

Essays on Copula-Based and Non-Linear Modeling in Regression and Time Series Problems

Dissertation

by

Sven Pappert

born in Gelsenkirchen

in partial fulfillment of
the requirements for the degree of
Doktor der Naturwissenschaften (Dr. rer. nat.)

Submitted: Dortmund, April 2026
Primary referee: Prof. Dr. Antonia Arsova
Secondary referee: Prof. Dr. Matei Demetrescu
Tertiary referee: Prof. Dr. Florian Ziel

Sven Pappert

**Essays on Copula-Based and Non-Linear Modeling in Regression and Time
Series Problems**

Dissertation
in partial fulfillment of
the requirements for the degree of
Doktor der Naturwissenschaften (Dr. rer. nat.)



Submitted to the
Department of Statistics of the
TU Dortmund University

Primary referee: Prof. Dr. Antonia Arsova
Secondary referee: Prof. Dr. Matei Demetrescu
Tertiary referee: Prof. Dr. Florian Ziel

Abstract

Non-linear relations are present in many real world data-sets. Although linear models often perform well, it is worthwhile to explore non-linear statistical modeling. This cumulative dissertation comprises five scientific articles that explore non-linear modeling in the time series and regression setting. In the multivariate time series setting copulas are used to model non-linear joint conditional dependencies as well as the non-linear joint dependencies of multivariate consecutive observations. These modeling approaches are used to forecast energy time series. Furthermore, methodological innovations, that generalize the copula-based modeling of univariate stationary time series, are presented and investigated. In the regression context, the role of risks, which are used to train models, are analyzed. The functions that minimize risks are derived. The particularly interesting case, where risks are penalized, is the focus of these derivations.

Zusammenfassung

Nicht-lineare Beziehungen erscheinen in vielen realen Datensets. Auch wenn lineare Modelle meist gut abschneiden, lohnt es sich nicht-lineares statistisches Modellieren zu untersuchen. Diese kumulative Dissertation besteht aus fünf wissenschaftlichen Arbeiten, die nicht-lineare Modellierung in der Zeitreihenanalyse und in der Regression untersuchen. In dem Kontext der multivariaten Zeitreihen werden Copulas sowohl dafür verwendet nicht-lineare gemeinsame bedingte Verteilungen zu modellieren als auch für die Modellierung der gemeinsamen Verteilung aufeinanderfolgender multivariater Beobachtungen. Diese Modellierungsansätze werden verwendet um Energie Zeitreihen vorherzusagen. Weiterhin werden methodologische Beiträge, die das Copula-basierte Modellieren univariater stationärer Zeitreihen generalisieren, präsentiert und untersucht. In dem Kontext der Regression wird die Rolle von Loss-Funktionen, die benutzt werden um Modelle zu trainieren, untersucht. Es werden die Funktionen, die die jeweiligen Loss-Funktionen minimieren hergeleitet. Dabei liegt der Fokus auf dem besonders interessanten Fall, bei dem die Loss-Funktion penalisiert ist.

List of publications

This cumulative dissertation is based on five articles. In the following these articles are listed.

Article 1:

Pappert, Sven and Antonia Arsova (2023). “Forecasting Natural Gas Prices with Spatio-Temporal Copula-Based Time Series Models”. In: Valenzuela, O., Rojas, F., Herrera, L.J., Pomares, H., Rojas, I. (eds) Theory and Applications of Time Series Analysis. ITISE 2022. Contributions to Statistics. Springer. Cham., pp. 221–236.
DOI: https://doi.org/10.1007/978-3-031-40209-8_15.

Contribution: I implemented the statistical methods, conducted the statistical analysis and interpreted the results. In the conception of the project and preparing the final manuscript both authors contributed equally.

Article 2:

Berrisch, Jonathan, Sven Pappert, Florian Ziel, and Antonia Arsova (2023). “Modeling volatility and dependence of European carbon and energy prices”. In: Finance Research Letters 52, 103503.
DOI: <https://doi.org/10.1016/j.frl.2022.103503>

Contribution: Considering all parts of the project, from the data work and model development to interpreting the results and writing the manuscript, my contribution was proportional.

Article 3:

Pappert, Sven (2024). “Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas)”. arXiv preprint arXiv:2402.01491v3.
DOI: <https://doi.org/10.48550/arXiv.2402.01491>

Contribution: Single-author.

Article 4:

Pappert, Sven and Harry Joe (2026). “Copula-Based Time Series for Non-Gaussian and Non-Markovian Stationary Processes”. arXiv preprint arXiv:2604.01500v1.

DOI: <https://doi.org/10.48550/arXiv.2604.01500>

Contribution: I took the lead in the conception of the project, the implementation of statistical methods, performing the statistical analyses, interpreting the results, conducting the theoretical analyses, doing the data work and drafting the manuscript.

Article 5:

Pappert, Sven (2025). “The Field Equations of Penalized non-Parametric Regression”. arXiv preprint arXiv:2503.14763v2.

DOI: <https://doi.org/10.48550/arXiv.2503.14763>

Contribution: Single-author.

The version of the paper included here is a version that was updated after submission of the thesis. The updates were purely editorial.

Additional Publications:

Here I list additional publications that I have worked on while being a PhD student. These publications are not part of the cumulative dissertation.

- Lebedev, Andreas, Abhinav Das, Sven Pappert, and Stephan Schlüter (2026). "Analyzing Uncertainty Quantification in Statistical and Deep Learning Models for Probabilistic Electricity Price Forecasting". In: IEEE Access, vol. 14, pp. 52162-52189, 2026.
DOI: <https://doi.org/10.1109/ACCESS.2026.3672716>
- Schlüter, Stephan, Sven Pappert, and Martin Neumann (2024). “Interval Forecasts for Gas Prices in the Face of Structural Breaks–Statistical Models vs. Neural Networks”. arXiv preprint arXiv:2407.16723.
DOI <https://doi.org/10.48550/arXiv.2407.16723>

Abbreviations

ACF	autocorrelation function
AR	autoregressive
ARCH	autoregressive conditional heteroskedasticity
ARMA	autoregressive moving average
CDF	cumulative distribution function
CLT	central limit theorem
DGP	data-generating process
D-vine	drawable vine
GARCH	generalized autoregressive conditional heteroskedasticity
<i>iid</i>	independent and identically distributed
ML	maximum likelihood
MLE	maximum likelihood estimator/estimation
MSE	mean squared error
$N(0, 1)$	standard normal distribution
PACF	partial autocorrelation function
PDE	partial differential equation
PIT	probability integral transformation
Φ	cumulative distribution function of the standard normal distribution
R-vine	regular vine
t_ν	cumulative distribution function of the student's t -distribution
$U(0, 1)$	standard uniform distribution
VECM	vector error correction model

Contents

Abstract	V
List of publications	IX
Abbreviations	XI
1 Introduction	1
2 Statistical Methods	5
2.1 Copula-Based Statistical Modeling	5
2.1.1 General Ideas of Copula-Based Modeling	5
2.1.2 Notions Related to Copulas	6
2.1.3 Construction of Copulas and Examples	7
2.1.4 Statistical Modeling with Copulas	8
2.1.5 Vine-Copulas	10
2.2 Time Series Analysis	12
2.2.1 Time Series Models	12
2.2.2 (Probabilistic) Forecasting	13
2.2.3 Risks in Non-Linear Regression	14
2.3 Copulas and Time Series	15
2.3.1 Modeling of Multivariate Time Series Across Lags	15
2.3.2 Modeling of Conditional Joint Distributions	16
2.3.3 Modeling of Univariate Time Series	16
3 Summary of the Articles	19
3.1 Article 1: Forecasting Natural Gas Prices with Spatio-Temporal Copula- Based Time Series Models	19
3.2 Article 2: Modeling volatility and dependence of European carbon and energy prices	20
3.3 Article 3: Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas)	21
3.4 Article 4: Copula-Based Time Series for Non-Gaussian and Non-Markovian Stationary Processes	23

3.5 Article 5: The Field Equations of Penalized non-Parametric Regression . .	24
4 Discussion and Outlook	27
References	29
Contributed Articles	39

1 Introduction

This cumulative dissertation is based on five articles. These articles explore non-linear statistical modeling. With one exception, they explore how copulas can be used to enhance, generalize or complement time series modeling and forecasting. The papers lie in the intersection of applied work and methodological contribution with the first and second paper (Pappert et al., 2023 and Berrisch et al., 2023) leaning more to the application and the third and fourth papers (Pappert, 2024 and Pappert et al., 2026) more to the methodological side. The fifth paper (Pappert, 2025), which will be introduced at the end of this section, is a theoretical paper in the distinct but related statistical regression context. Now the interplay between copulas and time series modeling is introduced.

Among the best known models for modeling the stationary stochastic component of a time series is the autoregressive (AR) and the autoregressive moving average model (ARMA), cf. Chapter 3 by Hamilton, 2020. The current observation is assumed to be fluctuating around a linear function of the past observations. While this model alone is powerful, several generalizations are conceivable. The AR model assumes a linear relation between the current and past observations. For some time series, the true data-generating process may be of a more general form, e.g. a non-linear AR(1)-process $\{Y_t\}_{t \in \mathbb{Z}}$ following the updating equation

$$Y_t = m(Y_{t-1}) + \varepsilon_t. \quad (1.1)$$

$\{Y_t\}$ is a real valued time series, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ are *iid* innovations and $m : \mathbb{R} \rightarrow \mathbb{R}$ an unknown mapping. While this expression already generalizes the linear AR(1) model, further generalizations are possible. The innovations may not enter in an additive way. Then it is not only the conditional mean that is modeled; other aspects of a conditional distribution can be modeled too. For e.g. ARCH processes the innovations are multiplicative giving rise to conditional variance modeling, cf. Engle, 1982. The corresponding model updating equation then writes as $Y_t = \tilde{m}(Y_{t-1}, \varepsilon_t)$, compare Chapter 4 by Douc et al., 2014. These authors call this type of equation stochastic recurrence equation and provide tools to establish desirable properties of $\{Y_t\}_{t \in \mathbb{Z}}$. The question for these non-linear generalizations is how to properly model the mapping m or $\tilde{m}$. A bivariate copula-based approach can solve this question for the non-linear AR(1) model with possibly non-additive innovations.

In this approach the joint distribution of consecutive observations is modeled by a copula and a model for the stationary distribution. Darsow et al., 1992 introduced this approach for Markov processes, establishing that Markov processes are characterized by their copula. Chen et al., 2006 use a copula and non-parametric estimation of the stationary distribution to capture non-linear dynamics of Markov(1) processes and establish asymptotic results for the respective estimators. In this copula-based approach the mapping $\tilde{m}$ is parameterized via a bivariate copula and the stationary distribution, compare Sect. 2.3.3.

This approach can be extended to a non-linear AR(p) models, i.e. $\{Y_t\}_{t \in \mathbb{Z}}$ following an updating equation

$$Y_t = m(Y_{t-1}, \dots, Y_{t-p}) + \varepsilon_t, \quad (1.2)$$

where the additional generalization to include the innovation ε_t in a mapping $\tilde{m}$ can also be made. This approach is discussed in e.g. Smith et al., 2010 and Brechmann et al., 2015. In order to assure stationarity, suitable conditions have to be enforced on the copula of $(Y_t, \dots, Y_{t-p})$ and D-vines are particularly useful in this regard, see Sect. 2.1.5 for D-vine copulas and Sect. 2.3.3 for enforcing stationarity. The copula approach to univariate time series modeling has several desirable properties. The decoupling of the marginal distribution and the temporal dependence allows for choosing any marginal distribution. This allows for modeling e.g. strictly positive time series with Gaussian temporal dependence. Additionally, any dependence structure that can be represented by a copula can be assigned as temporal dependence. Particularly relevant are heavy-tailed dependencies for volatile time series, compare Loaiza-Maya et al., 2018, Spanhel, 2015 and Pappert et al., 2023, and asymmetric tail dependencies, compare the applications in McNeil et al., 2022, Pappert, 2024 and Bladt et al., 2025. The assumptions of this modeling approach are that the to-be-modeled time series is (strictly) stationary and Markov(p). The loosening of the Markov(p) assumption is the context of the methodological contributions of the third and fourth paper of this cumulative dissertation. Notice that in the classical AR(p)-framework, the Markov(p) assumption can be loosened by the introduction of past innovations into the model updating equation, i.e. moving average (MA) terms. The resulting stationary and causal autoregressive moving average (ARMA) model has an AR(∞) representation with decaying weights. This means that all past observations have an actual (conditional) influence on the current observation. The same holds for the conditional variance in GARCH models, which is influenced by all past observations, cf. Bollerslev, 1986. This implies that a non-Markovian generalization of copula-based time series may also be viewed as a non-linear distributional generalization of the ARMA model. In the third and fourth paper of this work, this approach is translated in different ways to copula-based modeling by combining suitable MA and AR-like terms to construct non-linear and distributional copula-based generalizations of the ARMA model. The

moving average generalization is called moving aggregate ('MAG'). In the third paper the moving aggregate modified autoregressive copula-based time series model is introduced. In the fourth paper the copula ARMA extension by Joe, 2014 is discussed in detail. In Sect. 3, the contributions of these papers are discussed. In Sect. 2.3.3 the copula-based time series model is formally introduced. Now generalizations in different directions are discussed.

The ARMA model and its (non-linear) generalizations are all univariate time series models. A natural extension is to consider multivariate time series. The natural generalization of the $AR(p)$ model is the vector autoregressive model, see Kilian et al., 2017; Sims, 1980. In this model the (multivariate) conditional expectation is modeled as linearly dependent on past lags of all time series. Furthermore, the contemporaneous covariance matrix is modeled unconditionally. Several non-linear generalizations to this model are conceivable and necessary if the true data-generating process is not linear. In the multivariate context, the non-linearity can enter on several occasions. The contemporaneous dependence, captured by the unconditional covariance matrix, may be modeled conditionally. In the one-dimensional case conditional heteroskedasticity models, i.e. ARCH and GARCH, see Engle, 1982 and Bollerslev, 1986, are able to do this. In the multivariate case, constant and dynamical conditional covariance modeling, see Bollerslev, 1990 and Engle et al., 2001, can be employed. The copula approach goes beyond the modeling of the conditional covariance matrix and aims to model the conditional joint contemporaneous distribution of a multivariate time series. Notably the modeling of the conditional joint dependence structure is decoupled from the modeling of the conditional marginal distributions of the individual time series. Typically, an ARMA-GARCH model is employed for the individual modeling of each time series. The joint conditional dependence is then captured by a copula, compare Jondeau et al., 2006, Hu, 2006 and Aloui et al., 2014. This approach is often called "copula-GARCH", since the GARCH is a common building block. Note that the copula parameters can be modeled as constant as well as conditionally varying. This modeling approach is the context of the second paper of this dissertation. In the paper, European carbon allowances (EUA) are modeled and forecasted jointly with European energy futures. The individual conditional expectations and variances of each time series are modeled in a sophisticated way and the non-linear conditional dependence structure is modeled by a t -copula with conditionally varying correlation matrix. In the paper we perform an exhaustive forecasting study, experiment with transformations and interpret the resulting (conditionally) time-varying parameters. This copula-based modeling approach is revealed to be a powerful tool for the time series under investigation. See Sect. 3.2 for details.

Now, going beyond generalizing the conditional dependence modeling, a copula can be employed to capture not only the cross-sectional dependence but also the temporal and

cross-temporal dependencies. This can be done by finding the copula of consecutive observations of the multivariate time series. Various specifications are available, see e.g. Simard et al., 2015 for the incorporation of the t -copula, Beare et al., 2015, Brechmann et al., 2015 and Nagler et al., 2022a for vine-copula specifications. Forecasting can then be performed by extracting the conditional distribution. This approach may be regarded as a fully copula-based non-linear generalization of the VAR(p) time series. The modeling approach is further explained in Sect. 2.3.1. In the first paper of this thesis, we employ this modeling approach in order to forecast natural gas prices and related commodities. Furthermore we explore and exemplify how volatile dynamics are captured by this approach, see Sect. 3.1.

The presented papers successfully employ copula-based non-linear generalizations of time series modeling. Furthermore, methodological contributions are made in the modeling of univariate time series. The fifth paper of this dissertation also analyzes non-linear relations, but in a different context. The relations under investigation are in the regression setting. Assuming a non-linear relation between a target variable Y and covariates $\mathbf{X}$,

$$Y = m(\mathbf{X}) + \varepsilon, \quad (1.3)$$

the goal of non-parametric regression is to estimate the mapping g . Notice the similarity between Eqs. 1.3 and 1.2. The difference only lies in the structure of the covariates. While there are no assumptions on the distribution of $\mathbf{X}$ in this setting, the covariates in Eq. 1.2 are from a Markov(p) time series and hence have a corresponding dependence structure. In order to estimate the mapping m in Eq. 1.3, the distance between the estimated predictions and the actual observations needs to be minimized. In the fifth paper we explore how the choice of the distance, which is to be minimized, influences the function that an estimator is approximating. In Sect. 3.5 more details are provided.

In the next chapter, Chapter 2, we introduce the theoretical foundations, that are shared by the contributed articles. In Chapter 3, the contributed articles are summarized and in Chapter 4, the contributions as well as potential future research directions are discussed.

2 Statistical Methods

In this section, the statistical methods that are used in the papers are introduced. First, copulas and copula-based modeling is explained. Then time series analysis and forecasting is introduced. Subsequently, the usage of copulas in time series analysis is discussed. Regression modeling and the role of risks are briefly explained after introducing evaluation measures for probabilistic forecasting.

2.1 Copula-Based Statistical Modeling

In this section basics of copula modeling are introduced briefly. First, general ideas are presented, then copula construction methods and examples are given, proceeding with exemplary statistical modeling and vine copulas. For comprehensive introductions see Joe, 2014 and Nelsen, 2006.

2.1.1 General Ideas of Copula-Based Modeling

The key to copula-based modeling is the Theorem by Sklar (see Theorem 1.1 in Joe, 2014). The theorem states that for any distribution of random variables, say $(X_1, \dots, X_d)$, there is a distribution function with uniform margins $C : (0, 1)^d \rightarrow (0, 1)$ such that the CDF of $(X_1, \dots, X_d)$, $F_{X_1, \dots, X_d}$, can be decomposed as

$$F_{X_1, \dots, X_d}(x_1, \dots, x_d) = C[F_{X_1}(x_1), \dots, F_{X_d}(x_d)], \quad (2.1)$$

for all $(x_1, \dots, x_d) \in (0, 1)^d$. If the random variables are continuous, C is unique. In the following we assume the continuity of all random variables. $F_{X_1, \dots, X_d}$ is the joint distribution of the random vector $(X_1, \dots, X_d)$. F_{X_j} is the marginal distribution of the random variable X_j with $j \in \{1, \dots, d\}$, and C is the *copula* of the random vector $(X_1, \dots, X_d)$. Eq. 2.1 gives rise to interpretations of C . First, it can be interpreted as the mapping such that Eq. 2.1 is true. As such it acts as the coupling between the joint distribution and the marginal distributions. Second, from C fulfilling Eq. 2.1, it follows that C is the distribution function of the random vector $(F_{X_1}(X_1), \dots, F_{X_d}(X_d))$. Hence

C is the joint distribution of the probability integral-transformed (PIT) random variables. The PIT'ed random variables are uniformly distributed – any structure in the marginals is eliminated by the PIT. This gives rise to the interpretation of the copula as the dependence structure of the random vector $(X_1, \dots, X_d)$. In the following, related notions such as the copula density, conditional CDFs and conditional quantile functions are defined. Then, construction methods of copulas will be discussed and examples of copulas are given, followed by an introduction to statistical modeling with copulas.

2.1.2 Notions Related to Copulas

Assuming that a copula C is differentiable, the *copula density* can be derived as any density associated to a distribution function – namely by taking the derivative w.r.t. all arguments,

$$c(u_1, \dots, u_d) = \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d}. \quad (2.2)$$

The copula density is important for maximum likelihood estimation (MLE). Given observations $(\mathbf{X}_i)_{i=1}^n$, the estimated log-likelihood of marginal parameters θ and copula parameters φ can be derived as

$$\log L(\theta, \varphi) = \sum_{i=1}^n \log c(F_1(X_{1,i}; \theta_1), \dots, F_d(X_{d,i}; \theta_d); \varphi) + \sum_{i=1}^n \sum_{j=1}^d \log f_j(X_{j,i}; \theta_j),$$

where θ_j denotes the candidate parameters of the j -th marginal distribution F_j , with density f_j , which is assumed to exist. Joe et al., 1996 realize that minimizing the negative log-likelihood in two steps is computationally efficient and it yields a good approximation to the minimizing parameters. The first step is to minimize the marginal NLLs, $\sum_{i=1}^n [-\log f_j(X_{j,i}; \theta_j)]$, individually, yielding an estimate $\hat{\theta}_j$ for each marginal distribution. Then, in the second step, these estimates are plugged in the copula part of the NLL, yielding $NLL(\varphi) = \sum_{i=1}^n [-\log c(F_1(X_{1,i}; \hat{\theta}_1), \dots, F_d(X_{d,i}; \hat{\theta}_d); \varphi)]$. The minimization of this NLL is less complicated. This approach is followed in the papers of this cumulative dissertation, if not stated otherwise. In Sect. 2.1.4 this modeling approach is exemplified on simulated data.

The conditional distribution associated to a d -variate copula, which is written with its subscripts as $C_{d,\dots,1}$ in this part, is given as

$$C_{d|(d-1),\dots,1}(u_d|u_{d-1}, \dots, u_1) = \frac{\partial^{d-1} C(u_d, \dots, u_1)}{\partial u_{d-1} \dots \partial u_1}.$$

The corresponding (generalized) inverse, $C_{d|(d-1),\dots,1}^{-1}(u_d|u_{d-1},\dots,u_1)$, is the conditional quantile function associated to $C_{d,\dots,1}$. Crucially the following holds: Let $(U_1, U_2) \sim C_{12}$, then $U_1 \perp C_{2|1}(U_2|U_1)$. On the other hand, if $V_1, V_2 \stackrel{iid}{\sim} U(0,1)$ it holds that $(C_{2|1}^{-1}(V_2|V_1), V_1) \sim C_{21}$. These properties can be regarded as the multivariate extension of the PIT and the quantile transformation. Such transformations are called Rosenblatt and inverse Rosenblatt transformations and in general they are used to simulate from copulas, compare section 5 of Czado et al., 2022. The conditional quantile function is used extensively in the construction of copula-based time series models in Paper 3 and Paper 4. The conditional distribution function is used to estimate the residual time series in the MLE of the copula-based time series models.

2.1.3 Construction of Copulas and Examples

There are several ways to construct copulas. An important class of such constructions are *implicit copulas*, cf. Smith, 2021. These are copula, which are implied by a joint distribution, and can be constructed in a straight-forward way as (cf. Eq. 5 by Smith, 2021)

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)),$$

where F is a d -variate joint distribution and F_j^{-1} , $j \in \{1, \dots, d\}$ is the quantile function corresponding to the j -th marginal distribution implied by F . It can easily be seen that such construction solves $F(u_1, \dots, u_d) = C(F_1(u_1), \dots, F_d(u_d))$. This method allows to recover dependence structures implied by joint distribution models. Consider e.g. the multivariate normal distribution with standard normal marginals and correlation matrix Σ . The corresponding implicit copula is given as $C(u_1, \dots, u_d) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$ for all $(u_1, \dots, u_d) \in (0, 1)^d$. This construction is the *Gaussian* or *normal* copula and it allows for assigning a Gaussian dependence to random variables with any marginal distribution. A scatter plot of a sample from a bivariate Gaussian copula is visualized in the upper left panel of Fig. 1. Similar constructions can be derived from e.g. the multivariate t -distribution, resulting in the t -copula (see Demarta et al., 2005 for a detailed introduction to the t -copula). The copula implied by a multivariate t -distribution with correlation matrix Σ , degrees of freedom ν and marginal distributions t_{ν} is given as $C(u_1, \dots, u_d) = t_{\nu, \Sigma}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_d))$ for all $(u_1, \dots, u_d) \in (0, 1)^d$. The dependence encapsulated by this copula is a heavy-tailed dependence. Heavy-tailed dependence structures promote the joint occurrence of extreme events. The scatter plot of bivariate samples from the t -copula are visualized in the lower right panel of Fig. 1. This feature is what makes the t -copula especially relevant for (multivariate) volatile time series, where the joint occurrence of extreme events is the defining characteristic (compare also Sect. 3.1

and Loaiza-Maya et al., 2018). Constructing copulas from known joint distributions and letting the copula inherit the dependence structures is not the only way to construct copulas. Archimedean copulas, which are named after their relation to the Archimedean axiom (see Theorem 4.3.1 by Nelsen, 2006), can be constructed using a suitable *generator function*, $\psi : (0, 1) \rightarrow (0, \infty)$, which is continuous, strictly decreasing and it holds that $\psi(1) = 0$, cf. Def. 4.1.1. by Nelsen, 2006. The copula is recovered from the generator function as $C(u_1, \dots, u_d) = \psi^{[-1]}(\sum_{i=1}^d \psi(u_i))$, where $\psi^{[-1]} = \psi^{-1}(\cdot) \mathbb{1}\{0 \leq \cdot \leq \psi(0)\}$, cf. Def. 4.1.1. from Nelsen, 2006. This method can be used to construct the *Gumbel* and the *Clayton* Copula. The bivariate versions of these copulas are given as

$$\begin{aligned} \text{Gumbel: } C(x_1, x_2) &= e^{-[(-\log(x_1))^\theta + (-\log(x_2))^\theta]^{\frac{1}{\theta}}}, \\ \text{Clayton: } C(x_1, x_2) &= \left[\max\{x_1^{-\theta} + x_2^{-\theta} - 1, 0\} \right]^{-\frac{1}{\theta}}. \end{aligned}$$

Compare Table 4.1 in Nelsen, 2006 for these and other Archimedean copulas and their generator function. Scatter plots of bivariate samples from the Gumbel and Clayton copula are visualized in the lower left and upper right panel of Fig. 1. These two copulas exhibit asymmetric tail dependence, compare Sect. 2.13 and 2.14 by Joe, 2014. This means that e.g. for the Clayton copula it holds that $\lambda_l := \lim_{u \rightarrow 0+} C(u, u)/u > 0$ while $\lambda_u := \lim_{u \rightarrow 0+} [1 - 2(1 - u) + C(1 - u, 1 - u)]/u = 0$. For the Gumbel copula there is upper tail dependence but no lower tail dependence, compare Chapter 5.4 by Nelsen, 2006.

2.1.4 Statistical Modeling with Copulas

Copulas and the Theorem by Sklar enable sophisticated and flexible modeling of joint distributions. The copula approach to modeling joint distributions has two important features. The first feature is that the dependence structure, captured by the copula C , and the marginal distributions F_{X_j} can be modeled separately. This is in contrast to e.g. assuming a joint normal distribution, where implicitly every marginal distribution is normal. This enables joint modeling of random variables with distinct marginal distributions. Second, non-linear relations can be modeled by choosing suitable copulas. Consider as an example $X_1 \sim N(\mu, \sigma^2)$ and $X_2 \sim \text{Exp}(\lambda)$ with unknown joint distribution of (X_1, X_2) . While classical joint distribution approaches (e.g. the multivariate normal distribution) struggle to model such random vectors, copula-based modeling naturally solves this problem. Given a sample $(x_{1,j}, x_{2,j})_{j=1}^n$ of size $n \in \mathbb{N}$, following the approach by Joe et al., 1996, the modeling is performed by the following steps. First, the respective marginal distributions, say $\hat{F}_1$ and $\hat{F}_2$, must be estimated. This can be done parametrically or non-parametrically. Then these estimated marginal distributions are used to probability-integral-transform (PIT) the observations. The resulting observations are called *pseudo-*

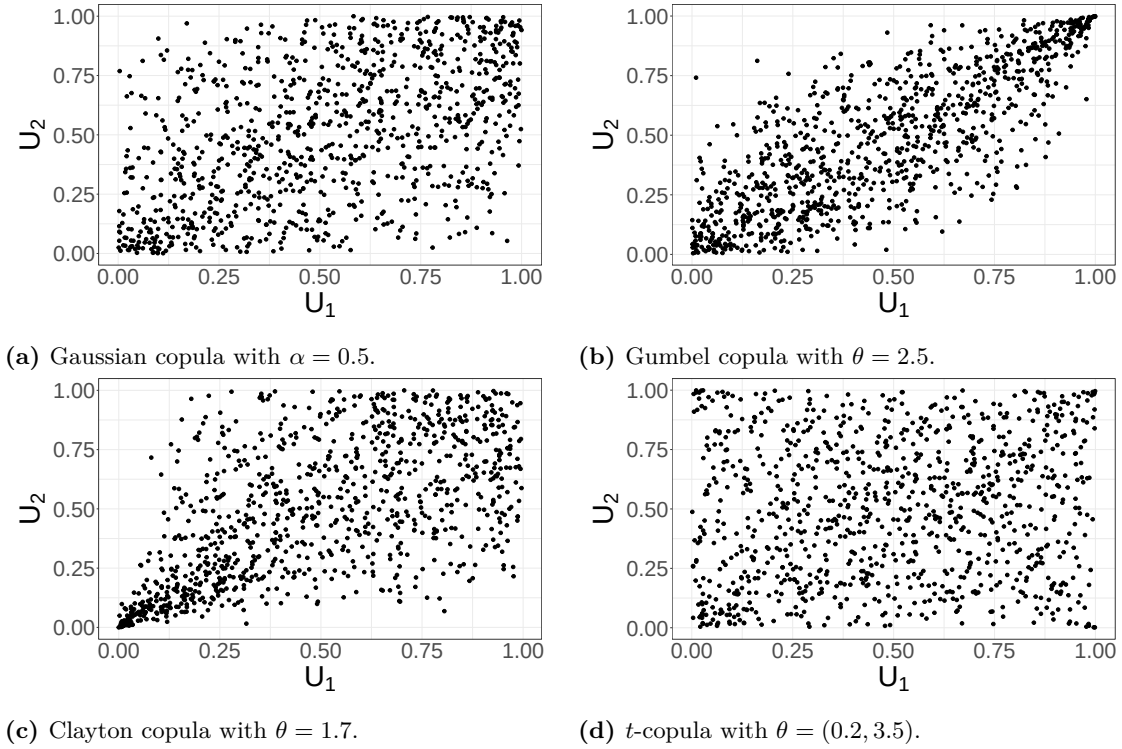


Figure 1: Scatter plots of $n = 1000$ draws from the respective bivariate copula models.

observations and are calculated as $u_{k,j} = \hat{F}_k(x_{k,j})$ for $k \in \{1, 2\}$ and $j \in \{1, \dots, n\}$. If the marginal distributions are well-specified, the pseudo-observations are approximately distributed standard-uniformly. The copula can then be estimated by e.g. MLE as the joint distribution of the pseudo-observation sample $(u_{1,j}, u_{2,j})_{j=1}^n$. In Fig. 2 and 3, this modeling scenario is exemplified. The plots show scatter plots and histograms of simulations from a random vector (X_1, X_2) , where $X_1 \sim N(0.6, 2)$ and $X_2 \sim \text{Exp}(2.7)$. Furthermore, the copula of (X_1, X_2) is the t -copula with parameters $\varphi = (\rho, \nu) = (0.3, 2.7)$, i.e. $\mathbb{P}(F_1(X_1) \leq x_1, F_2(X_2) \leq x_2) = t_{\rho, \nu}[t_\nu^{-1}(x_1), t_\nu^{-1}(x_2)]$, compare Sect. 2.1.3. The joint distribution of (X_1, X_2) is visualized as a scatter plot of samples in the lower left panel of Fig. 2. The upper right panel shows the mirrored scatter plot. Notice that this joint distribution can not be modeled properly by standard multivariate distributions. Now assume that the true distributions and parameter values are not known. Following the modeling approach from above, the joint distribution is decomposed into marginal distributions and the copula of (X_1, X_2) . The upper left and lower right panel of Fig. 2 show histograms of the respective marginal samples. The red lines are estimated densities. The upper left histograms density is assumed to be the normal distribution and the lower right histograms density as exponential distribution. After estimating the marginals, the first building block of the decomposition is obtained, and the second building block, the copula, can be estimated. To this end, the samples are transformed by the PIT w.r.t. the corresponding marginal distribution estimate. The histograms of the resulting transformed

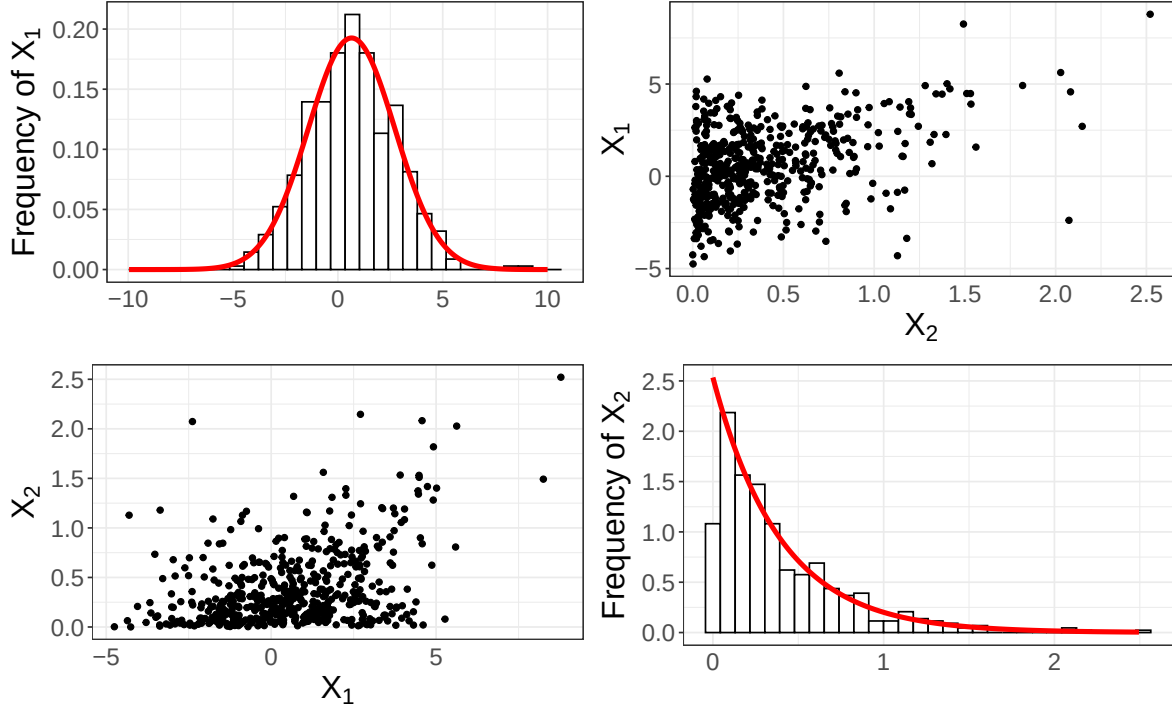


Figure 2: Histograms and scatter plots of simulated data. The red line in the histograms is an estimated density.

sample is displayed in the upper left and lower right panel of Fig. 3. In the lower left panel of Fig. 3, the scatter plot of the pseudo-observations is displayed. In the upper right panel, the mirrored scatter plot is shown. The samples are understood as samples from the underlying copula. Visual inspection of the corners indicates that the dependence structure is heavy tailed. In fact, comparing the plot with the scatter plot in the lower right panel of Fig. 1, which shows draws from the t -copula, indicates that the copula of (X_1, X_2) is the t -copula.

2.1.5 Vine-Copulas

The above exposition of copula modeling is mainly focused on bivariate copulas. While most copulas have d -dimensional versions, these versions often are not flexible enough to capture different dependence properties of real data. A flexible solution to d -dimensional dependence modeling is offered by the vine-copula approach. This approach is also especially relevant for time series modeling because a sub-class of vine copulas, the drawable vine ("D-vine"), is a natural choice for modeling univariate temporal dependencies. In the following the concept is introduced briefly and the D-vine density is derived. For a comprehensive introduction, see Czado et al., 2022.

Vine-copula modeling means decomposing a d -dimensional copula density, see Eq. 2.2,

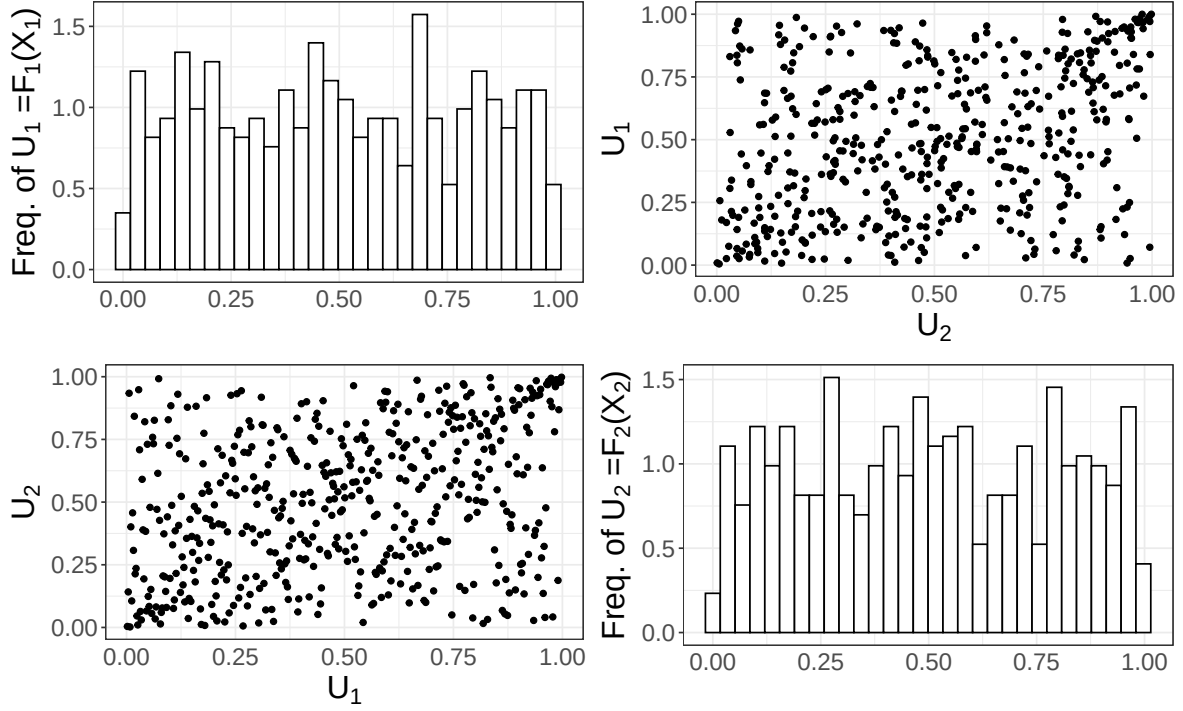


Figure 3: Histograms and scatter plots of probability integral transformed simulated data.

into a product of $d(d-1)/2$ bivariate (conditional) densities, compare Czado et al., 2022. Such decomposition is called *pair-copula construction*. This decomposition is not unique and there are $2^{(d-3)(d-2)}d!$ feasible decompositions, cf. Chang et al., 2019, Sect. 3.1.1. Which decompositions are permitted is determined by the conditions given in Def. 4.1 by Bedford et al., 2002. Decompositions that are permitted are called regular vine ("R-vine"). Every d -dimensional R-vine consists of $(d-1)$ trees. Each tree is made up of conditional densities, where the conditioning set contains the same amount of variables. For instance, there are $(d-1)$ unconditional bivariate copula densities in the pair-copula decomposition. These densities make up the first tree of the R-vine. The second tree contains the $(d-2)$ bivariate conditional copula densities, where each density is conditioned on one variable. This pattern continues up to the last tree, which contains one bivariate conditional density, which is conditioned on $(d-2)$ variables. In the following exposition the focus is on the D-vine because of its relevance for time series modeling. The D-vine structure, informally speaking, connects neighbouring variables. That means, given a random vector $(X_1, \dots, X_d)$, the first tree of the vine contains the copula densities $c_{12}, c_{23}, \dots, c_{(d-1)d}$, where the subscripts indicate which variables the copula density corresponds to. The second tree contains the conditional densities $c_{13|2}, c_{24|3}, \dots, c_{(d-2)d|(d-1)}$. This pattern is continued to the last tree, which contains the density $c_{1d|(2,\dots,(d-1))}$. This vine structure is natural to time series modeling because stationarity is easy to enforce given a D-vine. In Sect. 2.3.3, this will be elaborated on more. Given a D-vine, its d -variate density can be

written as

$$c(u_1, \dots, u_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|(i+1),\dots,(i+j-1)} \left(F_{i|(i+1),\dots,(i+j-1)}(u_i | u_{i+1}, \dots, u_{i+j-1}), \right. \\ \left. F_{i+j|(i+1),\dots,(i+j-1)}(u_{i+j} | u_{i+1}, \dots, u_{i+j-1}) \right),$$

compare Sect. 3 of Czado et al., 2022. The index j reflects the tree depth and the index i iterates through all bivariate (conditional) densities in the tree depth j . Notice that the copula density $c_{i,i+j|(i+1),\dots,(i+j-1)}$ implicitly depends on the variables $(u_{i+1}, \dots, u_{i+j-1})$. In statistical modeling it is common to omit this dependence. This is called the simplifying assumption, see Czado et al., 2022 and Mroz et al., 2021. The conditional CDFs, e.g. $F_{i|(i+1),\dots,(i+j-1)}$, can be calculated by using Eq. 19 by Czado et al., 2022.

2.2 Time Series Analysis

In this section important time series models are introduced, probabilistic forecasting is established and strictly proper scoring rules are discussed briefly. For general introduction to these topics, see Hamilton, 2020, Francq et al., 2019 and Gneiting et al., 2007. After introducing probabilistic forecast evaluation, the role of risks in non-parametric regression is discussed, because of the thematic similarity. This field is important for the fifth paper of this thesis.

2.2.1 Time Series Models

The model, which we aim to generalize in the third and fourth paper of this work, and which is also part of models used in the first and second paper, is the autoregressive moving average (ARMA) model. The model is defined in terms of its updating equations. Let $\{Y_t\}_{t \in \mathbb{Z}}$ follow a stationary ARMA(p, q) model then it holds that

$$Y_t = \varepsilon_t + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \quad \varepsilon_t \sim WN(0, \sigma^2),$$

for all $t \in \mathbb{Z}$, compare chapter 3 by Hamilton, 2020. The constant mean is omitted. If $p = 0$ the MA(q) model is recovered. If $q = 0$ the AR(p) model is recovered. If $\varepsilon_t \sim N(0, \sigma^2)$, the Gaussian-ARMA(p, q) model is recovered. In the following these models are discussed briefly w.r.t. its memory properties. For comprehensive introductions see Hamilton, 2020 or Shumway et al., 2000. For the discussion of memory properties, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) are

useful, compare Shumway et al., 2000, Chapter 3.3. The (P)ACF measures the (partial) correlation between instances of a time series. The measures are defined as $\rho : \mathbb{Z} \rightarrow [-1, 1]$ such that $\rho(h) = \text{Corr}(Y_t, Y_{t-h})$ for the ACF and similarly $\kappa : \mathbb{Z} \rightarrow [-1, 1]$ such that $\kappa(h) = \text{Corr}(Y_t, Y_{t-h} | Y_{t-1}, \dots, Y_{t-h+1})$ for the PACF. An $\text{AR}(p)$ time series is a $\text{Markov}(p)$ process. This property is manifested as a cut-off PACF after lag p . The ACF of the $\text{AR}(p)$ process decays geometrically. On the other hand, a $\text{MA}(q)$ process has a cut-off ACF after lag q , hence it is also called q -dependent. The PACF of an MA process decays geometrically. The ARMA process, which combines the AR and MA part, has a decaying ACF and PACF. The process can be written in a $\text{MA}(\infty)$ and in an $\text{AR}(\infty)$ representation, compare Chapter 3 by Shumway et al., 2000. The $\text{AR}(\infty)$ representation, with decaying but never vanishing weights, reveals that all past observations have an actual influence on the current observations. This is the behavior that a copula-based extension of the ARMA process aims to mimic. Hence combinations of MA and AR-like copula parts are discussed in the third and fourth paper.

Another important time series process, which is referenced in the fourth paper and used in the first and second paper is the generalized autoregressive conditional heteroskedasticity (GARCH) model. See Bollerslev et al., 1994 and Francq et al., 2019 for general introductions. Building upon the ARCH model, introduced by Engle, 1982, Bollerslev, 1986 introduced the GARCH model to capture volatility clusters. The model equation of the model with orders $p = q = 1$ is given as

$$\begin{aligned} Y_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \omega + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned} \tag{2.3}$$

Notice that the object that is modeled conditionally, i.e. the conditional variance σ_t^2 , enters multiplicatively in Eq. 2.3. This is in contrast to the ARMA model, where the conditional mean enters in an additive way. The first instance of such a model, the ARCH model only involves $\sigma_t^2 = \omega + \alpha_1 Y_{t-1}^2$. The generalization is introduced to allow for long term autoregressive effects. In contrast to the ARMA model, these effects take place in the conditional variance and not the conditional mean. In the fourth paper we are particularly interested in mimicking the dynamics of the GARCH(1, 1) model using copula-based MA and AR-like terms, see Sect. 3.4.

2.2.2 (Probabilistic) Forecasting

Predicting new observations based on old observations is a major task in time series modeling. The prediction is obtained by conditioning on past observations. In e.g. the ARMA model, the prediction is given by the conditional mean, i.e. $\hat{\mu}_t = \mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$, where $\mathcal{F}_{t-1}$ contains all information up to and including $(t-1)$.

The conditional mean forecast can also be extended to be a distributional forecast by assuming an explicit distribution for the innovations. If e.g. $\varepsilon_t \sim N(0, \sigma^2)$, the *probabilistic forecast* is given as $Y_t | \mathcal{F}_{t-1} \sim N(\mu_t, \sigma^2)$. Such a probabilistic forecast can be reported in terms of parameters of a distribution, if the distribution is known. Conditional densities, distribution functions or quantiles can also characterize a probabilistic forecast. It is also possible to access the probabilistic forecast by simulating from it. Depending on the model, different characterizations may be more suitable than others. In e.g. the first and second paper we chose to access the probabilistic forecast by simulating from the predictive distribution. In the fourth paper the probabilistic forecast is characterized by percentiles. The appropriate evaluation of probabilistic forecasts is a major field of study in the respective literature. The evaluation measure that is used in most of the contributed papers is the continuous ranked probability score (CRPS). The CRPS is the one-dimensional version of the Energy score, which is used in the second paper. See Gneiting et al., 2007 for these and other evaluation measures. Both scores can be derived from the energy distance. The energy distance of two d -variate distributions, say F and G , characterized by random variables $\mathbf{W}$ and $\mathbf{Z}$, is given as

$$d(F, G) = 2\mathbb{E}(\|\mathbf{W} - \mathbf{Z}\|) - \mathbb{E}(\|\mathbf{W} - \mathbf{W}'\|) - \mathbb{E}(\|\mathbf{Z} - \mathbf{Z}'\|), \quad (2.4)$$

compare Székely et al., 2013. The random variables $\mathbf{W}'$ and $\mathbf{Z}'$ are independent copies of the random variables $\mathbf{W}$ and $\mathbf{Z}$. In the probabilistic forecast evaluation context one distribution takes on the role of the predictive distribution at time t , i.e. $W = (Y_t | \mathcal{F}_{t-1})$ and the other distribution takes on the role of the empirical distribution function of the realized observation, $Z = y_t$. In the one-dimensional case the CRPS is recovered. If, furthermore, the forecast is not probabilistic, e.g. a conditional mean or median prediction, the mean absolute error $|\mu_t - y_t|$ is recovered. Related but different is the mean squared error, which is obtained by averaging $(\mu_t - y_t)^2$. These evaluation measures are strictly proper scoring rules, compare Gneiting et al., 2007. The expression in Eq. 2.4 being based on expectation values admits a direct simulation based estimation by averaging. The CRPS may also be approximated by aggregating the Pinball score values of the percentiles, compare Appendix D by Hirsch et al., 2024.

2.2.3 Risks in Non-Linear Regression

In the fifth paper of this cumulative dissertation such score functions are analyzed from the perspective of non-linear/non-parametric regression. From this perspective these scores, which in this context are called risks, are used to train models. Model parameters are chosen such that they minimize such a risk. In addition a penalty term may be added to the risk. For linear models classical choices are the ridge penalty, see Hoerl et al., 1970

or Chapter 3.4.1 by Hastie et al., 2009, and the LASSO-penalty, see Tibshirani, 1996. Non-linear generalizations of these penalties can be obtained by using gradient-based penalization, see e.g. Rosasco et al., 2013. In the fifth paper of this work, the minima of (penalized) risks are derived, see Sect. 3.5.

2.3 Copulas and Time Series

Copulas can be used to model joint distributions in the time series setting. There are three related but distinct ways to do this. First, the joint distribution of consecutive multivariate observations can be modeled using the copula approach. This approach is briefly introduced in Sect. 2.3.1. Second, the joint conditional distribution of a stable multivariate time series, given all past information, which is of interest for forecasting purposes, can be decomposed using copulas. This approach is discussed in Sect. 2.3.2. Third, the modeling of temporal dependencies of a univariate stationary time series is discussed in Sect. 2.3.3.

2.3.1 Modeling of Multivariate Time Series Across Lags

Let $\{\mathbf{X}\}_{t \in \mathbb{Z}}$ be a d -variate stable time series. The copula-based approach amounts to decomposing a finite dimensional joint distribution, associated to the time series, into marginal distributions and the copula, cf. Brechmann et al., 2015; Nagler et al., 2022a; Simard et al., 2015; Smith, 2015. Let F_j denote the univariate stationary distribution of $\{X_{j,t}\}_{t \in \mathbb{Z}}$ for $j \in \{1, \dots, d\}$. Furthermore let $F_{\mathbf{X}_t, \dots, \mathbf{X}_{t-p}}$ denote the joint distribution of $(p+1)$ consecutive multivariate observations. The copula-based approach, which we focus on in this section, aims to decompose this joint distribution into a copula and the marginal distributions. This means modeling the marginal distributions F_j and a copula C such that

$$F_{\mathbf{X}_t, \dots, \mathbf{X}_{t-p}} = C[F_1(X_{1,t}), \dots, F_d(X_{d,t}), \dots, F_1(X_{1,t-p}), \dots, F_d(X_{d,t-p})]. \quad (2.5)$$

Note that $C : (0, 1)^{(p+1) \times d} \rightarrow (0, 1)$ simultaneously captures temporal dynamics of the individual time series, the contemporaneous dependence and cross-temporal effects. Depending on the properties that are desired for the time series $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, certain restrictions can be put on $F_{\mathbf{X}_t, \dots, \mathbf{X}_{t-p}}$. To ensure stationarity, the marginal distributions should not vary over time. This is already intact in Eq. 2.5. Furthermore, restrictions have to be enforced on the copula. E.g. the d -dimensional marginal distribution representing the cross-sectional dependence of $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ should not vary with t , see also Simard et al., 2015. For C being a vine copula, see Sect. 2.1.5, Nagler et al., 2022a establish *translational*

invariance as a suitable condition. The estimation in this approach can be conducted by minimizing the likelihood given in Sect. 2.1.2. If stationarity is enforced, the copula has to be restricted or parameterized in a suitable way. For forecasting purposes, the conditional distribution $F_{\mathbf{X}_t|\mathbf{X}_{t-1},\dots,\mathbf{X}_{t-p}}$ needs to be calculated or simulated from, see e.g. Simard et al., 2015, who do this for the t -copula. In the first paper of this thesis this approach is used to forecast commodity prices with different copula specifications.

2.3.2 Modeling of Conditional Joint Distributions

The key result for the modeling of the conditional joint distribution of a multivariate time series is the conditional version of Sklar's Theorem. In simplified notation it states that

$$F_{Y,X|Z}(y, x) = C(F_{Y|Z}(y), F_{X|Z}(x)),$$

where C may depend on Z , see Sect. 2 of Nagler et al., 2025a. This conditional version of Sklar's Theorem is applied to the conditional distribution of a multivariate time series. That is $F_{\mathbf{X}_t|\mathcal{F}_{t-1}}$ is decomposed into the conditional marginal distributions $F_{X_{j,t}|\mathcal{F}_{t-1}}$ and the copula, which may depend on the information contained in $\mathcal{F}_{t-1}$. The model is often called copula-GARCH, see e.g. Aloui et al., 2014; Hu, 2006; Jondeau et al., 2006. The model can be estimated by minimizing the negative log-likelihood given in Sect. 2.1.2. In the second article, where this model is used, the estimation is performed in one step.

2.3.3 Modeling of Univariate Time Series

Here, the copula-based approach for modeling stationary univariate time series is discussed briefly. See e.g. Chen et al., 2006, Smith et al., 2010 or the third and fourth paper of this thesis for more detailed discussions. Let $\{Y_t\}_{t \in \mathbb{Z}}$ be a univariate Markov(p) time series. The joint distribution of $(p+1)$ consecutive observations specifies the temporal dynamics of $\{Y_t\}_{t \in \mathbb{Z}}$. In conjunction with the stationary distribution of $\{Y_t\}_{t \in \mathbb{Z}}$ the process is specified. The model might be thought of as a non-linear and probabilistic generalization of the AR(p) model. Viewing this modeling approach in light of the non-linear AR(1) model, Eq. 1.1, raises the question of what the relation is between the mapping $\tilde{m}$ and the copula of (Y_t, Y_{t-1}) . Using basic copula arguments, it becomes clear that

$$Y_t = F_Y^{-1}(C_{2|1}^{-1}(\varepsilon_t | F_Y(Y_{t-1}))), \quad (2.6)$$

compare e.g. Chapter 3.2 by Spanhel, 2015. F_Y is the stationary distribution of the time series and $C_{2|1}^{-1}$ the conditional quantile function corresponding to a copula C_{12} , compare Sect. 2.1.2. This is the updating equation of the copula-based non-linear AR(1)

generalization. Notice that $C_{2|1}^{-1}(\varepsilon_t|F_Y(Y_{t-1})) \sim U(0, 1)$ and hence $Y_t \sim F_Y$. Furthermore, Eq. 2.6 implies that the copula of (Y_t, Y_{t-1}) is C_{12} , compare Sect. 2.1.2. The extension to a copula-based $\text{AR}(p)$ is (almost) straight-forward:

$$Y_t = F_Y^{-1}(C_{(p+1)|p,\dots,1}^{-1}[\varepsilon_t|F_Y(Y_{t-1}), \dots, F_Y(Y_{t-p})]), \quad \varepsilon_t \stackrel{iid}{\sim} U(0, 1). \quad (2.7)$$

To assure stationarity additional conditions have to be put on the copula $C_{(p+1),\dots,1}$. If the copula is a D-vine copula, then it is sufficient that copulas in a tree are equal to each other. This fulfills translational invariance by Nagler et al., 2022a. This modeling approach differs from e.g. AR and ARCH approaches in the following way. The AR and ARCH approaches model conditional quantities directly. The AR models the conditional mean and ARCH models the conditional variance. In the copula approach the joint distribution of consecutive observations is modeled.

