\hat{P}_{t|T} = P_{t|t} + \hat{J}_t(\hat{P}_{t+1|T} - P_{t+1|t})\hat{J}_t'$. As in general $\hat{X}_{t|T} \neq \mathbb{E}(X_t|Z_{1:T}, X_{0|0})$ (see (3.1)–(3.2)), the formula of Hamilton (1994) cannot be directly applied to obtain the variance of $\hat{X}_{t|T}$. The smoother, $\hat{X}_{t|T}$ (Eq. (3.2)), can be rewritten as

$$\hat{X}_{t|T} = X_{t|t} + \sum_{j=t+1}^{T} \hat{J}_t \cdots \hat{J}_{j-1} K_j \tilde{Z}_j.$$

As the errors $\tilde{Z}_j$ are uncorrelated, the variance of $\hat{X}_{t|T}$ can be obtained as

$$P_{t|t} + \sum_{j=t+1}^{T} \left[\hat{J}_t \cdots \hat{J}_{j-1} K_j F_j K_j' \hat{J}_{j-1}' \cdots \hat{J}_t' - \hat{J}_t \cdots \hat{J}_{j-1} K_j \mathbb{E}(\tilde{Z}_j \tilde{X}_t') - \mathbb{E}(\tilde{X}_t \tilde{Z}_j') K_j' \hat{J}_{j-1}' \cdots \hat{J}_t' \right].$$

$$(4.16)$$

Inserting Eq. (4.10) into Eq. (4.16) results in the backward recursion

$$\mathbb{E}((X_t - \hat{X}_{t|T})(X_t - \hat{X}_{t|T})') = P_{t|t} + \hat{J}_t \hat{N}_t \hat{J}_t' - \hat{J}_t \hat{M}_t P_{t|t} - P_{t|t} \hat{M}_t' \hat{J}_t',$$
$$\hat{N}_t = K_{t+1} F_{t+1} K_{t+1}' + \hat{J}_{t+1} \hat{N}_{t+1} \hat{J}_{t+1}',$$
$$\hat{M}_t = K_{t+1} \tilde{D} + \hat{J}_{t+1} \hat{M}_{t+1} L_{t+1},$$

with initial conditions $\hat{N}_T = 0_{n \times n}$ and $\hat{M}_T = 0_{n \times n}$.

---

[3] A discussion of the comparative computational efficiency of the different smoother algorithms in the unmodified case can be found in Section 4.6.1–4.6.2 of Durbin and Koopman (2012) and in Koopman (1993).

5 Application: ARMA dynamics with measurement error

Data revisions are a typical phenomenon for economic time series. As a consequences researcher and decision maker have to rely on econometric models which are capable to allow or even explicitly model measurement errors. Jacobs and van Norden (2011) propose to use a state space model with a Kalman filter if the "true" signal can be described by a stochastic process, like an ARMA-process, and the signals can only be observed up to a measurement error.[4] In the following we will study ARMA(1,1)-processes, and as a special case a MA(1)-process, with measurement error. Like in Jacobs and van Norden (2011) let $\tilde{y}_t$ be the "true" unobserved value. We assume for $\tilde{y}_t$ an ARMA(1,1)-process

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t, \qquad\qquad \epsilon_t \sim N(0, \sigma_\epsilon^2),$$

and an additive measurement error $\delta_t \sim N(0, \sigma_\delta^2)$, so that the observed value is given by $y_t = \tilde{y}_t + \delta_t$. Using a specific state space representation of ARMA-processes given in Hamilton (1994, Eq. [13.1.22]–[13.1.23]) and by applying the theory of Nimark (2015), we get for y_t a state space representation with lagged state in the measurement equation of the form (1.1)

$$X_t = \phi_1 X_{t-1} + \sigma_\epsilon u_{1,t}, \qquad\qquad y_t = X_t + \theta_1 X_{t-1} + \sigma_\delta u_{2,t},$$

with a one-dimensional state variable, X_t, and a bivariate disturbance vector $u_t := (u_{1,t}, u_{2,t})'$ being independent standard normally distributed. As the parameters are not time-varying, we can numerically determine the steady state for the Kalman recursion and compute the MSE in the steady state for both smoothers.

W.l.o.g. let $\sigma_\epsilon^2 = 1$ be the variance of the disturbance of the ARMA(1,1)-signal, and the variance of the measurement error $\sigma_\delta^2 = \sigma_\epsilon^2 / q$ is implicitly given by the ratio q. For the AR(1)-parameter we consider the values $\phi_1 = -0.5$, 0, 0.9, for the MA(1)-parameter the range $\theta_1 \in [-0.99, 0.99]$ and for the ratio $q = \sigma_\epsilon^2 / \sigma_\delta^2$ the range $q \in [0.01, 3]$. In Figure 1, the left, middle and right plots correspond to the different AR(1)-parameters and the relative difference in the MSE of the two smoothers, $(\text{MSE}(\hat{X}_{t|T}) - \text{MSE}(X_{t|T}))/\text{MSE}(X_{t|T})$, is plotted against the MA(1)-parameter and the ratio $q = \sigma_\epsilon^2 / \sigma_\delta^2$. We clearly see that, if $\theta_1 = 0$ both smoothers have the same MSE. For all other considered scenarios the smoother of Nimark (2015), $\hat{X}_{t|T}$, has a larger MSE than the MSE-minimizing smoother, $X_{t|T}$. In