3 Summary of the Articles

3.1 Article 1: Forecasting Natural Gas Prices with Spatio-Temporal Copula-Based Time Series Models

Here, the first paper, "Forecasting Natural Gas Prices with Spatio-Temporal Copula-Based Time Series Models" (Pappert et al., 2023) is discussed. In this paper natural gas prices and related commodities are predicted in a forecasting study. The models that are used for prediction are copula-based time series models described in Sect. 2.3.3 and Sect. 2.3.1. In particular, a univariate t -copula based time series is compared with multivariate (Gaussian/ t)-copula based models and with a D-vine approach. Furthermore we include an ARMA-GARCH-type model, closely related to the model by Berrisch et al., 2022, as a benchmark. As these models naturally produce probabilistic forecasts, a proper scoring rule for such forecasts – the CRPS, see Sect. 2.2.2 – is used to evaluate the one-day-ahead forecasting performance. The forecasting study reveals that the benchmark model performs best w.r.t. the CRPS, however, the univariate copula-based time series models seem to be competitive.

In addition to comparing the probabilistic forecasting performances, the emergence of non-standard predictive distributions is discussed and contrasted to elliptic predictive distributions from ARMA-GARCH-type models. The heavy tails of the t -copula are suitable to model volatile time series, because the heavy tails imply that extreme events coincide, compare also Loaiza-Maya et al., 2018 who follow this approach. If the t -copula is used to model the dependence of consecutive observations, the predictive distributions can become bimodal. This is in contrast to the GARCH model, where the whole predictive distribution is just conditionally dispersed during extreme events. This underlying mechanism causes the bimodality of the predictive distributions from a t -copula based time series model. The mechanism is exemplified and visualized.

Furthermore, we argue that it is not obvious how to extract a suitable mean forecast from such copula-based predictive distributions. While it is true that the conditional expectation is the mapping that minimizes the (R)MSE functional, compare Sect. 3.5, there might be instances where the mode, the median or even other quantiles are more suitable. As an

experiment, an artificial neural network – a multilayer perceptron – is trained to predict which quantile of the predictive distribution minimizes the RMSE for each time-point. See Hastie et al., 2009, chapter 11, for an introduction to artificial neural networks. The mean forecasting performance of this approach is compared with other approaches and is seen to be competitive. In retrospect I think that simply using the expected value of the predictive distribution is the best choice, because it ultimately minimizes the RMSE. Still, it was worth experimenting with this approach at the time.

Finally, I would like to report and correct a typographical error. On page 2 and page 4 of the paper, the authors of the reference Loaiza-Maya et al., 2018 are mentioned by name as "Loazia-Maya et al." before citing the reference correctly. There is a typo in the naming and it should be "Loaiza-Maya et al."

3.2 Article 2: Modeling volatility and dependence of European carbon and energy prices

Here the second paper, "Modeling volatility and dependence of European carbon and energy prices" (Berrisch et al., 2023) is discussed. In this article we propose a sophisticated (econometric) model for the probabilistic forecasting and modeling of European carbon and energy prices. The model is an extension of the copula-GARCH model described in Sect. 2.3.2. This means that we predict the multivariate conditional distribution of the time series by decomposing it into a copula and the conditional marginal distributions.

The building blocks of this decomposition are the following. First, the conditional means are modeled (jointly) by employing a vector error correction (VECM) model, which incorporates co-integrating relationships, see Johansen, 1988 or chapter 3 by Kilian et al., 2017. Second, we model the conditional variances in a univariate fashion with (leverage) GARCH models. For each individual time series, the conditional marginal distributions are modeled as generalized non-central t -distribution. The conditional mean prediction by the VECM component and the conditional variance prediction by the GARCH model are the first two parameters of this distribution. The non-centrality parameters and the degrees-of-freedom parameters are modeled unconditionally. The copula that couples these conditional marginal distributions is chosen as the t -copula, compare Sect. 2.1, with a conditionally varying correlation matrix. The t -copula is suitable since it can capture the joint occurrence of extreme events. The conditionally varying correlation is employed by assigning a GARCH-like process, say $\{\xi_{ij,t}\}_{t \in \mathbb{Z}}$, for the conditional correlation of time series i and j at time t , that consists of a forcing term $z_{i,t-1}z_{j,t-1}$ and an autoregressive term containing the last estimated conditional correlation $\xi_{ij,t-1}$. The factors in the forcing

term are the standardized residuals of the respective time series at the respective times. In order to assure semi-positive definiteness, the estimated correlations are transformed by a suitable link function.

It is shown that the proposed model nests many relevant models. These models are considered in a forecasting study along with the full model with different specifications. All models are estimated in a one-step maximum likelihood estimation. In the forecasting study we also take preliminary log-transformations into account and explore how this affects the forecasting performance. Furthermore all time series are standardized by their respective carbon emissions, such that the scales are comparable. We use several strictly proper scoring rules to evaluate the (probabilistic) forecasting performance across different time horizons. Furthermore, the Diebold-Mariano test (see Diebold et al., 2002) is used to test whether the models significantly outperform simple random walks with unconditional variances and unconditional contemporaneous dependence. The results of the forecasting study are interpreted and the model which performs best in terms of the energy score is analyzed further.

The model that performs best in terms of the energy score has a conditionally varying correlation. Surprisingly it has no conditionally varying variances. We extract the estimated correlations and analyze the results. We observe a shock in the time series during the Russian invasion of Ukraine. In conclusion, sophisticated conditional dependence and marginal modeling enhances the probabilistic forecasting performance for European carbon and energy prices.

3.3 Article 3: Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas)

Here the third paper, "Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas)" (Pappert, 2024) is discussed. The MAGMAR-Copula model is an extension of the copula-based time series modeling approach for univariate stationary time series, see Sect. 2.3.3. The basic idea is to substitute the innovation in the updating equation of the copula-based $AR(p)$ model by a suitable moving average-type term. Since this term is a non-linear mapping, it is called *moving aggregate*. The updating equation is given in Eq. 2.6 for the $p = 1$ case and in Eq. 2.7 for the general case. The innovation in this updating equation is substituted by a suitable aggregate of the current and the last q observations, i.e. $(\varepsilon_t, \dots, \varepsilon_{t-q})$. The $(q + 1)$ dimensional mapping that is suitable is the conditional quantile function corresponding to a $(q + 1)$ -dimensional copula $K_{(q+1), \dots, 1}$. Compare also Chapter 3.1.4 by Joe, 2014, who calls this

process q -dependent extension. This mapping is suitable because it conserves the standard uniformity: The copula-MAG(q) time series $\{V_t\}_{t \in \mathbb{Z}}$, following the updating equation $V_t = K_{(q+1)|q, \dots, 1}^{-1}(\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_{t-q})$ is standard uniform, i.e. $V_t \sim U(0, 1)$ for all t . A key realization of the paper is that, while the standard uniformity holds for the MAG-part and the AR-part individually, it does not hold for the general MAGMAR-copula model. The unconditional distribution of the MAGMAR-copula model deviates from standard-uniformity. The reason for this can be exemplified in the updating equation of a MAGMAR(1, 1) time series $\{U_t\}_{t \in \mathbb{Z}}$ on the pseudo-observation scale,

$$U_t = C_{2|1}^{-1}(K_{2|1}^{-1}(\varepsilon_t | \varepsilon_{t-1}) | U_{t-1}), \quad \varepsilon_t \stackrel{iid}{\sim} U(0, 1).$$

The term $K_{2|1}^{-1}(\varepsilon_t | \varepsilon_{t-1})$, which is the MAG(1)-part and U_{t-1} are dependent through the innovation ε_{t-1} . This results in the distribution of U_t deviating from standard uniformity. This is a major problem, because the copula-based modeling approach, described in Sect. 2.3.3, heavily relies on the temporal dynamics taking place between standard uniform random variables. In the paper this problem is resolved by introducing an additional transformation, resulting in the *adjusted MAGMAR-copula model*, denoted by Ψ -MAGMAR(p, q). The time series $\{\tilde{U}_t\}_{t \in \mathbb{Z}}$ following a Ψ -MAGMAR(1, 1) model is defined by $\tilde{U}_t = \Psi(U_t)$ and has an updating equation,

$$\tilde{U}_t = \Psi[C_{2|1}^{-1}(K_{2|1}^{-1}(\varepsilon_t | \varepsilon_{t-1}) | \Psi^{-1}(\tilde{U}_{t-1}))].$$

If Ψ is chosen as the stationary distribution of $\{U_t\}_{t \in \mathbb{Z}}$, it holds that $\tilde{U}_t \sim U(0, 1)$ for all t . In application, the parameters of the MAGMAR-copula model and the transformation Ψ are estimated iteratively. During these iterations the MAGMAR-copula parameters are estimated by MLE and Ψ is estimated non-parametrically by kernel density estimation. This is the modeling approach that the paper follows and in the following it is implied that Ψ is the stationary distribution of $\{U_t\}_{t \in \mathbb{Z}}$. The existence of such stationary distribution is proven for the MAGMAR(1, 1) case. The proof is based on a contraction-on-average approach, compare Theorem 4.40 by Douc et al., 2014. For the AR-copula being the Gaussian copula and arbitrary MAG-copulas, the stationarity (and ergodicity) can also be established directly. The maximum likelihood estimation is discussed. Conditions under which the MLE for parameters of the MAG(1) process is consistent are derived. The conditions essentially come from general conditions of M-estimators in the time series context, compare Theorem 4.3 of Wooldridge, 1994. The crucial step is to establish the uniform weak law of large numbers (UWLLN) of the score time series. The key ingredient is to first establish the stationarity and ergodicity of the residual time series, again using contraction-on-average arguments by Douc et al., 2014. Then the UWLLN and consistency follows under suitable additional assumptions. Asymptotic properties of estimators are also investigated in a simulation study along with exemplifying the modeling approach.

The likelihood is derived for general model orders and the MAGMAR-copula model is used to model US inflation. The performance is competitive with benchmark models.

3.4 Article 4: Copula-Based Time Series for Non-Gaussian and Non-Markovian Stationary Processes

Here the fourth paper, "Copula-Based Time Series for Non-Gaussian and Non-Markovian Stationary Processes" (Pappert et al., 2026) is discussed. This article lies within the same context as the third paper (Pappert, 2024). We investigate the copula ARMA extension, proposed by Joe, 2014, in detail. This time series model also consists of an AR-like and a MA-like term. The difference is that the roles of the terms are reversed. For non-linear AR and MA-like terms it does make a difference which processes are modeled as latent. The model with model order p and q is given as

$$\begin{aligned} U_t &= K_{(q+1)|q, \dots, 1}^{-1}(\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_{t-q+1}, W_{t-q}), \quad t \in \mathbb{Z} \quad \text{where} \\ W_s &= C_{(p+1)|p, \dots, 1}^{-1}(\varepsilon_s | W_{s-1}, \dots, W_{s-p}), \quad s \in \mathbb{Z}. \end{aligned} \quad (3.1)$$

The latent process $\{W_s\}_{s \in \mathbb{Z}}$ is a copula AR(p) process and it is plugged into an MA-like term, yielding the time series $\{U_t\}_{t \in \mathbb{Z}}$. This approach has the crucial property that $U_t \sim U(0, 1)$ for all $t \in \mathbb{Z}$. The MAGMAR-copula approach from the third paper does not have this property and additional transformations are necessary. In the paper we investigate the model from Eq. 3.1 in detail.

We derive the relation to the Gaussian-ARMA model. Since the model in Eq. 3.1 is based on an AR(p) and a MAG(q) building blocks, it could be expected that for Gaussian copulas, the Φ -quantile transformed time series $\{\Phi^{-1}(U_t)\}_{t \in \mathbb{Z}}$ follows a Gaussian-ARMA(p, q) model. However, we show that for $p, q > 1$, $\{\Phi^{-1}(U_t)\}_{t \in \mathbb{Z}}$ follows a subset of a Gaussian-ARMA($p, q + p + 1$) model. This is an unexpected result and shows that the proposed model does not generalize all ARMA dynamics. We proceed to derive the relation to the Gaussian-GARCH(1, 1) model, compare Sect. 2.2.1. We derive the copulas K_{21} and C_{21} which are necessary such that the time series $\{U_t\}_{t \in \mathbb{Z}}$ from Eq. 3.1 has GARCH dynamics. This insight allows for assigning GARCH(1, 1) dynamics to time series with arbitrary stationary distributions. As the statistical modeling is not straight-forward, potential modeling approaches are discussed.

In a simulation study the dependence of consecutive observations of the MAG(1) and the model from Eq. 3.1 with $p = q = 1$ are investigated. Furthermore potential identifiability issues of the Gaussian-MAG(1) model are explored. It turns out that, similar to its linear pendant, the MA(1) process, the parameter space has to be restricted in order to assure

identifiability. The bound on the parameter space coincides with the bound that can be derived by following Proposition 7 from the third paper to establish the ergodicity of the estimated residual time series.

For the model with general model orders we discuss the MLE and provide an algorithm for estimation and forecasting based on the `rvinecopulib` package in R, see Nagler et al., 2022b. We use these algorithms to perform probabilistic forecasting studies, performing one-step ahead forecasting for US inflation and German wind power production. The forecasts are evaluated using strictly proper scoring rules. We observe that linear forecast, i.e. Gaussian-ARMA, are hard to beat in forecasting the time series of interest. Furthermore, copula-based models perform best when the marginal stationary distribution of the time series deviates from normality and a non-parametric approach is suitable.

3.5 Article 5: The Field Equations of Penalized non-Parametric Regression

Here the fifth paper, "The Field Equations of Penalized non-Parametric Regression" (Pappert, 2025), is discussed. The fifth article of this cumulative dissertation lies within the context of non-linear, or non-parametric, regression. Assuming an underlying relation between the target variable $Y \in \mathbb{R}$, the covariates $\mathbf{X} \in \Omega \subset \mathbb{R}^p$ and idiosyncratic innovations $\varepsilon \sim (0, \sigma^2)$ as

$$Y = m(\mathbf{X}) + \varepsilon,$$

it is investigated how the choice of the loss function/risk affects what an estimator for m approximates. Possible choices for losses/risks for training or evaluation are the (R)MSE, MAE or the CRPS, compare Sect. 2.2.2, possibly with penalizing terms. This is a distinct but related area of research to copula-based modeling in time series. Both research directions share the aspect of modeling non-linear relations. In the article, the general research question of which functions minimize a loss functional is posed. Examples for such functionals are (R)MSE, MAE or CRPS, compare also Sect. 2.2.2, where these functionals are introduced as evaluation measures for (probabilistic) forecasts. In the paper a general framework is introduced, which allows for systematic derivations of the minima of such risk functionals. This general framework follows from the calculus of variations, compare Rindler, 2018. It is well-known that the conditional expectation minimizes the MSE in the regression or time series context. Following Bishop, 2006 this and related results are reiterated.

While these derivations are interesting in of themselves, they do not involve any differentials. This is different if penalization is taken into account. The penalization that the paper focuses on is gradient-based penalization. This is a penalty, which involves the gradient of the regression function m , compare Rosasco et al., 2013. It turns out that the functions minimizing such functionals are subject to the Euler-Lagrange equations, compare Theorem 3.1 by Rindler, 2018. This insight offers an approach to derive the minima of (penalized) risk functionals. Then the statement "A penalized estimator is biased" can be shifted to "A penalized estimator is biased and it approximates the solution to the following field equation", where "the following field equation" has to be derived using the Euler-Lagrange equations. We exemplify this approach on the MSE, penalized by $\lambda \int \|\nabla g(x)\|_2^2 d\mathbb{P}^X(x)$. The field equations, which are second order partial differential equations, are derived for this example and a general result for gradient based penalties is derived.

4 Discussion and Outlook

In this cumulative dissertation five papers are presented. The papers are concerned with non-linear modeling in time series and regression. The core-part of the dissertation are the methodological contributions from the third and fourth paper (Pappert, 2024 and Pappert et al., 2026). These contributions explore how the copula-based modeling of stationary univariate time series can be extended to accommodate long-term autoregressive effects. Such an extension is equivalent to finding a copula-based model that completely generalizes the ARMA model. The papers discuss two related but different approaches, derive theoretical results and apply the extended models to real data in modeling and forecasting studies respectively. Both models have some theoretical challenges. The MAGMAR-copula model, presented in the third paper, needs an additional transformation to ensure that the stationary distribution is standard uniform. The model discussed in the fourth paper, which was originally introduced by Joe, 2014, automatically has a standard uniform stationary distribution. However, our analysis reveals that it can not be regarded as a full generalization of the ARMA model because not all models can be recovered. It remains an open question if there is a suitable combination of MA and AR-like copula-based terms that yields a full generalization of the $\text{ARMA}(p, q)$ model with arbitrary p and q , without the need for additional transformations. Nevertheless, the models can be used for modeling and (probabilistic) forecasting of univariate and stationary time series. An especially interesting potential future research direction is to implement the GARCH(1,1) specification of the model from the fourth paper and to experiment with the proposed modeling approach.

The second paper of this dissertation employs copula-based modeling not for the temporal modeling of a time series, but for the modeling of the conditional contemporaneous dependence of a multivariate time series, containing European carbon and energy prices. The article is in the spirit of the copula-GARCH model, cf. Jondeau et al., 2006. Combining the copula with sophisticated modeling of the conditional means and variances is revealed to be a powerful approach for the probabilistic forecasting of European allowances and commodity future prices. In the paper the multivariate time series contains 4 time series. Extending this approach to more time series gives rise to a drastic increase in the dependence parameters. Finding precise, efficient and parsimonious solutions in this regard is a potential future research direction. This holds true for vine copulas, but also

for any other copula-based approach. In the first paper of this cumulative dissertation, a copula is employed for the cross-sectional, temporal and cross-temporal joint distribution of multivariate consecutive observations. In this approach the number of parameters also drastically increases with the dimension. Parsimonious solutions would also be relevant in this approach.

The fifth paper of this work analyzes risk functionals. Some of the risk functionals are used as scoring rules in the other papers. In the paper it is explored, how the form of the risk influences the function that is learned when a learner is trained w.r.t. such risk. This question is answered for different risks and taking gradient-based penalization into account. The findings are purely of theoretical nature. In the future it will be interesting to see in what way these findings can be used to enhance statistical modeling in non-linear regression and in non-linear time series modeling.

References

- Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken (2009). “Pair-copula constructions of multiple dependence”. In: *Insurance: Mathematics and Economics* 44.2, pp. 182–198.
- Abdul Azees, Sheik Abdullah and Ramraj Sasikumar (2018). “Comparison Study on Exponential Smoothing and ARIMA Model for the Fuel Price”. In: *Logistics, Supply Chain and Financial Predictive Analytics: Theory and Practices*. Springer, pp. 33–41.
- Aldy, Joseph E, Matthew J Kotchen, Robert N Stavins, and James H Stock (2021). “Keep climate policy focused on the social cost of carbon”. In: *Science* 373.6557, pp. 850–852.
- Aloui, Riadh, Mohamed Safouane Ben Aïssa, Shawkat Hammoudeh, and Duc Khuong Nguyen (2014). “Dependence and extreme dependence of crude oil and natural gas prices with applications to risk management”. In: *Energy Economics* 42, pp. 332–342.
- Angus, John E (1994). “The probability integral transform and related results”. In: *SIAM review* 36.4, pp. 652–654.
- Anthoff, David and Richard SJ Tol (2013). “The uncertainty about the social cost of carbon: A decomposition analysis using fund”. In: *Climatic Change* 117.3, pp. 515–530.
- Beare, Brendan K (2010). “Copulas and Temporal Dependence”. In: *Econometrica* 78.1, pp. 395–410.
- Beare, Brendan K and Juwon Seo (2015). “Vine Copula Specifications for Stationary Multivariate Markov Chains”. In: *Journal of Time Series Analysis* 36.2, pp. 228–246.
- Becker, Raik and Daniela Thrän (2017). “Completion of wind turbine data sets for wind integration studies applying random forests and k-nearest neighbors”. In: *Applied Energy* 208, pp. 252–262.
- Bedford, Tim and Roger M Cooke (2001). “Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines”. In: *Annals of Mathematics and Artificial Intelligence* 32, pp. 245–268.
- Bedford, Tim and Roger M Cooke (2002). “Vines—a new graphical model for dependent random variables”. In: *The Annals of Statistics* 30.4, pp. 1031–1068.
- Benz, Eva and Stefan Trück (2009). “Modeling the price dynamics of CO₂ emission allowances”. In: *Energy Economics* 31.1, pp. 4–15.

- Berrisch, Jonathan, Sven Pappert, Florian Ziel, and Antonia Arsova (2023). “Modeling volatility and dependence of European carbon and energy prices”. In: *Finance Research Letters* 52, 103503.
- Berrisch, Jonathan and Florian Ziel (2022). “Distributional modeling and forecasting of natural gas prices”. In: *Journal of Forecasting* 41.6, pp. 1065–1086.
- Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9780471007104.
- Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.
- Bladt, Martin, Alexandra Dias, Jialing Han, and Alexander J McNeil (2025). “Semiparametric and parametric distributional forecasting of univariate time series using non-Gaussian ARMA models based on D-vines”. In: *Canadian Journal of Statistics*.
- Bollerslev, Tim (1986). “Generalized autoregressive conditional heteroskedasticity”. In: *Journal of Econometrics* 31.3, pp. 307–327.
- Bollerslev, Tim (1990). “Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model”. In: *The Review of Economics and Statistics*, pp. 498–505.
- Bollerslev, Tim, Robert F Engle, and Daniel B Nelson (1994). “ARCH models”. In: *Handbook of Econometrics* 4, pp. 2959–3038.
- Brechmann, Eike Christian and Claudia Czado (2015). “COPAR—multivariate time series modeling using the copula autoregressive model”. In: *Applied Stochastic Models in Business and Industry* 31.4, pp. 495–514.
- Bunn, Derek W. and Carlo Fezzi (2007). *Interaction of European Carbon Trading and Energy Prices*. Climate Change Modelling and Policy Working Papers 9092. Fondazione Eni Enrico Mattei (FEEM).
- Chang, Bo and Harry Joe (2019). “Prediction based on conditional distributions of vine copulas”. In: *Computational Statistics & Data Analysis* 139, pp. 45–63.
- Chen, Xiaohong and Yanqin Fan (2006). “Estimation of copula-based semiparametric time series models”. In: *Journal of Econometrics* 130.2, pp. 307–335.
- Chevallier, Julien, Duc Khuong Nguyen, and Juan Carlos Reboredo (2019). “A conditional dependence approach to CO2-energy price relationships”. In: *Energy Economics* 81, pp. 812–821.
- Czado, Claudia (2010). “Pair-copula constructions of multivariate copulas”. In: *Copula Theory and its Applications*. Springer, pp. 93–109.

- Czado, Claudia and Thomas Nagler (2022). “Vine copula based modeling”. In: *Annual Review of Statistics and Its Application* 9.1, pp. 453–477.
- Darsow, William F, Bao Nguyen, and Elwood T Olsen (1992). “Copulas and Markov processes”. In: *Illinois Journal of Mathematics* 36.4, pp. 600–642.
- Demarta, Stefano and Alexander J McNeil (2005). “The t Copula and Related Copulas”. In: *International Statistical Review* 73.1, pp. 111–129.
- Demetrescu, Matei, Vasył Golosnoy, and Anna Titova (2020). “Bias corrections for exponentially transformed forecasts: Are they worth the effort?” In: *International Journal of Forecasting* 36.3, pp. 761–780.
- Devore, Jay L, Kenneth N Berk, and Matthew A Carlton (2021). *Modern Mathematical Statistics with Applications*. 3rd ed. Springer Cham.
- Dias, Alexandra, Jialing Han, and Alexander J McNeil (2024). “GARCH copulas and GARCH-mimicking copulas”. In: *arXiv preprint arXiv:2408.07025*.
- Diebold, Francis X and Robert S Mariano (2002). “Comparing predictive accuracy”. In: *Journal of Business & Economic Statistics* 20.1, pp. 134–144.
- Ding, Zhao, Chenguang Duan, Yuling Jiao, and Jerry Zhijian Yang (2024). “Semi-Supervised Deep Sobolev Regression: Estimation, Variable Selection and Beyond”. In: *arXiv preprint arXiv:2401.04535*.
- Douc, Randal, Eric Moulines, and David Stoffer (2014). *Nonlinear Time Series: Theory, Methods and Applications with R Examples*. CRC press.
- Drucker, Harris and Yann Le Cun (1991). “Improving generalization performance in character recognition”. In: *Neural Networks for Signal Processing: Proceedings of the 1991 IEEE Workshop*.
- Drucker, Harris and Yann Le Cun (1992). “Improving generalization performance using double backpropagation”. In: *IEEE Transactions on Neural Networks* 3.6, pp. 991–997.
- Duan, Kun, Xiaohang Ren, Yukun Shi, Tapas Mishra, and Cheng Yan (2021). “The marginal impacts of energy prices on carbon price variations: evidence from a quantile-on-quantile approach”. In: *Energy Economics* 95, 105131.
- Dutta, Anupam, Naji Jalkh, Elie Bouri, and Probal Dutta (2020). “Assessing the risk of the European Union carbon allowance market: Structural breaks and forecasting performance”. In: *International Journal of Managerial Finance* 16.1, pp. 49–60.
- Eddelbuettel, Dirk and Romain François (2011). “Rcpp: Seamless R and C++ Integration”. In: *Journal of Statistical Software* 40, pp. 1–18.

- Embrechts, Paul, Marius Hofert, and Ruodu Wang (2016). “Bernoulli and tail-dependence compatibility”. In: *The Annals of Applied Probability* 26.3, pp. 1636–1658.
- Engle, Robert F (1982). “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”. In: *Econometrica*, pp. 987–1007.
- Engle, Robert F and Kevin Sheppard (2001). “Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH”. In: *NBER Working Papers*, 8554.
- Follain, Bertille and Francis Bach (2024). “Nonparametric linear feature learning in regression through regularisation”. In: *Electronic Journal of Statistics* 18.2, pp. 4075–4118.
- Francq, Christian and Jean-Michel Zakoian (2019). *GARCH models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons.
- Garcia, Agustin and Miguel A Jaramillo-Morán (2020). “Short-term European Union Allowance price forecasting with artificial neural networks”. In: *Entrepreneurship and Sustainability Issues* 8.1, pp. 261–275.
- Genest, Christian and Louis-Paul Rivest (1993). “Statistical Inference Procedures for Bivariate Archimedean Copulas”. In: *Journal of the American Statistical Association* 88.423, pp. 1034–1043.
- Giaquinta, Mariano and Stefan Hildebrandt (1996). *Calculus of Variations I*. Springer Berlin, Heidelberg.
- Gneiting, Tilmann and Adrian E Raftery (2007). “Strictly Proper Scoring Rules, Prediction, and Estimation”. In: *Journal of the American Statistical Association* 102.477, pp. 359–378.
- Groll, Andreas, Julien Hambuckers, Thomas Kneib, and Nikolaus Umlauf (2019). “LASSO-type penalization in the framework of generalized additive models for location, scale and shape”. In: *Computational Statistics & Data Analysis* 140, pp. 59–73.
- Hamilton, James D (2020). *Time Series Analysis*. Princeton University Press.
- Hanif, Waqas, Jose Arreola Hernandez, Walid Mensi, Sang Hoon Kang, Gazi Salah Uddin, and Seong-Min Yoon (2021). “Nonlinear dependence and connectedness between clean/renewable energy sector equity and European emission allowance prices”. In: *Energy Economics* 101, 105409.
- Harvey, David, Stephen Leybourne, and Paul Newbold (1997). “Testing the equality of prediction mean squared errors”. In: *International Journal of Forecasting* 13.2, pp. 281–291.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer New York, NY.
- Hayashi, Fumio (2011). *Econometrics*. Princeton University Press.

- Higham, Catherine F and Desmond J Higham (2019). “Deep learning: An introduction for applied mathematicians”. In: *Siam review* 61.4, pp. 860–891.
- Hinton, Geoffrey E, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov (2012). “Improving neural networks by preventing co-adaptation of feature detectors”. In: *arXiv preprint arXiv:1207.0580*.
- Hirsch, Simon, Jonathan Berrisch, and Florian Ziel (2024). “Online distributional regression”. In: *arXiv preprint arXiv:2407.08750*.
- Hoerl, Arthur E and Robert W Kennard (1970). “Ridge Regression: Applications to Nonorthogonal Problems”. In: *Technometrics* 12.1, pp. 69–82.
- Hofert, Marius (2008). “Sampling Archimedean copulas”. In: *Computational Statistics & Data Analysis* 52.12, pp. 5163–5174.
- Hornik, Kurt (1991). “Approximation capabilities of multilayer feedforward networks”. In: *Neural Networks* 4.2, pp. 251–257.
- Hu, Ling (2006). “Dependence patterns across financial markets: a mixed copula approach”. In: *Applied Financial Economics* 16.10, pp. 717–729.
- Hyndman, Rob J and Yeasmin Khandakar (2008). “Automatic Time Series Forecasting: The Forecast Package for R”. In: *Journal of Statistical Software* 27, pp. 1–22.
- Ibe, Oliver (2013). *Markov Processes for Stochastic Modeling*. Elsevier.
- Ibragimov, Rustam (2005). “Copula-Based Dependence Characterizations and Modeling for Time Series”. In: *Harvard Institute of Economic Research Discussion Paper* 2094.
- Joe, Harry (2014). *Dependence Modeling with Copulas*. CRC press, Boca Raton, FL.
- Joe, Harry and James Jianmeng Xu (1996). *The estimation method of inference functions for margins for multivariate models*. Tech. rep. 166. Vancouver, Canada: Department of Statistics, University of British Columbia.
- Johansen, Søren (1988). “Statistical analysis of cointegration vectors”. In: *Journal of Economic Dynamics and Control* 12.2-3, pp. 231–254.
- Jondeau, Eric and Michael Rockinger (2006). “The Copula-GARCH model of conditional dependencies: An international stock market application”. In: *Journal of International Money and Finance* 25.5, pp. 827–853.
- Jónsson, Tryggvi, Pierre Pinson, Henrik Aa Nielsen, and Henrik Madsen (2014). “Exponential smoothing approaches for prediction in real-time electricity markets”. In: *Energies* 7.6, pp. 3710–3732.
- Kilian, Lutz and Helmut Lütkepohl (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

- Kohler, Michael and Sophie Langer (2021). “On the rate of convergence of fully connected deep neural network regression estimates”. In: *The Annals of Statistics* 49.4, pp. 2231–2249.
- Loaiza-Maya, Rubén, Michael S Smith, and Worapree Maneesoonthorn (2018). “Time series copulas for heteroskedastic data”. In: *Journal of Applied Econometrics* 33.3, pp. 332–354.
- Lütkepohl, Helmut and Fang Xu (2012). “The role of the log transformation in forecasting economic variables”. In: *Empirical Economics* 42.3, pp. 619–638.
- McKenzie, Ed (1988). “Some ARMA Models for Dependent Sequences of Poisson Counts”. In: *Advances in Applied Probability* 20.4, pp. 822–835.
- McNeil, Alexander John and Martin Bladt (2022). “Time series with infinite-order partial copula dependence”. In: *Dependence Modeling*.
- Meier, Jan-Hendrik and Norman Voss (2020). “Do Commodities Determine the EU Emission Allowances Price?” In: *ICTERI Workshops*, pp. 423–437.
- Mroz, Thomas, Sebastian Fuchs, and Wolfgang Trutschnig (2021). “How simplifying and flexible is the simplifying assumption in pair-copula constructions—analytic answers in dimension three and a glimpse beyond”. In: *Electronic Journal of Statistics* 15, pp. 1951–1992.
- Müller, Alfred and Matthias Reuber (2022). “A copula-based time series model for global horizontal irradiation”. In: *International Journal of Forecasting*.
- Nagler, Thomas, Gerda Claeskens, and Irène Gijbels (2025a). “On dimension reduction in conditional dependence models”. In: *arXiv preprint arXiv:2505.01052*.
- Nagler, Thomas, Daniel Krüger, and Aleksey Min (2022a). “Stationary vine copula models for multivariate time series”. In: *Journal of Econometrics* 227.2, pp. 305–324.
- Nagler, Thomas, Ulf Schepsmeier, Jakob Stoeber, Eike Christian Brechmann, Benedikt Graeler, and Tobias Erhardt (2025b). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.6.1.
- Nagler, Thomas and Thibault Vatter (2022b). *rvinecopulib: High Performance Algorithms for Vine Copula Modeling*. R package version 0.6.2.1.1.
- Nagler, Thomas and Thibault Vatter (2024). *kde1d: Univariate Kernel Density Estimation*. R package version 1.1.0.
- Nelsen, Roger B (2006). *An Introduction to Copulas*. Springer New York, NY.
- Paolella, Marc S and Luca Taschini (2008). “An econometric analysis of emission allowance prices”. In: *Journal of Banking & Finance* 32.10, pp. 2022–2032.
- Pappert, Sven (2024). “Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas)”. In: *arXiv preprint arXiv:2402.01491v3*.

- Pappert, Sven (2025). “The Field Equations of Penalized non-Parametric Regression”. In: *arXiv preprint arXiv:2503.14763v2*.
- Pappert, Sven and Antonia Arsova (2023). “Forecasting Natural Gas Prices with Spatio-Temporal Copula-Based Time Series Models”. In: Valenzuela, O., Rojas, F., Herrera, L.J., Pomares, H., Rojas, I. (eds) *Theory and Applications of Time Series Analysis. ITISE 2022. Contributions to Statistics*. Springer. Cham., pp. 221–236.
- Pappert, Sven and Harry Joe (2026). “Copula-Based Time Series for Non-Gaussian and Non-Markovian Stationary Processes”. In: *arXiv preprint arXiv:2604.01500v1*.
- Patton, Andrew (2013). “Copula Methods for Forecasting Multivariate Time Series”. In: *Handbook of Economic Forecasting* 2, pp. 899–960.
- Patton, Andrew J (2009). “Copula-Based Models for Financial Time Series”. In: *Handbook of Financial Time Series*. Springer, pp. 767–785.
- Quaschnig, Volker (2014). *Understanding Renewable Energy Systems*. Routledge.
- Reboredo, Juan C (2014). “Volatility spillovers between the oil market and the European Union carbon emission market”. In: *Economic Modelling* 36, pp. 229–234.
- Rindler, Filip (2018). *Calculus of variations*. Vol. 5. Springer.
- Rosasco, Lorenzo, Silvia Villa, Sofia Mosci, Matteo Santoro, and Alessandro Verri (2013). “Nonparametric Sparsity and Regularization”. In: *Journal of Machine Learning Research* 14, pp. 1665–1714.
- Rosenblatt, Murray (1952). “Remarks on a Multivariate Transformation”. In: *The Annals of Mathematical Statistics* 23.3, pp. 470–472.
- Rudin, Leonid I, Stanley Osher, and Emad Fatemi (1992). “Nonlinear total variation based noise removal algorithms”. In: *Physica D: Nonlinear Phenomena* 60.1-4, pp. 259–268.
- Schepsmeier, Ulf, Jakob Stoeber, Eike Christian Brechmann, Benedikt Graeler, Thomas Nagler, Tobias Erhardt, Carlos Almeida, Aleksey Min, Claudia Czado, Mathias Hofmann, et al. (2015). “Package ‘vinecopula’”. In: *R package version* 2.5.
- Schervish, Mark J (2012). *Theory of Statistics*. New York, NY: Springer.
- Schmidt-Hieber, Johannes (2020). “Nonparametric regression using deep neural networks with ReLU activation function”. In: *The Annals of Statistics* 48.4, pp. 1875–1897.
- Shen, Xiaoxi and Jinghang Lin (2022). “Consistency of neural networks with regularization”. In: *arXiv preprint arXiv:2207.01538*.
- Shumway, Robert H, David S Stoffer, and David S Stoffer (2000). *Time series analysis and its applications*. Vol. 3. Springer.

- Simard, Clarence and Bruno Rémillard (2015). “Forecasting time series with multivariate copulas”. In: *Dependence Modeling* 3.1, pp. 59–82.
- Sims, Christopher A (1980). “Macroeconomics and reality”. In: *Econometrica* 48.1, pp. 1–48.
- Sklar, M (1959). “Fonctions de repartition an dimensions et leurs marges”. In: *Publ. inst. statist. univ. Paris* 8, pp. 229–231.
- Smith, Michael, Aleksey Min, Carlos Almeida, and Claudia Czado (2010). “Modeling Longitudinal Data Using a Pair-Copula Decomposition of Serial Dependence”. In: *Journal of the American Statistical Association* 105.492, pp. 1467–1479.
- Smith, Michael Stanley (2015). “Copula modelling of dependence in multivariate time series”. In: *International Journal of Forecasting* 31.3, pp. 815–833.
- Smith, Michael Stanley (2021). “Implicit copulas: An overview”. In: *Econometrics and Statistics*.
- Spanhel, Fabian (2015). “A copula-based approach to model serial dependence in financial time series”. PhD thesis. LMU.
- Spanhel, Fabian and Malte S Kurz (2016). “The partial copula: Properties and associated dependence measures”. In: *Statistics & Probability Letters* 119, pp. 76–83.
- Stone, Charles J (1982). “Optimal Global Rates of Convergence for Nonparametric Regression”. In: *The Annals of Statistics*, pp. 1040–1053.
- Svetunkov, Ivan, Huijing Chen, and John E Boylan (2023). “A new taxonomy for vector exponential smoothing and its application to seasonal time series”. In: *European Journal of Operational Research* 304.3, pp. 964–980.
- Székel, Gábor J and Maria L Rizzo (2013). “Energy statistics: A class of statistics based on distances”. In: *Journal of Statistical Planning and Inference* 143.8, pp. 1249–1272.
- Tao, Siyang (2025). “On Tridiagonal Correlation Matrices”. In: *Pan-American Journal of Mathematics* 4, 15.
- Tibshirani, Robert (1996). “Regression shrinkage and selection via the Lasso”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1, pp. 267–288.
- Tol, Richard SJ (2019). “A social cost of carbon for (almost) every country”. In: *Energy Economics* 83, pp. 555–566.
- Torres, Alejandro, Antonio Marquina, José A Font, and José M Ibáñez (2014). “Total-variation-based methods for gravitational wave denoising”. In: *Physical Review D* 90.8, 084029.
- Trabelsi, Nader and Aviral Kumar Tiwari (2023). “CO2 emission allowances risk prediction with GAS and GARCH models”. In: *Computational Economics* 61.2, pp. 775–805.
- Wahba, Grace (1990). *Spline models for observational data*. SIAM.

- Wooldridge, Jeffrey M (1994). “Estimation and inference for dependent processes”. In: *Handbook of Econometrics* 4, pp. 2639–2738.
- Yuan, Ming and Yi Lin (2006). “Model selection and estimation in regression with grouped variables”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68.1, pp. 49–67.
- Zhou, Yefan, Jianxu Liu, Jirakom Sirisrisakulchai, and Songsak Sriboonchitta (2020). “Measurements of the Conditional Dependence Structure Among Carbon, Fossil Energy and Renewable Energy Prices: Vine Copula Based GJR-GARCH Model”. In: *International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making*. Springer, pp. 322–334.
- Zou, Hui (2006). “The Adaptive Lasso and Its Oracle Properties”. In: *Journal of the American Statistical Association* 101.476, pp. 1418–1429.

Contributed Articles

The articles on which this cumulative dissertation is based on are appended as pre-prints or accepted manuscripts. Published articles have been reused with the permission of the copyright holder. Note that the version of the fifth paper included here is a version that was updated after submission of the thesis. The updates were purely editorial.

Forecasting Natural Gas Prices with Spatio-Temporal Copula-Based Time Series Models



Sven Pappert and Antonia Arsova

Abstract In this work, we model and forecast commodity price time series using multivariate copula-based time series models. In particular, we consider the Gaussian copula, the t-copula and D-vine copulas to model both the contemporaneous and the temporal dependencies of the series. We focus on daily short-term gas, coal, oil and carbon emission futures in the period from March 2010 to February 2021. Further, we examine how the copula-based model gives rise to non-elliptical conditional probabilistic forecasts. During volatile periods, the conditional forecast density can become bimodal. This can be problematic when not only a distributional but also a mean forecast is desirable. Because for such probabilistic forecasts, it is not clear what represents a good mean forecast. A possible solution is to augment the forecasts by an artificial neural network that predicts the best quantile to use as mean forecast based on past values and past best quantiles. In a forecasting study, the predictive performances of the models are examined and compared with a benchmark. The distributional forecasts are examined by the CRPS. The mean forecasts are examined by the RMSE. The copula-based models are competitive with benchmark models.

Keywords Copula · Dependence modeling · Deep neural network · Energy forecasting · Probabilistic forecasting

1 Introduction

Modeling and forecasting commodity prices is important for trading and political decision-making. In this work, we focus on modeling and forecasting short-term natural gas future prices jointly with related commodity prices. We model the time

S. Pappert (✉)

Department of Statistics, TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany

e-mail: pappert@statistik.tu-dortmund.de

A. Arsova

Department of Statistics, TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany & RWI - Leibniz Institute for Economic Research, Hohenzollernstr. 1-3, 45128 Essen, Germany

e-mail: arsova@statistik.tu-dortmund.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

221

O. Valenzuela et al. (eds.), *Theory and Applications of Time Series Analysis*, Contributions to Statistics, https://doi.org/10.1007/978-3-031-40209-8_15

series jointly to exploit the additional information carried by their mutual dependence. To this aim, we employ spatio-temporal copula-based time series models.

Copulas are popular choices to model the cross-sectional dependence in multivariate time series with the copula-GARCH approach in financial markets [16, 18], as well as in energy markets [2, 7]. In the copula-GARCH approach, the temporal dependence of each time series is modeled by univariate time series models, such as ARMA-GARCH. The cross-sectional dependence structure of the time series can be captured by the copula of the standardized residuals of the univariate time series model. However, such models are only able to allow for flexible dependence structures in the cross-sectional dimension. The mean process is modeled linearly.

On the other hand, temporal copula modeling (or ‘copula-based time series modeling’) as well as spatio-temporal copula time series modeling offers an alternative to classical linear time series approaches. The models are able to flexibly model contemporaneous dependencies across multiple time series, as well as temporal dependencies. There is emerging literature on the topic. Chen and Fan [9] investigate the estimation of univariate copula-based semiparametric time series models. The authors provide conditions for β -mixing and prove consistency as well as asymptotic normality of the model parameters. Beare [4] further investigates mixing conditions. Smith et al. [24] decompose the serial dependence of intraday electricity load using pair copula constructions. Simard and Rémillard [22] investigate the forecasting performance of the spatio-temporal t-copula, dependent on the strength and structure of the dependence as well as the marginal distributions. Loazia-Maya et al. [19] explore heteroskedasticity modeling using copulas. Beare and Seo [5] as well as Nagler et al. [20] examine spatio-temporal vine copula models.

Examples where flexible dependence modeling can be important are the following. The cross-sectional dependence between international stock markets can be asymmetric with dominant lower tail dependence, indicating the phenomenon of contagion in financial markets [16]. With regard to energy markets, it was found by Aloui et al. [2] that crude oil and gas markets comove rather during bullish periods, thereby also displaying an asymmetric cross-sectional dependence structure. A concise example for a possible emergence of non-linear dependence structures in the temporal domain and hence requiring sophisticated dependence modeling is given in the introduction of the work by Beare [4]: The continuous growth of financial time series contrasted with their sudden and quick decrease represents an asymmetric temporal relation. Thus, cross-sectional as well as temporal dependence modeling can be important in many fields.

In this paper, we explore the performance of spatio-temporal copula models for modeling energy commodity prices. The basic idea underlying spatio-temporal time series modeling with copulas is a decomposition of the joint distribution. Using Sklar’s theorem [23], the joint distribution of consecutive observations is decomposed into dependence and marginal structure, $F_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{a}, \mathbf{b}) = C[F_{\mathbf{X}_t}(\mathbf{a}), F_{\mathbf{X}_{t-1}}(\mathbf{b})]$. Various copula specifications can be employed. In this work, the t-copula (discussed in detail in [11]) and the Gaussian copula are considered. Vine copula models [1, 10] are also considered. Spatio-temporal forecasting with the t-copula was examined in

[22]. The spatio-temporal vine copula modeling of multivariate time series is, among others, explored in [5, 20, 24].

By employing conditional copulas, the models can be used for forecasting. The resulting probabilistic forecasts can be non-elliptical. It is not obvious what constitutes a sensible mean forecast in this case. The expectation value is a sensible mean forecast for elliptical or almost elliptical probabilistic forecasts. For non-elliptical forecasts, e.g. a bimodular probabilistic forecast, the expectation value predicts points that are unlikely. One possible solution to this problem would be to take the mode of the probabilistic forecast as the mean forecast. Another possibility is to augment the forecasting procedure by an artificial neural network (ANN).¹ The ANN predicts which quantile of the probabilistic forecast is optimal (in MSE sense) as mean forecast. The inputs of the ANN are past values of the times series and the last optimal quantiles. One advantage of the ANN-augmented forecast is that the ANN can be estimated and used for prediction completely independent from the probabilistic time series model used for forecasting. In this approach, the ANN mean forecasts are also equipped with an underlying probabilistic distribution, enabling the calculation of confidence intervals and other distributional properties. The models' performances are examined in a forecasting study. The ARMAX-AVTGARCH model, closely related to the model from [6], which was shown to outperform other popular models, is considered as a benchmark. We find that the spatio-temporal copula models with ANN-augmented mean forecasts are indeed competitive for natural gas and related commodity price forecasting.

The main contributions of this paper are two-fold. The first contribution is the application-oriented exploration of spatio-temporal copula-based time series models for predicting energy commodity prices. We evaluate the performance of the models in a forecasting study and find that they perform well. The second contribution is the methodological exploration of optimal mean forecasting from non-elliptical probabilistic forecasts by means of ANNs.

In the next Section, Sect. 2, the data used in this work is introduced briefly. Section 3 describes the statistical methods used in this work. The empirical results of the forecasting study are presented in Sect. 4, while Sect. 5 summarizes the results and outlines some avenues for future research.

2 Data Description

In this work, we analyze month-ahead natural gas futures (NGas) from the Netherlands (TTF Hub). The related commodities used for modeling are short-term carbon emission futures (CEF), short-term brent oil futures (oil) and short-term coal futures (coal). The analyzed time series are comprised of daily observations. The observation period spans from March 2010 to February 2021. In total, each time series comprises 2861 observations. Missing values, which occur especially during the holidays, are

¹ We refer to [14] for a concise introduction.

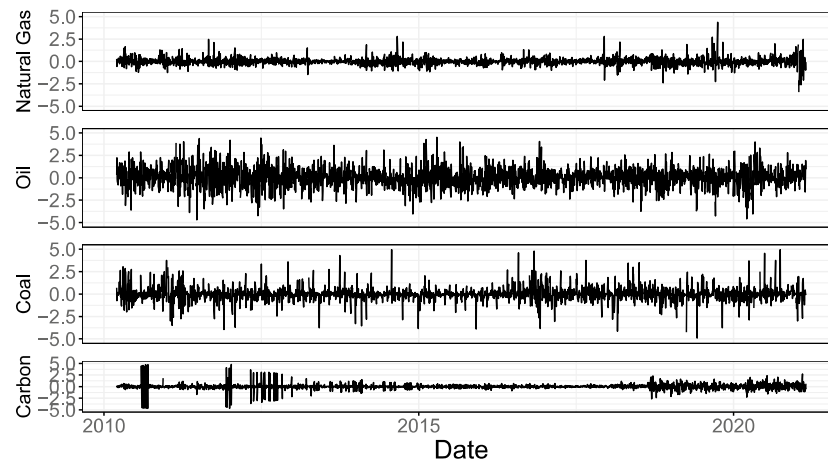


Fig. 1 First differences of the respective commodity price time series

trivially imputed as the last known value. The original time series are non-stationary. To obtain stationarity, which is necessary for the methods used in this report, the time series are differenced once. The differenced time series are displayed in Fig. 1. The hypothesis of non-stationarity in the first differences is rejected by the Dickey-Fuller test at the level $\alpha = 0.01$ for all time series.²

3 Statistical Methods

This section describes the methods used in this study. First copulas and related notions are introduced in Sect. 3.1. The copula specifications used in the analysis are presented. In particular vine copulas are presented in Sect. 3.2. The application of copulas to time series modeling is discussed in Sect. 3.3. The emergence of non-elliptical probabilistic forecasts is examined with regard to the t-copula in Sect. 3.4. It is shown how the t-copula may be used to model conditional heteroskedasticity. We note that Loazia-Maya et al. [19] also explored heteroskedasticity modeling based, among others, on the t-copula. The need for new mean forecast methods is presented.

3.1 Copulas

Copulas are distribution functions on the unit cube with uniform marginals:

² Results are omitted for brevity, but available upon request.

$$C : [0, 1]^d \rightarrow [0, 1]. \quad (1)$$

Copulas gain their relevance by Sklar's theorem [23]. It states that every multivariate distribution can be decomposed into a copula and marginal distributions. Let $X_1, \dots, X_d$ be real-valued random variables with joint distribution $F_{X_1, \dots, X_d}$ and marginal distributions $F_{X_1}, \dots, F_{X_d}$. Then it holds that there exists a copula C such that

$$F_{X_1, \dots, X_d}(x_1, \dots, x_d) = C_{U_1, \dots, U_d} [F_{X_1}(x_1), \dots, F_{X_d}(x_d)], \quad (2)$$

where $(U_1, \dots, U_d) := (F_{X_1}(X_1), \dots, F_{X_d}(X_d))$. In the following, the indices of the copula will be dropped. If the random variables $X_1, \dots, X_d$ are continuous, then the decomposition is unique [17]. The pseudo-observations $(u_1, \dots, u_d) := F_{X_1}(x_1), \dots, F_{X_d}(x_d)$ are, by virtue of the probability integral transformation, realizations from a uniform distribution, $U_i = F_{X_i}(X_i) \sim U[0, 1], i \in \{1, \dots, d\}$ [3]. This permits the copula to be interpreted as the dependence structure of the random variables $X_1, \dots, X_d$. The copula density c , which couples the joint density $f_{X_1, \dots, X_d}$ and marginal densities $f_{X_1}, \dots, f_{X_d}$, can be derived directly from Eq. 2 by taking derivatives (if the copula is differentiable) :

$$f_{X_1, \dots, X_d}(x_1, \dots, x_d) = c [F_{X_1}(x_1), \dots, F_{X_d}(x_d)] f_{X_1}(x_1) \dots f_{X_d}(x_d), \quad (3)$$

$$c[u_1, \dots, u_d] = \frac{\partial^d C[u_1, \dots, u_d]}{\partial u_1 \dots \partial u_d}. \quad (4)$$

The copula density is important for estimation via maximum likelihood as well as for the visualization of dependence structures. In this paper, the copula density will also be used to introduce the notion of vine copula models. Another important notion for dependence modeling is the conditional copula of $U_1, \dots, U_i$ given $U_{i+1}, \dots, U_d$, along with the respective conditional copula density. The conditional copula (density) can also be derived from Eq. 2. It is given as [22],

$$C[u_1, \dots, u_i | u_{i+1}, \dots, u_d] = \frac{\partial_{u_{i+1}} \dots \partial_{u_d} C[u_1, \dots, u_d]}{c[u_{i+1}, \dots, u_d]}, \quad (5)$$

$$c[u_1, \dots, u_i | u_{i+1}, \dots, u_d] = \frac{c[u_1, \dots, u_d]}{c[u_{i+1}, \dots, u_d]}. \quad (6)$$

Conditional copulas are especially relevant for conditional time series models as presented in this paper. The relation between the conditional density and the copula is as follows:

$$f_{X_1, \dots, X_i | X_{i+1} \dots X_d}(x_1, \dots, x_i | x_{i+1}, \dots, x_d) = \frac{c[F_{X_1}(x_1), \dots, F_{X_d}(x_d)]}{c[F_{X_{i+1}}(x_{i+1}), \dots, F_{X_d}(x_d)]} \quad (7)$$

$$\times f_{X_1}(x_1) \dots f_{X_i}(x_i).$$

The copula approach to multivariate modeling allows for separate modeling of marginal properties and dependence structure. This feature renders the approach far more flexible than standard multivariate modeling. Joint distributions such as the multivariate normal or Students t-distribution restrict the choice of marginal distributions. In the copula approach, marginal distributions can be arbitrary. Also the dependence structure of random variables can exhibit various features that have to be accounted for by choosing an appropriate copula specification. In this work, the Gaussian, Clayton, Gumbel and t-copula are utilized. In the following, they will be introduced briefly as the joint distribution of random variables $U_i \sim U[0, 1]$, $i \in \{1, \dots, d\}$. The Gaussian copula is a popular choice for modeling linear dependence structures. The Gaussian copula is constructed by extracting the dependence structure of the multivariate normal distribution and filtering the marginal influences [17],

$$C^{\text{Gaussian}}[u_1, \dots, u_d] = \Phi_{\Sigma}[\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)], \quad (8)$$

where ϕ is the cumulative distribution function of the standard normal distribution and Φ_{Σ} is the d -variate cumulative distribution function of the normal distribution with correlation matrix Σ . The correlation matrix $\Sigma \in [-1, 1]^{d \times d}$ contains $\frac{d(d-1)}{2}$ dependence parameters, $\rho_1, \dots, \rho_{\frac{d(d-1)}{2}}$, governing the linear dependencies among the random variables $U_1, \dots, U_d$. The density of a bivariate Gaussian copula with dependence parameter $\rho = 0.4$ is displayed in the upper left panel of Fig. 2. The density only displays a linear relation between the variables. Similar to recovering linear dependence structures from the multivariate normal distribution, heavy-tailed dependence structures can be recovered from the multivariate t-distribution using the t-copula

$$C^t[u_1, \dots, u_d] = t_{\Sigma, \nu}[t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_d)], \quad (9)$$

where t_{ν} is the cumulative distribution function of the t-distribution with degree of freedom ν and where $t_{\Sigma, \nu}$ is the cumulative distribution function of the multivariate t-distribution with correlation matrix Σ and degree of freedom ν . Incorporating the degree of freedom $\nu \in (2, \infty)$ permits heavy-tailed dependence structures. The heavy-tailedness can be interpreted as a higher probability of extreme events coinciding. A lower degree of freedom ν implies heavier tails. The density of a bivariate t-copula with dependence parameter $\rho = 0.4$ and degree of freedom $\nu = 4$ is displayed in the upper right panel of Fig. 2. The density displays the linear relation between the variables as well as the coincidence of extreme events. Another class of dependence structures can be described as asymmetric dependence structures. Two relevant copulas are the Gumbel and Clayton copula. The Gumbel copula exhibits dominant upper tail dependence while the Clayton copula exhibits dominant lower tail dependence. Their bivariate densities are displayed in the lower left, respectively

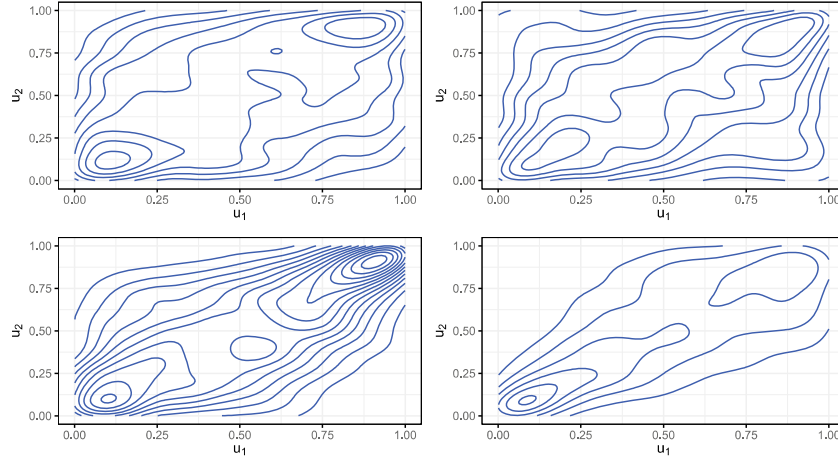


Fig. 2 Simulated density plots of four different two-dimensional copula specifications. The upper left plot shows the Gaussian copula density with dependence parameter set to $\rho = 0.4$. Upper right shows the t-copula density with dependence parameter $\rho = 0.4$ and degree of freedom $\nu = 4$. The lower left plot displays the Gumbel copula density with dependence parameter $\theta = 2$. The lower right plot displays the Clayton copula density with dependence parameter $\theta = 2$. All plots were created from simulations with 2000 samples

lower right panel of Fig. 2. Both copulas are part of the Archimedean copula family. They are constructed as $C[u_1, \dots, u_d] = \Psi^{-1}(\Psi(u_1) + \dots + \Psi(u_d))$ with a suitable generator function Ψ [12, 15]. The generators for the Gumbel, respectively Clayton copula, are given by

$$\Psi^{\text{Clayton}}(t) = (1 + t)^{-\frac{1}{\theta}}, \quad (10)$$

$$\Psi^{\text{Gumbel}}(t) = e^{-t^{\frac{1}{\theta}}}. \quad (11)$$

The dominant lower tail dependence of the Clayton copula can be interpreted as lower tail events coinciding more often than upper tail events and vice versa for the dominant upper tail dependence of the Gumbel copula. An even more flexible copula model is the vine copula model. Vine copula models will be explained next.

3.2 Vine Copulas

Vine copulas are special pair copula constructions. The idea of pair copula constructions is to decompose a d -variate dependence structure into a product of bivariate copulas. The joint density of d random variables can, by virtue of the law of total

probability, be decomposed into a product of conditional densities. Using the relation between conditional densities and copula densities (Eq. 7), one possible decomposition of a d -variate copula density can be derived as [1, 10]

$$c[u_1, \dots, u_d] = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c[u_i, u_{i+j} | u_{i+1}, \dots, u_{j-1}], \quad (12)$$

where indices of the copula densities were dropped again. Furthermore, we neglect the influence of conditioning on the copula density. In the literature, this model then is called simplified vine [20]. The decomposition is not unique. The decomposition in Eq. 12 is called drawable vine (D-vine). The unconditional copulas in the product all capture the dependence structure of neighboring variables, e.g. $c[u_i, u_{i+1}]$, $c[u_{i+1}, u_{i+2}]$ and so forth. In a graphical representation, the connection between the variables resembles a straight line, hence the name D-vine.³ The vine copula approach allows for flexible dependence modeling. It is advantageous in contexts where the bivariate dependence structures between variables can take different forms, e.g. the dependence structure between variables U_1 and U_2 may be linear whereas the dependence structure between U_2 and U_3 may be heavy-tailed and so forth. Vine copula models can be estimated by maximum likelihood; we refer to [1] for details.

3.3 Modeling Time Series with Spatio-Temporal Copulas

So far, our discussion on copulas referred to modeling dependence in a general multivariate setting. In this subsection, the copula-based time series models are reviewed and summarized. Further, it is explained how these models can be used for forecasting. First, the temporal copula modeling of univariate time series (see, for example, [4, 9]) is introduced. Next, the copula-based modeling of multivariate time series, the spatio-temporal copula modeling, is discussed. The exposition on the spatio-temporal t-copula is based on [22], while for vine copula modeling we refer to [5, 8, 20, 24]. It will be examined how conditional temporal copula models offer an alternative approach to conditional heteroskedasticity modeling, which was also explored in [19]. The emergence of non-elliptical conditional probabilistic forecasts will be exemplified. The consequences for forecasting and the need for new mean forecasting methods will be discussed.

Let X_t be a univariate stationary Markov(1) time series. The temporal evolution of the time series is completely specified by the joint distribution of random variables from consecutive time points, i.e. $F_{X_t, X_{t-1}}$. Using Sklar's theorem (Eq. 2), the joint

³ Another special class of decompositions are canonical vines (C-vine). In this decomposition, the unconditional dependence structures are all centered around one variable, e.g. $c[u_i, u_{i+1}]$, $c[u_i, u_{i+2}]$, $\dots$. In a graphical representation, the unconditional connection between variables resembles a star. In this study only D-vines are used.

distribution can be decomposed into copula and marginal distributions as

$$F_{X_t, X_{t-1}}(a, b) = C[F_{X_t}(a), F_{X_{t-1}}(b)]. \quad (13)$$

By the stationarity of X_t , $F_{X_t} = F_{X_{t-1}} =: F_X$. Hence, the model can be determined by choosing an appropriate marginal distribution F_X and an appropriate copula specification. Note that the marginal distribution F_X is the unconditional distribution of X_t . Conditional properties of the time series are completely determined by the conditional copula. The conditional density of $X_t|X_{t-1}$ is given by

$$f_{X_t|X_{t-1}}(a|b) = c[F_X(a), F_X(b)] f_X(a). \quad (14)$$

Hence, for forecasting, the conditional density of $(X_t|X_{t-1} = x_{t-1})$ can be used as a probabilistic forecast. This model can be understood as a generalization of the AR(1) model [24].⁴ The Gaussian autoregressive model can be recovered by choosing $C = C^{\text{Gaussian}}$ and $F_X = \Phi$. When allowing other dependence structures, any temporal dependency representable by a copula can be reproduced. The concept of the copula-based time series models can be further illustrated by its conditional model equation:

$$X_t|(X_{t-1} = x_{t-1}) = F_X^{-1}(C^{-1}[w_t|F_X(x_{t-1})]), \quad w_t \stackrel{iid}{\sim} U[0, 1]. \quad (15)$$

In this formulation, the non-linear connection between X_t and X_{t-1} becomes obvious. The generalization of the temporal copula time series model to d -variate time series, hence spatio-temporal time series models, is straightforward. Let $\mathbf{X}_t = (X_{1,t}, \dots, X_{d,t})$ be a stationary Markov(1) time series. The structure of the time series is completely captured by the joint distribution of $\mathbf{X}_t$ and $\mathbf{X}_{t-1}$,

$$F_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{a}, \mathbf{b}) = C[F_{X_1}(a_1), \dots, F_{X_d}(a_d), F_{X_1}(b_1), \dots, F_{X_d}(b_d)]. \quad (16)$$

The conditional density, given the observations from time point $(t-1)$, is as follows:

$$f_{\mathbf{X}_t|\mathbf{X}_{t-1}}(\mathbf{a}|\mathbf{b}) = \frac{c[F_{X_1}(a_1), \dots, F_{X_d}(a_d), F_{X_1}(b_1), \dots, F_{X_d}(b_d)]}{c[F_{X_1}(b_1), \dots, F_{X_d}(b_d)]} \times f_{X_1}(a_1) \cdots f_{X_d}(a_d). \quad (17)$$

To sample from the conditional distribution, as necessary for Monte Carlo approximations of conditional forecasts, the following procedure as in [22] is employed. First transform the observations at time $t-1$ to pseudo-observations. This is done by applying the probability integral transform to the observations, $(u_{t-1,1}, \dots, u_{t-1,d}) := (F_{X_1}(x_{t-1,1}), \dots, F_{X_d}(x_{t-1,d}))$. Then sample n d -dimensional realizations from

⁴ The generalization to $AR(p)$ models can be achieved by permitting the time series to be a Markov(p) process.

the conditional copula, Eq. 5 (Details on how to sample from the t-copula can be found in [22]. Details to sampling from vine copulas can be found in [1]). At last, the n realizations have to be quantile-transformed with their respective marginal distribution, yielding the n samples of the conditional distribution.

For our empirical study, namely forecasting energy commodity prices, we consider the following models: First, the spatio-temporal time series model where the copula is specified as the Gaussian copula (Eq. 8). The marginals are approximated non-parametrically by the empirical distribution:

$$F_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{a}, \mathbf{b}) = \Phi_{\Sigma}[\phi^{-1}(F_{X_1}^{\text{emp}}(a_1)), \dots, \phi^{-1}(F_{X_d}^{\text{emp}}(a_d)), \phi^{-1}(F_{X_1}^{\text{emp}}(b_1)), \dots, \phi^{-1}(F_{X_d}^{\text{emp}}(b_d))]. \quad (18)$$

This model is sensible to use when the dependence structure between the variables as well as the temporal dependence is linear. When the dependence structure exhibits heavy-tailedness, the spatio-temporal t-copula model with non-parametric marginals poses a viable alternative, which we consider as well,

$$F_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{a}, \mathbf{b}) = t_{\nu, \Sigma}[t_{\nu}^{-1}(F_{X_1}^{\text{emp}}(a_1)), \dots, t_{\nu}^{-1}(F_{X_d}^{\text{emp}}(a_d)), t_{\nu}^{-1}(F_{X_1}^{\text{emp}}(b_1)), \dots, t_{\nu}^{-1}(F_{X_d}^{\text{emp}}(b_d))]. \quad (19)$$

Third, to allow for more flexibility, we utilize the spatio-temporal D-vine copula with non-parametric marginals. For convenience, the model is presented in terms of its joint density and with variables $(\mathbf{a}, \mathbf{b}) =: \mathbf{p}$,

$$f_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{p}) = \prod_{j=1}^{2d-1} \prod_{i=1}^{2d-j} c[F_{X_i}(p_i), F_{X_{i+j}}(p_{i+j}) | F_{X_{i+1}}(p_{i+1}), \dots, F_{X_{j-1}}(p_{j-1})] \times f_{X_1}(p_1) \cdot \dots \cdot f_{X_d}(p_d) f_{X_1}(p_{d+1}) \cdot \dots \cdot f_{X_d}(p_{2d}). \quad (20)$$

This model is sensible to use when the dependence between variables differs in its structure or when the temporal dependence differs from the cross-sectional dependence.

Fourth, applicable solely for temporal modeling, the temporal t-copula is employed,

$$F_{X_t, X_{t-1}}(a, b) = t_{\nu, \Sigma}[t_{\nu}^{-1}(F_X^{\text{emp}}(a)), t_{\nu}^{-1}(F_X^{\text{emp}}(b))]. \quad (21)$$

The heavy-tailed temporal dependence that this model exhibits is suitable for conditional heteroskedasticity modeling as discussed in [19]. This, and the emergence of non-elliptical probabilistic forecasts, is what we focus on next.

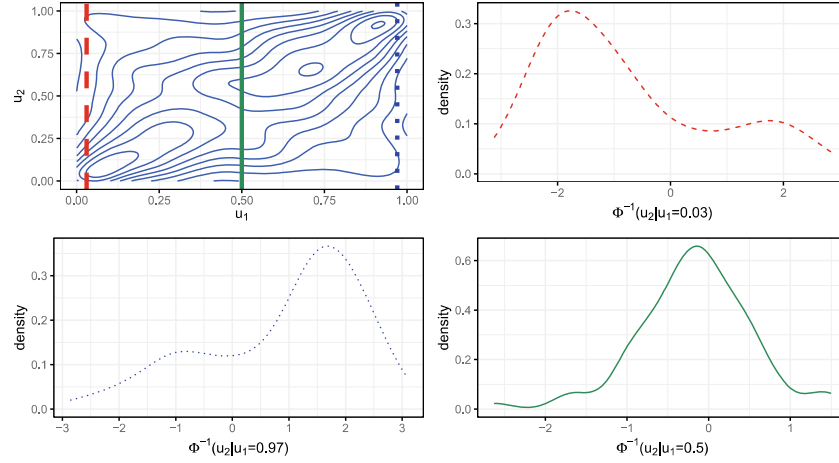


Fig. 3 Visualization of the conditional density structure depending on the value of the conditioning variable. The underlying model assumes the t-copula with dependence parameter $\rho = 0.4$ and degree of freedom $\nu = 2.1$. The marginal distribution is assumed as the standard normal distribution. The upper left panel shows the copula density of 2000 realizations of the before mentioned t-copula. The three lines indicate the three cases where the conditional density is examined. The conditional density is calculated by collecting all values in the neighborhood ($u_1 \pm 0.025$) of the conditioning variable and quantile-transforming them. The density in the upper right panel displays the conditional density given $u_1 = 0.03$. The lower panels display the conditional densities given $u_1 = 0.5$, respectively $u_1 = 0.97$

3.4 Emergence of Non-elliptical Probabilistic Forecasts

The conditional distributions, respectively probabilistic forecasts from (spatio)-temporal copula time series models, can be non-elliptical because of non-linear influences of the conditioning variable. In the following, the behavior of the conditional distributions will be examined with regard to the temporal t-copula with standard normal marginal distribution.⁵ The emergence of non-elliptical conditional densities from the heavy-tailed t-copula is visualized in Fig. 3. When the conditioning variable takes moderate values around $u_1 = 0.5$, the resulting conditional density is approximately elliptical. However, when the conditioning variable takes extreme values, e.g. $u_1 = 0.03$ and $u_1 = 0.97$, the conditional density becomes bimodal. Thus, depending on the value of the conditioning variable, the resulting conditional density can have fundamentally different structures. This behavior offers a new approach to conditional heteroskedasticity modeling. Instead of widening the conditional density as in GARCH models, the density gets bimodal. This can be viewed as a sensible approach to volatility because the extreme behavior in volatile phases is mirrored in this model: When the time series takes a very low value at time point $(t - 1)$, it

⁵ The choice of the standard normal distribution is just for convenience. The example would still be valid with other marginal distributions, e.g. students t-distribution.

can be expected that the value at time point t will either be also very low or very high. The variance at time point t is increased nevertheless, but the mechanism for the increase in variance is a new one.

The temporal t-copula approach to conditional volatility, however, holds a problem. When the conditional density is non-elliptical, it is not clear what constitutes a sensible mean forecast. The expectation value may not be suitable in extreme cases where the conditional density is bimodal because the expectation value will take a value which is less probable than, e.g. the modes. Taking the mode as mean forecast could be a solution. Another possible solution to the problem of mean forecasting is to augment the forecast by an artificial neural network (ANN). The ANN predicts which quantile of the conditional distribution is best (in terms of MSE) to use as mean forecast. The inputs of the ANN are past values of the time series and the last optimal quantiles. The ANN architecture used in this work is the basic multi-layer perceptron (MLP) structure. We refer to [14] for an introduction to the topic.

4 Forecasting Study

This section comprises the results of an expanding window forecasting study, investigating the one-step-ahead forecasting performance of different models. The first 1000 observations are used as initial training data set. The following models are considered for evaluation.

- (1) The temporal t-copula model with non-parametric marginals, Eq. 21, henceforth denoted by Tem-t,
- (2) The spatio-temporal D-vine copula model, Eq. 20, henceforth denoted by S-Tem D-vine,
- (3) The spatio-temporal t-copula model, Eq. 19, henceforth denoted by S-Tem-t,
- (4) The spatio-temporal Gaussian copula model, Eq. 18, henceforth denoted by S-Tem-Gaussian,
- (5) The autoregressive moving-average model with external regressors and absolute value, threshold generalized autoregressive conditional heteroskedasticity model, henceforth denoted by ARMAX-AVTGARCH (closely related to the newly proposed model of [6]).

The models' distributional forecasting performance is examined by the continuous ranked probability score (CRPS) [13]. Further, the ANN-assisted mean forecasts of the S-Tem D-vine model and the Tem-t model are compared with the mean forecasts from the ARMAX-AVTGARCH model by RMSE. For each time series, the ARMAX-AVTGARCH model is fitted individually. All models are estimated via Maximum Likelihood. However, the marginals of the copula models are estimated non-parametrically to avoid transmitting estimation errors caused by two-step estimation [21]. The order of the variables in the S-Tem D-vine copula model is fixed as

Table 1 Aggregated CRPS values of the competing models for their one-day-ahead probabilistic forecast for the four commodities. The CRPS is evaluated for the period 12/19/2013 to 02/23/2021, comprising 1861 observations

Model/commodity	S-Tem D-Vine	ARMAX- AVTGARCH	Tem-t	S-Tem-t	S-Tem- Gaussian
Natural gas	0.236	0.227	0.230	0.234	0.234
Oil	0.564	0.548	0.551	0.559	0.558
Coal	0.400	0.389	0.392	0.398	0.397
CEF	0.236	0.222	0.229	0.234	0.234

$$\text{CEF} - \text{coal} - \text{oil} - \text{NGas} - \text{NGas lag} - \text{oil lag} - \text{coal lag} - \text{CEF lag}. \quad (22)$$

This order is chosen to enable the lagged natural gas price to directly interact with the non-lagged natural gas price. The Gaussian, Gumbel, Clayton and t-copula are allowed as bivariate copulas in the D-vine decomposition (Eq. 12). The probabilistic forecasts of all models are approximated by Monte Carlo simulations with 1000 samples for each forecast. Table 1 displays the models' performances in terms of the CRPS. The ARMAX-AVTGARCH model performs best with regard to univariate distributional forecasting. However, the S-Tem D-vine model, the S-Tem-t and the Tem-t model are competitive. The performance of the copula models may be enhanced, when the marginal distributions are modeled parametrically. The empirical marginal distributions of the copula may not capture all marginal features of the time series. More versatile copula models could be used to enhance the forecast. The conditional dependence modeling may only be able to capture parts of the conditional effects.

The probabilistic forecasts from the Tem-t model during a volatile period are displayed in Fig. 4. During volatile times, the probabilistic forecasts are non-elliptical. During these times, the ANN-augmented mean forecasts can be valuable. The mean forecasting performance of the models can be found in Table 2. The evaluation starts at the 2001st observation, because the first 1000 probabilist forecasts are used to train the ANN. The hybrid, ANN-augmented S-Tem vine and the ANN-augmented Tem-t model generate the best mean forecasts. The mean forecasts of the ARMAX-AVTGARCH model are competitive though. Note that the ANN model used for forecasting is built according to the basic multi-layer perceptron architecture. It is not perfectly suitable for catching sequential patterns. Using recurrent neural networks, especially long short-term memory architectures could enhance the performance even more and could be subject to future research. Also incorporating a measure for the structure of the probabilistic forecast could enhance the performance. However, this would require more advanced architectures.

234

S. Pappert and A. Arsova

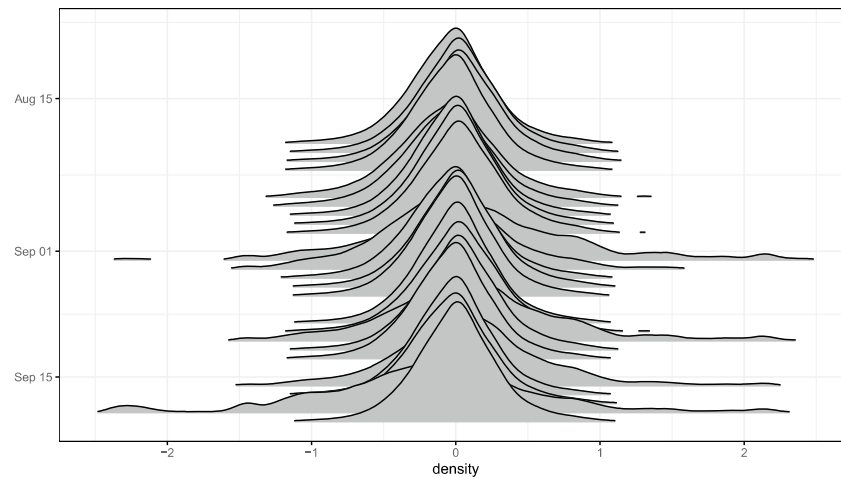


Fig. 4 Probabilistic forecasts for natural gas futures generated from the temporal t-copula model. The forecast densities can be non-elliptical during volatile times (August 2019–September 2019)

Table 2 Aggregated RMSE values of the competing mean forecasting procedures for the four commodities. The RMSE is evaluated for the period 10/23/2017 to 02/23/2021, comprising 861 observations

Model/commodity	S-Tem D-Vine ANN	ARMAX- AVTGARCH	S-Tem D-Vine mean	S-Tem D-Vine mode	Tem-t ANN
Gas	0.594	0.600	0.599	0.597	0.589
Oil	1.222	1.222	1.236	1.246	1.220
Coal	1.000	0.999	1.009	1.002	0.997
CEF	0.605	0.608	0.614	0.604	0.607

5 Conclusion

The application of copula-based time series models to natural gas and related commodity prices is explored in this work. An expanding window forecasting study is conducted. The time series comprises short-term future price series of natural gas, crude oil, coal and carbon emissions.

After introducing the basic notions of dependence modeling with copulas and the D-Vine copula, the copula-based time series models from the literature are reviewed. The emergence of non-elliptical probabilistic forecasts is exemplified using the temporal t-copula. It is visualized how the temporal t-copula offers an alternative approach to conditional heteroskedasticity modeling. It is not clear what constitutes a sensible mean forecast when the probabilistic forecast is non-elliptic. To this end, an artificial neural network is employed to predict what quantile of the probabilistic forecast is best to use as mean forecast. The inputs of the artificial neural network

are past values of the multivariate time series and the last best quantiles of the probabilistic forecast.

In a forecasting study, the predictive performance of the temporal t-copula, the spatio-temporal t-copula and the spatio-temporal D-Vine copula is examined. The marginal distributions are estimated by the respective empirical distribution. The performance is compared with the performance of an autoregressive moving-average model with external regressors and absolute value, threshold generalized autoregressive conditional heteroskedasticity model (ARMAX-AVTGARCH). A closely related model was recently shown to be the best model for natural gas forecasting. Hence, it is understood as a benchmark model. The distributional predictive performance is examined by the continuous ranked probability score (CRPS). We find that the copula-based time series models are competitive with the ARMAX-AVTGARCH model. The mean forecasts are evaluated by the root mean squared error (RMSE). The ANN-augmented mean forecasts perform best, although the forecasts from the ARMAX-AVTGARCH model are still competitive.

The performance of the copula-based time series models could be enhanced by modeling the marginal distributions parametrically. The non-parametric modeling may not catch all marginal features of the time series. However, this procedure requires the estimation to be conducted in one step to guarantee efficient estimation. Another possibility to enhance the performance is to consider more versatile copula models. The current modeling may not capture all conditional features of the time series. Another possibility, with regard to the vine copula model, is to consider other vine structures. In this work, the D-vine structure was imposed. Other structures may be able to capture the dependencies better. As for the mean forecasts, it was shown that the ANN-augmented forecasts perform well. Even though we choose to utilize the standard multi-layer perceptron architecture, which cannot model sequential information perfectly well, the precision was increased. Using more sophisticated architectures that are more suited to catch sequential information will be subject to future research. It would also be interesting to use other models to predict the best quantile for mean forecasting.

Acknowledgements The authors gratefully acknowledge the computing time provided on the Linux HPC cluster at Technical University Dortmund (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the German Research Foundation (DFG) as project 271512359.

References

1. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. *Insurance: Math. Econ.* **44**(2), 182–198 (2009)
2. Aloui, R., Aïssa, M.S.B., Hammoudeh, S., Nguyen, D.K.: Dependence and extreme dependence of crude oil and natural gas prices with applications to risk management. *Energy Econ.* **42**(332–243), 2014 (2014)
3. Angus, J.E.: The probability integral transform and related results. *SIAM Rev.* **36**(4), 652–654 (1994)
4. Beare, B.K.: Copulas and temporal dependence. *Econometrica* **78**(1), 395–410 (2010)

5. Beare, B.K., Seo, J.: Vine copula specifications for stationary multivariate Markov chains. *J. Time Ser. Anal.* **36**(2), 228–246 (2015)
6. Berrisch, J., Ziel, F.: Distributional modeling and forecasting of natural gas prices. *J. Forecast.* (2022)
7. Berrisch, J., Pappert, S., Ziel, F., Arsova, A.: Modeling volatility and dependence of European carbon and energy prices (2022). [arXiv:2208.14311](https://arxiv.org/abs/2208.14311)
8. Brechmann, E.C., Czado, C.: COPAR-multivariate time series modeling using the copula autoregressive model. *Appl. Stoch. Models Bus. Ind.* **31**(4), 495–514 (2015)
9. Chen, X., Fan, Y.: Estimation of copula-based semiparametric time series models. *J. Econom.* **130**(2), 307–335 (2006)
10. Czado, C.: Pair-copula constructions of multivariate copulas. In: *Copula Theory and its Applications*, pp. 93–109. Springer, Berlin, Heidelberg
11. Demarta, S., McNeil, A.J.: The t copula and related copulas. *Int. Stat. Rev.* **73**(1), 111–129 (2005)
12. Genest, C., Rivest, L.P.: Statistical inference procedures for bivariate Archimedean copulas. *J. Am. Stat. Assoc.* **88**(423), 1034–1043 (1993)
13. Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102**(477), 359–378 (2007)
14. Higham, C.F., Higham, D.J.: Deep learning: an introduction for applied mathematicians. *Siam Rev.* **61**(4), 860–891 (2019)
15. Hofert, M.: Sampling archimedean copulas. *Comput. Stat. Data Anal.* **52**(12), 5163–5174 (2008)
16. Hu, L.: Dependence patterns across financial markets: a mixed copula approach. *Appl. Financ. Econ.* **16**(10), 717–729 (2006)
17. Joe, H.: *Dependence Modeling with Copulas*. CRC Press (2014)
18. Jondeau, E., Rockinger, M.: The copula-Garch model of conditional dependencies: an international stock market application. *J. Int. Money Finance* **25**(5), 827–853 (2006)
19. Loaiza-Maya, R., Smith, M.S., Maneesoonthorn, W.: Time series copulas for heteroskedastic data. *J. Appl. Econom.* **33**(3), 332–354 (2018)
20. Nagler, T., Krüger, D., Min, A.: Stationary vine copula models for multivariate time series. *J. Econom.* **227**(2), 305–324 (2022)
21. Patton, A.: Copula methods for forecasting multivariate time series. In: *Handbook of Economic Forecasting*, vol. 2, pp. 899–960 (2013)
22. Simard, C., & Rémillard, B.: Forecasting time series with multivariate copulas. *Depend. Model.* **3**(1) (2015)
23. Sklar, M.: Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris* **8**, 229–231 (1959)
24. Smith, M., Min, A., Almeida, C., Czado, C.: Modeling longitudinal data using a pair-copula decomposition of serial dependence. *J. Am. Stat. Assoc.* **105**(492), 1467–1479 (2010)



Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Modeling volatility and dependence of European carbon and energy prices

Jonathan Berrisch^a, Sven Pappert^{b,*}, Florian Ziel^a, Antonia Arsova^{b,c}^a Chair of Environmental Economics, esp. Economics of Renewable Energy, University of Duisburg-Essen, Germany^b Chair of Econometrics, Department of Statistics, TU Dortmund University, Germany^c RWI – Leibniz Institute for Economic Research, Germany

ARTICLE INFO

JEL classification:

C21, Q41, Q5, C32, Q59, C58, G17

Keywords:

Carbon prices, Conditional volatility, Copula, Emission allowances, Energy markets, Forecasting, Multivariate modeling, Time series

ABSTRACT

We study the prices of European Emission Allowances (EUA), whereby we analyze their uncertainty and dependencies on related energy prices (natural gas, coal, and oil). We propose a probabilistic multivariate conditional time series model with a VECM-Copula-GARCH structure which exploits key characteristics of the data. Data are normalized with respect to inflation and carbon emissions to allow for proper cross-series evaluation. The forecasting performance is evaluated in an extensive rolling-window forecasting study, covering eight years out-of-sample. We discuss our findings for both levels- and log-transformed data, focusing on time-varying correlations, and in view of the Russian invasion of Ukraine.

1. Introduction

The European Union (EU) emissions trading system (ETS) allows polluters to trade emission allowances (European Emission Allowances, EUA). Their supply is fixed, while the demand is determined by the market participants. Understanding the underlying dynamics of this market is essential for policymakers, commodity trading, and sustainability planning. This motivates exploring the determinants of the EUA price and its statistical properties.

The EU-ETS was set up in 2005, making it the first ETS worldwide. Not much later, research picked up the market. The EU-ETS has significant implications for energy companies and the industrial sector. As a side-effect, the introduction of the EU-ETS has strengthened the link between gas and power prices, leading to undesirable geopolitical uncertainties in the natural gas price. More than fifteen years later, following the Russian invasion of Ukraine, the geopolitical uncertainties of natural gas supply are more pronounced than ever before. It is, therefore, plausible to assume that some dependence between energy assets and the EUA price exists. Early studies dealing with these relations support this view: Fezzi and Bunn (2009) employ a cointegrated VAR model and find gas prices to induce quick reactions of EUA prices, while electricity prices show a delayed reaction to shocks in EUA prices.

The dependence of EU-ETS on other energy markets (including coal, oil, natural gas, and electricity) has been analyzed more recently by Chevallier et al. (2019), who find a negative link between carbon prices and oil and gas prices. Meier and Voss (2020) also find that EUA prices are significantly influenced by fossil fuel prices, while Zhou et al. (2020) show evidence that the carbon price drives the price of renewable energies. Employing an artificial neural network, García and Jaramillo-Morán (2020) pronounce the effect of subjective economic and political decisions on the EUA price and reject the influence of the external variables considered (iron, electricity, and steel prices). Duan et al. (2021) find a negative impact of energy prices on EUA prices. The effect is asymmetric, i.e., it is more pronounced in the lower quantiles of the EUA distribution.

* Corresponding author.

E-mail addresses: jonathan.berrisch@uni-due.de (J. Berrisch), pappert@statistik.tu-dortmund.de (S. Pappert), florian.ziel@uni-due.de (F. Ziel), arsova@statistik.tu-dortmund.de (A. Arsova).

<https://doi.org/10.1016/j.frl.2022.103503>

Received 20 September 2022; Received in revised form 31 October 2022; Accepted 16 November 2022

Available online 21 November 2022

1544-6123/© 2022 Elsevier Inc. All rights reserved.

Some studies focus on forecasting EUA prices, which is particularly relevant for portfolio management. [Paolella and Taschini \(2008\)](#) look into forecastability and value-at-risk forecasting in the EU-ETS, employing an AR-GARCH model with generalized skewed student-t innovations. [Benz and Trück \(2009\)](#) use a Markov-switching framework including GARCH effects to model EUA prices, while [Trabelsi and Tiwari \(2022\)](#) compare the GARCH approach to Generalized Autoregressive Score models. Many of these studies find EUA prices to exhibit heavy tails, leverage effects, and asymmetries. Therefore student-t and skewed-student-t distributions are often used and have been shown to lead to superior performance in forecasting ([Trabelsi and Tiwari, 2022](#); [Benz and Trück, 2009](#); [Paolella and Taschini, 2008](#)).

Another strand of the literature deals with the volatility process of EUA prices and possible spillovers to other markets. [Dutta et al. \(2019\)](#) find structural breaks in the EUA volatility process. They show that the forecasting performance of GARCH models increases (while persistence decreases) when accounting for structural breaks. [Hanif et al. \(2021\)](#) investigate volatility spillovers between EUA prices and energy indices using copulas. Finally, [Reboredo \(2014\)](#) analyzes volatility spillovers between oil and EUA markets. Their results suggest leverage effects and volatility dynamics but do not confirm significant spillovers.

The brief literature review above highlights several key characteristics of EUA prices which have implications for modeling and forecasting. First, quite a few studies point towards a relationship to the prices of related commodities like oil, gas, and coal. Second, a normality assumption is likely to fail due to the excess kurtosis and possible skewness of the data. Third, there is evidence for volatility dynamics and even structural breaks.

In this paper, we focus on probabilistic multi-step ahead forecasting of EUA prices. Building upon the aforementioned findings of earlier studies, we propose a VECM-Copula-GARCH model to jointly model EUA, gas, oil and coal prices as a four-variate system. As a combination of three different components the model is able to account for possible long-run equilibrium relations between carbon and fuel prices, autoregressive effects, conditional heteroscedasticity in the individual series as well as for time-varying dependence structures between them in a simultaneous yet flexible way. In contrast to earlier combination models which are estimated in stages, we utilize a one-step estimation approach to improve efficiency.

The contributions of this paper are manifold. First and foremost, we jointly model EUA, oil, coal, and natural gas prices using a probabilistic approach. The obtained multivariate predictive distribution can be used for sampling future trajectories and therefore enables probabilistic multi-step forecasts which exploit the dependencies of EUA on other commodity markets. Second, we evaluate the proposed model in an extensive rolling-window forecasting study for both data in levels and after log-transformation, whose properties differ (e.g. variance, cointegration). Several strictly proper scoring rules are employed to evaluate the multivariate forecasts as well as the marginal forecasts. Moreover, we test for significant differences between the forecasts against competing standard benchmarks. Finally, we consider real price data, normalized by the carbon emissions of the respective commodity, to allow for scale-invariant evaluation.

The remainder of this paper is organized as follows. The data are presented in Section 2. Section 3 introduces the proposed VECM-Copula-GARCH model and discusses its one-step estimation by maximum likelihood (ML). The set-up for the forecasting study and forecast evaluation measures are outlined in Section 4, while Section 5 discusses the results. Finally, Section 6 concludes.

2. Data description and preliminary analysis

The data, plotted in [Fig. 1](#), comprise 3257 daily observations of short-term futures prices covering the period from March 15, 2010, until October 14, 2022, on four commodities, traded on the Intercontinental Exchange (ICE): European Allowances (carbon emissions), natural gas (Dutch TTF natural gas futures), coal (Rotterdam coal futures), and Brent crude oil. Oil and coal prices, originally given in USD, have been converted from USD into EUR. All depicted price series have been standardized by the CO₂ emissions of the respective commodity (for conversion factors see [Quaschnig \(2016\)](#)). This means that the emission-adjusted prices reflect one tonne of CO₂ emissions for each of the three commodities. EUA prices needed no adjustment since they already reflect one tonne of CO₂ emissions. Furthermore, we adjusted for inflation by Eurostat's HICP, excluding energy.¹ Thus the general price increase over time is accounted for without dampening the commodities' price volatility which is to be explained by the model.

Raw data have been transformed by taking natural logarithms following the literature convention to stabilize their variance. Unit root testing by the augmented Dickey-Fuller (ADF) test points to non-stationarity in the levels and logarithms of all four series, respectively, and stationarity in their first differences.² As for cointegration, Johansen's likelihood ratio trace test points to two cointegrating relationships in the system in levels, which vanish in the log-transformed data due to the non-linear nature of the transformation. Subsequently, both data in levels and in logarithms have been considered for the estimation and evaluation of the proposed model. This allows us to evaluate whether accounting for long-term equilibrium relationships contributes to better forecasting performance.

¹ Source: [Eurostat](#) (HICP - monthly data, Overall index excluding energy: TOT_X_NRG, Euro Area: EA, Base Year 2015: I15).

² Results of unit root and cointegration tests are omitted for brevity but are available upon request.

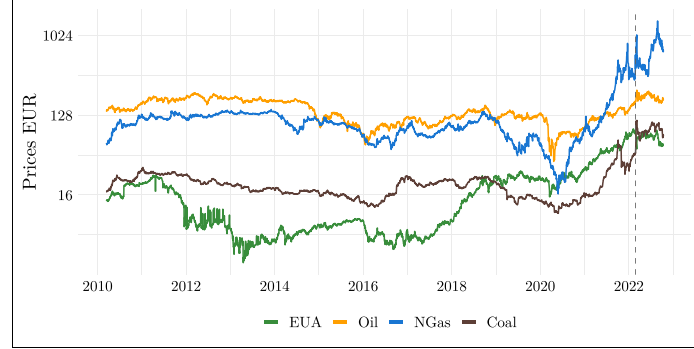


Fig. 1. Time series plot of the data in levels. The dashed line marks the Russian invasion of Ukraine (2022-02-24).

3. The VECM-Copula-GARCH model

3.1. Model design

The preliminary analysis in the previous section reveals several key characteristics of the data which we explicitly account for in our proposed model. First, the variables are non-stationary and cointegrated in the levels. While cointegration is not formally detected in the log-transformed data, we wish to explore whether allowing for it may nevertheless improve forecasting performance, so we employ a VECM for the mean process, and experiment with different cointegrating ranks, including rank zero, i.e. no cointegration. Second, consistent with the literature, we consider univariate GARCH models for the conditional heteroscedasticity of each series. Third, we model the conditional cross-sectional dependence structure of the individual time series by a copula, thereby accounting for both linear and non-linear dependence in a consistent manner. For more flexibility, the dependence is allowed to vary over time, which is implemented by modeling the copula parameters by GARCH-type models as in e.g. [Jondeau and Rockinger \(2006\)](#).

Denoting the K -dimensional vector of observable variables at time t by $\mathbf{X}_t$, our general model can be formulated in terms of the conditional joint distribution $F_{\mathbf{X}_t|\mathcal{F}_{t-1}}$, where $\mathcal{F}_t$ is the sigma field generated by all information available up to and including time t . Using Sklar's Theorem, we specify the joint conditional distribution $F_{\mathbf{X}_t|\mathcal{F}_{t-1}}$ of the model variables in terms of their marginal distributions $F_{X_{k,t}|\mathcal{F}_{t-1}}$ for $k = 1, \dots, K$, and the copula $C_{U_t|\mathcal{F}_{t-1}}$, as is popular in many financial applications ([Jondeau and Rockinger, 2006](#); [Hu, 2006](#)). Thus, for the realized values $\mathbf{x}_t = (x_{1,t}, \dots, x_{K,t})^\top$ it holds that

$$F_{\mathbf{X}_t|\mathcal{F}_{t-1}}(\mathbf{x}_t) = C_{U_t|\mathcal{F}_{t-1}}(u_t), \quad (1)$$

where $u_t = (u_{1,t}, \dots, u_{K,t})^\top$ and $u_{k,t} = F_{X_{k,t}|\mathcal{F}_{t-1}}(x_{k,t})$, $k = 1, \dots, K$. That is, we specify K marginal model components along with a copula that captures their dependency structure. In order to simplify the notation we drop the conditioning on $\mathcal{F}_{t-1}$ in Eq. (1) and write henceforth $F(\mathbf{x}_t) = C(u_t)$.

The model can thus be specified as follows:

$$F(\mathbf{x}_t) = C[\mathbf{F}(\mathbf{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\sigma}_t^2, \mathbf{v}, \boldsymbol{\lambda}); \Xi_t, \Theta], \quad (2)$$

where the $(K(K-1)/2)$ -dimensional vector Ξ_t comprises the time-varying dependence parameters, while the remaining time-invariant copula parameters are gathered in the vector Θ . Specifically, we take C as the t -copula to allow for heavy tails, in which case Ξ_t stands for the time-varying correlation matrix, while Θ denotes the degrees of freedom parameter. The individual marginal distributions, succinctly represented by the vector $\mathbf{F} = (F_1, \dots, F_K)^\top$ in Eq. (2), are specified in terms of the $(K \times 1)$ vectors of time-varying marginal means $\boldsymbol{\mu}_t = (\mu_{1,t}, \dots, \mu_{K,t})^\top$ and variances $\boldsymbol{\sigma}_t^2 = (\sigma_{1,t}^2, \dots, \sigma_{K,t}^2)^\top$, respectively. In order to allow for both skewness and heavy tails in the marginals, we consider the generalized non-central t -distribution for each F_i , $i = 1, \dots, K$, with individual degrees of freedom and non-centrality parameters gathered in $\mathbf{v} = (v_1, \dots, v_K)^\top$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K)^\top$. For estimation, each F_i has been parameterized so that its expectation and the variance are given by the parameters to be estimated.

We now turn our attention to the mean and variance processes $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t^2$. The temporal evolution of the means vector is modeled by a VECM model as

$$\Delta \boldsymbol{\mu}_t = \Pi \mathbf{x}_{t-1} + \Gamma \Delta \mathbf{x}_{t-1}, \quad (3)$$

where $\Pi = \alpha \beta^\top$, $\alpha \in \mathbb{R}^{K \times r}$ and $\beta \in \mathbb{R}^{K \times r}$ is the cointegrating matrix of rank r , $0 \leq r \leq K$.

The dynamics of the individual variances are modeled by univariate GARCH models:

$$\sigma_{i,t}^2 = \omega_i + \alpha_i^+ (\epsilon_{i,t-1}^+)^2 + \alpha_i^- (\epsilon_{i,t-1}^-)^2 + \beta_i \sigma_{i,t-1}^2,$$

where $\epsilon_{i,t-1}^+ = \max\{\epsilon_{i,t-1}, 0\}$ and $\epsilon_{i,t-1}^- = \min\{\epsilon_{i,t-1}, 0\}$, $i = 1, \dots, K$, respectively. Separating the coefficients for positive and negative innovations allows the model to capture leverage effects.

Finally, the temporal evolution of the $K(K-1)/2$ primary dependence parameters of the copula, gathered in the (scaled) correlation matrix Ξ_t , is modeled similarly to a GARCH process:

$$\begin{aligned}\Xi_t &= \Lambda(\xi_t), \\ \xi_{i,j,t} &= \eta_{0,ij} + \eta_{1,ij}\xi_{i,j,t-1} + \eta_{2,ij}z_{i,t-1}z_{j,t-1},\end{aligned}$$

where $\xi_{i,j,t}$ is a latent process, $z_{i,t}$ denotes the i th standardized residual from time series i at time point t , and $\Lambda(\cdot)$ is a link function. The latter ensures that Ξ_t does not exceed its support space and remains semi-positive definite by replacing negative values in the singular value decomposition of Ξ_t by 10^{-5} .

3.2. Relevant nested models

The proposed model (2) is the most general model incorporating cointegration, conditional heteroscedasticity, and time-varying dependence parameters. This model is denoted by $\text{VECM}_{\text{lev,ncp}}^{r,\sigma^2,\rho_t}$. Here the subscripts **lev** and **ncp** denote joint estimation along with the other model parameters of the leverage and non-centrality parameters, respectively, while the superscripts signify the assumed cointegrating rank (r), the time-varying conditional variance (σ_t^2) and the time-varying dependence parameters (ρ_t). The performance of this model will be compared to that of several nested models, following the same notational convention. **RW** will stand for a model with no cointegration such that the dynamics of the mean process can be thought of as a K -dimensional random walk. Assuming constant conditional variance and dependence parameters will be denoted by superscripts σ^2 and ρ , respectively, where the dependence on t is suppressed. Hence, e.g., $\text{VECM}^{r2,\sigma,\rho}$ will denote the VECM-Copula-GARCH model of cointegrating rank $r = 2$, where the dependence parameters and variances are constant in time, and no leverage or non-centrality parameters are estimated. Models estimated on log-transformed data are signified by the additional subscript **log**.

3.3. Estimation

Despite having three different model components, all parameters can be estimated jointly by maximum likelihood (ML). Using conditional independence, the joint likelihood can be factorized as

$$L = f_{X_1} \prod_{i=2}^T f_{X_i|F_{i-1}},$$

with the multivariate conditional density at any time point t given by

$$f_{X_t}(x_t|F_{t-1}) = c[F(x_t; \mu_t, \sigma_t^2, \nu, \lambda); \Xi_t, \Theta] \cdot \prod_{i=1}^K f_{X_{i,t}}(x_i; \mu_i, \sigma_i^2, \nu, \lambda).$$

Here the copula density c can be derived analytically from the copula as $c(u) = \partial_u C(u)$.

4. Forecasting study

4.1. Study design

To perform h -step ahead probabilistic forecasting from time point t the following procedure is employed. First the model is fitted and the values of all time-varying parameters, Ξ_t , μ_t and σ_t^2 are determined using the information contained in F_t . Then n samples are drawn from the K -dimensional copula, $C(\cdot; \Xi_t, \Theta)$. The i th marginal sample from the copula, $i = 1, \dots, K$, is quantile-transformed using the corresponding i th quantile function, $F_i^{-1}(\cdot; \mu_i, \sigma_i^2, \nu, \lambda)$. The point forecast is estimated by the mean. Using the point forecast, all time-varying parameters are updated and the procedure is repeated $h-1$ times. Thus the generated samples will have a cross-sectional dependence structure imposed by the copula and marginal distributions according to the univariate densities. Expected values, variances, and other (mixed) moments can be approximated with arbitrary accuracy, which is limited only by computational power.

4.2. Forecast evaluation

Evaluation of the probabilistic forecast is conducted using different evaluation measures. The probabilistic multivariate forecasts will be evaluated by the energy score (ES). The energy score of a multivariate probabilistic forecast with distribution F is given by

$$\text{ES}_t(F, \mathbf{x}_t) = \mathbb{E}_F(\|\tilde{\mathbf{X}}_t - \mathbf{x}_t\|_2) - \frac{1}{2} \mathbb{E}_F(\|\tilde{\mathbf{X}}_t - \tilde{\mathbf{X}}'_t\|_2),$$

where $\mathbf{x}_t$ is the observed K -dimensional realization and $\tilde{\mathbf{X}}_t$, respectively $\tilde{\mathbf{X}}'_t$ are independent random vectors distributed according to F (Gneiting and Raftery, 2007). A low value of the ES indicates a good probabilistic forecast. In the univariate case, the energy score becomes the Continuous Ranked Probability Score (CRPS). Mean forecasts are evaluated by the root mean squared error (RMSE). Significant differences in forecast performances are tested with the Diebold–Mariano test (Diebold and Mariano, 2002) with the adjustment by Harvey et al. (1997).

Table 1

Improvement in energy score of selected models relative to $RW^{\sigma,\phi}$ in % (higher = better). Colored according to the test statistic of a DM-Test comparing to $RW^{\sigma,\phi}$ (greener means lower test statistic i.e., better performance compared to $RW^{\sigma,\phi}$).

Model	ES ^{All} ₁₋₃₀	ES ^{EUA} ₁₋₃₀	ES ^{Oil} ₁₋₃₀	ES ^{Ngas} ₁₋₃₀	ES ^{Coal} ₁₋₃₀	ES ^{All} ₁	ES ^{All} ₅	ES ^{All} ₃₀
$RW^{\sigma,\phi}$	161.96	10.06	37.94	146.73	13.22	5.56	13.28	34.29
$RW^{r,\phi}$	9.40	3.75	-0.41	11.39	4.13	10.34	9.10	7.59
$RW^{\sigma,\phi}_{ncp,log}$	12.04	6.16	-0.56	14.33	7.35	9.22	9.82	10.02
$RW^{\sigma,\phi}_{log}$	12.10	6.25	-0.59	14.44	7.31	9.04	9.66	9.91
$VECM^{0,\sigma,\phi}_{ev,ncp}$	9.68	-0.72	0.32	11.74	3.70	10.82	10.50	8.21
$VECM^{0,\sigma,\phi}_{log}$	12.15	6.10	-0.70	14.57	7.80	8.05	9.99	10.04
$VECM^{1,\sigma,\phi}_{ev,ncp}$	10.65	-24.09	-4.81	13.37	-0.93	9.86	10.07	7.10
$VECM^{1,\sigma,\phi}_{ncp,log}$	8.38	-0.40	-4.65	11.44	8.66	6.62	7.98	1.42
$VECM^{2,\sigma,\phi}_{log}$	8.99	-13.79	-7.15	11.43	-0.51	9.86	10.87	4.32
$VECM^{2,\sigma,\phi}_{ev,ncp}$	8.93	-2.77	-8.07	12.44	6.58	7.09	9.42	4.42
$VECM^{3,\sigma,\phi}_{ev,ncp}$	9.12	-9.18	-3.68	11.70	-1.01	10.94	11.20	6.46
$VECM^{3,\sigma,\phi}_{log}$	0.05	-7.70	-6.51	2.43	4.31	5.09	5.40	-17.97
$VECM^{4,\sigma,\phi}_{ev,ncp}$	2.52	-30.98	-9.81	5.87	-2.04	9.00	7.29	-9.23
$VECM^{4,\sigma,\phi}_{ncp,log}$	6.21	-7.83	-15.69	10.43	6.83	5.54	7.28	-5.12
ETS^{σ}	9.94	5.75	0.08	13.05	7.83	6.96	7.74	6.21
ETS^{σ}_{log}	8.12	7.80	-0.51	11.17	8.54	5.05	6.14	2.66
VES^{σ}	5.50	-4.43	-3.22	6.29	4.68	-25.99	-2.42	3.07
VES^{σ}_{log}	7.68	3.31	-4.34	9.07	8.30	-22.11	1.07	4.32

Coloring w.r.t. test statistic: <-5 -4 -3 -2 -1 0 1 2 3 4 >5.

5. Results

We consider a rolling-window forecasting study with a window size of 1000 days which corresponds to about four years of data. As we have 3257 observations in total and compute 30-steps-ahead forecasts, this leaves 2227 potential starting points for the rolling-window study. To reduce computational costs, we sample $n = 250$ data points uniformly from those 2227 grid points for further evaluation. We report the multivariate distributional forecast of all considered models by providing $2^{12} = 2048$ samples from the predicted distribution.

We evaluate the full spectrum of the aforementioned VECM and RW models in conjunction with a simple exponential smoothing model, denoted by ETS^{σ} , and a vector exponential smoothing model (Svetunkov et al., 2023), denoted by VES^{σ} , which we include as benchmarks. We choose a univariate and a multivariate ETS model as benchmarks since they are commonly used for forecasting (Jónsson et al., 2014; Abdul Azees and Sasikumar, 2019; Berrisch and Ziel, 2022). For the sake of brevity, however, we report the performance only of selected ones, particularly the best-performing models. The performance of the models estimated on log-transformed data has been evaluated by taking the exponent of the forecasts without bias correction, in accordance with the recommendation of Demetrescu et al. (2020) for persistent data. However, the exponent of some log forecasts evaluates to infinity. We replace these values with a large real integer (10^5) to enable a valid evaluation of these forecasts. This issue affects less than 0.001% of all forecasts of the reported log models. We used 10^5 as a replacement because it can be considered an extreme scenario for the prices in question. Note that we deal with inflation-adjusted, emission-normalized prices here. Evaluating the research on the social costs of carbon confirms that 10^5 would be an extreme scenario, at least for EUA (Anthoff and Tol, 2013; Tol, 2019; Aldy et al., 2021). The latter also holds for the other commodities if other price components are negligible.

The performance of the selected models with respect to the ES can be found in Table 1. Their ES is compared to the ES of a simple random walk model with constant copula and no conditional heteroscedasticity modeling, $RW^{\sigma,\phi}$, estimated on the data in levels. ES^{All}_{1-30} denotes the score of the (4×30) -dimensional multivariate (4-dimensional) 1–30 days ahead forecast. The general form of the ES allows us to evaluate the effects of modeling both the temporal and the cross-sectional dependence. ES^{EUA}_{1-30} denotes the score of the univariate 30-dimensional 1–30 days ahead forecast for the EUA prices (analogous for oil, natural gas, and coal prices). In this formulation of the ES, only temporal modeling of the time series is evaluated. ES^{All}_1 denotes the score of the 4-dimensional multivariate one day-ahead forecast (analogously ES^{All}_5 denotes the score of the multivariate 5 day ahead forecast). This form of the ES evaluates the effect of modeling the cross-sectional dependence.

The results of Table 1 can be summarized as follows. No model dominates uniformly in terms of forecasting performance for all variables and all horizons. Models estimated on the log-transformed data generally perform better than their counterparts estimated on the data in levels, despite the naïve forecasts of the former being potentially biased. This result can be explained by the log-transformation successfully stabilizing the variance of the series (Lütkepohl and Xu, 2012). As no cointegration was found in the log-transformed data, the VECM models are over-parameterized and are hence not expected to deliver superior forecasts. Indeed, considering all variables and all forecast horizons altogether, the best-performing models yield an improvement of ca. 12% in ES^{All}_{1-30} over the simple RW model in levels. This is achieved by the $VECM^0$ and the RW models with constant volatility and log transformation. Allowing for time-varying dependence in the copula, or modeling the non-centrality parameter does not seem to influence their results much. These models also deliver the best forecasts at long horizons, and for natural gas prices in particular.