---

[4] In the context of data revisions, Jacobs and van Norden (2011) provide a detailed discussion of measurement errors and their components, like noise, news and spillover effects.
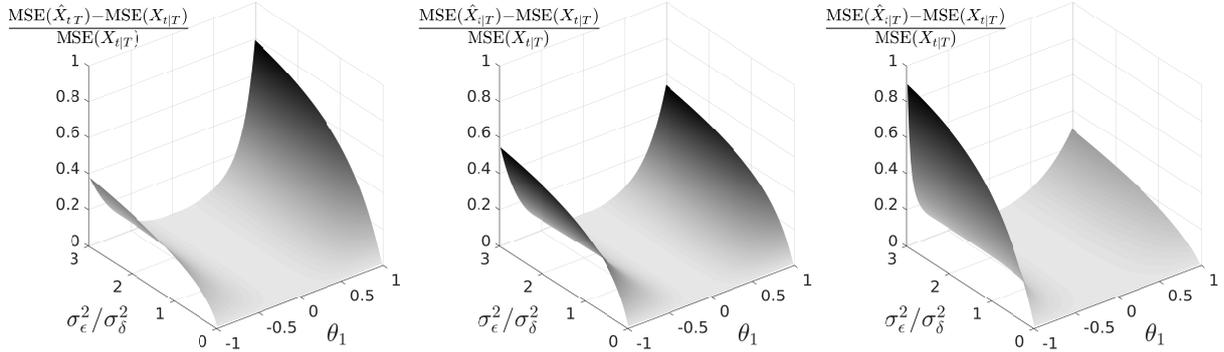
Figure 1: Relative increase in the MSE, $(\text{MSE}(\hat{X}_{t|T}) - \text{MSE}(X_{t|T}))/\text{MSE}(X_{t|T})$, as a function of the MA(1)-parameter, θ_1, and the ratio $q = \sigma_\epsilon^2/\sigma_\delta^2$. The left, middle and right plots show the result for the AR(1)-parameters $\phi_1 = -0.5, 0, 0.9$, respectively.

the most extreme shown case, $\phi_1 = 0.9$, $\theta_1 = -0.99$, $q = 3$, the MSE of the smoother $\hat{X}_{t|T}$ is 89.46% larger than for the MSE-minimizing smoother.

6 Conclusion

In this note we derive the MSE-minimizing smoother for the modified Kalman filter of Nimark (2015) for state space systems with a lagged state in the measurement equation. We demonstrate that the smoother of Nimark (2015) is not minimizing the MSE. Furthermore, we present computationally more efficient smoothing algorithms for the modified system. The MSE-minimizing smoother for the modified system can also be used in combination with the simulation smoother of Durbin and Koopman (2002) as suggested by Nimark (2015).

Accompanying `MATLAB` code is available at `https://github.com/MalteKurz/SSMwLS`.

Acknowledgments

The author thanks Stefan Mittnik for helpful discussions and he is particularly thankful to Kristoffer P. Nimark for valuable comments, suggestions and discussions which helped to greatly improve this paper.

References for Chapter VI

Anderson, B. D. O., and J. B. Moore. 1979. *Optimal Filtering.* Englewood Cliffs, N.J.: Prentice-Hall.

Durbin, J., and S. J. Koopman. 2002. "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis". *Biometrika* 89 (3): 603–615.

— . 2012. *Time Series Analysis by State Space Methods.* Oxford Statistical Science Series. OUP Oxford.

Hamilton, J. D. 1994. *Time Series Analysis.* Princeton, NJ: Princeton Univ. Press.

Jacobs, J. P. A. M., and S. van Norden. 2011. "Modeling data revisions: Measurement error and dynamics of "true" values". *Journal of Econometrics* 161 (2): 101–109.

de Jong, P. 1988. "A Cross-Validation Filter for Time Series Models". *Biometrika* 75 (3): 594–600.

— . 1989. "Smoothing and Interpolation with the State-Space Model". *Journal of the American Statistical Association* 84 (408): 1085–1088.

Kohn, R., and C. F. Ansley. 1989. "A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State Space Models". *Biometrika* 76 (1): 65–79.

Koopman, S. J. 1993. "Disturbance Smoother for State Space Models". *Biometrika* 80 (1): 117–126.

Nimark, K. P. 2015. "A low dimensional Kalman filter for systems with lagged states in the measurement equation". *Economics Letters* 127: 10–13.

Rauch, H. E., F Tung, and C. T. Striebel. 1965. "Maximum likelihood estimates of linear dynamic systems". *AIAA Journal* 3 (8): 1445–1450.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.2011, § 8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

_____

München, den 04.05.2018   Malte Simon Kurz