Table 2

Improvement in CRPS of selected models relative to $RW^{\sigma,\rho}$ in % (higher = better). Colored according to the test statistic of a DM-Test comparing to $RW^{\sigma,\rho}$ (greener means lower test statistic i.e., better performance compared to $RW^{\sigma,\rho}$).

Model	EUA			Oil			NGas			Coal		
	H1	H5	H30	H1	H5	H30	H1	H5	H30	H1	H5	H30
$RW^{\sigma,\rho}$	0.4	0.9	2.1	1.5	3.4	9.1	4.7	11.6	29.8	0.3	0.9	2.8
RW^{σ,ρ_1}	5.6	6.0	2.8	2.1	2.7	-0.8	12.6	10.5	9.6	10.7	6.5	2.1
$RW^{\sigma,\rho_1}_{ncp,log}$	5.1	8.7	5.0	0.7	0.8	-0.4	11.4	11.5	12.4	8.0	7.3	6.7
RW^{σ,ρ_1}_{log}	4.7	8.9	5.2	0.0	0.3	-0.6	11.2	11.4	12.4	7.7	7.5	6.6
$VECM^{r0,\sigma,\rho_1}_{lev,ncp}$	3.6	0.6	-1.6	2.7	3.0	0.0	13.1	12.2	10.4	11.8	7.2	1.5
$VECM^{r0,\sigma,\rho_1}_{log}$	4.2	8.9	5.1	0.2	0.4	-0.8	9.9	11.8	12.7	7.8	7.9	7.3
$VECM^{r1,\sigma,\rho_1}_{lev,ncp}$	-2.2	-2.6	-43.2	2.4	0.7	-8.0	12.0	12.3	10.4	6.9	2.4	-4.1
$VECM^{r1,\sigma,\rho_1}_{ncp,log}$	3.6	7.6	-6.4	0.3	-0.2	-6.6	8.3	10.1	4.4	7.8	8.9	6.5
$VECM^{r2,\sigma,\rho_1}_{log}$	-0.9	-0.9	-26.8	2.1	2.6	-12.2	11.8	12.8	8.3	8.3	1.0	-4.0
$VECM^{r3,\sigma,\rho_1}_{log}$	3.7	7.1	-11.6	0.8	-1.0	-11.6	8.8	11.8	8.4	7.5	8.2	4.8
$VECM^{r3,\sigma,\rho_1}_{lev,ncp}$	0.7	2.7	-19.0	1.3	1.6	-6.4	13.4	13.5	9.4	6.1	1.7	-4.1
$VECM^{r3,\sigma,\rho_1}_{log}$	3.1	8.6	-22.2	1.8	1.3	-10.3	6.3	7.1	-17.9	8.7	7.2	-1.2
$VECM^{r4,\sigma,\rho_1}_{lev,ncp}$	-2.8	-5.3	-52.7	1.1	-2.6	-11.5	11.2	9.8	-7.0	6.0	0.2	-7.4
$VECM^{r4,\sigma,\rho_1}_{ncp,log}$	2.3	6.5	-19.6	1.3	-2.6	-23.2	6.9	9.6	-0.8	8.2	8.9	3.6
ETS^{σ}	0.2	6.8	5.7	1.1	0.9	-0.2	10.9	11.3	10.9	7.5	6.7	5.6
ETS^{σ}_{log}	1.0	8.6	8.0	0.1	0.7	-0.6	8.9	9.4	7.1	7.3	7.8	6.7
VES^{σ}	-38.5	-6.4	-5.4	-33.3	-6.1	-2.4	-26.6	-2.6	3.6	-37.5	-5.5	4.7
VES^{σ}_{log}	-32.4	2.8	1.8	-30.4	-6.2	-3.2	-22.0	1.8	5.4	-27.0	2.3	6.4

Coloring w.r.t. test statistic: -5 -4 -3 -2 -1 0 1 2 3 4 5

VECM models in levels provide slightly worse overall forecasts than the RW in logarithms but seem to perform best of all at short horizons with $r = 3$. Nevertheless, accounting for cointegrating relations offers at most modest improvement over the corresponding RW models in levels. Rather, allowing for time variation in the conditional variance and dependence parameters in the levels-RW models improves the forecast significantly and comparably to considering a simple RW in logarithms for $H = 1, 5$. The contribution of the leverage effects and the non-centrality parameter for the models in levels is subordinate. The forecasts of the level VECMs, however, seem quite unsatisfactory for the EUA and oil price series. They are rather best forecast by RW and ETS: the EUA price in logarithms, while the oil price in levels.

Table 2 presents the CRPS of the selected models for every univariate time series. H1, H5, and H30 denote the score or improvement thereof for the 1-, 5- and 30-day-ahead probabilistic forecast, respectively. The CRPS gives information about the quality of the univariate modeling only, as modeling the cross-sectional dependence has no effect on the univariate density obtained by quantile transforming the samples generated from the copula. Even though the samples from the copula have a dependence structure imposed by the copula, the marginal distributions are distributed uniformly.

The results in Table 2 align with those of the ES evaluation: There is no uniformly superior model in general. The RW and $VECM^r$ models seem to perform quite well also in terms of modeling the marginal distributions: in levels for oil and gas prices, in logarithms for EUA prices. For natural gas prices, the levels-VECM models outperform the rest at short horizons, and accounting for leverage effects in the GARCH-part of the model is beneficial.

Table 3 presents the mean forecast performance in terms of the RMSE. Considering that no model significantly outperforms the level RW models we find evidence supporting the efficient markets hypothesis. All RW models perform similarly as distributional modeling has no influence on the mean forecast. Rather, the differences emerge from random simulation fluctuations. The models in logarithms, however, seem to perform significantly worse than the level ones, in particular for the EUA prices. This may be attributed to the few extreme trajectories that required capping, as discussed above.

We next evaluate the temporal evolution of the copula-implied pairwise correlations of the best-performing (in terms of ES) model $VECM^{r0,\sigma,\rho_1}_{log}$ presented in Fig. 2. The correlations between the prices of natural gas and coal, and EUA and coal are approximately constant over time. The correlations of oil with coal and gas fluctuate around a level of 0.2 featuring a temporary increase of up to 0.5 in the emerging Russian invasion. The most volatile correlation is that between EUA and natural gas prices. It fluctuates strongly around a level of approximately 0.3. Immediately after the Russian invasion of Ukraine, the correlation between EUA and gas prices changes dramatically, dropping from a level of about 0.3 to nearly -0.4. By mid-March 2022 the relation has stabilized.

Finally, Fig. 3 displays the predictive quantiles in percent extracted from the 2048 trajectories of the $VECM^{r0,\sigma,\rho_1}_{log}$ model starting right before the Russian invasion in Ukraine. Evidently, the prices' development in the first few days thereafter does not seem too extreme with the exception of the gas price peak on February 24 itself. However, coal, EUA, and gas prices drift well beyond the projected 99.8th percentile in the three weeks after the subsequent freeze of the foreign reserves of Russia's central bank, suggesting that this reaction had been largely unanticipated by the markets.

6. Conclusion

In this paper, we propose a VECM-Copula-GARCH model to jointly model EUA, oil, coal, and natural gas prices with the aim of producing short-term probabilistic forecasts. The price series are in real terms and standardized to reflect one tonne of CO₂

Table 3

Improvement in RMSE score of selected models relative to $RW^{\sigma,\rho}$ in % (higher = better). Colored according to the test statistic of a DM-Test comparing to $RW^{\sigma,\rho}$ (greener means lower test statistic i.e., better performance compared to $RW^{\sigma,\rho}$).

Model	EUA			Oil			NGas			Coal		
	H1	H5	H30	H1	H5	H30	H1	H5	H30	H1	H5	H30
$RW^{\sigma,\rho}$	0.9	2.0	5.0	2.9	6.4	16.7	17.8	42.8	85.4	0.9	2.9	7.0
$RW^{\sigma,\rho}_{lev}$	-0.1	-0.1	0.7	0.0	-0.3	-0.1	-0.2	0.3	1.3	-0.2	0.0	-1.8
$RW^{\sigma,\rho}_{ncp,log}$	-270.5	-154.1	-139.9	0.5	-0.5	-2.9	-0.8	0.7	-1.6	0.3	-31.2	-24.5
$RW^{\sigma,\rho}_{log}$	-705.0	-265.4	-125.2	0.6	0.2	-0.2	-0.4	0.1	-1.6	-0.9	-0.3	-8.3
$VECM^{r0,\sigma,\rho}_{lev,ncp}$	-0.9	0.2	0.5	0.5	0.2	0.0	-0.4	0.7	0.2	1.4	0.1	0.2
$VECM^{r0,\sigma,\rho}_{log}$	-271.5	-191.3	-114.3	1.7	-12.3	-3.6	-0.6	1.6	-4.1	0.0	-0.8	-6.7
$VECM^{r1,\sigma,\rho}_{lev,ncp}$	-7.7	-6.6	-64.7	1.2	-2.2	-8.4	-3.8	-1.2	-5.3	1.4	-2.0	-0.6
$VECM^{r1,\sigma,\rho}_{ncp,log}$	-2.7	-86.5	-108.7	0.3	-1.3	-6.5	-3.0	-0.5	-21.1	0.4	-1.1	-24.7
$VECM^{r2,\sigma,\rho}_{lev,ncp}$	-5.0	-2.3	-40.6	1.2	0.2	-33.7	0.2	3.9	1.6	1.0	0.5	-1.4
$VECM^{r2,\sigma,\rho}_{log}$	-3.8	-117.4	-141.4	1.6	-11.8	-18.8	-2.1	1.8	-10.4	0.1	-44.5	-12.9
$VECM^{r3,\sigma,\rho}_{lev,ncp}$	-4.3	-0.2	-29.8	-0.2	-1.1	-5.8	0.2	2.0	-0.1	1.0	-0.1	0.9
$VECM^{r3,\sigma,\rho}_{log}$	-4.5	-140.2	-877.4	1.5	-0.8	-19.3	-5.7	-6.1	-74.6	0.7	-1.6	-15.5
$VECM^{r4,\sigma,\rho}_{lev,ncp}$	-8.8	-12.1	-79.5	-0.7	-4.5	-10.0	-1.6	-2.1	-32.3	0.8	-0.7	-2.3
$VECM^{r4,\sigma,\rho}_{log}$	-4.1	-3.3	-40.9	1.6	-4.8	-25.1	-5.0	-2.5	-31.7	0.7	0.4	3.0
ETS^{σ}	-0.3	0.3	1.6	0.7	0.1	-0.1	0.1	-0.1	0.2	-2.4	-3.9	2.5
ETS^{σ}_{log}	-1.0	0.4	1.6	0.9	0.0	-0.1	-1.9	-1.9	-13.9	-0.3	-3.6	-1.8
VES^{σ}	-37.4	-8.9	-6.0	-27.9	-7.4	-2.8	-27.2	-9.5	-2.4	-41.7	-1.2	1.6
VES^{σ}_{log}	-37.6	-9.2	-7.8	-26.8	-7.3	-3.0	-27.0	-6.8	-3.5	-41.2	-2.2	-0.3

Coloring w.r.t. test statistic: <-5 -4 -3 -2 -1 0 1 2 3 4 >5

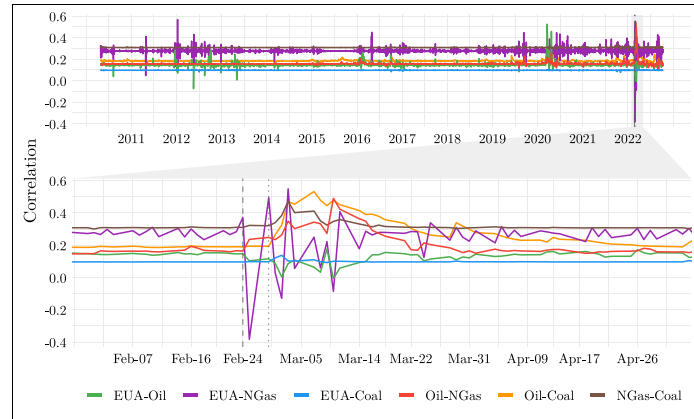


Fig. 2. Temporal evolution of the linear dependence as provided by time varying dependence parameter modeling in the $VECM^{r0,\sigma,\rho}_{log}$ model. The dashed line marks the Russian invasion of Ukraine (2022-02-24). The dotted line marks the freeze of foreign reserves of the central bank of Russia (2022-02-28).

emissions for proper comparison. Allowing time-varying dependence parameters in the copula and leverage effects in the model of the conditional heteroscedasticity allows us not only to utilize important data characteristics to improve the forecasts, but also to analyze the temporal evolution of the estimated parameters, providing insight into the linkages between EUA and fossil fuels markets. An extensive rolling-window forecasting study compares the proposed model with relevant benchmark models for both the data in levels and after log-transformation.

The best-performing model is $VECM^{r0,\sigma,\rho}_{log}$, i.e. a VAR with constant conditional variance and time-varying dependence parameters estimated on log-transformed data. In general, stabilizing the variance by the log-transformation and employing multivariate random walk models is more beneficial for the overall forecasting performance than accounting for long-term relationships between the variables in levels. The latter can be preferable for forecasts at short horizons, but only in conjunction with modeling the time variation in the conditional volatility, which accounts for the greatest share of improvement in the forecasts. The improvement contributed by time-varying copula dependence, leverage effects in the GARCH-part of the model and marginal distribution parameters is less pronounced.

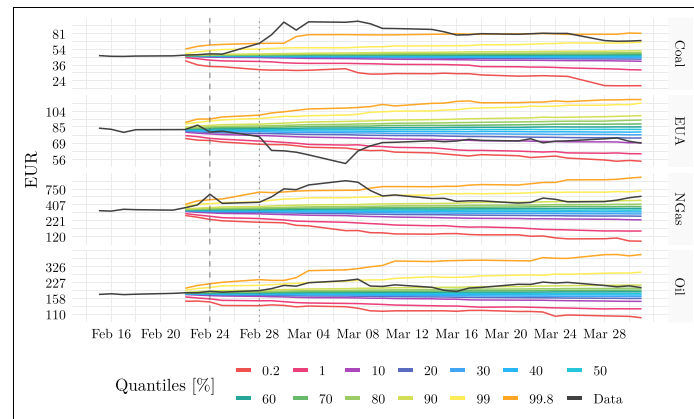


Fig. 3. Predictive quantiles in % extracted from 2048 trajectories predicted by the $VECM_{log}^{0,\sigma,\phi}$ model starting 2022-02-21. The plot depicts inflation-adjusted and emission-normalized data. The dashed line marks the Russian invasion of Ukraine (2022-02-24). The dotted line marks the freeze of foreign reserves of the central bank of Russia (2022-02-28).

CRediT authorship contribution statement

Jonathan Berrisch: Conception and design of study, Acquisition of data, Analysis of data, Interpretation of empirical results, Writing – original draft, Writing – review & editing. **Sven Pappert:** Conception and design of study, Analysis of data, Interpretation of empirical results, Writing – original draft, Writing – review & editing. **Florian Ziel:** Conception and design of study, Acquisition of data, Interpretation of empirical results, Writing – original draft, Writing – review & editing. **Antonia Arsova:** Conception and design of study, Interpretation of empirical results, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgment

All authors approved the version of the manuscript to be published.

References

- Abdul Azees, S.A., Sasikumar, R., 2019. Comparison study on exponential smoothing and ARIMA model for the fuel price. In: Logistics, Supply Chain and Financial Predictive Analytics. Springer, pp. 33–41. http://dx.doi.org/10.1007/978-981-13-0872-7_3.
- Aldy, J.E., Kotchen, M.J., Stavins, R.N., Stock, J.H., 2021. Keep climate policy focused on the social cost of carbon. *Science* 373 (6557), 850–852. <http://dx.doi.org/10.1126/science.abi7813>.
- Anthoff, D., Tol, R.S., 2013. The uncertainty about the social cost of carbon: A decomposition analysis using fund. *Clim. Change* 117 (3), 515–530. <http://dx.doi.org/10.1007/s10584-013-0706-7>.
- Benz, E., Trück, S., 2009. Modeling the price dynamics of CO2 emission allowances. *Energy Econ.* 31 (1), 4–15. <http://dx.doi.org/10.1016/j.eneco.2008.07.003>.
- Berrisch, J., Ziel, F., 2022. Distributional modeling and forecasting of natural gas prices. *J. Forecast.* <http://dx.doi.org/10.1002/for.2853>.
- Chevallier, J., Nguyen, D.K., Reboredo, J.C., 2019. A conditional dependence approach to CO2-energy price relationships. *Energy Econ.* 81, 812–821. <http://dx.doi.org/10.1016/j.eneco.2019.05.010>.
- Demetrescu, M., Golosnoy, V., Titova, A., 2020. Bias corrections for exponentially transformed forecasts: Are they worth the effort? *Int. J. Forecast.* 36, 761–780. <http://dx.doi.org/10.1016/j.ijforecast.2019.09.001>.
- Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. *J. Bus. Econom. Statist.* 20 (1), 134–144. <http://dx.doi.org/10.1198/073500102753410444>.
- Duan, K., Ren, X., Shi, Y., Mishra, T., Yan, C., 2021. The marginal impacts of energy prices on carbon price variations: Evidence from a quantile-on-quantile approach. *Energy Econ.* 95, 105131. <http://dx.doi.org/10.1016/j.eneco.2021.105131>.
- Dutta, A., Jalkh, N., Bouri, E., Dutta, P., 2019. Assessing the risk of the European Union carbon allowance market: Structural breaks and forecasting performance. *Int. J. Manage. Finance* 16 (1), 49–60. <http://dx.doi.org/10.1108/IJMF-01-2019-0045>.
- Feezzi, C., Bunn, D., 2009. Interaction of European carbon trading and energy prices. *JEM* 24, 53–69.
- García, A., Jaramillo-Morán, M.A., 2020. Short-term European Union Allowance price forecasting with artificial neural networks. *Entr. Sustain. Issues* 8 (1), 261. [http://dx.doi.org/10.9770/jesi.2020.8.1\(18\)](http://dx.doi.org/10.9770/jesi.2020.8.1(18)).
- Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* 102 (477), 359–378. <http://dx.doi.org/10.1198/016214506000001437>.
- Hanif, W., Hernandez, J.A., Mensi, W., Kang, S.H., Uddin, G.S., Yoon, S.-M., 2021. Nonlinear dependence and connectedness between clean/renewable energy sector equity and European emission allowance prices. *Energy Econ.* 101, 105409. <http://dx.doi.org/10.1016/j.eneco.2021.105409>.

- Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. *Int. J. Forecast.* 13 (2), 281–291. [http://dx.doi.org/10.1016/S0169-2070\(96\)00719-4](http://dx.doi.org/10.1016/S0169-2070(96)00719-4).
- Hu, L., 2006. Dependence patterns across financial markets: a mixed copula approach. *Appl. Financial Econ.* 16 (10), 717–729. <http://dx.doi.org/10.1080/09603100500426515>.
- Jondeau, E., Rockinger, M., 2006. The copula-garch model of conditional dependencies: An international stock market application. *J. Int. Money Finance* 25 (5), 827–853. <http://dx.doi.org/10.1016/j.jimonfin.2006.04.007>.
- Jónsson, T., Pinson, P., Nielsen, H.A., Madsen, H., 2014. Exponential smoothing approaches for prediction in real-time electricity markets. *Energies* 7 (6), 3710–3732. <http://dx.doi.org/10.3390/en7063710>.
- Lütkepohl, H., Xu, F., 2012. The role of the log transformation in forecasting economic variables. *Empir. Econ.* 42, 619–638. <http://dx.doi.org/10.1007/s00181-010-0440-1>.
- Meier, J.-H., Voss, N., 2020. Do commodities determine the EU emission allowances price? In: *ICTERI Workshops*. pp. 423–437.
- Paolella, M.S., Taschini, L., 2008. An econometric analysis of emission allowance prices. *J. Bank. Financ.* 32 (10), 2022–2032. <http://dx.doi.org/10.1016/j.jbankfin.2007.09.024>.
- Quaschnig, V., 2016. *Understanding Renewable Energy Systems*. Routledge.
- Reboredo, J.C., 2014. Volatility spillovers between the oil market and the European Union carbon emission market. *Econ. Model.* 36, 229–234.
- Svetunkov, I., Chen, H., Boylan, J.E., 2023. A new taxonomy for vector exponential smoothing and its application to seasonal time series. *European J. Oper. Res.* 304 (3), 964–980.
- Tol, R.S., 2019. A social cost of carbon for (almost) every country. *Energy Econ.* 83, 555–566.
- Trabelsi, N., Tiwari, A.K., 2022. CO2 emission allowances risk prediction with GAS and GARCH models. *Comput. Econ.* 1–31.
- Zhou, Y., Liu, J., Sirisrisakulchai, J., Sriboonchitta, S., 2020. Measurements of the conditional dependence structure among carbon, fossil energy and renewable energy prices: vine copula based GJR-GARCH model. In: *International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making*. Springer, pp. 322–334.

Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas)

Sven Pappert^{a,*}

^a*Chair of Econometrics, Department of Statistics, TU Dortmund University, Germany*

Abstract

Copula-based time series models can model univariate and stationary time series in a flexible way by decomposing the joint distribution of consecutive observations into a copula and the stationary distribution. Implicitly this approach assumes a finite Markov order. In reality a time series may not follow the Markov property. We modify the copula-based time series models by introducing a moving aggregate (MAG) part into the model updating equation. The functional form of the MAG-part is given as the conditional quantile function corresponding to a copula. The resulting MAG-modified Autoregressive Copula-Based Time Series model (MAGMAR-Copula) is discussed in detail and distributional properties are derived in a D-vine framework. We show that the stationary distribution implied by the model is not standard-uniform. Hence we propose an adjustment transformation that recovers the desired standard-uniformity. The model nests the classical ARMA model and can be interpreted as a non-linear generalization of the ARMA-model. The modeling performance is evaluated by modeling US inflation. Our model is competitive with benchmark models in terms of information criteria.

Keywords: Copula, Copula-based time series, Dependence Modeling, Non-linear time series, Persistency, Vine Copula

1. Introduction

Copula-based time series models are a class of univariate probabilistic autoregressive time series models which allow for flexible temporal dependence modeling as well as flexible marginal modeling. The idea amounts to modeling (possibly non-linear) autoregressive temporal dependencies by a copula while modeling marginal properties independently. Given a univariate, strictly stationary time series $\{X_t\}_{t \in \mathbb{Z}}$, the model decomposes the joint distribution, $F_{X_t, \dots, X_{t-p}}$ of $(p+1)$ consecutive random variables of the time series into their stationary marginal distribution, $F_X := F_{X_t} = F_{X_{t-1}} = \dots = F_{X_{t-p}}$, and their copula $C = F_{F_X(X_t), \dots, F_X(X_{t-p})}$. The marginal distribution, F_X , accounts for marginal properties such as mean, variance, skewness etc., while the copula accounts for the temporal dependence structure. The general idea was formulated by [Darsow et al. \(1992\)](#) for Markov chains and further investigated in (amongst others) [Ibragimov \(2005\)](#). A short review is given in [Patton \(2009\)](#). [Chen and Fan \(2006\)](#) and [Beare \(2010\)](#) developed the idea for time series. [Simard and Rémillard \(2015\)](#) explore multivariate copula-based time series forecasting with standard copulas while [Beare and Seo \(2015\)](#), [Brechmann and Czado \(2015\)](#)

*Corresponding author
Email address: pappert@statistik.tu-dortmund.de (Sven Pappert)

and Nagler et al. (2022) explore multivariate time series modeling using vine-copulas.

These models can be used to model non-linearities in temporal dependencies such as asymmetry; see Müller and Reuber (2022) for irradiation data and McNeil and Bladt (2022) for US inflation. For modeling of heavy-tailedness in the temporal dependence, see Loaiza-Maya et al. (2018) and Pappert and Arsova (2023). Linear temporal dependencies may also be modeled simply by employing the normal copula. Thus the model can be understood as a generalization to the AR model, which permits broader temporal dependencies and marginal distributions.

One caveat of the copula-based time series models is that, implicitly, the process is ascribed to be a Markov process of order p , i.e. the true data-generating process is of finite autoregressive order p . This means that the variable X_t may depend on a finite number of past values $X_{t-1}, \dots, X_{t-p}$ only. However, in reality the Markov assumption may fail depending on the time series. The time series that is to be modeled may display persistent behavior i.e. non cut-off autodependence. A popular example of this behavior is volatility clustering. There, persistency is apparent in the second moment of the time series. In principle it is possible to extend the copula-based time series model to arbitrary Markov order, however it becomes unfeasible as the number of parameters increases indefinitely. Hence, we think that there is a need for a copula-based time series model with infinite order partial dependence. To this end, we take inspiration from classic time series models. The problem of limited feasible autoregressive orders in time series models also occurs in the $AR(p)$ and $ARCH(p)$ models. The solution for these models is to introduce moving average terms, resulting in the $ARMA(p, q)$ and $GARCH(p, q)$ models, respectively (see Engle (1982), Bollerslev (1986), Bollerslev et al. (1994), Francq and Zakoian (2019) and Hamilton (2020)). By inverting the corresponding linear filter, $AR(\infty)$ and $ARCH(\infty)$ representations can be derived. Hence the time series can be understood as having no Markov restriction. This results in the time series models being more persistent in the mean in the case of ARMA and more persistent in the variance in the case of GARCH. The latter can be interpreted as the model being able to capture volatility clusters (Bollerslev et al., 1994). See also Fig. 1 and Fig. 2.

Inspired by these approaches, we modify the copula-based time series model by incorporating a moving aggregate (MAG) part in the updating equation. This allows for long-term autoregressive effects in the whole conditional distribution. The MAG part is a non-linear function taking the current and the past q innovations as arguments. Formally it is a mapping from $(0, 1)^{q+1}$ to $(0, 1)$. The functional form of the MAG-function is given by a conditional quantile function corresponding to a copula. We call it 'moving aggregate', because, similar to the moving average, it aggregates the current and past q innovations but it is not a linear aggregation, hence 'aggregate' instead of 'average'. Analogous to the classical ARMA and GARCH models, the MAG-part permits infinite partial dependence. This means that the partial copula (see Spanhel and Kurz, 2016) of X_t and X_{t-j} is in general different from the independence copula for all $j, t \in \mathbb{Z}$. We call the resulting model *MAG-Modified Autoregressive Copula-Based Time Series Model*, or in short: *MAGMAR-Copula*. In this paper we present the model and show that the classical copula-based time series model and the classical linear ARMA model are nested within the proposed model. We proceed to derive its distributional properties in a D-vine framework (see Aas et al. (2009); Czado (2010); Czado and Nagler (2022) and Sect. 2.1 for D-vine copulas). $AR(\infty)$ and

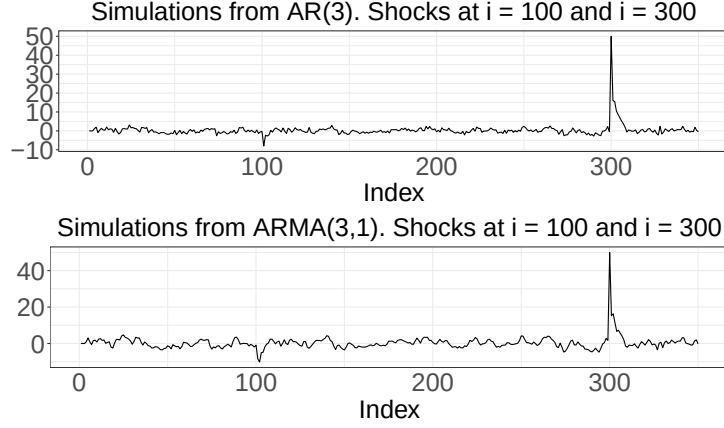


Figure 1: Simulations of an AR(3) and a ARMA(3,1) process using the same random number generation. The ARMA process exhibits long-term autoregressive effects in the mean, i.e. smoothness.

MAG(∞) representations are derived. Building upon the MAG(∞) representation, it is first proven that the MAGMAR(1,1)-Copula model, where the AR-copula is the normal copula, is stationary when excluding perfect comonotonicity and perfect countermonotonicity. Second, we establish stationarity and ergodicity under a contraction-on-average property of the time series. In general the stationary distribution of the time series is not uniform (as is desired). We introduce an auxiliary adjustment transformation to account for this problem. The transformation guarantees that the unconditional distribution of the time series can always be modeled properly. We derive the conditional distribution of $X_t | \mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ denotes the σ -field of all information up to and including $(t-1)$. Based on conditional independence, the likelihood function is derived. Similar to ARMA and GARCH models, the likelihood has to be calculated in an iterative scheme. The algorithm is presented and asymptotic properties of the estimation are derived. Since asymptotic results for the estimation of Markov copula-based time series models are available in the literature (see [Chen and Fan \(2006\)](#) or [Nagler et al. \(2022\)](#)), we focus on the consistency of the maximum-likelihood estimation of the MAG(1)-copula model. We proceed to systematically compare our proposed model with the generalizations of copula-based time series by [McNeil and Bladt \(2022\)](#) and [Joe \(2014\)](#) (chapter 3.14). These models also aim to overcome the Markov restriction/generalize the linear ARMA model. Last, we test the model in a simulation study and apply it to quarterly US inflation. We compare the modeling performance of the MAGMAR-Copula model to the model by [McNeil and Bladt \(2022\)](#) and to linear modeling (i.e. choosing normal copulas for the temporal dependence structure). We find that our model performs well in this setting and is competitive with the model from [McNeil and Bladt \(2022\)](#). In line with their results, we find that proper modeling of asymmetric temporal tail dependence as well as infinite partial dependence is beneficial. Furthermore, we find that incorporating heavy-tailed temporal dependence (i.e. incorporating the t -copula) enhances the performance, indicating not only asymmetric tail-dependence but also heavy-tailed dependencies in US inflation.

In the next section, Sect. 2, the copula-based time series model is introduced and reviewed with a focus on

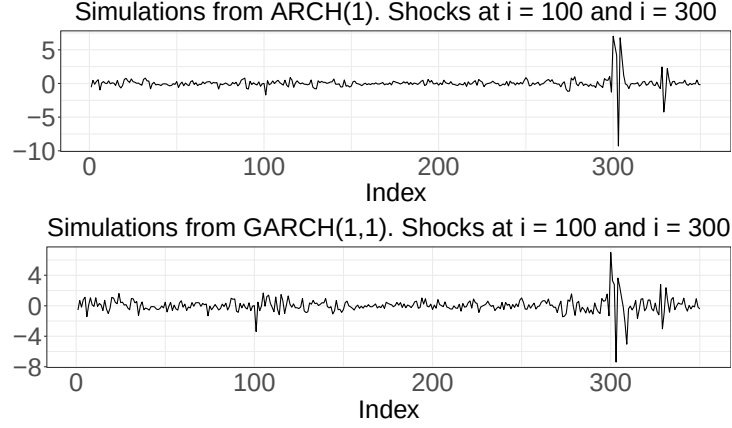


Figure 2: Simulations of an ARCH(1) and a GARCH(1,1) process using the same random number generation. The GARCH process exhibits long-term autoregressive effects in the variance, i.e. volatility clustering.

D-vines. Furthermore related generalizations of copula-based time series models are reviewed. In Sect. 3 the Moving Aggregate Modified Autoregressive Copula-Based Time Series Model (MAGMAR-Copula) is introduced and theoretical properties are derived. In Sect. 4, we explore the model in a simulation study, verifying theoretical findings, extending them and comparing the model with related ones. The MAGMAR-Copula model is applied to US Inflation in Sect. 5. Sect. 6 concludes the work and gives an outlook to future research.

2. Copula-based Time Series Models

In this section the notion of copula-based time series models is introduced. First, we give a short introduction to copulas. Then, general ideas of copula-based time series modeling are developed. Last, we discuss the approaches to generalize copula-based time series modeling by McNeil and Bladt (2022) and Joe (2014).

2.1. Copulas and D-vine Copulas

In this section we give a short introduction to copulas and the main ideas to multivariate modeling with copulas. For detailed and comprehensive presentations we refer to standard works: Joe (2014) and Nelsen (2006).

Consider p continuous random variables $X_1, \dots, X_p$. Let their joint distribution be given by $F_{X_1, \dots, X_p}$ and their marginal distributions by $F_{X_1}, \dots, F_{X_p}$, respectively. The general idea of multivariate copula modeling amounts to decomposing the joint distribution into marginal distributions and a copula,

$$F_{X_1, \dots, X_p}(x_1, \dots, x_p) = C[F_{X_1}(x_1), \dots, F_{X_p}(x_p)]. \quad (1)$$

Sklar's theorem states that there exists a copula that fulfills the statement from Eq. 1, which is unique if the random variables are continuous (Sklar, 1959; Joe, 2014). The copula C is the joint distribution

of probability integral transformed (PIT) and hence uniform random variables $F_{X_1}(X_1), \dots, F_{X_p}(X_p)$ (see [Angus, 1994](#), for the PIT), $F_{F_{X_1}(X_1), \dots, F_{X_p}(X_p)} = C$. The uniformity of $U_i := F_{X_i}(X_i)$, $i \in \{1, \dots, p\}$ allows the copula to be interpreted as the dependence structure of the random variables $X_1, \dots, X_p$. An important notion for visualization, estimation and for D-vine copulas is the copula density. The copula density of a copula C is given as ([Joe, 2014](#))

$$c[u_1, \dots, u_p] = \frac{\partial^p C[u_1, \dots, u_p]}{\partial u_1 \dots \partial u_p}. \quad (2)$$

Since the copula density is given by the partial derivative of the respective copula, Eq. 2 is only valid if the copula is differentiable. In applications, the copula and the marginal distributions have to be chosen appropriately. The normal, t , Gumbel and independence copulas are described in [Appendix C.1](#).

Now we introduce the notion of D-vine copulas. D-vine copulas are a special case of pair-copula constructions. A p -variate copula is decomposed into pair-copulas, i.e. bivariate copulas ([Bedford and Cooke, 2001, 2002](#); [Aas et al., 2009](#); [Czado, 2010](#); [Czado and Nagler, 2022](#)). The decomposition is based on the law of total probability and is not unique. The D-vine decomposition amounts to a special ordering of the variables in the decomposition. The decomposition in terms of the p -variate copula density is given as ([Aas et al., 2009](#)),

$$c[u_1, \dots, u_p] = \prod_{j=1}^{p-1} \prod_{i=1}^{n-j} c_{i, i+j|i+1, \dots, i+j-1} \left(F(u_i|u_{i+1}, \dots, u_{i+j-1}), F(u_{i+j}|u_{i+1}, \dots, u_{i+j-1}) \right), \quad (3)$$

where $c_{i, i+j|i+1, \dots, i+j-1}$ is the conditional copula density of U_i and U_{i+j} given the intermediate variables $U_{i+1}, \dots, U_{i+j-1}$. $F(u_i|u_{i+1}, \dots, u_{i+j-1})$ is the conditional distribution function of U_i given $U_{i+1}, \dots, U_{i+j-1}$ evaluated at u_i and conditioning variables $u_{i+1}, \dots, u_{i+j-1}$. Since the joint distribution of $U_i, U_{i+1}, \dots, U_{i+j-1}$ is a copula, $F(u_i|u_{i+1}, \dots, u_{i+j-1})$ may also be understood as a conditional distribution function, corresponding to a copula. The decomposition in Eq. 3 is called 'D-vine' for 'drawable vine' because the variables are ordered in a path like manner ([Czado and Nagler, 2022](#)). The vine copula approach allows for more flexibility when modeling multivariate dependencies because each bivariate dependence can be specified separately.

2.2. Copula-based Time Series with Autoregressive Order of One

Consider a strictly stationary time series $\{X_t\}_{t \in \mathbb{Z}}$. A time series being Markov(p) means that $\mathbb{P}(X_t \leq a | X_{t-1} = x_{t-1}, \dots, X_1 = x_1) = \mathbb{P}(X_t \leq a | X_{t-1} = x_{t-1}, \dots, X_{t-p} = x_{t-p})$, $t > p$ ([Ibe, 2013](#)). A stationary and univariate Markov(p) time series is characterized by the $(p+1)$ -variate joint distribution $F_{X_t, \dots, X_{t-p}}$ of consecutive random variables X_t to X_{t-p} . The idea is to decompose the joint distribution into a copula and marginal distributions. By Sklar's Theorem there always exists a copula $C := F_{F_X(X_t), \dots, F_X(X_{t-p})}$ such that the joint distribution can be decomposed into marginal distributions F_{X_t} to $F_{X_{t-p}}$ and the copula ([Sklar, 1959](#); [Joe, 2014](#)). If the random variables $\{X_t\}$ are continuous, the decomposition is unique. Due to strict stationarity we can write $F_{X_t} = \dots = F_{X_{t-p}} =: F_X$. Then the decomposition may be written

as

$$F_{X_t, \dots, X_{t-p}}(x_0, \dots, x_p) = C[F_X(x_0), \dots, F_X(x_p)].$$

Hence, the joint distribution is specified by a marginal distribution F_X and the copula C . Note that the copula is the joint distribution of the probability-integral transformed random variables $U_t := F_X(X_t)$ to $U_{t-p} := F_X(X_{t-p})$. The copula may be parametrized by dependence parameter(s) ϕ , so we write $C = C_\phi$. We introduce this notation to emphasize the analogy to the linear AR model and later on, in Sect. 3 to the linear ARMA model. Then, analogous to the linear AR model, the Markov(p) model can be formulated in terms of a updating equation as

$$X_t = F_X^{-1} \left(C_\phi^{-1}[w_t | F_X(X_{t-1}), \dots, F_X(X_{t-p})] \right), \quad w_t \stackrel{iid}{\sim} U(0, 1).$$

where C_ϕ^{-1} is the conditional quantile function corresponding to the copula C_ϕ . Formally $C_\phi^{-1}[x | u_1, \dots, u_p]$ is the solution to $C_\phi[x | u_1, \dots, u_p] = x$. We note that this or similar formulations are also used in other works under different names such as 'updating equation' in [McNeil and Bladt \(2022\)](#) or in descriptions on how to sample from copulas in [Aas et al. \(2009\)](#). This model can be regarded as a generalization of the Gaussian AR(p)-model. This can be seen by choosing C_ϕ as the normal copula and the marginal distribution F_X as the standard normal distribution. Although this is a well-known fact in the literature, see e.g. [Chen and Fan \(2006\)](#) or [Smith et al. \(2010\)](#), we give the derivation for $p = 1$ in [Appendix B.1](#) for completeness. In the following only the pseudo-observation time series, $\{U_t\}_{t=1}^T$, $U_t := F_X(X_t)$, is considered. The actual time series can be recovered by quantile transforming every pseudo-observation by F_X^{-1} . The dynamics of the time series are solely determined by the relation between the pseudo-observations. The model equation of the pseudo-observations is given as,

$$U_t = C_\phi^{-1}[w_t | U_{t-1}, \dots, U_{t-p}], \quad w_t \stackrel{iid}{\sim} U(0, 1). \quad (4)$$

In order to account for possibly different forms of dependency and to be flexible in modeling, D-vine copulas ([Bedford and Cooke, 2001, 2002](#); [Aas et al., 2009](#); [Czado, 2010](#); [Czado and Nagler, 2022](#)) can be employed for modeling the temporal dependence as is also done in [Nagler et al. \(2022\)](#) and [Smith et al. \(2010\)](#). In this approach the joint density of $(p + 1)$ consecutive pseudo-observations is decomposed into a product of bivariate copulas. The decomposition is given by ([Smith et al., 2010](#); [McNeil and Bladt, 2022](#); [Nagler et al., 2022](#))

$$c_{t, \dots, t-p}(u_t, \dots, u_{t-p}) = \prod_{k=t-p}^{t-1} \prod_{j=k+1}^t c_{j-k, j | S_{j-k, j}} \left[F_{j-k | S_{j-k, j}}(u_{j-k}), F_{j | S_{j-k, j}}(u_j) \right], \quad (5)$$

where the shorthand notation $c_{t, \dots, t-p} = c_{U_t, \dots, U_{t-p}}$ is used. Furthermore, $S_{j-k, j} = \{j - k + 1, \dots, j - 1\}$. As in the cited literature, only the simplified D-vine is considered, dropping the conditioning in the bivariate

copula densities, $c_{j-k,j|S_{j-k,j}} = c_{j-k,j}$. In the case where $c_{j-k,j} = c_{i-k,i} := c_k \forall j, i \in \{k+1, \dots, T\}$, the joint density of all random variables $\{U_t\}_{t \in \mathbb{Z}}$ admits translational invariance. Nagler et al. (2022) show that such processes are stationary. Hence we refer to such D-vine copulas as *stationary* D-vine copulas. They also may be called s-vine ('s' for stationary) (McNeil and Bladt, 2022). Furthermore, the bivariate copulas, $C_{j-k,j}$ may be parametrized by parameters $\phi_1, \dots, \phi_p$, where ϕ_1 governs the dependence of U_{t-1} and $U_t \forall t \in \mathbb{Z}$, ϕ_2 the dependence of $U_t|U_{t-1}$ and $U_{t-2}|U_{t-1}$ and so on. Note that depending on the chosen copula, each $\phi_i, i \in \{1, \dots, p\}$ may also be a vector containing more than one parameter.

It is common to denote conditional distributions of pseudo observations by the h -function (Aas et al., 2009; Czado, 2010; Czado and Nagler, 2022). The h -functions associated to a copula C are defined as

$$\begin{aligned} h^1(u_1, u_2) &= \frac{\partial C[u_1, u_2]}{\partial u_1}, \\ h^2(u_1, u_2) &= \frac{\partial C[u_1, u_2]}{\partial u_2}. \end{aligned} \quad (6)$$

If no superscript is given, then conditioning on the second variable is implied. Using this notation, the conditional distribution of e.g. $F_{U_t|U_{t-1}}$ may be rewritten as $F_{U_t|U_{t-1}}(u_t, u_{t-1}) = C_\phi(u_t|u_{t-1}) = h_{\phi_1}^2(u_t, u_{t-1})$. Explicit forms of h -functions and their inverses can be found in Smith et al. (2010) for the normal, t , Clayton, Gumbel and Galambos copula.) An advantage of using the h -functions is that all conditional distributions can be decomposed into a composition of h -functions, i.e. Rosenblatt (forward/backward) functions (McNeil and Bladt, 2022). The decomposition of $U_t|U_{t-1}, \dots, U_{t-p}$ can be called Rosenblatt backward and the decomposition of $U_{t-p}|U_{t-p+1}, \dots, U_t$ Rosenblatt forward function (inspired by the Rosenblatt transform from Rosenblatt (1952)). For $p = 1$ the Rosenblatt functions are just the corresponding h -functions, $R_1^1 = h_1^1, R_1^2 = h_1^2$. McNeil and Bladt (2022) define the Rosenblatt forward and backward functions for $p > 1$, $R_p : (0, 1)^p \times (0, 1) \rightarrow (0, 1)$, recursively as follows:

$$\begin{aligned} R_p^1(\mathbf{u}, x) &= h_p^1(R_{p-1}^2(\mathbf{u}_{-1}, u_1), R_{p-1}^1(\mathbf{u}_{-1}, x)), \\ R_p^2(\mathbf{u}, x) &= h_p^2(R_{p-1}^2(\mathbf{u}_{-p}, x), R_{p-1}^1(\mathbf{u}_{-p}, u_p)), \end{aligned} \quad (7)$$

where $\mathbf{u}_{-j} = (u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_k)$. The second argument of the Rosenblatt functions, x , is the variable that the Rosenblatt functions are evaluated at. The conditioning variables are contained in the vector $\mathbf{u}$. Using the newly introduced notation, the D-vine copula density (Eq. 5) can be rewritten as (McNeil and Bladt, 2022),

$$\begin{aligned} c_{t, \dots, t-p}(u_t, \dots, u_{t-p}) &= \prod_{k=t-p}^{t-1} \prod_{j=k+1}^t c_{j-k,j} \left[R_{k-1}^1((u_{j-k+1}, \dots, u_{j-1}), u_j), \right. \\ &\quad \left. R_{k-1}^2((u_{j-k+1}, \dots, u_{j-1}), u_{j-k}) \right]. \end{aligned}$$

In order to rewrite the time series updating equation (Eq. 4) in the D-vine framework, the inverse Rosenblatt transforms can be utilized:

$$U_t = R_p^{-1}((U_{t-p}, \dots, U_{t-1}), w_t), \quad w_t \stackrel{iid}{\sim} U(0, 1). \quad (8)$$

Here R_p^{-1} is the inverse Rosenblatt function corresponding to the $(p+1)$ -dimensional copula C_ϕ from Eq. 4. The inversion is w.r.t. the last argument. Formally, $R_p^{-1}(\mathbf{u}, x)$ is the solution to $R_p(\mathbf{u}, x) = x$. In this notation, the Markov(p) restriction becomes obvious. Next, we are going to discuss the extension of McNeil and Bladt (2022) and Joe (2014) to allow for infinite partial dependence.

2.3. Review of the extension by McNeil & Bladt

McNeil and Bladt (2022) allow for infinite partial dependence in copula-based time series models by letting the autoregressive order, p , be infinite and reparametrizing the joint density with a finite amount of parameters. The parameters are chosen in such a way that they represent a classical ARMA(r, s) process with finite r and s . To model a time series, the following steps are employed. First, the autoregressive and the moving average order, r and s , respectively, are determined by minimizing model selection criteria (e.g. AIC). Next, the partial autocorrelations of the chosen ARMA(r, s) process are parametrized in terms of the autoregressive and the moving average parameters, $\phi_1, \dots, \phi_r, \theta_1, \dots, \theta_s$. Using the relation between autocorrelations and Kendall's τ , the Kendall's partial autocorrelations are obtained. In a third step, all partial copulas in Eq. 5 are set to a chosen family (e.g. Gaussian or Gumbel) and are reparametrized by Kendall's τ and consecutively by the AR and MA parameters, which can then be estimated by maximum likelihood. Bladt et al. (2025) generalize this approach and consider a Gaussian D-vine, i.e. all copulas in Eq. 5 are Gaussian, and then replace a finite amount of selected copulas by non-Gaussian copulas. Henceforth their approach will be called 'IPD' for infinite partial dependence. The conceptual difference to our MAGMAR-copula model and to the model by Joe (2014) is that their approach is based on classical ARMA models. Furthermore, the partial copulas are modeled and not the AR- and MAG-copulas. We discuss the conceptual differences in more detail below.

2.4. Review of the extension by Joe

Now we want to briefly introduce the model of Joe (2014), which is to the best of our knowledge the earliest copula extension of the ARMA model. The model for a time series $\{U_t\}_{t \in \mathbb{Z}}$ with uniform marginals is given by the following updating equations:

$$W_s = h(\varepsilon_s; W_{s-1}, \dots, W_{s-p}), \quad s \in \mathbb{Z}, \quad (9)$$

$$U_t = g(\varepsilon_t; W_{t-q}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q+1}), \quad t \in \mathbb{Z}, \quad (10)$$

where $\varepsilon_t \stackrel{iid}{\sim} U(0, 1)$ are the common innovations, h is the conditional quantile function of a $(p+1)$ -variate copula and g is the conditional quantile function of a $(q+1)$ -variate copula. It is required that the copula corresponding to h is translationally invariant and that the q -variate marginal distribution of the copula corresponding to g is the independence copula. The idea is inspired by the Poisson ARMA

process of McKenzie (1988): First create a process corresponding to an AR process, Eq. 9, and then feed this process to an MA-like filter together with the common innovations, as in Eq. 10. In our proposed MAGMAR-copula model, we too attempt to mix AR-type and MA-type copula processes. The difference, however, is that we feed the MA-type process (what we call 'MAG') into the AR-type filter, similar to classical ARMA-type models .

2.5. Comparison of the Models with the Proposed Model

We briefly point out the main differences between the model by McNeil and Bladt (2022), the model by Joe (2014) and our proposed model. In Appendix A.7 we compare the models in detail w.r.t. flexibility, computational demand and modeling approach. For brevity we give the conclusions of the comparison here.

The MAGMAR and common-innovation model are similar in spirit. However, the MAGMAR-model has a more straight-forward modeling approach. In return, the common-innovations approach can be expected to be more tractable due to the stationary distribution naturally being standard-uniform. The additional adjustment transformation needed for the MAGMAR-model hinders theoretical tractability. The computational demand can be expected to be similar. The flexibility of the common-innovations approach is a potential direction for future research, since it is not obvious due to the AR-part being latent.

The modeling approach of the MAGMAR and IPD model are similar as both models can rely on e.g. autodependence and scatter plots. The main difference is that the partial copulas of $(U_t, U_{t-k} | \mathcal{F}_{t-1:t-k+1})$ result from the AR and MAG-copula in the MAGMAR model, and are typically non-standard, whereas in the IPD model, the partial copulas can be chosen directly. This results in the MAGMAR model being preferable if the long-term dependencies are non-Gaussian. If the long-term dependencies are Gaussian and only a few partial dependencies seem to be non-linear, the IPD model is more suitable. While both models have moderate computational demand, the IPD models demand is less.

3. MAGMAR-Copulas

In this section, we introduce and discuss the moving aggregate modification to copula-based time series models. For illustration, the autoregressive and moving aggregate will be set to one. Results for arbitrary orders p and q can be derived in a similar fashion but are beyond the scope of this paper; however, some results with arbitrary orders p and q are given in the Appendix.

3.1. Introduction to the model

Our proposed approach enabling the model to overcome the Markov restriction amounts to modifying the autoregressive updating equation, Eq. 8, by a moving aggregate part. The moving aggregate part is, similar to classical, linear ARMA processes, a function of the current and past innovations. It is given as a conditional quantile function, corresponding to a copula, which we parametrize with parameter(s) θ . So instead of w_t in the updating equation, Eq. 8, we have $(R_{\theta_1, \dots, \theta_q}^1)^{-1}((w_{t-q}, \dots, w_{t-1}), w_t)$. Where the inversion is again w.r.t. the last element. We call the model Moving Aggregate Modified Autoregressive

Copula-Based Time Series Model (MAGMAR-Copula model). Below we formally define the model with arbitrary model orders p and q . The exposition that follows focuses on the case $p = q = 1$. Furthermore, we assume that 1) all copulas are differentiable, 2) that they are within the Fréchet-Hoeffding bounds (see [Joe, 2014](#), Chapter 2.9) and 3) that the inverses of h -functions associated to all used copulas exist. Now we formally introduce the model.

Definition 1 (MAGMAR(p, q)-Copula Model). Let $R_{\phi_1, \dots, \phi_p} : (0, 1)^{p+1} \rightarrow (0, 1)$ be a Rosenblatt function associated to a stationary D-vine copula $C_{\phi_1, \dots, \phi_p}$, which is parametrized by parameters $\phi_1, \dots, \phi_p$. Furthermore, let $R_{\theta_1, \dots, \theta_q}^{-1} : (0, 1)^{q+1} \rightarrow (0, 1)$ be an inverse Rosenblatt function associated to a D-vine copula $C_{\theta_1, \dots, \theta_q}$ with independent margins, which is parametrized by parameters $\theta_1, \dots, \theta_q$. The time series $\{U_t\}_{t \in \mathbb{Z}}$ following the updating equation

$$R_{\phi_1, \dots, \phi_p}^1((U_{t-p}, \dots, U_{t-1}), U_t) = (R_{\theta_1, \dots, \theta_q}^1)^{-1}((w_{t-q}, \dots, w_{t-1}), w_t) \quad (11)$$

is a MAGMAR(p, q)-time series. The corresponding model is the MAGMAR(p, q)-Copula model. The copula $C_{\phi_1, \dots, \phi_p}$ is the AR-Copula. The copula $C_{\theta_1, \dots, \theta_q}$ is the MAG-Copula. The LHS of Eq. 11 is the AR-part and the RHS is the MAG-part. The time series following the updating equation

$$U_t = (R_{\theta_1, \dots, \theta_q}^1)^{-1}((w_{t-q}, \dots, w_{t-1}), w_t) \quad (12)$$

is the *MAG(q)-copula model*. Furthermore, we define the *adjusted MAGMAR(p, q)-copula time series* as $\tilde{U}_t = \Psi(U_t)$, $\forall t \in \mathbb{Z}$, where Ψ is a distribution function on $(0, 1)$.

The Rosenblatt functions in Eq. 11 ought to be understood as a composition of h -functions (as given in Eq. 7). The AR-Rosenblatt function, $R_{\phi_1, \dots, \phi_p}^1$ is composed of h -functions $h_{\phi_1}, \dots, h_{\phi_p}$ while the MAG-Rosenblatt function is composed of h -functions $h_{\theta_1}, \dots, h_{\theta_q}$. The AR-copula is a stationary D-vine copula, cf. Sect. 2.2 or [McNeil and Bladt \(2022\)](#). The MAG-copula has independent margins. This property is also imposed by [Joe \(2014\)](#). The family of the copulas (e.g. normal, t , Gumbel, etc.) has to be chosen. Parameters ϕ and θ have to be set/estimated. We note that the MAG-part, that is used here, was also derived in [McNeil and Bladt \(2022\)](#) as the 'causal representation' for the classical copula-based time series models. In [Joe \(2014\)](#), it is also given as the ' q -dependent copula extension'. The new contribution is that we use this representation to construct the novel MAGMAR-Copula model. The reason for the adjustment, using a distribution function Ψ , is the following: The unconditional stationary distribution of a MAGMAR(p, q)-copula time series, $\{U_t\}$ is in general not uniform if $p, q > 0$. The distribution that arises is a distribution on $(0, 1)$, and it depends on the AR and MAG parameters (ϕ, θ) . The probability integral transformation using Ψ then ensures that $\tilde{U}_t \sim U(0, 1)$. Notice that we did not prove that the stationary distribution of the MAGMAR-copula time series actually exists yet. The corresponding proposition is given in Sect. 3.3 for the MAGMAR(1, 1)-copula model. This is the reason for introducing Ψ as a distribution function on $(0, 1)$ and not as the stationary distribution of the process in Def. 1. In applications, we estimate Ψ with non-parametric methods in a complementary step. In the following, we mostly consider the unadjusted time series $\{U_t\}$, since the transformation does not affect most results.

When appropriate, we also give the results for the adjusted time series.

Before giving the relations to other models and deriving distributional properties, we proceed to discuss implications of the model equation and further explain the model for the case where $p = q = 1$. In this case the updating equation is given as

$$U_t = h_\phi^{-1}(h_\theta^{-1}(w_t, w_{t-1}), U_{t-1}). \quad (13)$$

First, we note that C_θ is **not** the copula of w_t and w_{t-1} . The innovations w_t and w_{t-1} are independent and their copula is the independence copula, $F_{w_t, w_{t-1}} = \Pi$. The expression $h_\theta(w_t, w_{t-1})$ is to be understood as a function of $(q+1)$ random variables w_t and w_{t-1} , just as $\varepsilon_t - \theta\varepsilon_{t-1}$ in the classical ARMA model is a function of two random variables ε_t and ε_{t-1} . Second, the copula C_ϕ is also **not** the copula of U_t and U_{t-1} , however it is connected to the copula of the two random variables. This is analogous to the classical ARMA(1,1) model, where the partial autocorrelation of X_{t-1} and X_t is not solely determined by the AR(1) parameter but also by the MA(1) parameter. (The influence of the MAG-Copula and the AR-Copula to the dependence of U_t and U_{t-1} becomes clear when we derive the joint conditional distribution of U_t and U_{t-1} given $\mathcal{F}_{t-2}$.) The expression $h_\phi(h_\theta^{-1}(w_t, w_{t-1}), U_{t-1})$ may also be understood as a function of random variables. To verify that the MAGMAR(1,1) model equation actually yields the expected behavior to have no Markov restriction i.e. an implicit infinite autoregressive order, the model equation can be iterated. To this aim, an explicit form for the innovations is necessary. It can be directly obtained from Eq. 13 as

$$w_t = h_\theta(h_\phi(U_t, U_{t-1}), w_{t-1}).$$

Iteratively substituting for w_{t-1} in Eq. 13 yields the AR(∞) representation. Utilizing the composition symbol, which acts as $\bigcirc_{i=1}^N f_i = f_1 \circ f_2 \circ \dots \circ f_N$, the representation can be written as

$$\begin{aligned} w_t &= h_\theta(h_\phi(U_t, U_{t-1}), \cdot) \circ h_\theta(h_\phi(U_{t-1}, U_{t-2}), \cdot) \circ \dots, \\ &= \bigcirc_{i=0}^{\infty} h_\theta(h_\phi(U_{t-i}, U_{t-i-1}), \cdot). \end{aligned} \quad (14)$$

Hence, in general, each past observation, $U_{t-i}, i = 1, 2, \dots$ has an actual influence on U_t . This is the desired property. Furthermore, the MAG(∞) representation can be derived as

$$\begin{aligned} U_t &= h_\phi^{-1}(h_\theta^{-1}(w_t, w_{t-1}), \cdot) \circ h_\phi^{-1}(h_\theta^{-1}(w_{t-1}, w_{t-2}), \cdot) \circ \dots, \\ &= \bigcirc_{i=0}^{\infty} h_\phi^{-1}(h_\theta^{-1}(w_{t-i}, w_{t-i-1}), \cdot). \end{aligned} \quad (15)$$

As noted before, there also exists a MAG(∞) representation for the case with no moving aggregate extension. The representation was derived in McNeil and Bladt (2022) as 'causal representation'. It is easy to see that for $C_\theta = \Pi$, the two representations are equal. This is in line with proposition 1, which proposes that for $C_\theta = \Pi$, the MAGMAR(1,1)-Copula model is equal to the classical copula-based time series model. The corresponding proposition is given in Sect. 3.2. In Sect. 3.3, it is explicitly shown

that the $\text{MAG}(\infty)$ representation of the MAGMAR-Copula time series converges for the case where the AR-copula, C_ϕ , is the normal copula. For other copulas, it is hard to establish the convergence of the $\text{MAG}(\infty)$ representation explicitly. However, we can use results from Douc et al. (2014), involving a contraction-on-average, to establish desired properties.

3.2. Relation to other models

Now we derive the relation of the proposed MAGMAR-Copula model to the classical copula-based time series model and the classical linear ARMA model. The results are given for the $\text{MAGMAR}(1, 1)$ case but can be naturally extended to the $\text{MAGMAR}(p, q)$ case. Furthermore we define the MAG-copula time series and show that it is stationary.

Proposition 1. *Let $\{U_t\}_{t \in \mathbb{Z}}$ be a $\text{MAGMAR}(1, 1)$ -Copula time series with AR-Copula C_ϕ and the MAG-Copula, C_θ . The following holds.*

- i) *If C_ϕ and C_θ are chosen as the normal copulas, the quantile transformed time series $\{\Phi^{-1}(U_t)\}_{t \in \mathbb{Z}}$ is the classical Gaussian and linear $\text{ARMA}(1, 1)$ model with AR-parameter ϕ , MA-parameter $\frac{\theta}{(1-\theta^2)^{\frac{1}{2}}}$ and independent normally distributed innovations with variance $(1-\phi^2)(1-\theta^2)$.*
- ii) *If the MAG-copula is the independence copula, $C_\theta = \Pi$, then the classical copula-based time series model (see Sect. 2.2) is recovered.*
- iii) *If the AR-copula is the independence copula, then the $\text{MAG}(1)$ -Copula time series is recovered.*
- iv) *If both the AR and MAG-copula are chosen as the independence copula, the time series $\{U_t\}_{t \in \mathbb{Z}}$ is an iid standard uniform time series.*

Proof. See Appendix Appendix B.1. □

The first two statements show that the classical linear ARMA model as well as the classical copula-based time series model is nested by the MAGMAR-Copula model. Part ii) also justifies describing the copula-based time series model as $\text{MAGMAR}(1, 0)$ -Copula model in further discussions. Next we define the MAG-Copula model.

Definition 2 ($\text{MAG}(1)$ -Copula model). Let $\{U_t\}_{t \in \mathbb{Z}}$ be a $\text{MAGMAR}(1, 1)$ -Copula time series. If the AR-copula is set to be the independence copula, $C_\phi = \Pi$, the resulting model is the $\text{MAG}(1)$ -Copula model.

We note again that the $\text{MAG}(1)$ -Copula model coincides with the 1-dependent model from Chapter 3.14 of Joe (2014). It is easy to see that a $\text{MAG}(1)$ -Copula time series, $\{V_t\}_{t \in \mathbb{Z}}$ follows the updating equation

$$V_t = h_\theta^{-1}(w_t, w_{t-1}), \quad w_t \stackrel{iid}{\sim} U(0, 1).$$

Now we establish the stationarity of the $\text{MAG}(1)$ -Copula model and show that the unconditional marginal distribution is given as standard uniform.

Proposition 2. *Let $\{U_t\}_{t \in \mathbb{Z}}$ be a MAG(1) time series. Then $\{U_t\}_{t \in \mathbb{Z}}$ is strictly stationary, ergodic and $U_t \sim U(0, 1)$, $\forall t \in \mathbb{Z}$.*

Proof. See [Appendix B.2](#) □

The models distributional properties are important for further work. They play an important role in establishing the stationarity of the MAGMAR(1,1)-Copula model with the normal AR-Copula, see [Sect. 3.3.1](#).

3.3. Distributional Properties of the model

In this section, the distributional properties of random variables $\{U_i\}_{i \in \mathbb{Z}}$, following the MAGMAR-Copula model, are derived. In order to keep the findings comprehensible, the ideas are developed for the MAGMAR(1,1) case. They can be generalized to the MAGMAR(p, q) case. First, we show that the process is stationary for the case where the AR-copula is the normal copula and the parameter fulfills $|\phi| < 1$. Second, we derive the conditional distribution of U_t given past observations. Third, using the law of total probability and conditional independence we derive the joint distribution of $\{U_i\}_{i \in \mathbb{Z}}$.

3.3.1. Stationarity

We first show that the process is stationary for the case where the AR-copula is the normal copula. Then we generalize the result for the case, where the AR-copula fulfills a contraction-on-average property. In order to establish stationarity for the normal AR-copula case, we consider the MAG representation of the MAGMAR(1,0)-normal-Copula time series and show that it does converge almost surely for all t . Together with the stationarity of the MAG(1)-Copula model, the stationarity of the MAGMAR(1,1)-Copula model will follow. The result is a special case of the more general result on MAGMAR(1,0) models, where the copula on average induces a contraction in [proposition 5](#). We still propose the result for the normal AR-copula for illustrational purpose

Proposition 3. *Let $\{U_t\}$ be a MAGMAR(1,0)-Copula time series where the AR-Copula is the normal copula, $C_\phi = C_\phi^{\text{normal}}$, with parameter $|\phi| < 1$. Then the MAG(∞) representation,*

$$A_{t,n} := \bigcirc_{i=0}^n h_\phi^{-1}(w_{t-i}, \circ)$$

does converge almost surely as $n \rightarrow \infty$ and $\lim_{n \rightarrow \infty} A_{t,n} = U_t$, $U_t \sim U(0, 1)$ almost surely.

Proof. See [Appendix B.3](#) □

Notice that the almost sure convergence also holds when instead of an *iid* sequence w_t , a dependent, but stationary, sequence is plugged in. This is in line with Theorem 36.4 from [Billingsley \(1995\)](#). Furthermore, from [Lemma 3](#) it follows that the MAGMAR(1,0)-Copula time series is strictly stationary. We want to note, that the MAG(∞)-representation of the MAGMAR(1,0)-Copula time series is analogous to the MA(∞) representation in the classical linear ARMA model (see [Hamilton, 2020](#)) and is equivalent to the representation derived in [McNeil and Bladt \(2022\)](#). In principle the MA(∞) representation of the classical linear ARMA model can also be written as a infinite composition. However, there the composed functions

are linear and, hence, also their composition is linear. We noted earlier that the *a.s.* convergence in proposition 3 also holds if the innovation time series $\{w_t\}_{t \in \mathbb{Z}}$ is not *iid* but merely stationary. In the next proposition we exploit this fact to show that the MAGMAR(1, 1)-Copula time series with the AR-Copula being the normal copula is stationary.

Proposition 4. *Let $\{U_t\}_{t \in \mathbb{Z}}$ be a MAGMAR(1, 1)-Copula time series where the AR-Copula is the normal copula, $C_\phi = C_\phi^{\text{normal}}$, with parameter $|\phi| < 1$. Then $\{U_t\}_{t \in \mathbb{Z}}$ is strictly stationary.*

Proof. See [Appendix B.4](#) □

Now we generalize the result from proposition 4 for general bivariate copulas that fulfill a contraction-on-average assumption.

Proposition 5 (AR-Copula contraction). *Let $\{U_t\}_{t \in \mathbb{Z}}$ be a MAGMAR(1, 1)-Copula time series. Denote the AR copula by C_α and the associated inverse h -function as $h_\alpha^{-1}(\cdot, \cdot)$. If there is a bijective mapping $\kappa : \mathbb{R} \rightarrow (0, 1)$ such that $f_z : (0, 1) \rightarrow \mathbb{R}$ with $f_z(x) = \kappa^{-1}[h_\alpha^{-1}(z, \kappa(x))]$ fulfills assumptions A4.37 and A4.38 from [Douc et al. \(2014\)](#), then $\{U_t\}$ is stationary and ergodic.*

Proof. See [Appendix B.5](#) □

The conditions have to be checked for each AR-copula individually. In particular, a suitable mapping κ has to be found. In [Appendix A.5](#) we check the conditions for the normal, t and Clayton copula. The generalization of the result for $q > 1$ is straight-forward, since the resulting MAG(q) process is still stationary and ergodic. The generalization for $p > 1$ is not straight-forward. The typical approach would be to extend the updating equation to a p -dimensional updating equation and to show that the corresponding mapping induces a contraction. This approach is not straight-forward for copula-based models because the inverse Rosenblatt transformation is hard to tract and it is not clear how to establish a contraction. This question will be the subject of future research. Note that the proposition now enables us to consider the adjustment transformation Ψ from Def. 1 as the unconditional stationary distribution function of the process, since we now proved such a distribution exists under suitable conditions. Now we proceed to derive the conditional distributions of the model.

3.3.2. Conditional distribution

The conditional distribution of $U_t | \mathcal{F}_{t-1}$, where $\mathcal{F}_s := \sigma(\{X_r\}_{r \leq s})$, can be derived by using the density transformation theorem (see e.g. [Devore et al., 2012](#), Sect. 4.7). We can interpret the updating equation of the MAGMAR(1, 1)-Copula model conditional on $\mathbf{U}_{t-1} := (U_{t-i})_{i=1}^\infty = (U_t, U_{t-1}, \dots)$ as follows: The random variable $w_t \sim U(0, 1)$ is transformed by a function $g_{\mathbf{U}_{t-1}} : (0, 1) \rightarrow (0, 1)$ to the random variable U_t , i.e. $U_t | \mathcal{F}_{t-1} = g_{\mathbf{U}_{t-1}}(w_t) | \mathcal{F}_{t-1}$. The other random variables, $\mathbf{U}_{t-1} = (U_{t-i})_{i=1}^\infty$, are regarded as deterministic due to conditioning. We give an explicit form for $g_{\mathbf{U}_{t-1}}$ in the proof of the following proposition.

Proposition 6. Let $\{U_t\}_{t \in \mathbb{Z}}$ be a MAGMAR(1,1)-Copula time series. Furthermore, let the AR-Copula, C_ϕ , and the MAG-Copula, C_θ , be differentiable. Then the conditional density of $U_t | \mathcal{F}_{t-1}$ is given as

$$f_{U_t | \mathcal{F}_{t-1}}(x) = c_\theta(h_\phi(x, U_{t-1}), w_{t-1}(\{U_{t-i}\}_{i=1}^\infty)) c_\phi(x, U_{t-1}). \quad (16)$$

where c_θ is the copula density associated to the MAG-Copula and c_ϕ is the density associated to the AR-Copula. The expression $w_{t-1}(\{U_{t-i}\}_{i=1}^\infty)$ is to be understood as a function of conditioning variables $(U_{t-i})_{i=1}^\infty$, which are regarded as constant in the conditional density. For the adjusted time series, $\{\tilde{U}_t\}_{t \in \mathbb{Z}} = \{\Psi(U_t)\}_{t \in \mathbb{Z}}$, the conditional density is given as

$$f_{\tilde{U}_t | \mathcal{F}_{t-1}}(x) = c_\theta(h_\phi(\Psi^{-1}(x), \Psi^{-1}(\tilde{U}_{t-1})), w_{t-1}(\{\tilde{U}_{t-i}\}_{i=1}^\infty)) c_\phi(\Psi^{-1}(x), \Psi^{-1}(\tilde{U}_{t-1})) \\ \times \frac{1}{\psi(\Psi^{-1}(x))},$$

where ψ is the density corresponding to the distribution function Ψ .

Proof. See [Appendix B.6](#) □

It is evident from the conditional density, Eq. 16, that $U_t | \mathcal{F}_{t-1} = U_t | (U_{t-i})_{i=1}^\infty = U_t | (U_{t-1}, w_{t-1})$. This means that conditioning on all past random variables $(U_{t-i})_{i=1}^\infty$ is equivalent to conditioning on (U_{t-1}, w_{t-1}) . This is the same property as in the classical linear ARMA model. Note that for the special case where c_θ is the independence copula density, $c_\theta = 1_{(0,1)^2}$, the conditional density of the copula-based time series model with autoregressive order of one is recovered. In the MAGMAR(p, q) case, the conditional density can be derived in a similar fashion. The random variable w_t is transformed to the random variable U_t by an appropriate transformation function $g_{U_{t-1}}$. The resulting density is a rather complicated expression which is given in [Appendix A.3](#). Furthermore we note that multivariate conditional densities such as $f_{U_t, U_{t-1} | \mathcal{F}_{t-2}}$ can be derived in a similar way. The derivation of joint conditional densities for the MAGMAR(1,1) model can be found in [Appendix A.4](#).

3.3.3. Joint distribution and Likelihood

To derive the joint distribution of t consecutive random variables from the time series $\{U_t\}_{t=1}^T$ following the MAGMAR-Copula model, conditional independence and the law of total probability ([Schervish, 2012, B.3.5](#)) is utilized. The joint distribution of the t random variables $(U_t, \dots, U_1)$ is given as

$$f_{U_t, \dots, U_1}(u_t, \dots, u_1) = f_{U_1}(u_1) \prod_{i=0}^{t-2} f_{U_{t-i} | \mathcal{F}_{t-i-1}}(u_{t-i} | \mathcal{F}_{t-i-1}). \quad (17)$$

For the MAGMAR(1,1)-Copula model, the conditional distribution from Eq. 16 can be plugged in for $f_{U_{t-i} | \mathcal{F}_{t-i-1}}$. For the MAGMAR(p, q)-model the conditional densities from [Appendix A.3](#), Eq. A.2 need to be used. In spite of the factorization, the infinite partial dependence is carried by the $\{w_i\}_{i=1}^T$. For maximum likelihood estimation purposes, the initial density, f_{U_1} , may be dismissed, hence resulting in pseudo-MLE. It is easy to verify (via integration by substitution) that $\int f_{U_t, \dots, U_1}(u_t, \dots, u_1) du_t = f_{U_{t-1}, \dots, U_1}$, which is to be expected of a joint density. Using the joint density derived in this section,

maximum likelihood estimation can be performed. This is discussed in the next section.

3.4. Estimation & Modeling

In this section we describe how a univariate, strictly stationary time series may be modeled with the MAGMAR-Copula model. First, we describe the (pseudo-)maximum likelihood estimation given fixed AR and MAG orders and fixed bivariate copulas. Second, we describe challenges in modeling and model selection.

Given a fixed AR order p , fixed MA order q , fixed bivariate copula specifications $C_{\phi_1}, \dots, C_{\phi_p}$ and $C_{\theta_1}, \dots, C_{\theta_q}$ for the respective D-vines and observations $(u_T, \dots, u_1)$, the estimation can be performed by (pseudo-)maximum likelihood estimation. The likelihood can be derived directly from the joint density (Eq. 17). The pseudo-log-likelihood of the MAGMAR(1, 1)-Copula model of T observations is given by

$$\log L(\phi, \theta) = \sum_{i=0}^{T-1} \left(\log c_{\theta}[h_{\phi}(u_{T-i}, u_{T-i-1}), w_{T-i-1}] + \log c_{\phi}[u_{T-i}, u_{T-i-1}] \right). \quad (18)$$

The likelihood of the MAGMAR(p, q)-Copula model is given in a completely analogous way using the Eq. A.2 from Appendix A.3 for the conditional densities. We note that this likelihood is only valid if the true data-generating process is a MAGMAR(p, q) process. When marginal distributions and the adjustment transformation Ψ are included, an additional term emerges in accordance with proposition 6. If the MAG-copula is set to the independence copula, the likelihood for the classical copula-based time series emerges. Chen and Fan (2006) establish desirable asymptotic properties (consistency and asymptotic normality) of the estimators for the case where the marginal distribution is estimated non-parametrically. More generally Nagler et al. (2022) establish these properties for multivariate stationary time series and arbitrary lags. The peculiarity of the likelihood in Eq. 18 is that it includes the term $\log c_{\theta}[h_{\phi}(u_{T-i}, u_{T-i-1}), w_{T-i-1}]$, originating from the MAG-copula. The following proposition establishes the consistency of ML-estimators for the MAG(1) model. The extension for the MAG(q) model follows naturally.

Proposition 7 (Consistency of MLE of MAG(1) copula process). *Let $\{U_t\}_{t \in \mathbb{Z}}$ be a MAG(1)-Copula time series with MAG-copula C_{θ_0} . Let the parameter space Θ be compact and the MAG-Copula continuously differentiable w.r.t. parameters $\theta \in \Theta$ and continuous in its arguments, $(u, v) \in (0, 1)^2$. Furthermore, let the time series of residuals, $\{\hat{w}_t\}_{t \in \mathbb{Z}}$ given as $\hat{w}_t = h_{\theta}(U_t, \hat{w}_{t-1})$, fulfill Assumptions A4.36-A4.38 from Douc et al. (2014). Let $l_t(\theta) = -\log c_{\theta}[U_t, \hat{w}_{t-1}]$, then the (pseudo) maximum-likelihood estimator*

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{t=2}^T l_t(\theta)$$

is a consistent estimator for the unique $\theta_0 = \arg \min_{\theta \in \Theta} \mathbb{E}(l_t(\theta))$.

Proof. See Appendix B.7 □

We note that the crucial step in establishing consistency is the stationarity and ergodicity of the time series of residuals, $\{\hat{w}_t\}$. To this end, theorem 4.40 from [Douc et al. \(2014\)](#) is used. The theorem is based on a contraction-on-average, which has to be established by verifying Assumptions A3.36-A3.38 for each copula individually. The calculations are similar to the ones for establishing stationarity and ergodicity of the MAGMAR(1, 1) time series model. The major difference is that the U_t , which act as innovations are not independent and that the updating equation involves the conditional CDF associated to the MAG-copula and not the conditional quantile function. A natural extension of proposition 7 would be asymptotic normality. However, to establish such result a uniform WLLN for the Hessian matrix is needed, cf. condition M.5(i) of Theorem 4.4 by [Wooldridge \(1994\)](#). This is beyond the scope of this paper. Nevertheless, the simulation study in Sect. 4.1 indicates that asymptotic normality holds for the MAGMAR(1, 1) model.

The likelihood has to be calculated using an iterative scheme and initial values have to be set. The algorithm for calculating the likelihood of a MAGMAR(p, q)-Copula time series with given AR & MAG orders, p & q , fixed bivariate copula specifications, $C_{\phi_1}, \dots, C_{\phi_p}$ and $C_{\theta_1}, \dots, C_{\theta_q}$ for the respective D-vines, fixed parameters $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ and observations $(u_T, \dots, u_1)$ is in spirit the same as for ARMA or GARCH models. For completion we provide the detailed steps in [Appendix A.1](#).

Now we turn our attention to challenges in general modeling. While the MLE is well-behaved, if it is assumed that the observed data is generated from a MAGMAR(p, q) time series model, there are some subtleties in modeling real data. When modeling a time series with copula-based approaches, the first step is to estimate the unconditional stationary distribution of the time series and transform the data to pseudo-observations, which are distributed standard-uniformly if the unconditional stationary distribution is correctly specified. The problem with this approach in the MAGMAR-context is that the unconditional stationary distribution of the regular MAGMAR model is not standard-uniform. This is why the adjusted MAGMAR(p, q) model is introduced in Def. 1. In principle it is possible to jointly estimate the adjustment transformation Ψ and the MAGMAR model, however, this approach is computationally very demanding. This is why we suggest the following approach: First we fit a regular MAGMAR-model on the pseudo-observation data. This will be misspecified. Nevertheless, we then use the estimated Model to obtain the implied non-standard-uniform stationary distribution and quantile-transform the pseudo-observation data w.r.t. this distribution. Then the estimation is repeated on this transformed data. This procedure can be repeated. In our simulation study we show that two iterations yield satisfactory results on a MAGMAR(1, 1) model.

Now regarding model selection, in contrast to linear models (i.e. ARMA), model selection in the MAGMAR-Copula context does not only amount to selecting appropriate model orders, i.e. AR & MAG orders, but also appropriate copulas. In popular vine copula software such as the [VineCopula](#) package in R ([Schepsmeier et al., 2015](#)), 38 bivariate copula specifications are available. Hence estimating all possible models is computationally expensive. In order to be able to perform sensible model selection, we suggest not solely relying on information criteria but also doing pre-selection based on descriptive statistics i.e. pair plots, acf, pacf or kpacf (see [McNeil and Bladt, 2022](#), for the kpacf) and tail dependencies. Similar to the (p)acf for classical linear ARMA models, (partial) auto-dependence measures can give information

about sensible lag orders for the MAGMAR-copula model. If the partial auto-dependence measures are cut-off after some lag order, a MAG part is not necessary, because a cutoff partial auto-dependence is equivalent to the process being Markov, assuming that the relations are monotone. We formally prove this in [Appendix A.6](#). When the partial auto-dependence measures are not cut-off, a long auto-dependence is indicated hence requiring inclusion of a MAG part in the model. The tail dependence measures can give information about sensible bivariate copula choices. No tail dependence indicates linear relations, while symmetric tail dependence indicates heavy-tailed dependence. Asymmetric tail dependence naturally indicates asymmetric copulas.

For further use, we introduce new notation. First, we abbreviate the t, normal, Gumbel, Joe and independence copulas by 't', 'n', 'g', 'j' and 'i', respectively. By the following example, we explain the new notation: We let $\text{MAGMAR}(3,2)\text{-(n,g,t)-(i,t)}$ denote the MAGMAR-Copula model with 3 AR-copulas and 2 MAG-copulas. The first AR-copula, governing the dependence of U_t and U_{t-1} , is the normal copula, indicated by 'n'. The second AR-copula is the Gumbel copula indicated by 'g' and the third AR-copula is the t-copula, indicated by 't'. The first MAG-copula is the independence copula, indicated by 'i' and the second MAG-copula is the t-copula, indicated by 't'. In the next section we model the US inflation with the MAGMAR-Copula model using the likelihood derived in this section.

4. Simulation Study

In order to test the proposed model, we conduct the following simulation studies. First, we explore asymptotic properties of the estimators. To this end we re-run the MLE and use QQ-plots to assess the validity of asymptotic normality and consistency. Second, we exemplify the modeling approach on a toy example and last we compare the approach with other copula-based approaches.

4.1. Asymptotics

We are interested in consistency and asymptotic normality of the maximum likelihood estimators in the MAGMAR-model. To this end, we simulate $m = 200$ realizations of a $\text{MAGMAR}(1,1)\text{-(g)-(n)}$ process with length $n = 1000$. The parameters are chosen as $\phi = 1.4$ for the Gumbel AR-copula and $\theta = 0.3$ for the Gaussian MAG-copula. Based on each of the m realizations of the process, we estimate the parameters of the above model by MLE, leaving us with an ensemble of $m = 200$ realizations of the AR and MAG parameter estimators $\hat{\phi}$ and $\hat{\theta}$. Based on these estimators consistency and asymptotic normality may be inspected. To this end the QQ-plots of both CLT statistics of the estimators are inspected. The CLT statistics are calculated as $Z_{i,n}^{(\phi)} = \frac{\hat{\phi}_i - \phi_0}{\text{sd}(\hat{\phi})}$ (and analog for θ). The resulting QQ-plots are given in [Fig. 3](#). We observe that 1) the estimators in fact seem to be fluctuating around the true underlying value and 2) the fluctuations scale with $\sqrt{n}$ and approach normality for large-sample sizes, as is expected. As a side effect of repeating the estimation $m = 200$ times for the $\text{MAGMAR}(1,1)$ model, we can also estimate the average computational demand for estimating the parameters of such model. The estimated average run time is roughly 1min per estimation.

Next we investigate whether the iterative fitting approach described above leads to asymptotic normality. To this end we simulate from an adjusted MAGMAR-(g)-(t) model with $\phi = 1.4$ and $(\rho, \nu) = (0.3, 4.2)$

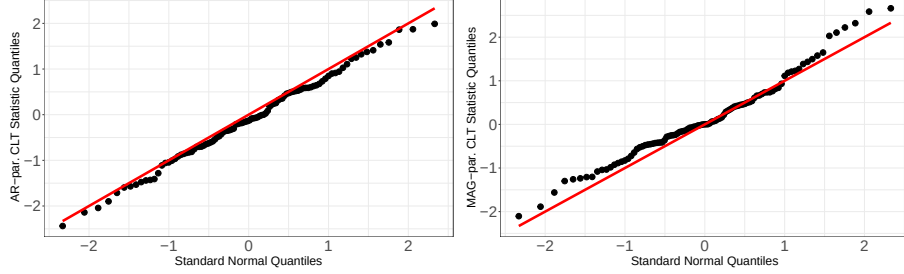


Figure 3: QQ-plot for the CLT statistic of the AR and MAG-parameters of a MAGMAR(1,1) model. The quantiles are obtained by 1) simulating $m = 100$ MAGMAR(1,1) processes of length $n = 1000$, 2) estimating the parameters $m = 100$ times, 3) estimating the standard deviation of the estimator based on the $m = 100$ estimates 4) calculating the CLT statistic $Z_{i,n} = \frac{\hat{\phi}_i - \bar{\phi}}{\text{sd}(\hat{\phi})}$, where $i \in \{1, \dots, m\}$ 5) calculating the sample quantiles.

and quantile-transform the simulated time series w.r.t. the standard normal quantile function. This is a more realistic modeling scenario than the one above. To estimate the adjusted MAGMAR-copula model, we first transform the time series $\{Y_t\}$ to pseudo-observations $\{U_t\}$ by applying the PIT w.r.t. the empirical CDF. Then we perform two iterations of estimating the copula parameters and the stationary distribution, as described in Sect. 3.3.3. This procedure is repeated $m = 200$ times. We observe that the first-iteration estimators are in fact biased. The average of e.g. $\{\hat{\phi}_i\}_{i=1}^m$ is $\bar{\phi} = 1.56$ with the true value being $\phi = 1.4$. This is expected because the marginal distribution of the pseudo-observations and the one of the MAGMAR model differ. The estimators of the second iteration, however, are well-behaved with average estimates $(\bar{\phi}, \bar{\rho}, \bar{\nu}) = (1.41, 0.29, 4.50)$. In Fig. D.6 in Appendix D we also provide the QQ-plots of the respective CLT-statistics, which indicate asymptotic normality for the estimators $\hat{\phi}$ and $\hat{\rho}$. Asymptotic normality for $\hat{\nu}$ does not seem to hold.

4.2. Modeling Toy Example

In this section we exemplify our modeling approach for the MAGMAR-Copula model. To this end we simulate $n = 1000$ observations $\{\tilde{U}_t\}_{t=1}^n \sim \Psi\text{-MAGMAR}(1,1)\text{-(g)-(t)}$ with $\phi = 1.4$ and $\theta = (0.3, 4)$, and quantile transform to $Y_t = \Phi^{-1}(\tilde{U}_t)$ for all t . Then we follow the modeling approach laid out in Sect. 3.3.3. The exemplary time series is given in the upper panel of Fig. 4. Now we treat the time series as if we do not know the true data generating process and try to model it with the MAGMAR-copula model. As the first step, we transform the time series to pseudo-observations by using the empirical CDF, yielding the time series $\{V_t\}_{t=1}^n$. Second, we run automated model selection of the model orders p and q using the `auto.arima` command from the `forecast`-package in R, see Hyndman and Khandakar (2008). This yields $p = q = 1$. Third we perform automatic model selection for the copula of (U_t, U_{t-1}) to get an idea of suitable copulas. Model selection by BIC yields a Gumbel copula. We note that this copula is distinct from the AR-copula, however we can still use it as starting point. Model selection by AIC yields a t -copula for (U_t, U_{t-1}) . Hence, we proceed to fit an adjusted-MAGMAR(1,1)-(g)-(t) and an adjusted-MAGMAR(1,1)-(t)-(t) model to the data. We recall that the purpose of the adjustment transformation Ψ is to ensure that the resulting process has a standard uniform stationary distribution. For this we follow the procedure proposed in Sect. 3.3.3 and estimate the parameters and the adjustment

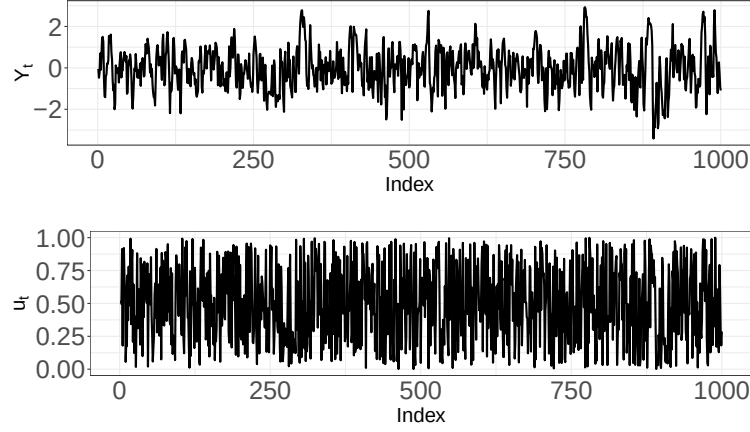


Figure 4: The simulated time series $\{Y_t\}$ and residuals of a MAGMAR(1,1) model fitted to the time series.

transformation iteratively with two iterations. Comparison of information criteria then shows that the $\Psi^{(2)}$ -MAGMAR(1,1)-(g)-(t) model is more suitable. In the lower panel of Fig. 4 the residuals of this model are plotted. We observe that the MAGMAR(1,1)-(g)-(t) model successfully captures the dynamics.

4.3. Comparison with other Non-Markovian Copula-Based Models

We compare the proposed model with Markovian copula-based time series and the infinite partial dependence model by McNeil and Bladt (2022). We use the simulated time series from Sect. 4.2 and fit these models to the data. For the standard copula-based time series, we perform order selection again using the `auto.arima` command (cf. Hyndman and Khandakar (2008)). This yields $p = 3$. We then select the copula of (U_t, U_{t-1}) automatically by BIC, calculate the conditional variables $(U_t|U_{t-1})$ and $(U_{t-2}|U_{t-1})$ and proceed with automatic model selection. We also take the fourth autoregressive copula into account. The AIC of these two models is -669.4702 and -667.4702 , respectively. The model by McNeil and Bladt (2022) offers two potential versions for the given time series. First, replacing every copula in the infinite D-vine by a Gumbel copula, or taking all copulas as normal and only replacing a fixed amount by suitable copulas. We test both version obtaining AIC values of -623.3599 and -677.0362 , respectively. For the MAGMAR-model we follow the procedure described above in Sect. 4.2 and obtain an AIC of -819.0715 . This is expected since the DGP is the (adjusted) MAGMAR-copula model. Nevertheless, this simulation study confirms that the MAGMAR-copula model can produce complicated temporal dependencies, which can only be properly modeled using the correct model specification. Furthermore the simulation study showcases that a sparse model such as the $\Psi^{(2)}$ -MAGMAR(1,1) model in fact performs better than copula-based models with high autoregressive orders.

Table 1: Number of Parameters, negative log-likelihood and Information criterion values of the mentioned models fitted to quarterly US inflation data (244 observations). The 'IPD-' models are the models for infinite partial copula dependence from McNeil and Bladt (2022). The MAGMAR(p, q) model is the proposed model. The letters after the model orders indicate which copulas are used at the respective order, as is described in Sect. 3.3.3. The prefix ' $\Psi^{(k)}$ -' indicates that the adjustment transformation is conducted and k iterations of estimation were performed. The highlighted values are the lowest ones in the respective column.

Model	# pars	nLL	AIC	BIC
IPD-normal	6	-98.31	-184.62	-163.64
IPD-Gumbel	6	-110.64	-209.28	-188.30
IPD-(j,t,n,g)	5	-111.12	-210.25	-189.26
MAGMAR(4, 0)-(g,g,t,g)	5	-111.05	-212.10	-194.61
MAGMAR(5, 0)-(g,g,t,g,g)	6	-111.27	-210.54	-189.56
MAGMAR(4, 1)-(n,n,n,n)-(n)	5	-93.48	-176.95	-159.47
MAGMAR(4, 1)-(g,i,n,g)-(t)	5	-112.16	-214.31	-196.83
$\Psi^{(2)}$ -MAGMAR(4, 1)-(n,n,n,n)-(n)	5	-103.74	-197.47	-179.99
$\Psi^{(2)}$ -MAGMAR(4, 1)-(g,i,n,g)-(t)	5	-116.80	-223.60	-206.12

5. Application to US Inflation

In this section, the modeling performance of the MAGMAR-Copula model is investigated. We compare its performance with the performance of the copula-based time series model with infinite partial dependence proposed by McNeil and Bladt (2022). Just as in McNeil and Bladt (2022), quarterly US inflation is modeled. We use the consumer price index data, `cpi` from the R package `tscopula` provided by the authors, and calculate the inflation as the growth rates of the `cpi`. The time series comprises 244 observations spanning from 1960Q1 to 2020Q4. For implementation the R package `rvinecopulib` by Nagler and Vatter (2025b) is used. The package allows for fast vine-copula modeling using C++ via `rcpp`, see Eddelbuettel and François (2011). We consider two versions of the MAGMAR-Copula model. First the pure model without any adjustment transformation. Second, we consider the adjusted model, where we estimate the adjustment transformation and the parameters in an iterative way as described in Sect. 3.3.3. For this model we add the prefix ' $\Psi^{(k)}$ -', where $k \in \mathbb{N}$ denotes the number of iterations. In order to estimate the adjustment transformation Ψ , the R package `kde1d` by Nagler and Vatter (2025a) is used. Before estimating the respective models, the time series is transformed to pseudo-observations by applying the rescaled empirical CDF to the data. The models' performances are assessed with the AIC and BIC. The results can be found in Table 1. The models from McNeil and Bladt (2022) are called 'IPD-name' if all copulas are chosen as the same copula. 'Name' is a placeholder for the copula, e.g. 'Gumbel'. If only certain copulas are replaced by non-Gaussian copulas, we will follow the same naming scheme as for the MAGMAR-copula model. The results for the IPD-normal and IPD-Gumbel model are directly taken from their paper. The MAGMAR-copula models follow the naming-scheme from Sect. 3.3.3. The $\Psi^{(2)}$ -MAGMAR(4,1)-(g,i,n,g)-(t) model minimizes the AIC, as well as the BIC. The necessity of including the Gumbel copula indicates asymmetric temporal dependence while the inclusion of the t -copula indicates heavy-tailedness of the dependence structure. All of the best performing models incorporate 4 autoregressive orders. This result could indicate stochastic seasonality in the quarterly inflation. The necessity of including a MAG part indicates that the data generating process of US inflation may not fulfill the Markov assumption. This is also reflected by the fact that increasing the model order to $p = 5$ in the Markovian models does not enhance the fit. Furthermore, we observe that the

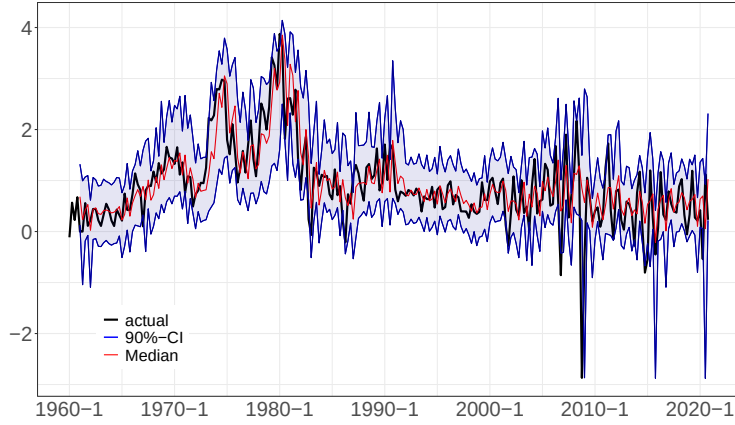


Figure 5: In-sample fit of the KDE-quantile-transformed $\Psi^{(2)}$ -MAGMAR-(g,i,n,g)-(t) model to US inflation.

adjusted MAGMAR-copula models perform better than their unadjusted counterparts. This is expected since the model is misspecified without adjustment. Our proposed model minimizes the AIC/BIC in comparison with the IPD model by [McNeil and Bladt \(2022\)](#). This could indicate that the long-term temporal dependence structures have a complex form, which are best captured by the MAGMAR-model. However, we note that the model by [McNeil and Bladt \(2022\)](#) is still competitive. In Fig. 5 we show the in-sample fit of the best-performing model to the data. Since the model naturally produces probabilistic in-sample predictions, we report the conditional median and the conditional 90%-confidence interval. The fit indicates that the $\Psi^{(2)}$ -MAGMAR(4, 1)-(g,i,n,g)-(t) model seems to capture the dynamics of US inflation quite well. In Fig. D.7 in [Appendix D](#) we also provide the residual time series transformed to the z -scale and its ACF/PACF. There seems to be not much structure left, which confirms that the model captures the dynamics well.

6. Conclusion

Non-linear temporal dependencies as well as deviations from normal or elliptical distributions are apparent in many univariate time series. A popular approach to capture these effects is to employ a copula C_ϕ for the temporal dependence structure equipped with a suitable marginal distribution F_X . The resulting model is the well-known copula-based time series model. Alongside stationarity, the model assumes a finite Markov order p . This means that the distributional properties of a random variable of the time series can only be conditionally influenced by the preceding p values. This results in the model not being able to properly capture long-term autoregressive effects, e.g. persistency in the mean or the variance. Properly modeling persistency in the mean and the variance was a major improvement for 'classical' time series modeling. To this end, the moving average term (MA) was introduced to autoregressive (AR) models resulting in the ARMA model. The generalization of the autoregressive conditional heteroskedasticity models (ARCH) was introduced to account for persistency in the variance, i.e. volatility clustering. Inspired by this approach

we introduce a moving aggregate (MAG) part to the model equation of copula-based time series models. The moving aggregate is a non-linear function of the current and past innovations. The resulting model is the Moving Aggregate Modified Autoregressive Copula-Based Time Series Model (MAGMAR-Copula time series model). After introducing the model we discuss the model components and nested models. We derive both the $MAG(\infty)$ and the $AR(\infty)$ representations and continue to derive distributional properties such as the conditional marginal distribution and the likelihood function. Furthermore, we prove that the $MAGMAR(1, 1)$ -Copula model is stationary when the AR-copula fulfills a contraction-on-average property. We exemplify the modeling approach in a simulation study. In particular, we suggest to incorporate an adjustment transformation that guarantees that the models marginal distribution is well specified. In an application to US inflation we compare our model with the model for infinite order partial dependence of [McNeil and Bladt \(2022\)](#). Their model is also able to capture the infinite memory of a time series using copulas. The comparison in the application is conducted by using information criteria (AIC and BIC). We find that our model is competitive.

Although our proposed MAGMAR-Copula model is quite general, it is a model for univariate time series. The incorporation of other variables as covariates is in principle possible but may not necessarily enhance the performance of the model in its current form. For the modeling of multivariate time series, we suggest to use copula-GARCH type-models and employ MAGMAR-Copulas for temporal modeling instead of ARMA-GARCH models ([Hu \(2006\)](#); [Jondeau and Rockinger \(2006\)](#); [Aloui et al. \(2014\)](#) are standard references for the copula-GARCH model while [Berrisch et al. \(2023\)](#) is a recent application). Furthermore, we would like to explore the forecasting performance of our model. As mentioned earlier, our model is able to capture volatility clusters. It will be interesting to see how our model performs in comparison with ARCH-models. From a theoretical point of view, further distributional properties are of interest. Especially interesting are asymptotic normality for the MLE of $MAG(1)$ processes and the stationarity of higher order processes. Further exploring theoretical properties of the common-innovations model by [Joe \(2014\)](#) and also testing it in simulations and on real data is of high interest.

Data Availability Statement

The data is publicly available through the R package `tscopula`.

Funding Statement

This research was (partially) funded in the course of TRR 391 *Spatio-temporal Statistics for the Transition of Energy and Transport* (520388526) by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

Associated Key Words

Copula, Copula-based time series, Dependence Modeling, Non-linear time series, Persistency, Vine Copula.

Mathematics Subject Classification

62-02, 62H05, 62M09, 62M10, 62P20, 60G10

References

- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.
- Aloui, R., Aïssa, M. S. B., Hammoudeh, S., and Nguyen, D. K. (2014). Dependence and extreme dependence of crude oil and natural gas prices with applications to risk management. *Energy Economics*, 42:332–342.
- Angus, J. E. (1994). The probability integral transform and related results. *SIAM review*, 36(4):652–654.
- Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica*, 78(1):395–410.
- Beare, B. K. and Seo, J. (2015). Vine copula specifications for stationary multivariate markov chains. *Journal of Time Series Analysis*, 36(2):228–246.
- Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32:245–268.
- Bedford, T. and Cooke, R. M. (2002). Vines—a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068.
- Berrisch, J., Pappert, S., Ziel, F., and Arsova, A. (2023). Modeling volatility and dependence of european carbon and energy prices. *Finance Research Letters*, 52:103503.
- Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley.
- Bladt, M., Dias, A., Han, J., and McNeil, A. J. (2025). Semiparametric and parametric distributional forecasting of univariate time series using non-gaussian arma models based on d-vines. *Canadian Journal of Statistics*.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.
- Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). Arch models. *Handbook of econometrics*, 4:2959–3038.
- Brechmann, E. C. and Czado, C. (2015). Copar—multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry*, 31(4):495–514.
- Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335.
- Czado, C. (2010). Pair-copula constructions of multivariate copulas. In *Copula theory and its applications*, pages 93–109. Springer.

- Czado, C. and Nagler, T. (2022). Vine copula based modeling. *Annual Review of Statistics and Its Application*, 9(1):453–477.
- Darsow, W. F., Nguyen, B., and Olsen, E. T. (1992). Copulas and markov processes. *Illinois journal of mathematics*, 36(4):600–642.
- Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International statistical review*, 73(1):111–129.
- Devore, J. L., Berk, K. N., and Carlton, M. A. (2012). *Modern mathematical statistics with applications*, volume 285. Springer.
- Douc, R., Moulines, E., and Stoffer, D. (2014). *Nonlinear time series: Theory, methods and applications with R examples*. CRC press.
- Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless r and c++ integration. *Journal of statistical software*, 40:1–18.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007.
- Franqc, C. and Zakoian, J.-M. (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.
- Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043.
- Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.
- Hofert, M. (2008). Sampling archimedean copulas. *Computational Statistics & Data Analysis*, 52(12):5163–5174.
- Hu, L. (2006). Dependence patterns across financial markets: a mixed copula approach. *Applied financial economics*, 16(10):717–729.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *Journal of statistical software*, 27:1–22.
- Ibe, O. (2013). *Markov processes for stochastic modeling*. Newnes.
- Ibragimov, R. (2005). Copula-based dependence characterizations and modeling for time series. *Harvard Institute of Economic Research Discussion Paper*, (2094).
- Joe, H. (2014). *Dependence modeling with copulas*. CRC press.
- Jondeau, E. and Rockinger, M. (2006). The copula-garch model of conditional dependencies: An international stock market application. *Journal of international money and finance*, 25(5):827–853.

- Loaiza-Maya, R., Smith, M. S., and Maneesoonthorn, W. (2018). Time series copulas for heteroskedastic data. *Journal of Applied Econometrics*, 33(3):332–354.
- McKenzie, E. (1988). Some arma models for dependent sequences of poisson counts. *Advances in Applied Probability*, 20(4):822–835.
- McNeil, A. J. and Bladt, M. (2022). Time series with infinite-order partial copula dependence. *Dependence Modeling*.
- Müller, A. and Reuber, M. (2022). A copula-based time series model for global horizontal irradiation. *International Journal of Forecasting*.
- Nagler, T., Krüger, D., and Min, A. (2022). Stationary vine copula models for multivariate time series. *Journal of Econometrics*, 227(2):305–324.
- Nagler, T. and Vatter, T. (2025a). Package ‘kde1d’.
- Nagler, T. and Vatter, T. (2025b). Package ‘rvinecopulib’.
- Nelsen, R. B. (2006). *An introduction to copulas*. Springer.
- Pappert, S. and Arsova, A. (2023). Forecasting natural gas prices with spatio-temporal copula-based time series models. *arXiv preprint arXiv:2301.03328*.
- Patton, A. J. (2009). Copula-based models for financial time series. In *Handbook of financial time series*, pages 767–785. Springer.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3):470–472.
- Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M., et al. (2015). Package ‘vinecopula’. *R package version*, 2(5).
- Schervish, M. J. (2012). *Theory of statistics*. Springer Science & Business Media.
- Simard, C. and Rémillard, B. (2015). Forecasting time series with multivariate copulas. *Dependence modeling*, 3(1).
- Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.
- Smith, M., Min, A., Almeida, C., and Czado, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492):1467–1479.
- Smith, M. S. (2021). Implicit copulas: An overview. *Econometrics and Statistics*.
- Spanhel, F. and Kurz, M. S. (2016). The partial copula: Properties and associated dependence measures. *Statistics & Probability Letters*, 119:76–83.
- Wooldridge, J. M. (1994). Estimation and inference for dependent processes. *Handbook of econometrics*, 4:2639–2738.

Appendix A. Supporting Results

Appendix A.1. Likelihood Calculation for MAGMAR(p, q)

The following calculation steps have to be performed to evaluate the likelihood of the MAGMAR(p, q)-Copula model.

- 1) Set initial values for the first $s = \max\{p, q\}$ observations. (A natural choice for initial values in this context is 0.5 because for symmetric marginal distributions this corresponds to the expectation value. This is also the choice of initial values that we choose for our empirical application.)
- 2) For $i \in \{s+1, \dots, T\}$ calculate $R_{\phi_1, \dots, \phi_p}^1((u_{i-1}, \dots, u_{i-p}), u_i)$ iteratively.
- 3) For $i \in \{s+1, \dots, T\}$ calculate

$$w_i = (R_{\theta_1, \dots, \theta_q}^1((w_{i-1}, \dots, w_{i-q}), R_{\phi_1, \dots, \phi_p}^1((u_{i-1}, \dots, u_{i-p}), u_i))$$

iteratively.

- 4) For $i \in \{s+1, \dots, T\}$ calculate $f_{U_i|\mathcal{F}_{i-1}}(u_i)$ (Eq. A.2 for MAGMAR(p, q) and Eq. 16 for MAGMAR(1, 1)).
- 5) Take the product of all conditional densities for the likelihood or the negative sum of log-densities for the negative log-likelihood: $L = \prod_{i=s+1}^T f_{U_i|\mathcal{F}_{i-1}}(u_i)$.

Appendix A.2. AR and MAG representation of the MAGMAR(p, q) model

The AR representation of the MAGMAR(p, q)-model (Eq. 11) can be derived by substituting for all $\{w_{t-i}\}_{i=1}^\infty$ iteratively. We start with the following expression:

$$w_t = R_{\theta_1, \dots, \theta_q}^1((w_{t-1}, \dots, w_{t-q}), R_{\phi_1, \dots, \phi_p}^1((U_{t-1}, \dots, U_{t-p}), U_t)).$$

Now, $w_{t-1}, \dots, w_{t-q}$ have to be substituted again. Formally this amounts to composing the function $R_{\theta_1, \dots, \theta_q}^1(\circ, R_{\phi_1, \dots, \phi_p}^1((U_{t-1}, \dots, U_{t-p}), U_t)) : (0, 1)^q \rightarrow (0, 1)$ with a vector of functions:

$$w_t = R_{\theta_1, \dots, \theta_q}^1(\circ, R_{\phi_1, \dots, \phi_p}^1((U_{t-1}, \dots, U_{t-p}), U_t)) \circ \left(\begin{array}{c} R_{\theta_1, \dots, \theta_q}^1(\circ, R_{\phi_1, \dots, \phi_p}^1((U_{t-2}, \dots, U_{t-p-1}), U_{t-1})) \\ \vdots \\ R_{\theta_1, \dots, \theta_q}^1(\circ, R_{\phi_1, \dots, \phi_p}^1((U_{t-q-1}, \dots, U_{t-p-q}), U_{t-q})) \end{array} \right) \circ \dots$$

Now the $\{w_{t-i}\}_{i=2}^\infty$ in each vector has to be substituted for again. The input for each element of the vector is again a q -dimensional vector. The total dimension of the input vector at the third level hence is given by q^2 . We shorten notation by introducing the $(\vec{\bigcirc}_d)_{i \in \mathbb{N}_0}$ symbol which acts as follows for a sequence

of functions $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$, $i \in \mathbb{N}_0$

$$(\vec{\odot}_d)_{i \in \mathbb{N}_0} f_i(\circ) = f_0 \circ \begin{pmatrix} f_1 \\ \vdots \\ f_d \end{pmatrix} \circ \begin{pmatrix} f_{d+1} \\ \vdots \\ f_{d+d^2} \end{pmatrix} \circ \dots$$

Using the newly introduced notation, the AR representation may be written as

$$w_t = (\vec{\odot}_q)_{i \in \mathbb{N}_0} R_{\theta_1, \dots, \theta_q}^1(\circ, R_{\phi_1, \dots, \phi_p}((U_{t-1-i}, \dots, U_{t-p-i}), U_{t-i})) \quad (\text{A.1})$$

Using the same notation, the MAG(∞)-representation can be written as

$$\begin{aligned} U_t &= R_{\phi_1, \dots, \phi_p}^{-1}(\circ, R_{\theta_1, \dots, \theta_q}^{-1}((w_{t-1}, \dots, w_{t-q}), w_t)) \\ &\circ \begin{pmatrix} R_{\phi_1, \dots, \phi_p}^{-1}(\circ, R_{\theta_1, \dots, \theta_q}^{-1}((w_{t-2}, \dots, w_{t-q-1}), w_{t-1})) \\ \vdots \\ R_{\phi_1, \dots, \phi_p}^{-1}(\circ, R_{\theta_1, \dots, \theta_q}^{-1}((w_{t-p-1}, \dots, w_{t-p-q}), w_{t-p})) \end{pmatrix} \circ \dots \\ &= (\vec{\odot}_p)_{i \in \mathbb{N}_0} R_{\phi_1, \dots, \phi_p}^{-1}(\circ, R_{\theta_1, \dots, \theta_q}^{-1}((w_{t-1-i}, \dots, w_{t-q-i}), w_{t-i})) \end{aligned}$$

Appendix A.3. Conditional Distribution MAGMAR(p, q)

The conditional distribution of U_t given all past observations is given by:

$$\begin{aligned} f_{U_t | \mathcal{F}_{t-1}}(x) &= c_{\theta_q} \left[R_{\phi_1, \dots, \phi_p}^1((u_{t-1}, \dots, u_{t-p}), x), R_{\theta_1, \dots, \theta_{q-1}}^1((w_{t-2}, \dots, w_{t-q}), w_{t-1}) \right] \\ &\prod_{i=0}^{p-1} c_{\phi_{p-i}} \left[R_{\phi_1, \dots, \phi_{p-1-i}}^1((u_{t-1}, \dots, u_{t-p+1+i}), x), \right. \\ &\left. R_{\phi_1, \dots, \phi_{p-1-i}}^2((u_{t-1}, \dots, u_{t-p+1+i}), u_{t-p+i}) \right]. \end{aligned} \quad (\text{A.2})$$

The proof is analog to the proof for the MAGMAR(1,1)-Copula case.

Appendix A.4. Joint Conditional Distribution of the MAGMAR model

The joint conditional distribution of $(U_t, U_{t-1}) | \mathcal{F}_{t-2}$ is derived. The density transform theorem for multivariate transformations (see [Devore et al., 2012](#), 5.6) is utilized. The two random variables w_t and w_{t-1} are transformed by the function G to the random variables U_t and U_{t-1} . The appropriate choice for

G and the corresponding inverse and Jacobian are given by,

$$\begin{aligned}
\begin{pmatrix} z \\ w \end{pmatrix} &:= G((a, b)^T) = \begin{pmatrix} h_\phi^{-1}(h_\theta^{-1}(a, w_{t-2}), u_{t-2}) \\ h_\phi^{-1}(h_\theta^{-1}(b, a), h_\phi^{-1}(h_\theta^{-1}(a, w_{t-2}), u_{t-2})) \end{pmatrix}, \\
&= \begin{pmatrix} h_\phi^{-1}(h_\theta^{-1}(a, w_{t-2}), u_{t-2}) \\ h_\phi^{-1}(h_\theta^{-1}(b, a), z) \end{pmatrix}. \\
\begin{pmatrix} a \\ b \end{pmatrix} &:= G^{-1}(z, w) = \begin{pmatrix} h_\theta(h_\phi(z, u_{t-2}), w_{t-2}) \\ h_\theta(h_\phi(w, z), h_\theta(h_\phi(z, u_{t-2}), w_{t-2})) \end{pmatrix}, \\
&= \begin{pmatrix} h_\theta(h_\phi(z, u_{t-2}), w_{t-2}) \\ h_\theta(h_\phi(w, z), a) \end{pmatrix} \\
\mathcal{J} &= \begin{pmatrix} c_\theta[h_\phi(z, u_{t-2}), w_{t-2}]c_\phi[z, u_{t-2}] & 0 \\ \partial_z(\dots) & c_\theta[h_\phi(w, z), h_\theta(h_\phi(z, u_{t-2}), w_{t-2})]c_\phi[w, z] \end{pmatrix}, \\
|\mathcal{J}| &= c_\theta[h_\phi(z, u_{t-2}), w_{t-2}]c_\phi[z, u_{t-2}]c_\theta[h_\phi(w, z), h_\theta(h_\phi(z, u_{t-2}), w_{t-2})]c_\phi[w, z].
\end{aligned}$$

Here u_{t-2} and w_{t-2} ought to be understood as (deterministic by conditioning) constants. Since w_t and w_{t-1} are independent and uniformly distributed their contribution to the conditional density is just $1_{(0,1)^2}$. The density can be formulated as

$$\begin{aligned}
f_{U_t, U_{t-1} | \mathcal{F}_{t-2}}(z, w) &= f_{w_t, w_{t-1}}(z, w) |\mathcal{J}| \\
&= c_\theta[h_\phi(z, u_{t-2}), w_{t-2}]c_\phi[z, u_{t-2}]c_\theta[h_\phi(w, z), h_\theta(h_\phi(z, u_{t-2}), w_{t-2})]c_\phi[w, z].
\end{aligned}$$

The partial conditional density of $U_t, U_{t-k} | U_{t-1}, \dots, U_{t-k+1}, U_{t-k-1}, \dots$ can again be derived by using the density transformation theorem. The random variables w_t and w_{t-k} are transformed to the random variables U_t and U_{t-k} . The appropriate choice of the transformation G is,

$$\begin{aligned}
G((a, b)^T) &= \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} h_\phi^{-1}(h_\theta^{-1}(a, w_{t-k-1}), u_{t-k-1}) \\ h_\phi^{-1}(h_\theta^{-1}(b, w_{t-1}(a)), u_{t-1}) \end{pmatrix} \\
G^{-1}(z, w) &= \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} h_\theta(h_\phi(z, u_{t-k-1}), w_{t-k-1}) \\ h_\theta(h_\phi(w, u_{t-1}), w_{t-1}(z)) \end{pmatrix} \\
\mathcal{J} &= \begin{pmatrix} c_\theta[h_\phi(z, u_{t-k-1}), w_{t-k-1}]c_\phi[z, u_{t-k-1}] & 0 \\ \partial_z(\dots) & c_\theta[h_\phi(w, u_{t-1}), w_{t-1}(z)]c_\phi[w, u_{t-1}] \end{pmatrix} \\
|\mathcal{J}| &= c_\theta[h_\phi(z, u_{t-k-1}), w_{t-k-1}]c_\phi[z, u_{t-k-1}]c_\theta[h_\phi(w, u_{t-1}), w_{t-1}(z)]c_\phi[w, u_{t-1}].
\end{aligned}$$

Here $w_{t-1}(z)$ means evaluating the $\text{AR}(k-1)$ representation of w_{t-1} at $u_{t-1}, \dots, u_{t-k+1}$ and $w_{t-k} = h_\theta(h_\phi(z, u_{t-k-1}), w_{t-k-1})$ and $U_{t-k} = z$.

Appendix A.5. Verifying conditions for stationarity for several copulas.

Here we verify the assumptions from Douc et al. (2014) to establish the stationarity and ergodicity of the MAGMAR(1, 1) time series for the normal, t and Clayton Copula. For the sake of completeness, we

give the assumptions from Douc et al. (2014) here too. The assumptions revolve around a stochastic recurrence equation, $X_t = f_{Z_t}(X_{t-1})$. The assumptions are as follows:

Assumption A4.36: Z_t is strictly stationary and ergodic.

Assumption A4.37: There is a measurable function K_z such that $d(f_z(x), f_z(y)) \leq K_z d(x, y)$, with $\mathbb{E}(\log^+(K_Z)) \leq \infty$ and $\mathbb{E}(\log(K_Z)) < 0$

Assumption A4.38: There is a $x \in \mathbb{R}$ s.t. $\mathbb{E}(\log^+(d(x, f_Z(x)))) < \infty$.

In order to use the result from Douc et al. (2014), we recognize that the MAGMAR(1, 1) updating equation can be interpreted as stochastic recurrence equation. The MAGMAR(1, 1) updating equation may be written as $U_t = f_{Z_t}(U_{t-1})$, where Z_t is the stationary and ergodic MAG(1) process and $f_z(x) = h_\alpha^{-1}(z, x)$. Now we verify the conditions.

Appendix A.5.1. Normal copula as AR-Copula

Let $\{U_t\}$ be a MAGMAR(1, 1)-Copula time series where the AR-Copula is the normal Copula with parameter $|\alpha| < 1$. The conditions for the theorem are **not** fulfilled for $\kappa = \text{id}$. However, we are able to show that there is a mapping $\kappa : \mathbb{R} \rightarrow (0, 1)$ such that the theorem is applicable to the time series $\{Y_t\}$ with $Y_t = \kappa^{-1}(U_t)$. The corresponding stochastic recurrence equation is specified by $\tilde{f}_z(x) = \kappa^{-1}(h_\alpha^{-1}(z, \kappa(x)))$.

¹ Now we can check the assumptions.

Assumption A4.36: $Z_t = V_t = h_\theta^{-1}(w_t, w_{t-1})$ is stationary and 1-dependent, and hence ergodic.

Assumption A4.37: We choose $d(x, y) = |x - y|$. Then by the mean value theorem, it follows that

$$|\tilde{f}_z(x) - \tilde{f}_z(y)| \leq \max_{\zeta \in (x, y)} \tilde{f}'_z(\zeta) |x - y| \quad (\text{A.3})$$

$$\leq \max_{\zeta \in \mathbb{R}} \tilde{f}'_z(\zeta) |x - y| \quad (\text{A.4})$$

where

$$\tilde{f}'_z(\zeta) = \alpha \frac{\phi[\alpha \Phi^{-1}(\kappa(x)) + (1 - \alpha^2)^{\frac{1}{2}} \Phi^{-1}(z)]}{\phi[\Phi^{-1}(\kappa(x))]} \frac{\kappa'(x)}{\kappa'(\kappa^{-1}(\Phi(\alpha \Phi(\kappa(x)) + (1 - \alpha^2)^{\frac{1}{2}} \Phi^{-1}(z)))}.$$

We notice that for $\kappa = \Phi$ we obtain $\tilde{f}'_z(\zeta) = \alpha < 1$. Hence, choosing $\kappa = \Phi$ and $K_z = \alpha$ satisfies the requirement.

Assumption A4.38: There is a $x \in \mathbb{R}$ s.t. $\mathbb{E}(\log^+(d(x, \tilde{f}_Z(x)))) < \infty$. Choose e.g. again $\kappa = \Phi$, then $\mathbb{E}(\log^+(d(x, \tilde{f}_Z(x)))) = \log^+(\alpha) < \infty$.

From $\{Y_t\}$ being stationary it follows that $\{\kappa(Y_t)\} = \{U_t\}$ is also stationary.

¹ $U_t = f_{Z_t}(U_{t-1}) \iff Y_t = \kappa^{-1}[U_t] = \kappa^{-1}[f_{Z_t}(\kappa[\kappa^{-1}[U_{t-1}]])] = \kappa^{-1}[f_{Z_t}(\kappa[Y_{t-1}])]$. As an example, the reader may think of κ as $\kappa = \Phi$, which will also be used in the following.

Appendix A.5.2. t-copula as AR-copula:

Let C_α be given as the t -copula with parameters $\alpha = (\rho, \nu)$. The inverse h -function is given as (cf. [Smith et al. \(2010\)](#))

$$h_{\rho, \nu}^{-1}(z, x) = t_\nu \left(t_{\nu+1}^{-1}(z) \left[\frac{(\nu + (t_\nu^{-1}(x))^2)(1 - \rho^2)}{\nu + 1} \right]^{\frac{1}{2}} + \rho t_\nu^{-1}(x) \right)$$

Choosing $\kappa = t_\nu$, $\tilde{f}_z$ is given as

$$\tilde{f}_z(x) = t_{\nu+1}^{-1}(z) \left[\frac{(\nu + x^2)(1 - \rho^2)}{\nu + 1} \right]^{\frac{1}{2}} + \rho x$$

Assumption A4.37: We again write $|\tilde{f}_z(x) - \tilde{f}_z(y)| \leq K_z|x - y|$ where $K_z = \max_{\zeta \in \mathbb{R}} |\tilde{f}_z'(\zeta)|$. The first derivative of $\tilde{f}_z$ is given as

$$\tilde{f}_z'(x) = t_{\nu+1}^{-1}(z) \left[\frac{(\nu + x^2)(1 - \rho^2)}{\nu + 1} \right]^{-\frac{1}{2}} \frac{x(1 - \rho^2)}{\nu + 1} + \rho.$$

This expression is monotone increasing and obtains its absolute maximum at $x \rightarrow \pm\infty$. Hence $K_z = |t_{\nu+1}^{-1}(z) \left(\frac{1 - \rho^2}{\nu + 1} \right)^{\frac{1}{2}} + \rho|$. Then $\mathbb{E}(\log |K_{V_t}|)$ is the expectation of the logarithm of a folded unstandardized t -distribution. The expression can be evaluated numerically by e.g. Monte-Carlo simulation to show that $\mathbb{E}(\log |K_{V_t}|) < 0$. In fact we observe that only for ν very close to zero, $0 < \nu < 0.2$, the expression is positive. Hence part a) from condition A4.37 from [Douc et al. \(2014\)](#) is satisfied. Condition b) amounts to showing that $\mathbb{E}(\log^+ K_{V_t}) < \infty$. We may write $\mathbb{E}(\log^+ K_{V_t}) = \mathbb{E}(\log K_{V_t} \mathbb{1}_{\{K_{V_t} > 1\}}) = \mathbb{E}(\log |Y \left(\frac{1 - \rho^2}{\nu + 1} \right)^{\frac{1}{2}} + \rho| \mathbb{1}_{\{Y > K_{\nu, \rho}\}})$ where $Y \sim t_{\nu+1}$. Now there are two thresholds s_1 and s_2 depending on ν and ρ such that we can write the expectation as $\mathbb{E}(\log |Y \left(\frac{1 - \rho^2}{\nu + 1} \right)^{\frac{1}{2}} + \rho| \mathbb{1}_{\{Y > K_{\nu, \rho}\}}) = \int_{-\infty}^{s_1} \log[-y \left(\frac{1 - \rho^2}{\nu + 1} \right)^{\frac{1}{2}} + \rho] f_{t_{\nu+1}}(y) dy + \int_{s_2}^{\infty} \log[y \left(\frac{1 - \rho^2}{\nu + 1} \right)^{\frac{1}{2}} + \rho] f_{t_{\nu+1}}(y) dy$. We observe that integrands are logarithmic, weighted by the t -distribution density. Since linear functions dominate the logarithm asymptotically and the first moment of the t -distribution exists for $\nu > 2$, $\mathbb{E}(\log^+ K_{V_t})$ is finite as long as $\nu > 2$.

Assumption A4.38: The assumption can be shown to be satisfied for e.g. $x = 0$. Then $d(x, \tilde{f}_z(x)) = t_{\nu+1}^{-1}(z) \left[\frac{\nu(1 - \rho^2)}{\nu + 1} \right]^{\frac{1}{2}}$. The existence follows again by the existence of the first moment of the t -distribution for $\nu > 2$.

Appendix A.5.3. The Clayton Copula as AR-Copula:

The inverse h -function of the Clayton copula with parameter θ is given as,

$$h^{-1}(z, x) = ([zx^{\theta+1}]^{-\frac{\theta}{\theta+1}} + 1 - x^{-\theta})^{-\frac{1}{\theta}}$$

We choose κ such that $\kappa(x) = x^{-\frac{1}{\theta}}$ and $\kappa^{-1}(x) = x^{-\theta}$ respectively. Then $\tilde{f}_z$ is given such that $\tilde{f}_z(x) = z^{-\frac{\theta}{\theta+1}}x + 1 - x = 1 + x(z^{-\frac{\theta}{\theta+1}} - 1)$. We can immediately see that an appropriate choice for K_z is given as $K_z = |(z^{-\frac{\theta}{\theta+1}} - 1)|$.

Assumption A4.37: $\mathbb{E}(\log K_V) = \int_0^1 \log |y^{-\frac{\theta}{\theta+1}} - 1| dy < \infty$ and $\mathbb{E}(\log^+ K_V)$ can be verified numerically.

Assumption A4.38: For $x = 0$, $\mathbb{E}(\log^+(d(x, f_Z(x)))) = 1 < \infty$.

Appendix A.6. The Autodependence of Markov and q -Dependent Processes

Here we provide the proofs of the following two claims.

Proposition 8. $\{X_t\}_{t \in \mathbb{Z}}$ is a Markov(p) process $\iff$ for $r > p$

$$F(x_0, x_r) = \mathbb{P}(X_t \leq x_0, X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1}) \quad (\text{A.5})$$

$$= \mathbb{P}(X_t \leq x_0 | X_{t-1}, \dots, X_{t-r+1}) \times \mathbb{P}(X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1}) \quad (\text{A.6})$$

This can be seen by considering the following scenario. Let $\{X_t\}_{t \in \mathbb{Z}}$ be a Markov(p) process. Let $r > p$, then the partial dependence of $(X_t, X_{t-r} | X_{t-1}, \dots, X_{t-r})$ is independent:

$$\begin{aligned} F(x_0, x_r) &= \mathbb{P}(X_t \leq x_0, X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1}) \\ &= \mathbb{P}(X_t \leq x_0 | X_{t-1}, \dots, X_{t-r+1}, X_{t-r} \leq x_r) \times \mathbb{P}(X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1}) \\ &= \mathbb{P}(X_t \leq x_0 | X_{t-1}, \dots, X_{t-r+1}) \times \mathbb{P}(X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1}) = F(x_0) \times F(x_r) \end{aligned} \quad (\text{A.7})$$

where in the first line we have used the law of total probability and in the second line that $\{X_t\}_{t \in \mathbb{Z}}$ is a Markov(p) process and $r > p$, hence conditioning on X_{t-r} can be dropped in the first probability measure. For the other direction, let $F_{X_t, X_{t-r} | \mathcal{F}_{t-r+1}^{t-1}}$ be independent for all $r > p$. Then

$$\begin{aligned} F_{X_t | \mathcal{F}_{t-r}^{t-1}}(x_0) &= \mathbb{P}(X_t \leq x_0 | X_{t-1}, \dots, X_{t-r}) \\ &= \frac{\mathbb{P}(X_t \leq x_0 | X_{t-1}, \dots, X_{t-r}) \times \mathbb{P}(X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1})}{\mathbb{P}(X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1})} \end{aligned} \quad (\text{A.8})$$

$$= \frac{\mathbb{P}(X_t \leq x_0, X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1})}{\mathbb{P}(X_{t-r} \leq x_r | X_{t-1}, \dots, X_{t-r+1})} \quad (\text{A.9})$$

Appendix A.7. Conceptual Comparison to Similar Approaches

We compare our proposed MAGMAR-model with the common-innovations copula model by [Joe \(2014\)](#) and the infinite partial dependence model by [McNeil and Bladt \(2022\)](#). The common-innovations model has not yet been examined in detail, and there is no implementation. Therefore, some comparisons are not yet fully possible. A more in-depth examination and an implementation is a potential direction for future research. Nevertheless, we compare the models with regards to the following criteria:

Flexibility: We discuss flexibility in terms of (partial) dependencies that the models can produce. That is, how flexible is each model in constructing $(U_t, U_{t-k} | \mathcal{F}_{t-1})$. The partial densities of the MAGMAR-model are derived in [Appendix A.4](#). Each partial density is a product of MAG and AR-copula densities. Given flexibility in these densities, the partial dependencies can be modeled flexibly too. Typically, however, a partial copula density of the (adjusted) MAGMAR-model will not be a 'standard' copula density. This means that complex dependencies can be modeled in arbitrary lags. The IPD approach is also very flexible in modeling partial dependencies since they are modeled directly. For each $k \in \mathbb{Z}$, the copula of

$(U_t, U_{t-k} | \mathcal{F}_{t-1:t-k+1})$ can be chosen. In practice, either only one copula family is allowed for all lags k or all partial copulas are Gaussian except for a finite amount of non-Gaussian copulas for specified lags k (see [Bladt et al. \(2025\)](#)). This approach can be arbitrarily flexible in the copula families. However, eventually, the partial copulas are Gaussian with increasing lag order k . Hence the main differences between the MAGMAR and the IPD model in terms of flexibility are that the long-range dependencies can differ. If long-range dependencies seem to approach Gaussianity and only some partial dependencies are non-linear, the IPD model is suitable. If long-range dependencies remain non-Gaussian (e.g. because of heavy-tailedness) the MAGMAR-Copula approach is suitable. The common-innovations model is similar to the MAGMAR model and similar behavior can be expected.

Computational demand: The computational demand of all models is moderate with MAGMAR-copula models needing *1min* on average to estimate a MAGMAR(1,1) model with 1000 observations. In order to estimate an adjusted MAGMAR(1,1) model with two iterations, approximately *2min* are necessary. The common-innovations model can be expected to have similar computational demand, however in this model no adjustment transformation is needed. The IPD model takes a few seconds for a similar estimation if all copulas are replaced. If only certain copulas are replaced the estimation takes around 15s.

Modeling Approach: The suggested modeling approach for the presented adjusted-MAGMAR copula model is to rely on descriptive statistics, such as autodependence and scatter plots. A similar approach can be pursued for the IPD model. For the common-innovations model, however, it is hard to perform model selection directly from descriptive statistics. This is because the AR-model is latent. Approaches for model selection in the common-innovations model may be explored in future research.

Appendix B. Proofs

In this section we provide proofs for the propositions made in the manuscript.

Appendix B.1. Proof of Proposition 1: Relation to other models

Proof. We show the results step-by-step:

- i) Let $X_t = \Phi^{-1}[h_\alpha^{-1}(h_\beta^{-1}(w_t, w_{t-1}), \Phi(X_{t-1}))]$ for all t . We take the explicit form of the (inverse) h -function associated to the normal copula from [Smith et al. \(2010\)](#):

$$\begin{aligned} h_\alpha(u_1, u_2) &= \Phi([\Phi^{-1}(u_1) - \alpha\Phi^{-1}(u_2)](1 - \alpha^2)^{-\frac{1}{2}}), \\ (h_\alpha^{-1}(u_1, u_2)) &= \Phi(\Phi^{-1}(u_1)(1 - \alpha^2)^{\frac{1}{2}} + \alpha\Phi^{-1}(u_2)), \end{aligned}$$

where Φ is the CDF of the standard normal distribution. Now we set

$$\begin{aligned} \text{LHS} &= h_\alpha(\Phi(X_t), \Phi(X_{t-1})) \\ &= \Phi([\Phi^{-1}(\Phi(X_t)) - \alpha\Phi^{-1}(\Phi(X_{t-1}))](1 - \alpha^2)^{-\frac{1}{2}}) \\ &= \Phi([X_t - \alpha X_{t-1}](1 - \alpha^2)^{-\frac{1}{2}}), \\ \text{RHS} &= \Phi(\Phi^{-1}(w_t)(1 - \beta^2)^{\frac{1}{2}} + \beta\Phi^{-1}(w_{t-1})). \end{aligned}$$

Setting LHS=RHS and $\Phi^{-1}(w_t) =: \varepsilon_t, \forall t \in \{1, \dots, T\}$, where due to the quantile transformation $\varepsilon_t \stackrel{iid}{\sim} N(0, 1)$, we obtain

$$\begin{aligned} (X_t - \alpha X_{t-1})(1 - \alpha^2)^{-\frac{1}{2}} &= \varepsilon_t(1 - \beta^2)^{\frac{1}{2}} + \beta \varepsilon_{t-1}, \\ \iff X_t - \alpha X_{t-1} &= (1 - \alpha^2)^{\frac{1}{2}}(1 - \beta^2)^{\frac{1}{2}} \varepsilon_t + \beta(1 - \alpha^2)^{\frac{1}{2}} \varepsilon_{t-1}, \\ \iff X_t - \alpha X_{t-1} &= \tilde{\varepsilon}_t + \frac{\beta}{(1 - \beta^2)^{\frac{1}{2}}} \tilde{\varepsilon}_{t-1}. \end{aligned}$$

This is the model equation for the ARMA(1, 1) model with AR-parameter α , MA-parameter $\frac{\beta}{(1 - \beta^2)^{\frac{1}{2}}}$ and white noise normal innovations with variance $(1 - \alpha^2)(1 - \beta^2)$.

- ii) The h -function associated to the independence copula is $h_{\Pi}(u_1, u_2) = \frac{\partial(u_1 u_2)}{\partial u_2} = u_1$. The corresponding inverse h -function is $h_{\Pi}^{-1}(u_1, u_2) = u_1$. It follows that $h_{\theta}^{-1}(w_t, w_{t-1}) = w_t$, so the updating equation becomes $U_t = h_{\phi}^{-1}(w_t, U_{t-1}), w_t \stackrel{iid}{\sim} U(0, 1)$, which is the updating equation of the classical copula-based time series model.
- iii) We follow the same steps as for ii) and see that $\{U_t\}_{t \in \mathbb{Z}}$ follows the updating equation $U_t = h_{\beta}(w_t, w_{t-1})$, which is the MAG(1) updating equation.
- iv) Again we use that the inverse h -function associated to the independence copula is $h_{\Pi}^{-1}(u_1, u_2) = u_1$. From this it follows that the model updating equation is $U_t = w_t$, where $w_t \stackrel{iid}{\sim} U(0, 1)$.

□

Appendix B.2. Proof of Proposition 2: Distributional Properties of the MAG(1)-Copula Process

The MAG(1) model is given by

$$U_t = h_{\theta}^{-1}(w_t, w_{t-1}), \quad w_t \stackrel{iid}{\sim} U(0, 1)$$

The marginal distribution is standard uniform and the time series is stationary and ergodic.

Proof. For $a \in (0, 1)$, it holds that

$$\begin{aligned} \mathbb{P}(U_t \leq a) &= \mathbb{P}(h_{\theta}^{-1}(w_t, w_{t-1}) \leq a), \\ &= \mathbb{E}(h_{\theta}(a, w_{t-1})), \\ &= \int_0^1 h_{\theta}(a, w_{t-1}) dw_{t-1} = a. \end{aligned}$$

For $a \leq 0$, $\mathbb{P}(U_t \leq a) = 0$ and for $a \geq 1$, $\mathbb{P}(U_t \leq a) = 1$ since $U_t \in (0, 1)$. So $\mathbb{P}(U_t \leq a)$ is the distribution function of the uniform distribution. This result would also follow directly from the consideration that U_t results from quantile transforming a uniform random variable w.r.t. a quantile function corresponding to a copula, conditioning on an independent uniform random variable. It is well-known that the resulting random variable is uniform. Stationarity follows directly because the marginal distribution and the

updating equation are time invariant. Ergodicity follows directly from the fact that the process is 1-dependent. The joint distribution of (U_t, U_{t-1}) can be calculated as follows

$$\begin{aligned}
F_{U_t, U_{t-1}}(a, b) &= \mathbb{P}(U_t \leq a, U_{t-1} \leq b) \\
&= \mathbb{P}(h_\theta^{-1}(w_t, w_{t-1}) \leq a, h_\theta^{-1}(w_{t-1}, w_{t-2}) \leq b) \\
&= \mathbb{E}(h_\theta(a, w_{t-1}) 1_{\{w_{t-1} \leq h_\theta(b, w_{t-2})\}}) \\
&= \int_0^1 \int_0^{h_\theta(b, w_{t-2})} h_\theta(a, w_{t-1}) dw_{t-1} dw_{t-2} \\
&= \int_0^1 C_\theta[a, h_\theta(b, w_{t-2})] dw_{t-2}
\end{aligned}$$

Note that implicitly it was used that the marginal distribution of U_t is standard uniform and time independent. Now using that the joint distribution of (U_t, U_{t-1}) is time independent, it can be shown that also $F_{U_t, U_{t-1}, U_{t-2}}$ is time independent. This procedure can be generalized and then by induction it follows that the process is strictly stationary. (Furthermore it can be seen that $F_{U_t, U_{t-1}}(a, b)$ is increasing in a and b , the triangle inequality is fulfilled and $F_{U_t, U_{t-1}}(1, b) = b$ as well as $F_{U_t, U_{t-1}}(a, 1) = a$. Hence $F_{U_t, U_{t-1}}(a, b)$ is a copula.) To establish ergodicity, notice that the process is 1-dependent. Then ergodicity follows immediately. $\square$

Appendix B.3. Proof of Proposition 3

Proof. The inverse h -function associated to a normal copula with parameter ϕ is given as $h_\phi^{-1}(u_1, u_2) = \Phi(\Phi^{-1}(u_1)(1 - \phi^2)^{\frac{1}{2}} + \phi\Phi^{-1}(u_2))$ (Smith et al., 2010). Then $A_{t,n}$ can be written as

$$A_{t,n} = \Phi \left((1 - \phi^2)^{\frac{1}{2}} \sum_{i=0}^n \phi^i \Phi^{-1}(w_{t-i}) \right).$$

Since $w_t \stackrel{iid}{\sim} U[0, 1]$ and ϕ^i is decreasing geometrically in i , the variance of the sum $\sum_{i=0}^n \phi^i \Phi^{-1}(w_{t-i})$ is bounded and it converges almost surely (Billingsley, 1995). The limiting random variable is distributed as $N(0, \frac{1}{1-\phi^2})$. By the continuous mapping theorem and the continuity of Φ as well as the continuity of the multiplication with $(1 - \phi^2)^{\frac{1}{2}}$ it follows that $A_{t,n}$ also converges almost surely. Since $(1 - \phi^2)^{\frac{1}{2}} \sum_{i=0}^n \phi^i \Phi^{-1}(w_{t-i}) \sim N(0, 1)$, the limiting random variable is standard uniformly distributed. $\square$

Appendix B.4. Proof of Proposition 4: Strict Stationarity of MAGMAR(1, 1) with normal AR-Copula

Proof. Similar to the reasoning in the proof of proposition 3, the MAG(∞) representation of $\{U_t\}$ can be written as

$$A_{t,n} = \Phi \left((1 - \phi^2)^{\frac{1}{2}} \sum_{i=0}^n \phi^i \Phi^{-1}(h_\theta^{-1}(w_{t-i}, w_{t-i-1})) \right).$$

Now notice that $\{V_t, V_t := h_\theta^{-1}(w_t, w_{t-1})\}_{t \in \mathbb{Z}}$ is a MAG(1) time series. From proposition 2 it follows that $\{V_t\}_{t \in \mathbb{Z}}$ is strictly stationary, so $\{\Phi^{-1}(V_t)\}_{t \in \mathbb{Z}}$ is also strictly stationary. Since by proposition 3 converges

almost surely and thereby constitutes a measurable mapping, by Theorem 36.4 of Billingsley (1995) it follows that the process is strictly stationary. $\square$

Appendix B.5. Proof of Proposition 5: Stationarity of MAGMAR(1,1) based on a Contraction

Proof. The stationarity and ergodicity of the time series $\{\kappa^{-1}(U_t)\}_{t \in \mathbb{Z}}$ follows directly from theorem 4.40 from Douc et al. (2014), since assumptions A4.37 and A4.38 are fulfilled by assumption, and assumption A4.36 is fulfilled since the MAG(1) time series is stationary and ergodic. Since κ is a bijective and contemporaneous transformation, $\{U_t\}$ is also stationary. $\square$

Appendix B.6. Proof of Proposition 6: Conditional distribution of MAGMAR(1,1)

Let $g_{\mathbf{U}_{t-1}} : (0, 1) \rightarrow (0, 1)$ be given as

$$g_{\mathbf{U}_{t-1}}(x) = h_{\phi}^{-1}(h_{\theta}^{-1}(x, w_{t-1}), U_{t-1}) \quad \text{where } U_{t-1}, w_{t-1} \text{ are deterministic,}$$

It holds that $U_t | \mathcal{F}_{t-1} = g_{\mathbf{U}_{t-1}}(w_t) | \mathcal{F}_{t-1}$. Application of the density transformation theorem (Devore et al., 2012, Sect. 4.7) yields the result. The Jacobian of g can be calculated as

$$\begin{aligned} &\Rightarrow g^{-1}(x) = h_{\theta}(h_{\phi}(x, U_{t-1}), w_{t-1}), \\ &\Rightarrow \left| \frac{d}{dx} g^{-1}(x) \right| = c_{\theta}(h_{\phi}(x, U_{t-1}), w_{t-1}) c_{\phi}(x, U_{t-1}). \end{aligned}$$

Using that $w_t \sim U(0, 1)$ and hence $f_{w_t}(x) = 1_{(0,1)}(x)$, the resulting conditional density for $U_t | \mathcal{F}_{t-1}$ is

$$f_{U_t | \mathcal{F}_{t-1}}(x) = c_{\theta}(h_{\phi}(x, U_{t-1}), w_{t-1}) c_{\phi}(x, U_{t-1}).$$

For the adjusted MAGMAR-copula model, we follow the same procedure. In this case the function $g_{\mathbf{U}_{t-1}}$ is given as

$$g_{\tilde{\mathbf{U}}_{t-1}}(x) = \Psi[h_{\phi}^{-1}(h_{\theta}^{-1}(x, w_{t-1}), \Psi^{-1}(\tilde{U}_{t-1}))] \quad (\text{B.1})$$

and consequently

$$g^{-1}(x) = h_{\theta}(h_{\phi}(\Psi^{-1}(x), \Psi^{-1}(\tilde{U}_{t-1})), w_{t-1}), \quad (\text{B.2})$$

$$\Rightarrow \left| \frac{d}{dx} g^{-1}(x) \right| = c_{\theta}(h_{\phi}(\Psi^{-1}(x), \Psi^{-1}(\tilde{U}_{t-1})), w_{t-1}) \times c_{\phi}(\Psi^{-1}(x), \Psi^{-1}(\tilde{U}_{t-1})) \times \frac{1}{\psi(\Psi^{-1}(x))} \quad (\text{B.3})$$

Appendix B.7. Proof of proposition 7

Proof. We make use of the results from Wooldridge (1994) for the asymptotic behavior of M-estimators. First we use Theorem 4.2 to establish the uniform law of large numbers for the score-time series, $\{l_t\}$, then we use Theorem 4.3 to establish the consistency of the ML-estimators.

- 1) We check the conditions for Theorem 4.2 from Wooldridge (1994):

- i) $\Theta \subset \mathbb{R}^p$ is compact. This is true by assumption.
- ii) $l_t(\theta; u, v) = \log c_\theta[u, v]$ is continuous in u, v and θ and hence measurable.
- iii) $l_t(\theta; (U_s)_{s \leq t})$ fulfills the WLLN. In order to establish the WLLN, we first establish the ergodicity of $\{\hat{w}_t\}$. Since the time series $\{\hat{w}_t\}$ fulfills assumptions from A4.36-A4.38 from Douc et al. (2014), theorem 4.40 from the same reference is applicable. The theorem establishes the stationarity and ergodicity of $\{\hat{w}_t\}$. Additionally, $\{U_t\}$ is 1-dependent. Then it follows that $\{l_t\}$ is stationary and ergodic and hence fulfills the WLLN.
- iv) There is $c_t((U_s)_{s \leq t})$ such that $|l_t(\theta_1; (U_s)_{s \leq t}) - l_t(\theta_2; (U_s)_{s \leq t})| \leq c_t((U_s)_{s \leq t}) \|\theta_1 - \theta_2\|$ and $c_t((U_s)_{s \leq t})$ is subject to the WLLN. To see this notice that by i) and ii) it follows that $l_t(\theta; (U_s)_{s \leq t})$ is Lipschitz-continuous in θ , i.e. there is $L > 0$ such that $|l_t(\theta_1; (U_s)_{s \leq t}) - l_t(\theta_2; (U_s)_{s \leq t})| \leq L \|\theta_1 - \theta_2\|$. Since L is deterministic it trivially fulfills the WLLN.

All assumptions are met, it follows that $l_t(\theta; (U_s)_{s \leq t})$ fulfills the UWLLN.

- 2) We use Theorem 4.3 from Wooldridge (1994) to establish the consistency of the ML-estimator $\hat{\theta}$. Note that in this set-up, there are no nuisance parameters, so the application of the theorem is simplified. Again, we check the conditions:

- i) Θ is compact,
- ii) l_t is continuous and measurable.
- iii) l_t fulfills the the UWLLN (as shown above).
- iv) $\theta_0 = \arg \min_{\theta \in \Theta} \mathbb{E}(l_t(\theta))$ is unique by assumption.

It follows that the maximum likelihood estimator is consistent.

□

Appendix C. Additional Background Information

Appendix C.1. Copulas used in this Work

Normal Copula: The Normal copula models linear dependencies. It is constructed by the inversion method using the multivariate normal distribution (Smith, 2021; Nelsen, 2006),

$$C_\Sigma^{\text{normal}}[u_1, \dots, u_p] = \Phi_\Sigma(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)). \quad (\text{C.1})$$

Here Φ_Σ denotes the multivariate normal distribution function with correlation matrix $\Sigma \in [-1, 1]^{p \times p}$, symmetric and positive definite. Φ^{-1} denotes the quantile function of the univariate standard normal distribution.

t-copula: The t-copula models linear dependencies with heavy tails. Similar to the normal copula it is also constructed by the inversion method. It is given as (Demarta and McNeil, 2005),

$$C_{\Sigma, \nu}^t[u_1, \dots, u_p] = t_{\Sigma, \nu}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_p)), \quad (\text{C.2})$$

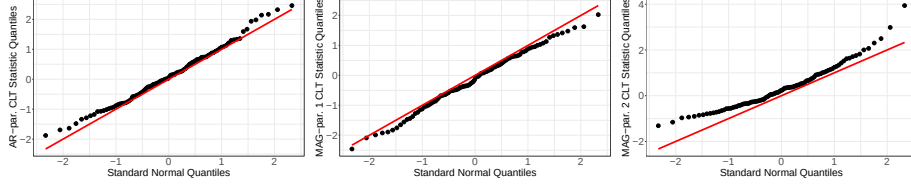


Figure D.6: QQ-plot of the CLT-statistic of the first MAG-parameter estimator in the $\Psi^{(2)}$ -MAGMAR(1,1)-(g)-(t) model from the simulation study in Sect. 4.1.

where $t_{\Sigma, \nu}$ denotes the multivariate t -distribution with correlation matrix $\Sigma \in [-1, 1]^{p \times p}$, symmetric and positive definite, degrees of freedom $\nu > 0$ and t_{ν}^{-1} is the quantile function of the standard t -distribution with degrees-of-freedom ν . For $\nu \rightarrow \infty$, the t -copula approaches the normal copula. For small ν , the dependence structure given by the t -copula is heavy-tailed, meaning that extreme events coincide.

Gumbel Copula: The Gumbel copula is an Archimedean copula (see Hofert, 2008; Genest and Rivest, 1993), hence it is constructed via $C[u_1, \dots, u_p] = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_p))$ by choosing a continuous and non-increasing generator function ψ . The generator function for the Gumbel copula is given as $\psi(t) = \exp(-t^{\frac{1}{\rho}})$ (Hofert, 2008). The resulting copula is

$$C_{\rho}^{\text{Gumbel}} = \exp(-((- \ln(u_1))^{\rho} + \dots + (- \ln(u_p))^{\rho})^{\frac{1}{\rho}}). \quad (\text{C.3})$$

The Gumbel copula has non-zero upper tail dependence while having no lower tail dependence (Nelsen, 2006, Ex. 5. 22.), hence displaying an asymmetric dependence structure.

Independence Copula: The independence copula simply models independence. Accordingly it is given by the product of its arguments (Nelsen, 2006),

$$\Pi[u_1, \dots, u_p] = \prod_{i=1}^p u_i. \quad (\text{C.4})$$

Choosing Π as the copula in Eq. 1, it can be seen, that the joint distribution is just the product of marginals. Hence, random variables having the independence copula as their copula implies independence and vice versa. For $\Sigma = I_{p \times p}$, the normal copula becomes the independence copula. For $\Sigma = I_{p \times p}$ and $\nu \rightarrow \infty$ the t -copula becomes the independence copula and for $\rho = 1$, the Gumbel copula becomes the independence copula.

Appendix D. Additional Plots

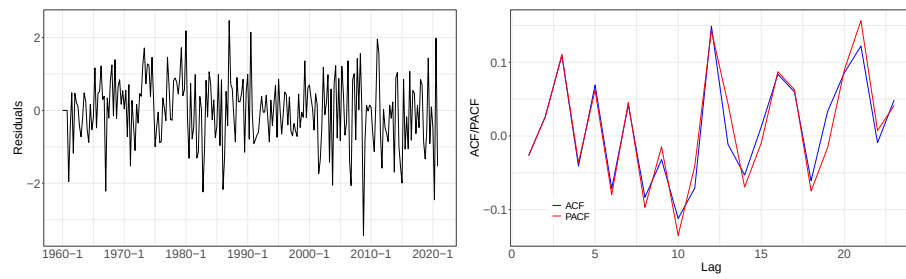


Figure D.7: Residuals and their ACF/PACF of the $\Psi^{(2)}$ -MAGMAR(4,1)-(g,i,n,g)-(t) model on US inflation. The residuals were transformed to the z -scale by applying the standard normal quantile function.

Copula-Based Time Series for Non-Gaussian and Non-Markovian Stationary Processes

Sven Pappert^{a,*}, Harry Joe^b

^aDepartment of Statistics, TU Dortmund University, Germany

^bDepartment of Statistics, University of British Columbia, Canada

Abstract

In the copula-based approach to univariate time series modeling, the finite dimensional temporal dependence of a stationary time series is captured by a copula. Recent studies investigate how copula-based time series models can be generalized to have long-term autoregressive effects. We study a generalization that comes from a Markov sequence of order p and a q -dependent sequence. We derive the relation of the model to Gaussian-ARMA models and to the Gaussian-GARCH(1, 1) model. We investigate distributional properties of the process and discuss the maximum likelihood estimation (MLE). Additionally we analyze the copula moving aggregate process of order one, or MAG(1), as it is a basic building block. Last we test the model in probabilistic forecasting studies on US inflation and German wind energy production.

Keywords: Copula, Dependence modeling, Identifiability, Probabilistic Forecasting, Tail dependence, Time series analysis

1. Introduction

Copula-based time series models have been used for non-Gaussian univariate time series. The joint distribution of $(p+1)$ consecutive observations is decomposed into the copula and the unconditional stationary distribution. If the underlying process is a Markov process of order p , the copula captures the serial dependence of the process. The model allows for non-linear serial dependence and can model clustering of extreme events properly by assigning a copula with suitable tail dependence. Additionally, the modeling of the unconditional stationary distribution is completely decoupled from the modeling of serial dependence, resulting in more flexibility. Copula-based time series models (or multivariate version thereof) are subject to extensive research and are used for applications. To this end, the model is successfully employed for forecasting of volatile time series by e.g. [Loaiza-Maya et al. \(2018\)](#) and [Pappert and Arsova \(2023\)](#). [Chen and Fan \(2006\)](#); [Nagler et al. \(2022\)](#) derive asymptotic properties of (semi-parametric) estimation in the univariate and the multivariate case respectively. [Simard and Rémillard \(2015\)](#); [Brechmann and Czado \(2015\)](#); [Nagler et al. \(2022\)](#); [Beare and Seo \(2015\)](#) consider multivariate extensions. Last, there is an emerging research brand on generalizing the model to have long-term serial dependence by [McNeil and Bladt \(2022\)](#); [Pappert \(2024\)](#); [Bladt et al. \(2025\)](#). This direction of research is also the context in which this work lies is. The purpose of this work is to explore a generalization that allows the copula-based time series model to have long-term autoregressive or non-Markov effects. As it turns out, this generalization

*Corresponding author

Email addresses: pappert@statistik.tu-dortmund.de (Sven Pappert), Harry.joe@ubc.ca (Harry Joe)

is equivalent to finding a copula-based generalization of the autoregressive moving average process (ARMA). We are interested in this generalization for the following reason. When the process that is to be modeled is not a Markov process of order p , modeling the distribution of $(p + 1)$ consecutive observations is not sufficient to capture the serial dependence. In fact, if the true data-generating process has asymptotically decreasing autocorrelation, e.g. ARMA(1, 1), no finite dimensional distribution is sufficient to capture the serial dependence. Additionally, we may be able to model processes with large p , that is a large autoregressive order, in a more parsimonious way. Now we discuss the existing approaches.

In the related literature three attempts for generalizing copula-based time series models are available. First, Joe (2014), in Sect. 3.14, proposes an approach inspired by the Poisson-ARMA model by McKenzie (1988). The basic idea is to model a time series by combining a $AR(p)$ copula-based time series model with a q -dependent copula-based time series model. The AR time series is a latent process in this model and serves as an innovation for the q -dependent time series. Second, McNeil and Bladt (2022) and Bladt et al. (2025) propose to consider an $AR(\infty)$ copula-based time series model and to reparametrize in terms of linear ARMA(p, q) parameters, where $p, q < \infty$. The authors conjecture the well-behavedness of the emerging time series based on the well-behavedness of the underlying linear ARMA model. Third, Pappert (2024) proposes the moving aggregate modified autoregressive copula-based time series model, which similarly to Joe (2014) introduces a q -dependent copula-based time series model.¹ The model is also combined with an AR copula-based time series model. In contrast to the model by Joe (2014), the q -dependent process is the latent process and is used as innovation process for the AR process. A major problem of this model is that, in its pure form, it does not have a uniform distribution as unconditional stationary distribution. This problem essentially prevents proper marginal modeling. The solution proposed by Pappert (2024) is to introduce an auxiliary transformation that ensures uniformity. However, it is cumbersome to properly employ this transformation. In summary, all of these models enable copula-based time series models to have infinite memory. Additionally, all these models can be understood as non-linear and probabilistic generalizations of the classic linear ARMA model.

The purpose of this paper is to examine and advance the theory for the model proposed by Joe (2014) and related processes. In its original publication, the model is introduced, the joint distribution of the time series and the innovations is derived and the likelihood is calculated. However, no dependence properties are obtained, and the link to Gaussian-ARMA is claimed without details. The model is given in terms of the following time series updating equations,

$$U_t = h(\varepsilon_t, \dots, \varepsilon_{t-q+1}, W_{t-q}), \quad W_t = g(\varepsilon_t, W_{t-1}, \dots, W_{t-p}), \quad \varepsilon_t \stackrel{iid}{\sim} U(0, 1). \quad (1)$$

The sequence $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ contains *i.i.d.* innovations; ε_t is independent of random variables with subscript $(t - 1)$ or less. The latent time series $\{W_t\}_{t \in \mathbb{Z}}$ follows an autoregressive copula-based time series model of order p , (compare Chen and Fan (2006), Beare (2010)). The mapping $g : (0, 1)^{(p+1)} \rightarrow (0, 1)$ is the conditional quantile function corresponding to some $(p + 1)$ -variate copula C . In the following we may call this copula the 'AR-copula'. It holds that $W_t \sim U(0, 1) \forall t \in \mathbb{Z}$ and that the copula of $(W_t, \dots, W_{t-p})$ is C . The time series $\{U_t\}_{t \in \mathbb{Z}}$ is obtained

¹In the paper, a q -dependent process is called 'moving aggregate'. We will use both names in this manuscript.

by plugging in W_{t-q} jointly with the innovations into the mapping $h : (0, 1)^{(q+1)} \rightarrow (0, 1)$. The mapping h again is a conditional quantile function corresponding to some $(q + 1)$ -dimensional copula K . In the following, K will be called the MAG-copula. As an example for a quantile function corresponding to a copula, we give the quantile function of a bivariate Gaussian copula. But first, we give the conditional cumulative distribution function (CDF). The conditional CDF corresponding to a bivariate Gaussian copula A with parameter α is given as (cf. [Smith et al. \(2010\)](#))

$$A_{2|1}(u|v) = \Phi\left(\frac{\Phi^{-1}(u) - \alpha\Phi^{-1}(v)}{\sqrt{1 - \alpha^2}}\right), \quad u, v \in (0, 1).$$

The conditional quantile function can be derived by inverting as

$$A_{2|1}^{-1}(p|v) = \Phi\left(\sqrt{1 - \alpha^2}\Phi^{-1}(p) + \alpha\Phi^{-1}(v)\right), \quad p, v \in (0, 1)$$

This form will play an important role for the Gaussian-MAG(1) process and the relation of the model under investigation to Gaussian-ARMA models discussed later. Since $(W_{t-q}, \varepsilon_{t-q+1}, \dots, \varepsilon_t)$ in Eq. 1 are independent and standard uniformly distributed, it follows that $U_t \sim U(0, 1)$. Consequently, any desired unconditional stationary distribution can be obtained by quantile-transforming $\{U_t\}$. This property is an advantage to the model by [Pappert \(2024\)](#), which needs an additional transformation to ensure uniformity. These features motivate us to further investigate the process defined by Eq. 1. We make the following contributions:

- a) First, we examine in detail the 1-dependent copula-based time series model, i.e. $\{V_t\}_{t \in \mathbb{Z}}$ given as $V_t = h(\varepsilon_t, \varepsilon_{t-1})$. This process is the building block of the model under investigation and also the MAGMAR-copula model by [Pappert \(2024\)](#). We show that time series inherits some dependence properties from the MAG(1)-copula, such as positive quadrant dependence (PQD) and stochastic increasingness (SI'ness). However, it seems that the tail dependence coefficients of (V_t, V_{t-1}) is restricted. A known result on the form of tail dependence matrices by [Embrechts et al. \(2016\)](#) gives an absolute upper bound as $\frac{1}{2}$. Using numerical approximations and simulations we observe that most MAG(1)-copula time series only exhibit a little to potentially no tail dependence in the joint distribution of consecutive observations.
- b) We show that a (subset of) Gaussian-ARMA process is recovered if the AR and MAG copulas are chosen as Gaussian and the time series is quantile transformed by the Gaussian quantile function. However, there are subtleties. For general model orders $p, q > 1$, a subset of a Gaussian-ARMA($p, q + p - 1$) is recovered. Due to the structure of the process additional terms arise that cause the distorted MA-order.
- c) We derive the copulas necessary to recover the ARCH and GARCH process. The copulas of a GARCH process are also investigated by [Dias et al. \(2024\)](#). The difference to our approach is that their approach captures the serial dependence employing the infinite partial dependence model by [McNeil and Bladt \(2022\)](#). We capture the dependence by an appropriate choice of the AR and MAG copulas. For the case where $q = 0$, i.e. the ARCH-model, our results coincide with the one from [Dias et al. \(2024\)](#) (albeit, we give a different derivation).
- d) We derive the copula of consecutive observations from the model in Eq. 1 with model orders $(p, q) \in$

$$\{(1, 1), (1, 2), (2, 1), (2, 2)\}.$$

- e) With regards to estimation, we first discuss the (near) non-identifiability of the classic linear MA(1) and our copula-based MAG(1). We show that, similar to the classic MA(1), the copula version also has two representations. The two representation can be obtained by permuting the order of the innovations in the updating equation. We explore potential identifiability problems.
- f) We discuss conditions for the consistency of the maximum-likelihood estimators of the process from Eq. 1, following the approach by Pappert (2024). The crucial step is to find conditions to establish the ergodicity of the score-time series based on the theory of stochastic recurrence equations, cf. Douc et al. (2014). Then general results for M-estimators for dependent data can be used, cf. Wooldridge (1994). Currently the conditions are on a high-level and future research directions could be to translate these high-level conditions to conditions on the copula parameters.
- g) Last, we model and forecast time series with different features using the model from Eq. 1. These time series are i) US-inflation and ii) German wind-power production. The R code for estimation, simulation and forecasting is available upon request. The implementation is based on the `rvinecopulib` package (see Nagler and Vatter (2022)).

In remainder of this manuscript, we introduce the model in Sect. 2. We proceed to derive the relation to other models, theoretical properties and discuss the iterative estimation. In Sect. 3 we use simulations and numerical approximations to explore properties of the MAG(1) and the model in Eq. 1. In Sect. 4, we use the model for probabilistic forecasting and in Sect. 5 we discuss our findings.

2. Copula-ARMA generalization

In Eq. 1 the mappings g and h are conditional quantile functions corresponding to the AR- and MAG copulas, C and K , respectively. The copula C is a D-vine copula where all copulas in a tree are set to be same copula. This condition is placed in order to assure stationarity. This condition may also be called *translational invariance*, cf. Nagler et al. (2022). We may call vine-copulas of this form *stationary D-vine* or *Toeplitz vine*. Due to the D-vine parametrization, the parameters governing the influences from different lags are algebraically independent. For illustration, we show the contour plot of such vine-copula in Fig. 1. Note that in this D-vine copula-based approach to time series modeling, the conditional dependencies are modeled. The copula K is set to be a D-vine copula with independent margins. This means that K is the copula of $(U_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q+1}, W_{t-q})$ with $(\varepsilon_{t-1}, \dots, \varepsilon_{t-q+1}, W_{t-q})$ being independent. K consists of q bivariate copulas, which govern the dependence between (U_t, ε_{t-1}) , $(U_t, \varepsilon_{t-2}|\varepsilon_{t-1})$ up to $(U_t, W_{t-q}|\varepsilon_{t-1}, \dots, \varepsilon_{t-q+1})$ with algebraically independent parameters. Such vine copulas may be called *MAG D-vine* or *q-dependent D-vine*. For illustration, we show the contour of such a vine-copula in Fig. 2. For the special case where $K = C^\perp$, the time series simplifies to $U_t = \varepsilon_t$, i.e. a standard uniform *i.i.d.* process. For $p = q = 1$ and $K = C^+$, we see that $U_t = W_{t-1} = C_{2|1}^{-1}(\varepsilon_{t-1}|W_{t-2}) = C_{2|1}^{-1}(\varepsilon_{t-1}|U_{t-1})$. This is the copula AR process with shifted innovations. The shift in the innovations does not affect the properties of the time series at all. In this section, we first derive the relation between the model from Eq. 1 and the linear Gaussian-ARMA and the Gaussian-GARCH(1,1) model. Then we proceed to investigate distributional

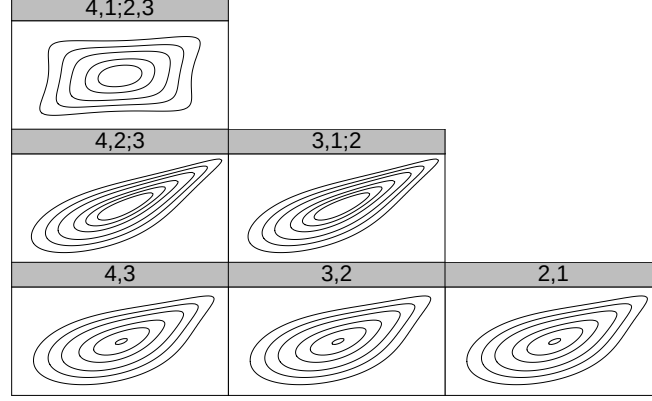


Figure 1: Contour plots of a stationary D-vine corresponding to an autoregressive copula-based time series model of order $p = 3$.

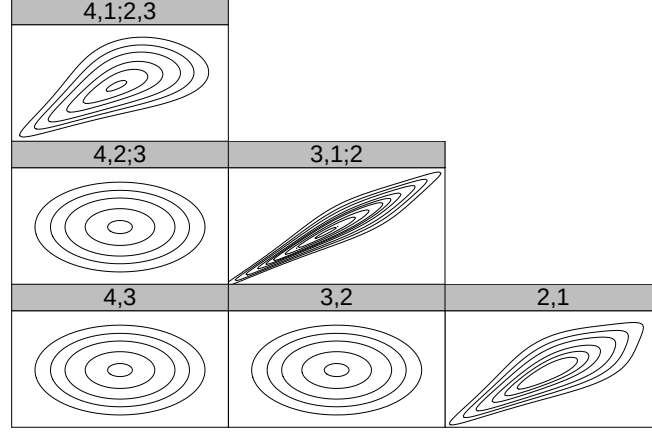


Figure 2: Contour plots of a MAG D-vine corresponding to a q -dependent copula-based time series model of order $q = 3$. Only dependencies involving variable 1 are different from independence.

properties such as stationarity, ergodicity, joint and conditional distributions, tail dependencies and measures of association. and last, we discuss the maximum likelihood estimation.

2.1. Relation to Other Models

In this section, we derive the relation between the model in Eq. 1 and popular time series models. In particular, we first show that for the AR and the MAG copula being the Gaussian copula, the transformed time series $\{\Phi^{-1}(U_t)\}_{t \in \mathbb{Z}}$ follows a Gaussian ARMA process. The relation between the parameters of the copula and the parameters of the emerging Gaussian ARMA is not trivial. Second, we show that we can recover the GARCH model with a suitable choice of copulas. We formalize the relation to the Gaussian ARMA process in the following proposition. We note that part a) is not new, but we include still for the sake of completeness.

Proposition 1 (Recovering Gaussian ARMA). *Let $\{U_t\}_{t \in \mathbb{Z}}$ be defined through Eq. 1 with latent process $W_t \sim U(0, 1)$ and innovations $\varepsilon_t \stackrel{iid}{\sim} U(0, 1)$. Let the copula corresponding to the AR-component be given as a*

$(p+1)$ -dimensional stationary Gaussian D-vine copula, which has a bivariate Gaussian copula with parameter α_j for lag $j \in \{1, \dots, p\}$. Let the copula corresponding to the MA-component be given as a $(q+1)$ -dimensional Gaussian MAG D-vine, which has a bivariate copula with parameter β_j for lag $j \in \{1, \dots, q\}$. Furthermore, let the time series $\{Y_t\}_{t \in \mathbb{Z}}$ be given as $Y_t = \Phi^{-1}(U_t)$ and the time series $\{X_t\}_{t \in \mathbb{Z}}$ as $X_t = \Phi^{-1}(W_t)$. Then the following holds.

a) The time series $\{X_t\}_{t \in \mathbb{Z}}$ is a Gaussian AR(p) process following the updating equation

$$X_t = \sum_{j=1}^p \phi_j^{(p)} X_{t-j} + \xi_t,$$

where $\xi_t = \sigma_\alpha \Phi^{-1}(\varepsilon_t) \stackrel{iid}{\sim} N(0, \prod_j (1 - \alpha_j^2))$ with $\sigma_\alpha = \prod_j (1 - \alpha_j^2)^{\frac{1}{2}}$ being the standard deviation of ξ_t so that $\text{Var}(X_t) = 1$ for all t . The parameters $\phi_j^{(p)}$ are defined recursively as,

$$\phi_j^{(p)} = \begin{cases} \phi_j^{(p-1)} - \alpha_p \phi_{p-j}^{(p-1)}, & j \in \{1, \dots, p-1\}, \\ \alpha_p, & j = p. \end{cases} \quad (2)$$

Compare Eq. 17 and 18 in [McNeil and Bladt \(2022\)](#), which also provide this result. We still include it for the sake of completeness.

b) If $p = 0$, the process $\{Y_t\}_{t \in \mathbb{Z}} = \{\Phi^{-1}(U_t)\}_{t \in \mathbb{Z}}$ is a Gaussian MA(q)-process and follows the updating equation

$$Y_t = \eta_t + \sum_{i=1}^q \theta_i^{(q)} \eta_{t-i},$$

where $\eta_t = \sigma_\beta \Phi^{-1}(\varepsilon_t) \stackrel{iid}{\sim} N(0, \prod_{j=1}^q (1 - \beta_j^2))$ with $\sigma_\beta = \prod_{j=1}^q (1 - \beta_j^2)^{\frac{1}{2}}$, which is the standard deviation of η_t such that $\text{Var}(Y_t) = 1$ for all t . The parameters $\theta_i^{(q)}$ are given as

$$\theta_i^{(q)} = \sigma_\beta^{-1} \left[\prod_{j=0}^{i-1} (1 - \beta_j^2)^{\frac{1}{2}} \right] \beta_i, \quad (3)$$

where we set $\beta_0 = 0$.

c) For arbitrary $p, q \in \mathbb{N}$, the time series $\{Y_t\}_{t \in \mathbb{Z}}$ follows the updating equation,

$$Y_t = \eta_t + \sum_{j=1}^p \phi_j^{(p)} Y_{t-j} + \sum_{j=1}^{p+q-1} \Psi_j^{(p,q)} \eta_{t-j}, \quad (4)$$

where the parameters $\phi_j^{(p)}$, $i \in \{1, \dots, p\}$ are given as Eq. 2 and the parameters $\Psi_j^{(p,q)}$, $j \in \{1, \dots, p+q-1\}$

are given as

$$\begin{aligned}
\Psi_j^{(p,q)} &= \tilde{\theta}_j^{(q)} - \tilde{\phi}_j^{(p)} - \zeta_j^{(p,q)}, \\
\tilde{\theta}_j^{(q)} &= \theta_j^{(q)} \times \mathbb{1}\{j \leq q \wedge q > 1\} + \theta_q^{(q)} \sigma_\alpha \times \mathbb{1}\{j = q\}, \\
\tilde{\phi}_j^{(p)} &= \phi_j^{(p)} \times \mathbb{1}\{j \leq p\}, \\
\zeta_j^{(p,q)} &= \left(\sum_{(r,s) \in S_j^{(p,q)}} \phi_r^{(p)} \theta_s^{(q)} \right) \times \mathbb{1}\{q > 1\}, \\
S_j^{(p,q)} &= \{(r,s) \in \mathbb{N}^2 : r \leq p, s \leq q, r+s = j\},
\end{aligned}$$

with $\theta_j^{(q)}$ as in Eq. 3.

Proof. The proof is given in [Appendix B.1](#). □

Now we give some examples where we calculate the coefficients from Eqs. 2-4 explicitly and do sanity checks by showing that $\text{Var}(Y_t) = 1$ for the respective time series.

Example 1 (Gaussian-MAG(1)). Consider $Y_t = \Phi^{-1}(V_t) = \Phi^{-1}(K_{2|1}^{-1}(\varepsilon_t | \varepsilon_{t-1}))$, where $K_{2|1}$, which is the copula of (V_t, ε_{t-1}) , is a Gaussian copula with parameter β_1 and $\varepsilon_t \stackrel{iid}{\sim} U(0, 1)$. Then $q = 1$ and $\{Y_t\}_{t \in \mathbb{Z}}$ follows the updating equation $Y_t = \theta_1^{(1)} \eta_{t-1} + \eta_t$. We have $\sqrt{\text{Var}(\eta_t)} = \sigma_\beta = (1 - \beta_1^2)^{\frac{1}{2}}$ and

$$\theta_1^{(1)} = (1 - \beta_1^2)^{-\frac{1}{2}} [1 - \beta_0^2]^{\frac{1}{2}} \beta_1 = \frac{\beta_1}{(1 - \beta_1^2)^{\frac{1}{2}}}.$$

As a sanity check, we test whether it holds true that $\text{Var}(Y_t) = 1$. In fact

$$\text{Var}(Y_t) = \text{Var}(\eta_t) + \left(\theta_1^{(1)} \right)^2 \text{Var}(\eta_{t-1}) = (1 - \beta_1^2) + \frac{\beta_1^2 (1 - \beta_1^2)}{1 - \beta_1^2} = 1.$$

This verifies that the process V_t is in fact $U(0, 1)$ and $Y_t \sim N(0, 1)$.

Example 2 (Gaussian-MAG(2)). Consider $Y_t = \Phi^{-1}(V_t) = \Phi^{-1}(K_{3|21}^{-1}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}))$ where $\varepsilon_t \stackrel{iid}{\sim} U(0, 1)$ and $K_{3|21}$ is a 3-dimensional Gaussian MAG D-vine copula with parameters β_1 and β_2 . Then the random vector (V_t, ε_{t-1}) has a Gaussian copula with parameter β_1 , the (conditional) random vector $(V_t, \varepsilon_{t-2} | \varepsilon_{t-1})$ has a Gaussian copula with parameter β_2 and $(\varepsilon_{t-1}, \varepsilon_{t-2})$ are independent. We have $q = 2$ and $\{Y_t\}_{t \in \mathbb{Z}}$ follows the updating equation $Y_t = \theta_1^{(2)} \eta_{t-1} + \theta_2^{(2)} \eta_{t-2} + \eta_t$ where $\sqrt{\text{Var}(\eta_t)} = \sigma_\beta = (1 - \beta_1^2)^{\frac{1}{2}} (1 - \beta_2^2)^{\frac{1}{2}}$. Furthermore

$$\begin{aligned}
\theta_1^{(2)} &= (1 - \beta_1^2)^{-\frac{1}{2}} (1 - \beta_2^2)^{-\frac{1}{2}} \beta_1 = \frac{\beta_1}{(1 - \beta_1^2)^{\frac{1}{2}} (1 - \beta_2^2)^{\frac{1}{2}}} \quad \text{and} \\
\theta_2^{(2)} &= (1 - \beta_1^2)^{-\frac{1}{2}} (1 - \beta_2^2)^{-\frac{1}{2}} [1 - \beta_1^2] \beta_2 = \frac{\beta_2}{(1 - \beta_2^2)^{\frac{1}{2}}}.
\end{aligned}$$

Then Y_t may be represented as

$$Y_t = \eta_t + \frac{\beta_1}{(1 - \beta_1)^{\frac{1}{2}}(1 - \beta_2)^{\frac{1}{2}}} \eta_{t-1} + \frac{\beta_2}{(1 - \beta_2)^{\frac{1}{2}}} \eta_{t-2}$$

We again check whether $\text{Var}(Y_t) = 1$,

$$\text{Var}(Y_t) = (1 - \beta_1^2)(1 - \beta_2^2) + \beta_1^2 + \beta_2^2(1 - \beta_1^2) = 1.$$

Example 3 (Gaussian copulas and $p = 1, q = 1$ in Eq. 1). Let $\{U_t\}_{t \in \mathbb{Z}}$ follow the model in Eq. 1 with $p = q = 1$. Let the copula corresponding to the AR component be given as bivariate Gaussian with parameter α and let the copula corresponding to the MA component be given as bivariate Gaussian with parameter β . Then $p + q - 1 = 1$ and $Y_t = \Phi^{-1}(U_t)$ follows the updating equation

$$Y_t = \phi_1^{(1)} Y_{t-1} + \Psi_1^{(1,1)} \eta_{t-1} + \eta_t,$$

where $\sqrt{\text{Var}(\eta_t)} = (1 - \beta^2)^{\frac{1}{2}}$ and the AR parameter is given as $\phi_1^{(1)} = \alpha$. The MA parameter is comprised of $\tilde{\theta}_1^{(1)} = \theta_1^{(1)} \sigma_\alpha = \frac{\beta(1-\alpha^2)^{\frac{1}{2}}}{(1-\beta^2)^{\frac{1}{2}}}$ and $\zeta_1^{(1,1)} = 0$ as well as $\tilde{\phi}_1^{(1)} = \phi_1^{(1)} = \alpha$. Consequently $\Psi_1^{(1,1)} = \frac{\beta(1-\alpha^2)^{\frac{1}{2}}}{(1-\beta^2)^{\frac{1}{2}}} - \alpha$. We again check whether it holds that $\text{Var}(Y_t) = 1$. To this end we note that the variance of an ARMA(1, 1) process, say Z_t , with parameters (ϕ, θ, σ) is given as

$$\text{Var}(Z_t) = \frac{1 + 2\phi\theta + \theta^2}{1 - \phi^2} \sigma^2.$$

Then it holds that

$$\begin{aligned} \text{Var}(Y_t) &= \frac{(1 + 2\phi_1^{(1)}\Psi_1^{(1,1)} + [\Psi_1^{(1,1)}]^2)}{1 - [\phi_1^{(1)}]^2} \sigma^2 \\ &= \frac{1 + 2\alpha \left[\frac{(1-\alpha^2)^{\frac{1}{2}}\beta}{(1-\beta^2)^{\frac{1}{2}}} - \alpha \right] + \left[\frac{(1-\alpha^2)^{\frac{1}{2}}\beta}{(1-\beta^2)^{\frac{1}{2}}} - \alpha \right]^2}{1 - \alpha^2} \sigma^2 \\ &= \frac{1 + (-2\alpha^2) + 2\alpha \frac{(1-\alpha^2)^{\frac{1}{2}}\beta}{(1-\beta^2)^{\frac{1}{2}}} + \frac{(1-\alpha^2)\beta^2}{(1-\beta^2)} - 2\alpha \frac{(1-\alpha^2)^{\frac{1}{2}}\beta}{(1-\beta^2)^{\frac{1}{2}}} + \alpha^2}{1 - \alpha^2} \sigma^2 \\ &= \frac{1 - \alpha^2 + \frac{(1-\alpha^2)\beta^2}{(1-\beta^2)}}{1 - \alpha^2} \sigma^2 \\ &= \left(1 + \frac{\beta^2}{1 - \beta^2}\right) \sigma^2 \\ &= \frac{1}{1 - \beta^2} \sigma^2 = 1 \end{aligned}$$

Example 4 (Gaussian copulas and $p = 2, q = 1$ in Eq. 1). Let $\{U_t\}_{t \in \mathbb{Z}}$ follow the model in Eq. 1 with $p = 2$ and $q = 1$. Let the copula corresponding to the AR component, C_{321} , be given as Gaussian with parameters (α_1, α_2) and let the copula corresponding to the MA component be given as bivariate Gaussian with parameter β . Then

we have $p + q - 1 = 2$ and $\{Y_t\}_{t \in \mathbb{Z}} = \{\Phi^{-1}(U_t)\}_{t \in \mathbb{Z}}$ has updating equation

$$Y_t = \phi_1^{(2)} Y_{t-1} + \phi_2^{(2)} Y_{t-2} + \Psi_1^{(2,1)} \eta_{t-1} + \Psi_2^{(2,1)} \eta_{t-2} + \eta_t,$$

where the AR parameters are given as $\phi_1^{(2)} = \alpha_1 - \alpha_2 \alpha_1 = \alpha_1(1 - \alpha_2)$ and $\phi_2^{(2)} = \alpha_2$. For $\Psi_1^{(2,1)}$ we see that $\tilde{\phi}_1^{(2)} = \phi_1^{(2)} = \alpha_1(1 - \alpha_2)$ and $\tilde{\theta}_1^{(1)} = \theta_1^{(1)} \sigma_\alpha = (1 - \beta_1^2)^{-\frac{1}{2}} \beta_1 (1 - \alpha_1^2)^{\frac{1}{2}} (1 - \alpha_2^2)^{\frac{1}{2}}$. Furthermore $\mathbb{1}\{q > 1\} = 0$ and hence $\zeta_{1/2}^{(2,1)} = 0$. Now for $\Psi_2^{(2,1)}$ we have $\tilde{\phi}_2^{(2)} = \phi_2^{(2)} = \alpha_2$ and $\tilde{\theta}_2^{(1)} = 0$. In total we have

$$Y_t = [\alpha_1(1 - \alpha_2)]Y_{t-1} + \alpha_2 Y_{t-2} + \left[\frac{\beta_1(1 - \alpha_1^2)^{\frac{1}{2}}(1 - \alpha_2^2)^{\frac{1}{2}}}{(1 - \beta_1^2)^{\frac{1}{2}}} - \alpha_1(1 - \alpha_2) \right] \eta_{t-1} + [-\alpha_2] \eta_{t-2} + \eta_t.$$

Now we proceed to derive copulas for which the model in Eq. 1 can recover ARCH and GARCH like processes. Specifically for the ARCH process we obtain a similar result to the one from [Dias et al. \(2024\)](#). However for the GARCH process, where [Dias et al. \(2024\)](#) use the IPD approach by [McNeil and Bladt \(2022\)](#), we use the model from Eq. 1, which is fundamentally different.

Proposition 2 (Recovering ARCH(1), cf. [Dias et al. \(2024\)](#)). *Let $\{W_t\}_{t \in \mathbb{Z}}$ be a copula-AR model with updating equation $W_t = C_{2|1}^{-1}(\varepsilon_t, W_{t-1})$ with $\varepsilon_t \stackrel{iid}{\sim} U(0, 1)$. Let Ψ be the unconditional marginal distribution associated to a stationary Gaussian ARCH(1) process. Then it holds that $\Psi^{-1}(W_t)$ follows a Gaussian ARCH(1) process if*

$$C_{21}(u_1, u_2) = \int_0^{u_2} \Phi \left[\frac{\Psi^{-1}(u_1)}{\sqrt{\alpha_0 + \alpha_1(\Psi^{-1}(u'_2))^2}} \right] du'_2,$$

where (α_0, α_1) are the parameters of the stationary Gaussian ARCH(1) process.

Proof. A proof is given in Chapter 3.2 of [Dias et al. \(2024\)](#). In Appendix [Appendix B.2](#) we give an alternative proof. $\square$

Proposition 3 (Recovering GARCH(1, 1)). *Let $\{U_t\}_{t \in \mathbb{Z}}$ follow the model in Eq. 1 with $p = q = 1$, i.e. $U_t = K_{2|1}^{-1}(\varepsilon_t, W_{t-1})$, where $W_t = C_{2|1}^{-1}(\varepsilon_t, W_{t-1})$ and $\varepsilon_t \stackrel{iid}{\sim} U(0, 1)$. Let Ψ be the unconditional marginal distribution associated to a stationary Gaussian-GARCH(1, 1) process. Then it holds that $\Psi^{-1}(U_t)$ follows a Gaussian-GARCH(1, 1) process if*

$$C_{21}(u_1, u_2) = \int_0^{u_2} \Phi \left[\sqrt{\frac{\Gamma^{-1}(u_1)^2}{\alpha_0 \alpha_1 + \alpha_1 \Gamma^{-1}(u'_2)^2}} - \frac{\beta}{\alpha_1} \right] du'_2,$$

$$K_{21}(u_1, u_2) = \int_0^{u_2} \Phi \left[\frac{\Psi^{-1}(u_1)}{\sqrt{\alpha_0 + \Gamma^{-1}(u'_2)^2}} \right] du'_2$$

where $(\alpha_0, \alpha_1, \beta_1)$ are the parameters of the stationary Gaussian-GARCH(1, 1) process and Γ is the CDF of the modified-ARCH process defined in the proof.

Proof. The proof is given in [Appendix B.3](#). $\square$

Remarks 1. *Remarks on recovering GARCH(1, 1)*

- a) The proposition regards recovering the Gaussian-GARCH(1,1) process. The conditional distribution is assumed to be Gaussian in this case. The generalization to other conditional distributions (e.g. t or JSU) is straight-forward by replacing Φ with a suitable distribution.
- b) The proposition assures that a standard uniform process with GARCH(1,1)-dynamics can be constructed. This allows for processes with arbitrary stationary distributions and GARCH(1,1) dynamics.
- c) The proposition only derives the theoretical form of the copulas needed to mimic GARCH dynamics. In order to actually model a time series an efficient estimation scheme is necessary. Since the copulas both involve the (hard to tract) stationary distribution of a Gaussian-GARCH(1,1) model, standard estimation (e.g. MLE) is computationally demanding. A suitable strategy to overcome this problem could be to transform the time series of interest first to the pseudo-observation scale by applying e.g. the (rescaled) empirical CDF, and then to quantile transform w.r.t. an initial choice of $\Psi = \Psi^{(0)}$. Then the standard GARCH(1,1) estimation can be performed. The estimators will be biased due to the mismatch between the stationary distribution and the temporal dynamics. However, we can use these parameters to update the stationary distribution, i.e. $\Psi^{(1)}$. Then this distribution can be used to quantile transform the pseudo-observation time series and the standard Gaussian-GARCH(1,1) estimation can be conducted. This process can be iterated. A similar strategy is used by Pappert (2024) for the estimation of a similar copula-based time series. There up to two iterations are sufficient.

Now we proceed to derive properties of the model in Eq. 1 and the MAG(1)-copula model.

2.2. Properties of the model

Before giving new results, we recall the properties already derived by Joe (2014). First, the joint distribution of $(U_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q})$ in the q -dependent model is shown to be given by the MAG-copula K . Then the likelihood function (respectively, the joint density) is derived. Second, the likelihood of the whole model, compare Eq. 1, is derived for different combinations of model orders p and q . The distribution of U_t is standard uniform for all $t \in \mathbb{Z}$, if g and h in Eq. 1 are appropriate copula conditional quantile functions. Furthermore, the AR copula-based time series model is stationary and the MAG copula-based time series model is stationary and obviously q -dependent and hence ergodic. The full model, i.e. Eq. 1, is stationary since the unconditional distribution of U_t is the same for all t and the updating equation is not time dependent (in terms of Markov processes this property may also be called time-homogeneity). For the 1-dependent process $\{V_t\}_{t \in \mathbb{Z}}$, following the updating equation

$$V_t = h(\varepsilon_t, \varepsilon_{t-1}), \quad \varepsilon_t \stackrel{iid}{\sim} U(0, 1),$$

the following holds. For K being the (co-/counter-) monotonicity copula, $K = C^+$ or $K = C^-$, respectively, the process $V_t = \varepsilon_{t-1}$ or $V_t = -\varepsilon_{t-1}$ is recovered. For K being the independence copula, $K = C^\perp$, we obtain $V_t = \varepsilon_t$, compare also Sect. 3.2 of Pappert (2024). Now we start to state additional properties. First, we are interested in the MAG(1) process. The following distributional properties hold.

Proposition 4 (Properties of MAG(1)). *Let $\{V_t\}_{t \in \mathbb{Z}}$ be a time series with updating equation $V_t = h(\varepsilon_t, \varepsilon_{t-1})$ and $\varepsilon_t \stackrel{iid}{\sim} U(0, 1)$, where $h(\varepsilon_t, \varepsilon_{t-1}) = K_{2|1}^{-1}(\varepsilon_t | \varepsilon_{t-1})$ and K_{21} is the copula of (V_t, ε_{t-1}) . It holds that*

- a) The time series $\{V_t\}_{t \in \mathbb{Z}}$ is stationary and ergodic.
- b) The joint distribution of consecutive observations is given as (cf. Appendix B.2 from [Pappert \(2024\)](#))
- $$F_{V_t, V_{t-1}}(u, v) = \int_0^1 K_{12}(K_{2|1}(u|s), v) ds. \quad (5)$$
- c) If K is positive quadrant dependent (PQD), then $F_{V_t, V_{t-1}}$ is also PQD. If K is negative quadrant dependent (NQD), then $F_{V_t, V_{t-1}}$ is NQD.
- d) If $K_{2|1}$ is stochastically increasing (SI) in its second argument, then V_t is SI in V_{t-1} .
- e) The Spearman's ρ associated to the joint distribution $F_{V_t, V_{t-1}}$ is bounded in absolute value by $\frac{1}{2}$ (cf. [Tao \(2025\)](#)).
- f) The lower and upper tail dependence coefficients, λ_l and λ_u , associated to $F_{V_t, V_{t-1}}$ are bounded in absolute value by $\frac{1}{2}$ (cf. [Embrechts et al. \(2016\)](#)).
- g) Let $\{\tilde{V}_t\}_{t \in \mathbb{Z}}$ be a time series with updating equation $\tilde{V}_t = h(\varepsilon_{t-1}, \varepsilon_t)$, i.e. the MAG(1) updating equation with reversed order. Furthermore let the MAG(1)-copula be the Gaussian Copula. Then it holds that $(V_t, \dots, V_{t-r}) \stackrel{d}{=} (\tilde{V}_t, \dots, \tilde{V}_{t-r})$, $\forall r \in \mathbb{Z}$.

Proof. The proof is given in [Appendix B.4](#) □

Remarks 2. Remarks on the properties of the MAG(1) model.

- a) Part g) of the proposition states that a Gaussian MAG(1) process has two representations. In particular, the process $\{V_t\}_{t \in \mathbb{Z}}$, defined as $V_t = \Phi[(1 - \alpha^2)^{\frac{1}{2}} \Phi^{-1}(\varepsilon_{t-1}) + \alpha \Phi^{-1}(\varepsilon_t)]$ with i.i.d. innovations $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ can be represented as $V_t = \Phi[(1 - \alpha^2)^{\frac{1}{2}} \Phi^{-1}(\eta_{t-1}) + \alpha \Phi^{-1}(\eta_t)]$ with i.i.d. innovations $\{\eta_t\}_{t \in \mathbb{Z}}$. From the proposition it is clear that $\eta_t = \varepsilon_{t-1}$ and $\eta_{t-1} = \varepsilon_t$. Since there is no temporal structure in the innovations, we are free to permute them and obtain the two representations. We see that there is a critical point at which there is only one representation. This point is reached when $\alpha = (1 - \alpha^2)^{\frac{1}{2}}$, which is solved by $\alpha = \pm 1/\sqrt{2}$. Furthermore, given the representation $\{V_t\}$ with parameter α , the parameter of $\{\tilde{V}_t\}$ has to be chosen as $\alpha^\dagger = (1 - \alpha^2)^{\frac{1}{2}}$ such that $\{\tilde{V}_t\}$ is a representation of $\{V_t\}$. We may call $\alpha^\dagger$ the reciprocal to the value α . This property is the pendant to the two representations of a Gaussian MA(1) process. In Gaussian MA(1) models, the two representations are

$$\begin{aligned} Y_t &= \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim WN(0, \sigma^2) \\ \tilde{Y}_t &= \tilde{\varepsilon}_t + \frac{1}{\theta} \tilde{\varepsilon}_{t-1}, \quad \tilde{\varepsilon}_t \sim WN(0, \sigma^2 \theta^2), \end{aligned}$$

compare e.g. [Hamilton \(2020\)](#). While the equivalence of both representations is usually proven by showing that their autocovariance functions coincide, the property could also be shown by using the permutation argument we presented. The result from our proposition may be seen as a complement of the MA(1) result. While the classic result holds for MA(1) processes with elliptic white noise, our result holds for any i.i.d.

innovation process. The two representations give rise to a non-identifiability issue in MLE. We address this issue in Sect. 2.3 and investigate the likelihood by simulations in Sect. 3.

- b) The Spearman's rho being bounded in absolute value by $\frac{1}{2}$ is similar to the property of classic MA(1) processes. In fact we used the classic property in our proof. The statement generalizes the classic result to any 1-dependent process. The intuition for this result becomes clear when considering the underlying process again: $V_t = K_{2|1}^{-1}(\varepsilon_t | \varepsilon_{t-1})$.
- c) The absolute upper bound for the tail dependence coefficients being $\frac{1}{2}$ follows from a general argument on the form of tail dependence matrices by Embrechts et al. (2016). However, numerical evaluations and estimations on simulations indicate a sharper bound for 'standard' copulas such as the Gumbel, t or Clayton copula (see Sect. 3). Furthermore we show that the tail dependence coefficients for the Fréchet copula are bounded by $\frac{1}{4}$. It appears that the MAG(1) construction with an absolutely continuous copula K_{21} might not be able to result in a time series with serial tail dependence. There could be other 1-dependent time series constructions that can lead to a tail dependence coefficient between $\frac{1}{4}$ and $\frac{1}{2}$.

Now we proceed with a special toy example for a MAG(1) process. We consider the Fréchet copula as MAG-copula and investigate the induced tail dependence on consecutive observations from the resulting time series.

Example 5 (MAG(1)-Fréchet Copula process). We consider a MAG(1) process $\{V_t\}_{t \in \mathbb{Z}}$, where the MAG-copula is given as the Fréchet copula,

$$\begin{aligned} K_{21}(u_1, u_2) &= (1 - \alpha)C^-(u_1, u_2) + \alpha C^+(u_1, u_2), \\ &= (1 - \alpha)u_1 u_2 + \alpha \min\{u_1, u_2\}. \end{aligned}$$

Sometimes the countermonotonicity copula, C^- is also added to the convex combination in the Fréchet copula. Here we omit this component. The conditional distribution and quantile function of K_{12} are given as

$$K_{2|1}(u_1 | u_2) = (1 - \alpha)u_1 + \alpha \mathbb{1}\{u_2 \leq u_1\}$$

almost everywhere. Now we calculate the implied tail dependence coefficients of the joint distribution of (V_t, V_{t-1}) using Eq. 5,

$$\begin{aligned} F_{V_t, V_{t-1}}(x_1, x_2) &= \int_0^1 [\alpha \min\{K_{2|1}(x_1 | s), x_2\} + (1 - \alpha)x_2 K_{2|1}(x_1 | s)] ds \\ &= \int_0^1 (1 - \alpha)x_2[(1 - \alpha)x_1 + \alpha \mathbb{1}\{s \leq x_1\}] ds + \int_0^1 \alpha \min\{(1 - \alpha)x_1 + \alpha \mathbb{1}\{s \leq x_1\}, x_2\} ds \\ &= (1 - \alpha)^2 x_1 x_2 \int_0^1 ds + (1 - \alpha)\alpha x_2 \int_0^{x_1} ds \\ &\quad + \int_0^{x_1} \alpha \min\{(1 - \alpha)x_1 + \alpha, x_2\} ds + \int_{x_1}^1 \alpha \min\{(1 - \alpha)x_1, x_2\} ds \\ &= (1 - \alpha)^2 x_1 x_2 + (1 - \alpha)\alpha x_1 x_2 + x_1 \alpha \min\{(1 - \alpha)x_1 + \alpha, x_2\} + (1 - x_1)\alpha \min\{(1 - \alpha)x_1, x_2\} \end{aligned}$$

Now we consider $\lim_{u \rightarrow 0} F_{V_t, V_{t-1}}(u, u)/u$. For small u relative to α , $F_{V_t, V_{t-1}}(u, u)$ simplifies to

$$F_{V_t, V_{t-1}}(u, u) = (1 - \alpha)^2 u^2 + (1 - \alpha)\alpha u^2 + \alpha u^2 + (1 - u)\alpha(1 - \alpha)u.$$

Then it is clear that $\lim_{u \rightarrow 0} F_{V_t, V_{t-1}}(u, u)/u = \alpha(1 - \alpha)$. This expression is maximal at $\alpha = \frac{1}{2}$ with a value of $\frac{1}{4}$.

Now we proceed with the model from Eq. 1 with model orders $p = 1$ and $q = 1$. In the following proposition, we again state distributional properties.

Proposition 5. Let $\{U_t\}_{t \in \mathbb{Z}}$ follow the model from Eq. 1 with $p = q = 1$, i.e. $\{U_t\}_{t \in \mathbb{Z}}$ follows the updating equation, $U_t = h(\varepsilon_t, W_{t-1})$, with $W_s = g(\varepsilon_s, W_{s-1})$ and $\varepsilon_t \stackrel{iid}{\sim} U(0, 1)$. The mapping $g(\cdot, \cdot) = C_{2|1}(\cdot | \cdot)$ is the conditional quantile function corresponding to the bivariate copulas C_{21} , which by this construction is the copula of (W_s, W_{s-1}) for all s . The mapping $h(\cdot, \cdot) = K_{2|1}(\cdot | \cdot)$ is the conditional quantile function corresponding to the bivariate copulas K_{21} , which is the copula of (U_t, W_{t-1}) for all t . The following holds

- a) If $\{W_s\}_{s \in \mathbb{Z}}$ is stationary and ergodic, the time series $\{U_t\}_{t \in \mathbb{Z}}$ is stationary and ergodic.
- b) Assuming C_{21} is absolutely continuous with density c_{21} , the joint distribution of (U_t, U_{t-1}) is given as

$$F_{U_t, U_{t-1}}(x_1, x_2) = \int_0^1 \int_0^{C_{2|1}^{-1}(K_{2|1}(x_2 | w_2) | w_2)} K_{2|1}(x_1 | w_1) c_{21}(w_1, w_2) dw_1 dw_2 \quad (6)$$

Proof. The proof is given in Appendix B.5. □

Remarks 3. Remarks on properties of the model from Eq. 1 with $p = q = 1$.

- a) For $K_{21} = C^\perp$ it holds that $F_{U_t, U_{t-1}} = C^\perp$ irrespective of the choice of C_{21} . For $K_{21} = C^+$ we see that $F_{U_t, U_{t-1}}$ equals the AR-copula C_{21} . Last, for $C_{21} = C^\perp$, we recover the MAG(1) joint distribution from Eq. 5. In Appendix A.1, we derive these results in detail.

Proposition 6. Let $\{U_t\}_{t \in \mathbb{Z}}$ follow the model from Eq. 1 with AR copula $K_{(p+1), \dots, 1}$, MAG copula $C_{(q+1), \dots, 1}$, latent standard uniform process $\{W_s\}_{s \in \mathbb{Z}}$, i.i.d. standard uniform innovations $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ and model orders p and q . Then it holds that

- a) For $p = 1, q = 2$, where C_{21} is the copula of (W_s, W_{s-1}) for all s and K_{321} is the copula of the random vector $(U_t, \varepsilon_{t-1}, W_{t-1})$ for all t , the joint distribution of consecutive observations is given as

$$F_{U_t, U_{t-1}}(x_1, x_2) = \int_{(0,1)^3} K_{3|21}(x_1 | e_1, w_2) \times c_{12}(w_2, w_3) \times \mathbb{1}\{e_1 \leq K_{3|21}(x_2 | C_{2|1}(w_2 | w_3), w_3)\} dw_2 dw_3 de_1,$$

assuming C_{12} is absolutely continuous.

- b) For $p = 2, q = 1$, where C_{321} is the copula of (W_s, W_{s-1}, W_{s-2}) for all s and K_{21} is the copula of (U_t, W_{t-1}) for all t , the joint distribution of consecutive observations is given as

$$F_{U_t, U_{t-1}}(x_1, x_2) = \int_{(0,1)^3} K_{2|1}(x_1 | w_1) \times c_{321}(w_1, w_2, w_3) \times \mathbb{1}\{w_1 \leq C_{3|21}^{-1}(K_{2|1}(x_2 | w_2) | w_2, w_3)\} dw_1 dw_2 dw_3,$$

assuming C_{321} is absolutely continuous.

- c) For $p = 2, q = 2$, where C_{321} is the copula of (W_s, W_{s-1}, W_{s-2}) for all s and K_{321} is the copula of $(U_t, \varepsilon_{t-1}, W_{t-2})$ for all t , the joint distribution of consecutive observations is given as

$$F(x_1, x_2) = \int_{(0,1)^4} K_{3|21}(x_1|e_1, w_2) \times c_{321}(w_2, w_3, w_4) \\ \times \mathbf{1}\{e_1 \leq K_{3|21}(x_2|C_{3|21}(w_2|w_3, w_4), w_3)\} de_1 dw_2 dw_3 dw_4,$$

assuming C_{321} is absolutely continuous.

Proof. The proof is given in [Appendix B.6](#). □

Now we state the likelihood of the model from Eq. 1 and provide an algorithm that calculates the likelihood iteratively.

2.3. Estimation

In this section we discuss the maximum likelihood estimation (MLE). [Joe \(2014\)](#) has a general result for the likelihood, with a Jacobian and specific details for $(p, q) \in \{(0, 1), (0, 2), (1, 2), (2, 1)\}$. We note that estimation properties of the MAG(1) process are also discussed by [Pappert \(2024\)](#).

In order to derive the likelihood, the conditional density of $U_t|\mathcal{F}_{t-1}$ is necessary, where $\mathcal{F}_s$ denotes all information up to and including $s \in \mathbb{Z}$. Assuming $K_{(q+1),\dots,1}$ is absolutely continuous, the conditional density of $U_t|\mathcal{F}_{t-1}$ from the model in Eq. 1 with arbitrary orders p and q is given as

$$f_{U_t|\mathcal{F}_{t-1}}(x) = k_{(q+1),\dots,1}(x, \varepsilon_{t-1}, \dots, \varepsilon_{t-q+1}, W_{t-q}),$$

where $k_{(q+1),\dots,1}$ is the $(q+1)$ -variate density of the MAG-copula, $(\varepsilon_{t-1}, \dots, \varepsilon_{t-q+1})$ are the innovations and W_{t-q} the latent AR-process. Given a set of candidate AR and MAG copulas, which we may denote by $\hat{K}_{(q+1),\dots,1}$ and $\hat{C}_{(p+1),\dots,1}$, with candidate parameter values, for which the likelihood ought to be estimated, the innovation at time s can be estimated as $\hat{\varepsilon}_s = \hat{K}_{(q+1)|q,\dots,1}(U_s|\hat{\varepsilon}_{s-1}, \dots, \hat{\varepsilon}_{s-q+1}, \hat{W}_{t-q})$, given the preceding values. Similarly the latent AR-process $\hat{W}_s$ at time s can be estimated as $\hat{W}_s = \hat{C}_{(p+1)|p,\dots,1}^{-1}(\hat{\varepsilon}_s|\hat{W}_{s-1}, \dots, \hat{W}_{s-p})$ given the parameters and preceding values. It is evident that $\{\hat{\varepsilon}_s\}$ and $\{\hat{W}_s\}$ need to be calculated iteratively. In algorithm 1 we show how the likelihood can be calculated using the `rvinecopulib`-package in R. Given the algorithm for calculating the likelihood, numerical optimization can be utilized to find the minimum of the negative log-likelihood. In order to establish consistency, the stationarity and ergodicity of the estimated latent processes $\{\hat{\varepsilon}_s\}$ and $\{\hat{W}_s\}$ is needed for all feasible parameter values. Assuming that the data-generating process is the model from Eq. 1, the stationarity and ergodicity of the observed process $\{U_t\}_{t \in \mathbb{Z}}$ follows. If the copulas are chosen such that $\{\hat{W}_t\}$ fulfills the conditions from Theorem 4.40 by [Douc et al. \(2014\)](#), the stationarity and ergodicity of $\{\hat{W}_t\}$ and $\{\hat{\varepsilon}_t\}$ follows. Then, under additional assumptions, Theorem 4.2 from [Wooldridge \(1994\)](#) can be applied to establish the uniform weak law of large numbers for the log-likelihood estimator. Then Theorem 4.3 by [Wooldridge \(1994\)](#) can be applied to establish consistency (again under additional

assumptions). Establishing stationarity and ergodicity of the estimated latent processes seems to be the crucial step. Currently it not clear how these conditions can be translated to conditions on the AR and MAG copulas.

Algorithm 1: likelihood calculation for the model from Eq. 1 based on the R-package `rvinecopulib`.

The objects `model_AR` and `model_MA` contain all information about the AR copula $C_{(p+1),\dots,1}$ and the MAG copula $K_{(q+1),\dots,1}$ respectively.

```

input :  $x$ ; model_AR, model_MA,  $p$  and  $q$ 
output: Negative Log-Likelihood

 $l \leftarrow \text{length}(x)$ 
 $d, e, W \leftarrow \text{vector}()$ 
 $r \leftarrow \max(p, q)$ 
 $d_{1:r}, e_{1:r}, W_{1:r} \leftarrow (1, 0.5, 0.5)$ 
for  $t \leftarrow (r + 1)$  to  $l$  do
     $e_t^\dagger \leftarrow \text{inverse\_rosenblatt}(u = (W_{t-q}, e_{(t-q+1):(t-1)}, 0.5), \text{model\_MA})$ 
     $e_t \leftarrow \text{rosenblatt}(x = (e_{1:q}^\dagger, x_t), \text{model\_MA})_{(q+1)}$ 
     $W_t^\dagger \leftarrow \text{rosenblatt}(x = (W_{(t-p):(t-1)}, 0.5), \text{model\_AR})$ 
     $W_t \leftarrow \text{inverse\_rosenblatt}((W_{1:p}^\dagger, e_t), \text{model\_AR})_{(p+1)}$ 
     $d_t \leftarrow \text{dvinecop}(u = (x_t, e_{(t-1):(t-q+1)}, W_{(t-q)}), \text{model\_MA})$ 
end
 $Q \leftarrow \sum_{t=1}^l [-\log d_t]$ 
return  $Q$ 

```

3. Simulation and Numerical Experiments

We evaluate dependence measures of the model from Eq. 1 for lag orders $(p, q) \in \{(0, 1), (1, 1)\}$. In particular we are interested in Spearman's rho, upper/lower tail dependence coefficients and the tail order. We refer to Sect. 2.13 and 2.16 by Joe (2014) for tail dependence and the tail order, respectively. We use the joint distributions derived in the respective propositions, cf. prop. 4 and 5, and evaluate the integrals numerically. For the numerical approximation of integrals, we use Gaussian quadrature. Furthermore, the tail dependence coefficients are approximated by setting $u = 0.05$ in $F_{U_t, U_{t-1}}(u, u)/u$ and $[1 - 2(1 - u) + F_{U_t, U_{t-1}}(1 - u, 1 - u)]/u$, respectively. The tail order is also approximated by setting $u = 0.05$ in $\log F_{U_t, U_{t-1}}(1 - u, 1 - u)/\log u$ and analogous for the upper tail dependence. These approximations are called 5%-tail dependence coefficients and 5%-tail order, respectively. For Spearman's rho, we also estimate the dependence measures on simulated time series with $n = 1000$ observations.

In Fig. 3, the estimated dependence measure of consecutive observations from the Gaussian MAG(1) process are displayed. First, the implied Spearman's rho, then the lower tail dependence coefficient/order and the upper tail dependence coefficient/order. The results are expected. First, the numerical approximation and simulation show Spearman's rho associated to the joint distribution of (V_t, V_{t-1}) is bounded in absolute value by $\frac{1}{2}$. This is in line with part e) of proposition 4. As the Gaussian copula exhibits no tail dependence, we also do not expect

tail dependence in the joint distribution of consecutive observations from a Gaussian-MAG(1) process. This expectation is backed by the insights from proposition 1: The Φ -quantile transformed Gaussian-MAG(1) is a Gaussian-MA(1) process, which exhibits no tail dependence. Since a contemporaneous monotone transformation (i.e. the Φ -quantile transformation) does not affect dependence properties, we can conclude that Gaussian-MAG(1) does not have tail dependence. This is what we can observe in the numerical approximations. The approximated 5%-tail dependence coefficient is close to zero with a few fluctuations. The 5%-tail order is above the critical value of 1. For the Frank copula we observe similar behavior in Fig. 7. Now for the Clayton and Gumbel copula in Fig. 5 and Fig. 4, respectively, we first observe that the absolute bound of $\frac{1}{2}$ for Spearman's rho is confirmed, again in line with proposition 4. For the tail dependence coefficients the conclusion is not as clear. We observe that the tail order stays above the critical limit of 1 for all copulas under consideration. Furthermore, the known absolute bound of $\frac{1}{2}$, cf. Embrechts et al. (2016) or part f) of proposition 4, is not attained. Compare the comment in Remark 2 part c).

Now we consider simulations from the model from Eq. 1 with model orders $p = q = 1$. We want to demonstrate that the dependence measures are not as restricted as in the MAG(1)-case. Due to computational demand reasons, we rely on correlations estimated on simulations and not on integral approximations. We examine the dependence of a model with a Gumbel AR-copula and a Gaussian MAG-copula. In Fig. 8, the estimated Spearman's rho values are displayed in dependence of the AR parameter. The MAG parameter for these simulations is fixed to $\beta \in \{0, 0.25, 0.9375\}$ respectively. For $\theta = 0$ we observe that the Spearman's rho value is zero irrespective of the AR parameter value. This is expected since the process simplifies to *i.i.d.* uniform noise for the MAG-copula being independence (cf. Sect. 1). Now for $\theta = 0.9375$, which is close to comonotonicity, The AR-copula's implied Spearman's rho is almost recovered. This is also not surprising because the process approaches a copula AR process for the MAG-copula being close to comonotonicity. For an intermediate value of the MAG parameter, $\theta = 0.25$, the Spearman's rho value is bounded. A numerical approximation/estimation on simulated data of the tail-dependence coefficients would also be interesting, however, due to the instability of the approximation and estimation in this case, we can not report reliable results.

Another research question, which we investigate by simulation, is related to part g) of proposition 4. Recall that it is shown that for a Gaussian-MAG(1) process $\{V_t\}_{t \in \mathbb{Z}}$, there is a process $\{\tilde{V}_t\}_{t \in \mathbb{Z}}$ which is equal in distribution to $\{V_t\}_{t \in \mathbb{Z}}$. This result in the process $\{V_t\}$ having two representations. We explore how these two representations affect MLE. To this end we simulate a copula-MAG(1) time series, calculate the likelihood for different parameter values and check whether the minimum of the negative log-likelihood (NLL) is at the true parameter value, which is used in simulation. First, we check the behavior of the NLL of the Gaussian-MAG(1). In Fig. 9 the estimated NLL is plotted against the copula parameter for three different values of the true parameter. The three different values are $\alpha \in \{0.25, 0.5, 0.875\}$. The true value is marked by the red dashed line. Additionally, the blue dashed line marks the reciprocal value to the true parameter value. This is the parameter value that needs to be chosen in the construction of $\{\tilde{V}_t\}_{t \in \mathbb{Z}}$ such that it is a representation of $\{V_t\}_{t \in \mathbb{Z}}$, compare Remark 2, part a). For the Gaussian copula the reciprocal value can be calculated as $\alpha^\dagger = \sqrt{1 - \alpha^2}$. We also recall that the critical value for the two representations of the Gaussian-MAG(1) process is $1/\sqrt{2} \approx 0.7071$. Given these information, the plots in Fig. 9 can be interpreted. First, for $\alpha \in \{0.25, 0.5\}$ we observe that the minimum of the NLL is attained at the true parameter value. This is expected. However, for $\alpha = 0.875$, which is above

the critical threshold $1/\sqrt{2}$ we observe that the minimum of the NLL is attained at the reciprocal value. Since the MAG(1) model with the reciprocal value is just a different representation of the data-generating process, the MLE does not give rise to bias. We conclude that the NLL is minimized by values which lie below the critical value. The reason for this behavior can be found when considering proposition 7 by Pappert (2024). The proposition concerns the consistency of MLE of the MAG(1) process. A crucial step in the derivation is that the time series of estimated innovations $\{\hat{e}_t\}_{t \in \mathbb{Z}}$ which is defined by the updating equation, $\hat{e}_t = \hat{K}_{2|1}(V_t|\hat{e}_{t-1})$, where $\{V_t\}$ is the observed process, is stationary and ergodic. Pappert (2024) assures this by assuming that $\hat{K}_{2|1}$ induces a contraction-on-average in the sense of Theorem 4.40 by Douc et al. (2014). For the Gaussian copula to induce a contraction-on-average, the parameter has to fulfill $|\hat{\alpha}| < 1/\sqrt{2}$.² If $\hat{\alpha}$ is outside the bound, then the estimated innovation time series is not stationary and ergodic. This is exemplified in the left plot of Fig. 11. After a burn-in period, the time series only oscillates. For the reciprocal of the true value, the residuals are stationary and ergodic. This is exemplified in the right plot of Fig. 11. Notice that the critical value for $\{\hat{e}_t\}_{t \in \mathbb{Z}}$ to be stationary and ergodic coincides with the critical value of the two representations of the Gaussian-MAG(1) process. We note that this behavior of the residuals and likelihood is similar to the behavior of the classical Gaussian MA(1). As discussed in Remark 2, part a), this model also has two representations. For estimation purposes the parameter is restricted to the *invertible region* $(-1, 1)$ (compare Sect. 3.7 by Hamilton (2020)). When relating the Gaussian-MAG(1) to the classical Gaussian MA(1) by Φ -quantile transforming, see Example 1, we see that the relation between the MAG-parameter β_1 and the MA-parameter θ_1 is given as $\theta_1 = \beta_1/(1 - \beta_1^2)^{\frac{1}{2}}$. Plugging in the critical value for β_1 we see that $\theta_1 = \pm 1$ is recovered which is the critical value of the classical MA(1) process. These findings imply that the MLE can not be consistent for $|\hat{\alpha}| \geq 1/\sqrt{2}$. This is in line with proposition 7 by Pappert (2024). In order to assure consistency, the parameter space may be restricted accordingly. This is not a problem, since $1/\sqrt{2}$ is also the critical value for the two representations. Hence a Gaussian-MAG(1) with estimated $|\hat{\alpha}| < 1/\sqrt{2}$ is not restricted in its temporal dependence by the bounds on $\hat{\alpha}$. We also investigate the NLL for the Gumbel-MAG(1) process. In Fig. 10 we again show the estimated NLL in dependence of parameter values. The red-dashed line again marks the true value and the blue-dashed line an approximate reciprocal value. The approximate reciprocal value is calculated as follows: We calculate the Kendall's tau value implied by the parameter of the Gumbel copula α , then we calculate the parameter of a Gaussian copula with such Kendall's tau value. Denote this value by ρ^* . We then calculate $\rho^\dagger = \sqrt{1 - \rho^{*2}}$, obtain the Kendall's tau value implied by this parameter and last calculate the parameter of the Gumbel copula, which has the same Kendall's tau value. This is only an approximation. In fact it is not even proven that there are two representations for the Gumbel-MAG(1) process. But to explore the process, we make use of the approximation. Now for $\alpha = 1.25$, the NLL is minimized at the true value. For $\alpha = 2.1$, the NLL attains its minimum at the approximate reciprocal value. However, for $\alpha = 4.5$ neither the true value nor the approximate reciprocal value minimize the likelihood, although the approximate reciprocal value is close. The results indicate that there may be identifiability for Gumbel-MAG(1) process, unlike with the Gaussian copula.

²This can be seen by considering the stochastic recurrence equation function $\tilde{f}_z(x) = \Phi^{-1}(\hat{K}_{2|1}(z|\Phi(x)))$. This mapping induces a contraction if the parameter of $\hat{K}_{2|1}$, say $\hat{\alpha}$, fulfills $|\hat{\alpha}/(1 - \hat{\alpha}^2)^{\frac{1}{2}}| < 1$. This is fulfilled if $|\alpha| < 1/\sqrt{2}$.

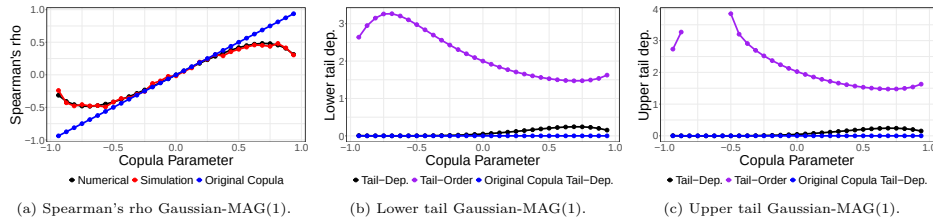


Figure 3: Numerical and simulation-based Spearman's rho, upper and lower 5%-tail dependence coefficients and order of the copula of consecutive observations from a Gaussian-MAG(1), along with the dependence measures of the original copula K_{21} .

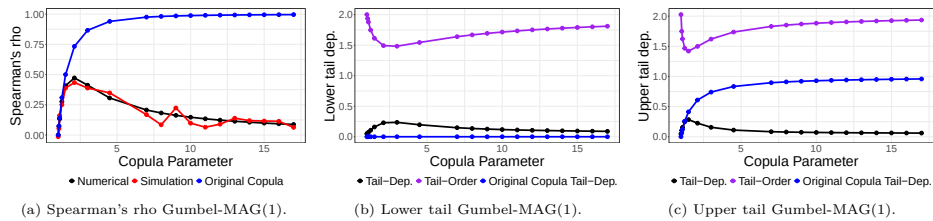


Figure 4: Numerical and simulation-based Spearman's rho, upper and lower 5%-tail dependence coefficients and order of the copula of consecutive observations from a Gumbel-MAG(1), along with the dependence measures of the original copula K_{21} .

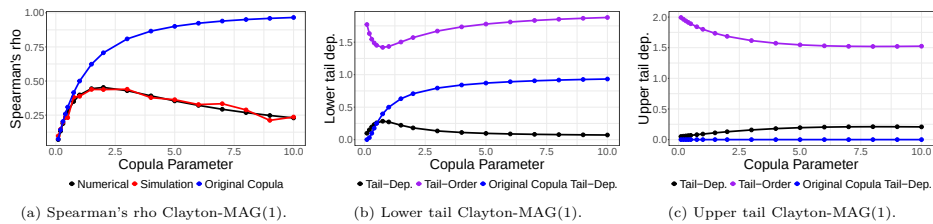


Figure 5: Numerical and simulation-based Spearman's rho, upper and lower 5%-tail dependence coefficients and order of the copula of consecutive observations from a Clayton-MAG(1), along with the dependence measures of the original copula K_{21} .

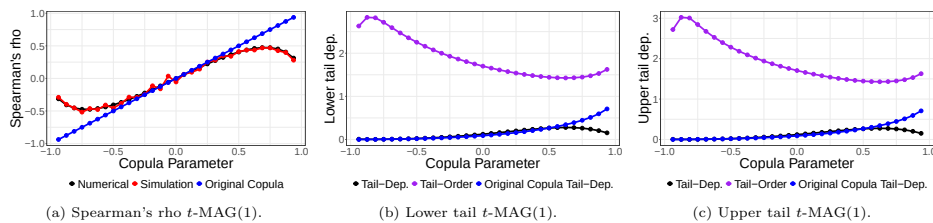


Figure 6: Numerical and simulation-based Spearman's rho, upper and lower 5%-tail dependence coefficients and order of the copula of consecutive observations from a t -MAG(1), along with the dependence measures of the original copula K_{21} .

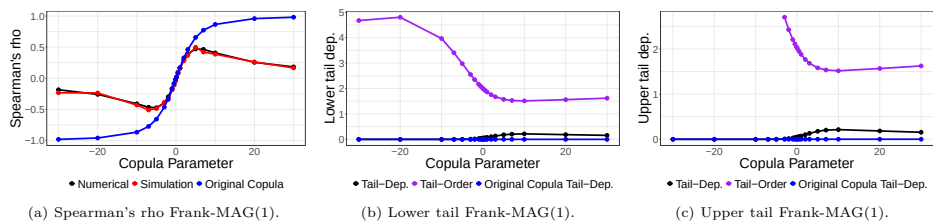


Figure 7: Numerical and simulation-based Spearman's rho, upper and lower 5%-tail dependence coefficients and order of the copula of consecutive observations from a Frank-MAG(1), along with the dependence measures of the original copula K_{21} .

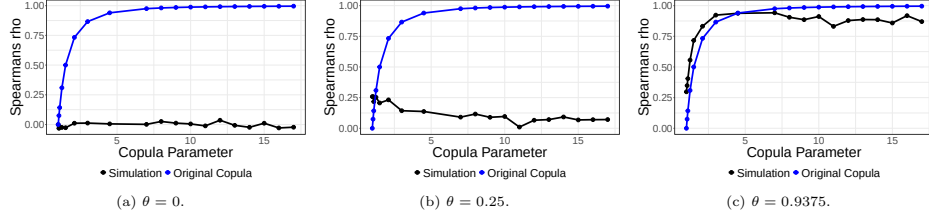


Figure 8: Spearman's rho estimated from simulations of the model in Eq. 1 with AR Gumbel, whose parameter is on the x -axis and MAG Gaussian copula with parameter θ , along with the dependence measure values of the original copula C_{21} .

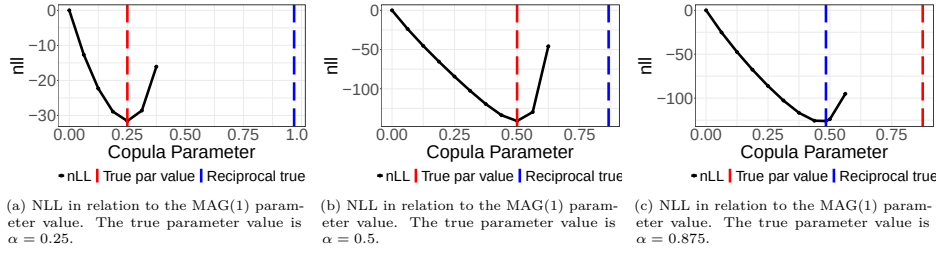


Figure 9: The negative log-likelihood of a Gaussian-MAG(1) model in dependence of the copula parameter. The black line is the NLL-value. The red line marks the true parameter value and the blue line the reciprocal true value, which is calculated as $\alpha^\dagger = \sqrt{1 - \alpha^2}$.

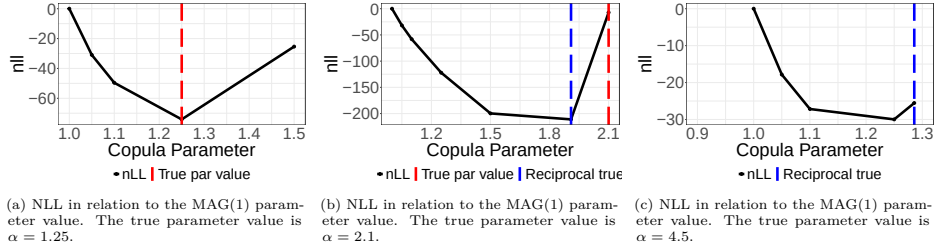


Figure 10: The negative log-likelihood of a Gumbel-MAG(1) model in dependence of the copula parameter. The black line is the NLL-value. The red line marks the true parameter value and the blue line the approximate reciprocal true value defined above.

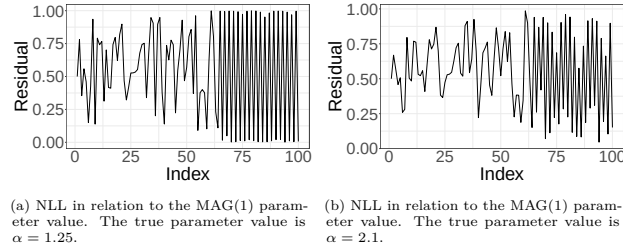


Figure 11: Residuals of the Gaussian-MAG(1) process. The left plot shows the residuals calculated with the true parameter value, which lies above the critical threshold. The left plot shows the residuals under the reciprocal of the true value.

4. Forecasting Real Data

We use the model from Eq. 1 to obtain probabilistic forecasts for univariate time series. Similar to the calculation of the likelihood, the construction of forecasts needs iterative calculations. In Algorithm 2, we present an algorithm that produces one-step ahead percentile forecasts for new data points. The iterative estimation of the latent processes $\{W_t\}$ and $\{\varepsilon_t\}$ is the same as for the likelihood calculation. Given these estimates the percentiles can be obtained by applying the conditional quantile function corresponding to the MAG-copula $K_{(p+1),\dots,1}$. We forecast two real time series. The first time series is quarterly US inflation. We take the `cpi` time series from the `tscopula` package in R and transform it to growth rates to calculate the inflation. The package is provided by McNeil and Bladt (2022). US inflation is also modeled with copula-based time series by McNeil and Bladt (2022) and Pappert (2024). Bladt et al. (2025) also forecast the time series. The second time series is German wind power production. The original data set by Becker and Thrän (2017) comprises observations from different wind power farms in Germany with a 15min frequency. We aggregate these time series to daily production for the whole of Germany. In both forecasting studies we split the time series into training, validation and test sets. The Gaussian-ARMA models, which serve as benchmarks, are selected by AIC using the `auto.arima` function from the `forecast` package in R, see Hyndman and Khandakar (2008). Selection is performed by grid-search and by stepwise selection. Additionally we use the Gaussian-ARMA(1, 1) and Gaussian-ARMA(0, 0) models as benchmarks. Additionally we use the classic copula-based time series as benchmark. For the selection of model order we again rely on automatic selection. We experiment with different copulas and chose the combination which performs best on the validation data. The same holds for the model from Eq. 1. The naming scheme of the models follows the naming by Pappert (2024). First, the marginal specification is stated, "kde" for kernel-density estimation (from the R-package `kde1d`) and "n" for the normal distribution. Then $\text{AR}(p)$ denotes the copula-based time series with model order p and " $\text{CoARMA}(p, q)$ " is short for the model from Eq. 1 with model order p and q . The letters in the parenthesis indicate which copulas are used. E.g. (c180,i,n,c180,n) means that the $\text{AR}(1)$ copula is the 180° -rotated Clayton copula, the $\text{AR}(2)$ copula is the independence copula and so on. "n" stands for the Gaussian (or normal) copula, "g" for the Gumbel copula and "t" for the t -copula. We note that the copula-based models are estimated in a two-step estimation scheme. First, the marginal distribution is estimated and second, the copula parameters are estimated on the probability integral transformed observations.

Algorithm 2: Forecasting with the model from Eq. 1 based on `rvinecopulib`. `olddata` is the data that the model is trained on and new `newdata` the data on which the forecasting study is conducted. F is a CDF estimated on the training data.

```

input : olddata, newdata, model_AR, model_MA,  $p$ ,  $q$  and  $F$ 
output: matrix  $\hat{U}$ 
 $e, x, W \leftarrow \text{vector}()$ 
 $r \leftarrow \max(p, q)$ 
 $l, n \leftarrow \text{length}(\text{olddata}), \text{length}(\text{newdata})$ 
 $x_{1:l} \leftarrow \text{pobs}(\text{olddata})$ 
 $x_{(l+1):(l+n)} \leftarrow F(\text{newdata})$ 
 $e_{1:r}, W_{1:r} \leftarrow (0.5, 0.5)$ 
 $\hat{U} = \text{matrix}(\text{ncol} = n, \text{nrow} = 99)$ 
for  $t \leftarrow (r + 1)$  to  $(l + n)$  do
     $e^\dagger \leftarrow \text{inverse\_rosenblatt}(u = (W_{t-q}, e_{(t-q+1):(t-1)}), 0.5), \text{model\_MA})$ 
     $e_t \leftarrow \text{rosenblatt}(x = (e_{1:q}^\dagger, x_t), \text{model\_MA})_{(q+1)}$ 
     $W^\dagger \leftarrow \text{rosenblatt}(x = (W_{(t-p):(t-1)}), 0.5), \text{model\_AR})$ 
     $W_t \leftarrow \text{inverse\_rosenblatt}((W_{1:p}^\dagger, e_t), \text{model\_AR})_{(p+1)}$ 
end
for  $t \leftarrow (l + 1)$  to  $(n + l)$  do
     $\alpha \leftarrow 1:99 / 100$ 
     $U^\dagger \leftarrow \text{rosenblatt}(x = (W_{(t-q)}, e_{(t-q+1):(t-1)}), 0.5), \text{model\_MA})$ 
     $\hat{U}_{(t-l), (1:m)} \leftarrow \text{inverse\_rosenblatt}((U_{(1:q)}^\dagger, \alpha), \text{model\_MA})_{(q+1)}$ 
end
return  $\hat{U}$ 

```

As evaluation measures we use the continuous ranked probability score (CRPS), the predictive negative log-likelihood (NLL), the 5/95-% pinball score (PBS), the root mean squared error (RMSE), and the mean absolute error (MAE). The CRPS evaluates the whole probabilistic forecast. It estimates the integrated squared distance between the predicted CDF and the empirical CDF of the realized observation, compare [Gneiting and Raftery \(2007\)](#). We approximate the CRPS by aggregating the pinball score values of all percentiles as is common in practice. Usually the CRPS is used for forecast evaluation only, we also report the in-sample CRPS to have a potential reference value. The NLL also evaluates the whole predictive distribution. Note that the models are trained w.r.t. the NLL. The PBS evaluates quantile forecasts, the RMSE mean forecasts and the MAE median forecasts. For the copula models, the mean prediction is obtained by averaging all percentiles. The median prediction is just the 50%-percentile. For the Gaussian-ARMA model mean and median prediction coincide. For all evaluation measures it holds that lower values mean better performance.

We first discuss the forecasting study on US inflation. The time series contains 244 observations. We split the time series into 0.7-train, 0.15-validation and 0.15-test set. The results of the in-sample fit are given in Table 1. We first recognize that the trivial model, Gaussian-ARMA(0, 0) is outperformed by all other models on the train data. This implies that the time series is predictable (this is expected for US inflation but nevertheless it is worth checking). The best fit is provided by the kde-AR(5)-(g,n,n,n,n) model, followed by the Gaussian-ARMA

models and the $(\text{kde}/n)\text{-CoARMA}(3,1)\text{-(n,n,n)-n}$ model. The kde-model seems to have a better distributional fit, while the n-model minimizes the in-sample RMSE and MAE among the CoARMA-models. We note that the Gaussian-ARMA(1,1) and the n-CoARMA(1,1)-(n)-(n) are the same model, compare proposition 1. The small differences in the fit-values are due to the two-step estimation of the copula-based model. The out-of-sample one-step ahead forecasting performance is evaluated on the validation data, see Table 2. Surprisingly the trivial model minimizes the NLL. This result hints at general problems in predicting US inflation. The temporal dependence seems to change over time. Hence the results of the forecasting study on US inflation have to be interpreted with caution. Nevertheless, we proceed to select the best models of each class on the validation data and then evaluate them on the test set. The results are given in Table 3. The results are inconclusive. While the Gaussian-ARMA model clearly minimizes the NLL and the 5%-PBS, the CRPS, MAE, RMSE and 95%-PBS values of the Gaussian-ARMA and the copula-AR model do not differ much. The CoARMA model does not minimize any of the criteria. We reiterate that these results have to be interpreted carefully because of the seemingly changing temporal dependence of US inflation. We also report that models whose stationary distribution is approximated by KDE suffer from the relatively low count of observations in quarterly US inflation (244 in total). If a new observation in the validation or test set lies beyond the points observed in the training set, the likelihood can be very small, in an extreme case it even vanishes. In this case we substitute by 10^{-10} . Additionally, the PIT of such extreme observation yields a value very close to zero (or one), which can go beyond the machine precision in R. This leads to biased predictions after such extreme value.

Now we discuss the forecasting study on German wind power production. The original aggregated time series comprises 16 years of daily observations of German wind power production. Before testing the copula-based and ARMA models, which are supposed to model the stochastic component of a time series, we eliminate trend and cyclical components from the time series. We then take the first two years as train data, the third year as validation and the fourth year as test set. The in-sample fitting results are given in Table 4. Among the Gaussian-ARMA models the ARMA(1,3) model fits best on the train data. The kde-AR(3)-(g,n,n) model is the best Markovian copula-based model, which is also the best fitting model on the train data. The kde-CoARMA(3,1)-(n,n,n)-(n) model fits best among the CoARMA models. On the validation data the kde-CoARMA(3,1)-(n,n,n)-(n) model is best among the model class and among all models. The best models from each model class are then compared on the test data. We observe that the Markovian copula-based and the model from Eq. 1 outperform the Gaussian-ARMA model. The Markovian model seems to predict wind power slightly better. We conclude that estimating the stationary distribution with KDE enhances the performance. This is also confirmed by comparing the Gaussian NLL for the marginal distribution with the KDE NLL on train, validation and test data. On all instances the KDE NLL is smaller. Furthermore we conclude that wind power is dominated by linear relations, as is indicated by the normal copula performing well.

Table 1: **In sample** mean CRPS, NLL, RMSE and MAE values of fits on quarterly US-inflation. The time series contains 244 observations. The values displayed here show the fit on the first 171 observations, which act as train data. The highlighted values are the ones which are best in the respective model class.

Model	CRPS	NLL	RMSE	MAE
Gaussian-ARMA Models				
Gaussian-ARMA(0,0)	0.429	199.8	0.779	0.580
Gaussian-ARMA(1,1)	0.244	102.2	0.438	0.332
Gaussian-ARMA(4,1)	0.223	86.9	0.400	0.306
Gaussian-ARMA(2,3)	0.221	83.9	0.393	0.306
Copula Markov models				
kde-AR(1)-(n)	0.253	102.0	0.463	0.358
kde-AR(5)-(g,n,n,n,n)	0.216	77.7	0.397	0.300
kde-AR(5)-(c180,i,n,c180,n)	0.258	105.8	0.482	0.359
CoARMA-models				
n-CoARMA(4,1)-(g,i,n,g)-(t)	0.244	101.2	0.436	0.332
n-CoARMA(1,1)-(n)-(n)	0.245	103.1	0.440	0.333
n-CoARMA(3,1)-(n,n,n)-(n)	0.227	91.6	0.408	0.312
kde-CoARMA(1,1)-(n)-(n)	0.243	97.3	0.446	0.3375
kde-CoARMA(3,1)-(n,n,n)-(n)	0.227	86.6	0.417	0.319

Table 2: **Validation** mean CRPS, NLL, PBS 5% & 95%, RMSE, MAE and NLL values of forecasts of quarterly US-inflation. The time series contains 244 observations. The values displayed here show the evaluation on 37 validation-observations. The highlighted values are the ones which are best in the respective model class.

Model	CRPS	NLL	PBS-5%	PBS-95%	RMSE	MAE
Gaussian-ARMA Models						
Gaussian-ARMA(0,0)	0.503	53.3	0.275	0.174	0.968	0.690
Gaussian-ARMA(1,1)	0.552	100.4	0.351	0.237	1.011	0.699
Gaussian-ARMA(4,1)	0.478	85.3	0.274	0.207	0.870	0.620
Gaussian-ARMA(2,3)	0.486	89.0	0.284	0.189	0.876	0.638
Copula Markov models						
kde-AR(1)-(n)	0.546	127.9	0.367	0.248	0.995	0.675
kde-AR(5)-(g,n,n,n,n)	0.466	79.7	0.304	0.204	0.873	0.592
kde-AR(5)-(c180,i,n,c180,n)	0.525	117.3	0.331	0.297	0.992	0.663
CoARMA-models						
n-CoARMA(4,1)-(g,i,n,g)-(t)	0.579	80.4	0.384	0.2515	1.023	0.723
n-CoARMA(1,1)-(n)-(n)	0.543	99.3	0.354	0.223	0.999	0.685
n-CoARMA(3,1)-(n,n,n)-(n)	0.466	96.3	0.300	0.178	0.851	0.602
kde-CoARMA(1,1)-(n)-(n)	0.540	133.3	0.353	0.285	0.984	0.662
kde-CoARMA(3,1)-(n,n,n)-(n)	0.443	115.1	0.304	0.232	0.855	0.541

Table 3: **Test** mean CRPS, NLL, PBS 5% & 95%, RMSE and MAE values of forecasts of quarterly US-inflation. The time series contains 244 observations. The values displayed here show the evaluation on 36 test-observations. The models shown here are the ones which performed best in their respective model class on the validation data. The highlighted values are the lowest ones in the respective column.

Model	CRPS	NLL	PBS-5%	PBS-95%	RMSE	MAE
Gaussian-ARMA(4,1)	0.332	36.7	0.124	0.175	0.576	0.444
kde-AR(5)-(g,n,n,n,n)	0.338	47.8	0.161	0.175	0.563	0.439
n-CoARMA(3,1)-(n,n,n)-(n)	0.353	48.0	0.139	0.177	0.601	0.466

Table 4: **In sample** mean CRPS, NLL, RMSE and MAE values of fits on German wind power production. The time series contains 16 years of observations. The values displayed here show the fit on observations from the first two years (730 observations), which act as train data. The highlighted values are the ones which are best in the respective model class.

Model	CRPS	NLL	RMSE	MAE
Gaussian-ARMA Models				
Gaussian-ARMA(0,0)	0.497	923.9	0.858	0.713
Gaussian-ARMA(1,1)	0.403	775.9	0.700	0.575
Gaussian-ARMA(3,1)	0.401	773.6	0.698	0.571
Gaussian-ARMA(1,3)	0.401	773.5	0.698	0.571
Copula Markov models				
kde-AR(3)-(g,n,n)	0.399	751.9	0.697	0.567
kde-AR(3)-(t,n,n)	0.401	756.7	0.700	0.570
CoARMA-models				
kde-CoARMA(3,1)-(n,n,n)-(n)	0.401	757.1	0.699	0.570
kde-CoARMA(3,1)-(g,n,n)-(n)	0.402	759.4	0.700	0.572
kde-CoARMA(1,1)-(g)-(t)	0.402	758.1	0.700	0.570
kde-CoARMA(2,2)-(g,n)-(n,t)	0.441	855.6	0.764	0.634

Table 5: **Validation** mean CRPS, NLL, Pinball 5% & 95%, RMSE and MAE values of forecasts of German wind power. The evaluation is performed on one year of data (365 observations). The highlighted values are the ones which are best in the respective model class.

Model	CRPS	NLL	PBS-5%	PBS-95%	RMSE	MAE
Gaussian-ARMA Models						
Gaussian-ARMA(0,0)	0.535	489.5	0.191	0.159	0.920	0.762
Gaussian-ARMA(1,1)	0.409	393.4	0.151	0.131	0.711	0.578
Gaussian-ARMA(3,1)	0.406	391.3	0.152	0.129	0.707	0.574
Gaussian-ARMA(1,3)	0.407	391.5	0.152	0.129	0.707	0.574
Copula Markov models						
kde-AR(3)-(g,n,n)	0.408	388.9	0.153	0.132	0.715	0.574
kde-AR(3)-(t,n,n)	0.406	385.2	0.155	0.128	0.709	0.570
CoARMA-models						
kde-CoARMA(3,1)-(n,n,n)-(n)	0.406	384.8	0.152	0.128	0.709	0.571
kde-CoARMA(3,1)-(g,n,n)-(n)	0.407	385.7	0.154	0.128	0.711	0.573
kde-CoARMA(1,1)-(g)-(t)	0.407	386.7	0.158	0.130	0.713	0.571
kde-CoARMA(2,2)-(g,n)-(n,t)	0.463	443.8	0.173	0.139	0.805	0.658

Table 6: **Test** mean CRPS, NLL, Pinball 5% & 95%, RMSE and MAE values of forecasts of German wind power production. The evaluation is performed on one year of test data (365 observations). The models shown here are the ones which performed best in their respective model class on the validation data. The highlighted values are the lowest ones in the respective column.

Model	CRPS	NLL	PBS-5%	PBS-95%	RMSE	MAE
Gaussian-ARMA(3,1)	0.391	379.7	0.140	0.140	0.684	0.548
kde-AR(3)-(t,n,n)	0.389	372.2	0.139	0.136	0.682	0.546
kde-CoARMA(3,1)-(n,n,n)-(n)	0.389	372.7	0.138	0.138	0.682	0.545

5. Discussion

We analyze the copula ARMA extension, Eq. 1, proposed by Joe (2014). The main feature of this process is that it is a copula generalization of the ARMA model with uniform stationary distribution. We show that the model nests some Gaussian-ARMA processes in Proposition 1. It is revealed that a Φ -quantile transformed process from Eq. 1 of order p and q , with AR and MAG copulas being the Gaussian copula, is a subset of a Gaussian-ARMA($p, q + p - 1$) process. The additional MA-parameters are all completely determined by the copula parameters. This result is surprising since the process involves a Markov copula of order p and a q -dependent copula. The derivation reveals that there are additional terms. The additional terms are due to the structure of the process. For $p = 1$, there are no additional terms though. This finding brings up the question whether there is a process composed of MA-type and AR-type parts that yields a Gaussian ARMA(p, q) as a special case for all $(p, q) \in \mathbb{N}^2$. The MAGMAR-copula model by Pappert (2024) can generalize ARMA models with any order, however, in this approach an additional transformation is necessary to ensure a uniform stationary distribution. This is a drawback as theoretical properties are not as easy to derive and the computational demand is increased.

We also show that, with a suitable choice of AR and MAG copulas, the Gaussian-GARCH(1, 1) process can be recovered in proposition 3. In principle this enables modeling of time series with GARCH(1, 1) dynamics and arbitrary marginal distributions and complements the research on GARCH-mimicking copulas by Dias et al. (2024). The estimation of such process, however, is not straight-forward. We discuss a computationally friendly approach in Remark 1. The essence of the approach is to use iterative estimations. Implementing and testing this approach on real time series is a future research direction. We proceed to investigate the copula MAG(1) process in detail. This process is a basic building block of the model under investigation and also of the MAGMAR-copula model by Pappert (2024). Simulations, numerical approximations and exemplary calculations on the Fréchet-MAG(1) process reveal that the dependence of the joint distribution of consecutive observations from a copula MAG(1) process is limited. We regularly find that the 5%-tail dependence coefficients associated to this joint distribution seem to be bounded by at least $\frac{1}{4}$. Numerical approximation of the tail order indicates that the tail dependence coefficients may also vanish for "standard"-copulas (Gumbel, Clayton and t). For the Fréchet-MAG(1) process, the upper bound of $\frac{1}{4}$ can also be found analytically.

Furthermore we find that the Gaussian-MAG(1) process, similar to the classic MA(1) process, has two equivalent representations. The two representations can be obtained by permuting the innovations. This gives rise to potential near non-identifiability problems. We argue that restricting the parameter space to avoid stronger dependence parameters is not problematic. This bound of the parameter space coincides with the bound that is needed such that the estimation is consistent (cf. Pappert (2024)). It is not clear if other MAG(1) processes have two equivalent representations. We proceed to examine the dependence of consecutive observations of the model from Eq. 1 with model orders $p = q = 1$ and show that the copulas can be chosen such that any correlation is attainable. Furthermore, we discuss distributional, dependence and estimation properties of the resulting process. A crucial condition for consistency is the stationarity and ergodicity of the estimated latent processes. This condition is on a high-level. Currently it is not clear how to translate this condition to properties of the AR and MAG copula. This question offers potential for further investigations.

For the general model with arbitrary model orders, we provide an algorithm, which calculates the likelihood

iteratively. Additionally we provide an algorithm that performs one-step ahead probabilistic forecasting.³ Using these algorithms we fit models to US inflation and German wind power production and conduct a forecasting study. Conducting a forecasting study on US inflation turns out to be challenging due to suspected changes in the temporal dependence over time. This is emphasized by the trivial model minimizing the predictive negative log-likelihood score on the validation data. Taking this problem into account, we nevertheless conclude – with some caution – that the Gaussian-ARMA(4, 1) model is the most suitable model among the models under consideration to forecast US inflation. One major advantage of copula-based models is that the stationary distribution can be chosen freely. For US inflation this advantage does not seem to help much. This could be caused by the relatively small number of observations, which additionally have a relatively strong temporal dependence. For German wind power production, more observations are available and we observe that estimating the stationary distribution with kernel-density estimation in fact enhances the models performances in comparison with models with normal stationary distributions. For both time series, linear temporal dependencies are revealed to dominate in terms of forecasting. In future research it would be interesting to apply the models to temporal dependencies, which are dominated by truly non-linear dynamics.

Acknowledgement

This research was (partially) funded in the course of TRR 391 *Spatio-temporal Statistics for the Transition of Energy and Transport* (520388526) by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

³The implemented R-code for estimation, simulation and forecasting is available upon request.

References

- Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica*, 78(1):395–410.
- Beare, B. K. and Seo, J. (2015). Vine copula specifications for stationary multivariate Markov chains. *Journal of Time Series Analysis*, 36(2):228–246.
- Becker, R. and Thrän, D. (2017). Completion of wind turbine data sets for wind integration studies applying random forests and k-nearest neighbors. *Applied Energy*, 208:252–262.
- Bladt, M., Dias, A., Han, J., and McNeil, A. J. (2025). Semiparametric and parametric distributional forecasting of univariate time series using non-gaussian arma models based on d-vines. *The Canadian Journal of Statistics*.
- Brechmann, E. C. and Czado, C. (2015). COPAR—multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry*, 31(4):495–514.
- Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335.
- Dias, A., Han, J., and McNeil, A. J. (2024). GARCH copulas and GARCH-mimicking copulas. *arXiv preprint arXiv:2408.07025*.
- Douc, R., Moulines, E., and Stoffer, D. (2014). *Nonlinear Time Series: Theory, Methods and Applications with R Examples*. CRC press.
- Embrechts, P., Hofert, M., and Wang, R. (2016). Bernoulli and tail-dependence compatibility. *The Annals of Applied Probability*, 26(3):1636–1658.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.
- Hamilton, J. D. (2020). *Time Series Analysis*. Princeton University Press.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27:1–22.
- Joe, H. (2014). *Dependence Modeling with Copulas*. CRC press, Boca Raton, FL.
- Loaiza-Maya, R., Smith, M. S., and Maneesoonthorn, W. (2018). Time series copulas for heteroskedastic data. *Journal of Applied Econometrics*, 33(3):332–354.
- McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, 20(4):822–835.
- McNeil, A. J. and Bladt, M. (2022). Time series with infinite-order partial copula dependence. *Dependence Modeling*.
- Nagler, T., Krüger, D., and Min, A. (2022). Stationary vine copula models for multivariate time series. *Journal of Econometrics*, 227(2):305–324.

- Nagler, T. and Vatter, T. (2022). *rvinecopulib: High Performance Algorithms for Vine Copula Modeling*. R package version 0.6.2.1.1.
- Pappert, S. (2024). Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas). *arXiv preprint arXiv:2402.01491v3*.
- Pappert, S. and Arsova, A. (2023). Forecasting natural gas prices with spatio-temporal copula-based time series models. *arXiv preprint arXiv:2301.03328*.
- Simard, C. and Rémillard, B. (2015). Forecasting time series with multivariate copulas. *Dependence Modeling*, 3(1).
- Smith, M., Min, A., Almeida, C., and Czado, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492):1467–1479.
- Tao, S. (2025). On tridiagonal correlation matrices. *Pan-American Journal of Mathematics*, 4:15.
- Wooldridge, J. M. (1994). Estimation and inference for dependent processes. *Handbook of Econometrics*, 4:2639–2738.

Appendix A. Supporting Results

Appendix A.1. Special cases of the CoARMA(1, 1) process

In proposition 5, Eq. 6 the joint distribution of consecutive observations of the CoARMA(1, 1) process is derived as

$$F_{U_t, U_{t-1}}(x_1, x_2) = \int_0^1 \int_0^{C_{2|1}^{-1}(K_{2|1}(x_2|w_2)|w_2)} K_{2|1}(x_1|w_1) c_{21}(w_1, w_2) dw_1 dw_2.$$

We now consider special cases.

- a) Let $K_{21} = C^\perp$, then $K_{2|1}(u|v) = u$ and

$$\begin{aligned} F_{U_t, U_{t-1}}(x_1, x_2) &= \int_0^1 \int_0^{C_{2|1}^{-1}(K_{2|1}(x_2|w_2)|w_2)} K_{2|1}(x_1|w_1) c_{21}(w_1, w_2) dw_1 dw_2 \\ &= \int_0^1 \int_0^{C_{2|1}^{-1}(x_2|w_2)} x_1 c_{21}(w_1, w_2) dw_1 dw_2 \\ &= \int_0^1 x_1 C_{2|1}(C_{2|1}^{-1}(x_2|w_2)|w_2) dw_2 \\ &= x_1 x_2 \end{aligned}$$

This result is expected since $U_t = \varepsilon_t$ for all t for $K_{21} = C^\perp$.

b) Let $K_{21} = C^+$, then $K_{2|1}(u|v) = \mathbb{1}\{u \geq v\}$ and

$$\begin{aligned}
F_{U_t, U_{t-1}}(x_1, x_2) &= \int_0^1 \int_0^{C_{2|1}^{-1}(K_{2|1}(x_2|w_2)|w_2)} K_{2|1}(x_1|w_1) c_{21}(w_1, w_2) dw_1 dw_2 \\
&= \int_0^1 \int_0^{C_{2|1}^{-1}(\mathbb{1}\{x_2 \geq w_2\}|w_2)} \mathbb{1}\{x_1 \geq w_2\} c_{21}(w_1, w_2) dw_1 dw_2 \\
&= \int_0^1 \int_0^{\min\{C_{2|1}^{-1}(\mathbb{1}\{x_2 \geq w_2\}|w_2), x_1\}} c_{21}(w_1, w_2) dw_1 dw_2 \\
&= \int_0^1 C_{2|1} \left(\min \left\{ C_{2|1}^{-1}(\mathbb{1}\{x_2 \geq w_2\}|w_2), x_1 \right\} \middle| w_2 \right) dw_2 \\
&= \int_0^1 \mathbb{1}\{w_2 \leq x_2\} C_{2|1}(x_1|w_2) dw_2 \\
&= C_{21}(x_1, x_2)
\end{aligned}$$

This result is also expected.

c) Let $C_{21} = C^\perp$, then

$$\begin{aligned}
F_{U_t, U_{t-1}}(x_1, x_2) &= \int_0^1 \int_0^{K_{2|1}(x_2|w_2)} K_{2|1}(x_1|w_1) dw_1 dw_2 \\
&= \int_0^1 K_{21}(x_1, K_{2|1}(x_2|w_2)) dw_2
\end{aligned}$$

which is the copula given in proposition 4.

Appendix B. Proofs

Appendix B.1. Proof of Proposition 1

Proof. We go step-by-step

a) We use the expression from [McNeil and Bladt \(2022\)](#) for the last element of the Rosenblatt transform associated to a stationary D-vine copula with Gaussian pair copulas. Using their result, we may write

$$h^{-1}(x; u_1, \dots, u_p) = \Phi \left(\frac{\Phi^{-1}(x) - \sum_{j=1}^p \phi_j^{(p)} \Phi^{-1}(u_j)}{\sigma_\alpha} \right) \quad (\text{B.1})$$

where $\sigma_\alpha^2 = \prod_{j=1}^p (1 - \alpha_j^2)$ and $\phi_j^{(p)}, j \in \{1, \dots, p\}$ are given as in part a) of proposition 1. It immediately follows that

$$\Phi^{-1}(U_t) = \sum_{j=1}^p \phi_j^{(p)} \Phi^{-1}(U_{t-j}) + \sigma_\alpha \varepsilon_t,$$

which proves the claim. Again we note that this result is well-known in the literature. The result follows directly from [McNeil and Bladt \(2022\)](#). Compare also [Smith et al. \(2010\)](#). The proof is given here for the sake of completeness.

- b) For the MAG-process, the underlying vine is not a stationary D-vine, but a vine with independent margins. The last element of the Rosenblatt transform associated to such a vine is given as

$$g^{-1}(x; u_1, \dots, u_q) = h_{\beta_q}(\cdot; u_q) \circ \dots \circ h_{\beta_1}(x; u_1). \quad (\text{B.2})$$

For Gaussian pair copulas, we observe that the following structure emerges,

$$g^{-1}(x; u_1, \dots, u_q) = \Phi \left(\sigma_{\beta}^{-1} \Phi^{-1}(x) - \sum_{i=0}^{q-1} \left[\prod_{j=0}^i (1 - \beta_{q-j}^2)^{-\frac{1}{2}} \right] \beta_{q-i} \Phi^{-1}(u_{i+1}) \right). \quad (\text{B.3})$$

Then we see

$$U_t = g(\varepsilon_t; \varepsilon_{t-1}, \dots, \varepsilon_{t-q}) \quad (\text{B.4})$$

$$\iff \varepsilon_t = g^{-1}(U_t; \varepsilon_{t-1}, \dots, \varepsilon_{t-q}) \quad (\text{B.5})$$

$$\iff \sigma_{\beta} \Phi^{-1}(\varepsilon_t) = \Phi^{-1}(U_t) - \sigma_{\beta} \sum_{i=0}^{q-1} \left[\prod_{j=0}^i (1 - \beta_{q-j}^2)^{-\frac{1}{2}} \right] \beta_{q-i} \Phi^{-1}(\varepsilon_{t-q+i}) \quad (\text{B.6})$$

$$\iff Y_t = \eta_t + \sum_{i=0}^{q-1} \left[\prod_{j=0}^i (1 - \beta_{q-j}^2)^{-\frac{1}{2}} \right] \beta_{q-i} \eta_{t-q+i} \quad (\text{B.7})$$

$$= \eta_t + \sum_{j=1}^q \theta_j^{(q)} \eta_{t-j} \quad (\text{B.8})$$

where $\theta_j^{(q)} = \sigma_{\beta}^{-1} \left[\prod_{k=0}^{j-1} (1 - \beta_{q-k}^2)^{\frac{1}{2}} \right] \beta_j$ and $\eta_t = \sigma_{\beta} \Phi^{-1}(\varepsilon_t)$.

- c) Following the procedure from part b), we have

$$Y_t = \Phi^{-1}(U_t) \quad (\text{B.9})$$

$$= \eta_t + \sum_{i=1}^{q-1} \theta_i^{(q)} \eta_{t-i} + \theta_q^{(q)} \sigma_{\beta} X_{t-q}. \quad (\text{B.10})$$

Where it is understood that the sum $\sum_{i=1}^{q-1} \theta_i^{(q)} \eta_{t-i}$ is zero for $q = 1$. Now we see that $X_{t-q} = \sum_{j=1}^p \phi_j^{(p)} X_{t-q-j} + \sigma_{\alpha} \varepsilon_{t-q}$ and for each $(X_{t-q-1}, \dots, X_{t-q-p})$ it holds that

$$\theta_q^{(q)} \sigma_{\beta} X_{t-q-j} = Y_{t-j} - \eta_{t-j} - \sum_{i=1}^{q-1} \theta_i^{(q)} \eta_{t-j-i} \quad (\text{B.11})$$

Consequently

$$Y_t = \eta_t + \sum_{i=1}^{q-1} \theta_i^{(q)} \eta_{t-i} + \sum_{j=1}^p \theta_j^{(q)} \sigma_\beta \phi_j^{(p)} X_{t-q-j} + \theta_q^{(q)} \sigma_\beta \sigma_\alpha \varepsilon_{t-q} \quad (\text{B.12})$$

$$= \eta_t + \sum_{i=1}^{q-1} \theta_i^{(q)} \eta_{t-i} + \theta_q^{(q)} \sigma_\alpha \eta_{t-q} \quad (\text{B.13})$$

$$+ \sum_{j=1}^p \phi_j^{(p)} [Y_{t-j} - \eta_{t-j} - \sum_{i=1}^{q-1} \theta_i^{(q)} \eta_{t-j-i}] \quad (\text{B.14})$$

$$= \eta_t + \sum_{j=1}^p \phi_j^{(p)} Y_{t-j} + \sum_{i=1}^{q-1} \theta_i^{(q)} \eta_{t-i} + \theta_q^{(q)} \sigma_\alpha \eta_{t-q} - \sum_{j=1}^p \phi_j^{(p)} \eta_{t-j} - \sum_{j=1}^p \sum_{i=1}^{q-1} \phi_j^{(p)} \theta_i^{(q)} \eta_{t-j-i} \quad (\text{B.15})$$

$$= \eta_t + \sum_{j=1}^p \phi_j^{(p)} Y_{t-j} + \sum_{j=1}^{p+q-1} \Psi_j^{(p,q)} \eta_{t-j} \quad (\text{B.16})$$

where

$$\Psi_j^{(p,q)} = \tilde{\theta}_j^{(q)} - \tilde{\phi}_j^{(p)} - \zeta_j^{(p,q)} \quad (\text{B.17})$$

$$\tilde{\theta}_j^{(q)} = \theta_j^{(q)} \times \mathbb{1}\{j \leq q \wedge q > 1\} + \theta_q \sigma_\alpha \times \mathbb{1}\{j = q\} \quad (\text{B.18})$$

$$\tilde{\phi}_j^{(p)} = \phi_j^{(p)} \times \mathbb{1}\{j \leq p\} \quad (\text{B.19})$$

$$\zeta_j^{(p,q)} = \left(\sum_{(r,s) \in S_j^{(p,q)}} \phi_r^{(p)} \theta_s^{(q)} \right) \times \mathbb{1}\{q > 1\} \quad (\text{B.20})$$

$$S_j^{(p,q)} = \{(r, s) \in \mathbb{N}^2 : r \leq p, s \leq q, r + s = j\} \quad (\text{B.21})$$

□

Appendix B.2. Proof of proposition 2

We derive the copula directly from the ARCH(1) process,

$$Y_t = \sigma_t \eta_t, \quad \eta_t \stackrel{iid}{\sim} N(0, 1) \quad (\text{B.22})$$

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2}. \quad (\text{B.23})$$

The unconditional distribution function of Y_t is denoted by Ψ and the corresponding PIT is denoted by $U_t = \Psi(Y_t)$.⁴ Then we can write

$$Y_t = \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2} \eta_t \quad (\text{B.24})$$

$$U_t = \Psi \left[\sqrt{\alpha_0 + \alpha_1 \Psi^{-1}(U_{t-1})^2} \Phi^{-1}(\varepsilon_t) \right] \quad (\text{B.25})$$

$$= g(\varepsilon_t, U_{t-1}) \quad (\text{B.26})$$

⁴The explicit form of Ψ (or its derivative, the density) is derived in e.g. [Dias et al. \(2024\)](#).

Now we identify the mapping g as the conditional quantile function corresponding to a copula C_{21} , which is evaluated at ε_t , conditioned on U_{t-1} , i.e. $U_t = g(\varepsilon_t, U_{t-1}) = C_{21}^{-1}(\varepsilon_t | U_{t-1})$. It follows that C_{21}^{-1} is given as

$$C_{21}^{-1}(u_1 | u_2) = \Psi \left[\sqrt{\alpha_0 + \alpha_1 \Psi^{-1}(u_2)^2} \Phi^{-1}(u_1) \right]. \quad (\text{B.27})$$

Then we can immediately see that the conditional CDF corresponding to C_{12} is given as

$$C_{21}(u_1 | u_2) = \Phi \left[\frac{\Psi^{-1}(u_1)}{\sqrt{\alpha_0 + \alpha_1 \Psi^{-1}(u_2)^2}} \right] \quad (\text{B.28})$$

and the expression for the C_{12} is obtained by integrating over the second argument. Similarly the copula density c_{12} can be derived by taking the derivative of C_{21} w.r.t. its first argument. Then the expression given in Eq. 9 of [Dias et al. \(2024\)](#) can be recovered.

Appendix B.3. Proof of proposition 3

We follow the same strategy as in the derivation for the ARCH-copula in Sect. [Appendix B.2](#). The main difference here is that the MAG and the AR part need to be identified. Let Ψ be the unconditional CDF corresponding to a Gaussian-GARCH(1,1) process and let $\{X_t\}_{t \in \mathbb{Z}}$ follow the modified-ARCH(1) model given by the following updating equation

$$X_t = \sqrt{(\alpha_0 + X_{t-1}^2)(\alpha_1 \eta_{t-1}^2 + \beta)}. \quad (\text{B.29})$$

Using contraction arguments by [Douc et al. \(2014\)](#) we see that $\{X_t\}_{t \in \mathbb{Z}}$ is stationary if $\frac{1}{2} \mathbb{E}[\log(\alpha_1 \eta_0^2 + \beta)] < 0$, which is automatically fulfilled if the assumed GARCH(1,1) is stationary. Now let Γ be the unconditional CDF of $\{X_t\}_{t \in \mathbb{Z}}$. The GARCH(1,1) model is given as

$$Y_t = \sigma_t \eta_t \quad (\text{B.30})$$

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2} \quad (\text{B.31})$$

Now we define

$$X_t := \sqrt{\alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2} \quad (\text{B.32})$$

$$= \sqrt{\alpha_1 Y_{t-1}^2 + \beta_1 \frac{Y_{t-1}^2}{\eta_{t-1}^2}} \quad (\text{B.33})$$

and hence

$$Y_t = \sqrt{\alpha_0 + X_{t-1}^2} \eta_t \quad (\text{B.34})$$

Now we can substitute for X_t in Eq. B.33 and obtain

$$X_t = \sqrt{\alpha_1(\alpha_0 + X_{t-1}^2)\eta_t^2 + \beta_1(\alpha_0 + X_{t-1}^2)} \quad (\text{B.35})$$

$$= \sqrt{(\alpha_0 + X_{t-1}^2)(\alpha_1\eta_t^2 + \beta_1)} \quad (\text{B.36})$$

Note that X_t follows the modified-ARCH process with unconditional CDF Γ . Now we can identify that X_t corresponds to the latent AR process and extract the copula associated to the process by applying the PIT. To this end we define $W_t = \Gamma(X_t)$ and we write

$$W_t = \Gamma \left[\sqrt{(\alpha_0 + \Gamma^{-1}(W_{t-1})^2)(\alpha_1\Phi^{-1}(\varepsilon_t)^2 + \beta_1)} \right]. \quad (\text{B.37})$$

Following the procedure from proof Appendix B.2, we identify the above mapping as a conditional quantile function corresponding to a bivariate copula C_{21} , i.e.

$$C_{2|1}^{-1}(u|v) = \Gamma \left[\sqrt{(\alpha_0 + \Gamma^{-1}(v)^2)(\alpha_1\Phi^{-1}(u)^2 + \beta_1)} \right], \quad (\text{B.38})$$

$$C_{2|1}(u|v) = \Phi \left[\sqrt{\frac{\Gamma^{-1}(u)^2}{\alpha_1\alpha_0 + \alpha_1\Gamma^{-1}(v)^2} - \frac{\beta_1}{\alpha_1}} \right] \quad (\text{B.39})$$

Integration over v yields the desired result. Now for the MAG-copula, we again consider the relation between Y_t and X_t and define $U_t = \Psi(Y_t)$. Then we can write

$$U_t = \Psi \left[\sqrt{\alpha_0 + \Gamma^{-1}(W_{t-1})^2}\Phi^{-1}(\varepsilon_t) \right]. \quad (\text{B.40})$$

Again we identify the above mapping as a conditional quantile function corresponding to a copula K_{21} . That is

$$K_{2|1}^{-1}(u|v) = \Psi \left[\sqrt{\alpha_0 + \Gamma^{-1}(v)^2}\Phi^{-1}(u) \right] \quad \text{and} \quad (\text{B.41})$$

$$K_{2|1}(u|v) = \Phi \left[\frac{\Psi^{-1}(u)}{\sqrt{\alpha_0 + \Gamma^{-1}(v)^2}} \right]. \quad (\text{B.42})$$

Again integration over v yields the desired result and the proof is complete.

Appendix B.4. Proof of proposition 4

Proof. We show properties step-by-step.

- a) The process $\{V_t\}$ given as $V_t = K_{2|1}^{-1}(\varepsilon_t|\varepsilon_{t-1})$ is 1-dependent with *i.i.d.* innovations and hence stationary and ergodic. See also Joe (2014) and Pappert (2024).

b) The derivation is given in [Pappert \(2024\)](#). We still derive $F_{U_t, U_{t-1}}$ explicitly for the sake of completeness.

$$F_{V_t, V_{t-1}}(x_1, x_2) = \mathbb{P}(V_t \leq x_1, V_{t-1} \leq x_2) \quad (\text{B.43})$$

$$= \mathbb{P}(K_{2|1}^{-1}(\varepsilon_t, \varepsilon_{t-1}) \leq x_1, K_{2|1}^{-1}(\varepsilon_{t-1}, \varepsilon_{t-2}) \leq x_2) \quad (\text{B.44})$$

$$= \int_0^1 \int_0^1 \mathbb{P}(K_{2|1}^{-1}(\varepsilon_t, w_1) \leq x_1, K_{2|1}^{-1}(w_1, w_2) \leq x_2) dw_1 dw_2 \quad (\text{B.45})$$

$$= \int_0^1 \int_0^1 K_{2|1}(x_1, w_1) \mathbb{1}\{w_1 \leq K_{2|1}(x_2 | w_2)\} dw_1 dw_2 \quad (\text{B.46})$$

$$= \int_0^1 K_{21}(x_1, K_{2|1}(x_2 | w_2)) dw_2 \quad (\text{B.47})$$

c) We recall that K is PQD, if $K(x_1, x_2) \geq x_1 x_2, \forall x_1, x_2 \in (0, 1)$, cf. [Joe \(2014\)](#), Sect. 2.8.1. Then we see that

$$\begin{aligned} F_{V_t, V_{t-1}}(x_1, x_2) &= \int_0^1 K_{12}(K_{2|1}(x_1 | s), x_1) ds \\ &\geq \int_0^1 K_{2|1}(x_2 | s) x_1 ds \\ &= x_1 x_2. \end{aligned}$$

In the first line we used the PQD of K and in the second line we solved the integral. For K being NQD, we can show that $F_{V_t, V_{t-1}}$ is NQD in the same way.

d) We recall that $K_{2|1}$ is SI, in its first argument, if $K_{2|1}(x_2 | x_1)$ is decreasing in x_1 . Alternatively, consider the random vector (Z_1, Z_2) which we define in such a way that $(Z_1, Z_2) \sim K$. Then K being SI in its first argument is equivalent to Z_2 being SI in Z_1 . We first realize that if $K_{2|1}(x_2 | x_1)$ is decreasing in x_1 , then $K_{2|1}^{-1}(y | x_1)$ is increasing in x_1 . Then we see that $V_t = K_{2|1}^{-1}(\varepsilon_t, \varepsilon_{t-1})$ is SI in ε_{t-1} . Additionally, $\varepsilon_{t-1} = K_{2|1}(V_{t-1} | \varepsilon_{t-2})$ is increasing in V_{t-1} . In combination, V_t is increasing in V_{t-1} .

e) Let R be the infinite dimensional correlation matrix for $\{U_t\}_{t \in \mathbb{Z}}$. It holds that $R_{ii} = 1, \forall i \in \mathbb{Z}$ and $R_{i, i \pm 1} = \rho$. All elements that are not on the main and secondary diagonal are zero since the process is one-dependent. Hence R is a tridiagonal matrix and we can use the result from [Tao \(2025\)](#) that establishes that $|\rho| \leq \frac{1}{2}$ has to hold in order for R to be a correlation matrix.

f) We use the result from [Embrechts et al. \(2016\)](#) which establishes that $|\lambda_u|, |\lambda_l| \leq \frac{1}{2}$ for general 1-dependent processes.

g) For $Y_t = \{\Phi^{-1}(V_t)\}$ it holds that $\{Y_t\}_{t \in \mathbb{Z}}$ follows the updating equation (compare proposition 1),

$$Y_t = \eta_t + \frac{\beta_1}{(1 - \beta_1)^{\frac{1}{2}}} \eta_{t-1}, \quad \eta_t \sim N(0, (1 - \beta_1^2))$$

For $\tilde{Y}_t = \{\Phi^{-1}(\tilde{V}_t)\}$ it holds that

$$\tilde{Y}_t = \tilde{\eta}_t + \frac{(1 - \beta_1^2)^{\frac{1}{2}}}{\beta_1} \tilde{\eta}_{t-1}, \quad \tilde{\eta}_t \sim N(0, \beta_1^2).$$

It is well known that $\{Y_t\}$ and $\{\tilde{Y}_t\}$ are two representations of the same MA(1) process. The proposition follows immediately. $\square$

Appendix B.5. Proof of proposition 5

Proof. We go step-by-step

- a) The time series $\{W_t\}$ follows an copula AR(1) process, with C_{21} being the copula of (W_t, W_{t-1}) for all t . Assume that C_{21} is chosen such that the process is stationary and ergodic (this is true under mild conditions). Then it follows that $\{U_t\}$ is also stationary and ergodic since it is calculated by aggregating the independent random variables ε_t and W_{t-1} , i.e. $U_t = K_{2|1}^{-1}(\varepsilon_t|W_{t-1})$, implying that K_{21} is the copula of (U_t, W_{t-1}) .
- b) We consider the joint distribution of (U_t, U_{t-1}) ,

$$\begin{aligned}
 F_{U_t, U_{t-1}}(x_1, x_2) &= \mathbb{P}(U_t \leq x_1, U_{t-1} \leq x_2) \\
 &= \mathbb{P}(h(\varepsilon_t, W_{t-1}) \leq x_1, h(\varepsilon_{t-1}, W_{t-2}) \leq x_2) \\
 &= \mathbb{P}(\varepsilon_t \leq K_{2|1}(x_1|W_{t-1}), \varepsilon_{t-1} \leq K_{2|1}(x_2|W_{t-2})) \\
 &= \mathbb{P}(\varepsilon_t \leq K_{2|1}(x_1|W_{t-1}), W_{t-1} \leq C_{2|1}^{-1}(K_{2|1}(x_2|W_{t-2})|W_{t-2})) \\
 &= \int_0^1 \int_0^1 c_{21}(w_1, w_2) \mathbb{P}(\varepsilon_t \leq K_{2|1}(x_1|w_1), w_1 \leq C_{2|1}^{-1}(K_{2|1}(x_2|w_2)|w_2)) dw_1 dw_2 \\
 &= \int_0^1 \int_0^1 c_{21}(w_1, w_2) K_{2|1}(x_1|w_1) 1_{\{w_1 \leq C_{2|1}^{-1}(K_{2|1}(x_2|w_2)|w_2)\}} dw_1 dw_2 \\
 &= \int_0^1 \int_0^{C_{2|1}^{-1}(K_{2|1}(x_2|w_2)|w_2)} c_{21}(w_1, w_2) K_{2|1}(x_1|w_1) dw_1 dw_2,
 \end{aligned}$$

where we used that the joint density of (W_t, W_{t-1}) is given by the copula density c_{21} . $\square$

Appendix B.6. Proof of proposition 6

Proof. For $p = 1, q = 2$ we have that C_{21} is the copula of (W_t, W_{t-1}) and K_{321} the copula of $(U_t, \varepsilon_{t-1}, W_{t-2})$. The joint distribution of consecutive observations can be calculated as follows.

$$\begin{aligned}
 F_{U_t, U_{t-1}}(x_1, x_2) &= \mathbb{P}(U_t \leq x_1, U_{t-1} \leq x_2) \\
 &= \mathbb{P}(K_{3|21}^{-1}(\varepsilon_t|\varepsilon_{t-1}, W_{t-2}) \leq x_1, K_{3|21}^{-1}(\varepsilon_{t-1}|\varepsilon_{t-2}, W_{t-3}) \leq x_2) \\
 &= \mathbb{P}(\varepsilon_t \leq K_{3|21}(x_1|\varepsilon_{t-1}, W_{t-2}), \varepsilon_{t-1} \leq K_{3|21}(x_2|\varepsilon_{t-2}, W_{t-3}))
 \end{aligned}$$

Now we condition on suitable random variables. It is important that the joint distribution of the random variables which the expression is conditioned on is known. Here we condition on $(\varepsilon_{t-1}, W_{t-2}, W_{t-3})$. The joint density of

this random vector is $f_{\varepsilon_{t-1}, W_{t-2}, W_{t-3}}(u_1, u_2, u_3) = c_{21}(u_2, u_3)$ by construction, because ε_{t-1} is independent of (W_{t-2}, W_{t-3}) . Hence we can continue the calculation as follows

$$\begin{aligned} F_{U_t, U_{t-1}}(x_1, x_2) &= \int_0^1 \int_0^1 \int_0^1 \mathbb{P}(\varepsilon_t \leq K_{3|21}(x_1|e_1, w_2), e_1 \leq K_{3|21}(x_2|e_2, w_3)) c_{21}(w_2, w_3) de_1 dw_2 dw_3 \\ &= \int_0^1 \int_0^1 \int_0^1 K_{3|21}(x_1|e_1, w_2) \times \mathbb{1}\{e_1 \leq K_{3|21}(x_2|e_2, w_3)\} c_{21}(w_2, w_3) de_1 dw_2 dw_3. \end{aligned}$$

Now for $p = 2, q = 1$ we have that C_{321} is the copula of (W_t, W_{t-1}, W_{t-2}) and K_{21} the copula of (U_t, W_{t-1}) . The joint distribution of consecutive observations can be calculated as follows.

$$\begin{aligned} F_{U_t, U_{t-1}}(x_1, x_2) &= \mathbb{P}(U_t \leq x_1, U_{t-1} \leq x_2) \\ &= \mathbb{P}(K_{2|1}^{-1}(\varepsilon_t|W_{t-1}) \leq x_1, K_{2|1}^{-1}(\varepsilon_{t-1}|W_{t-2}) \leq x_2) \\ &= \mathbb{P}(\varepsilon_t \leq K_{2|1}(x_1|W_{t-1}), \varepsilon_{t-1} \leq K_{2|1}(x_2|W_{t-2})) \\ &= \mathbb{P}(\varepsilon_t \leq K_{2|1}(x_1|W_{t-1}), W_{t-1} \leq C_{3|21}^{-1}(K_{2|1}(x_2|W_{t-2}))|W_{t-2}, W_{t-3}) \end{aligned}$$

Now we condition on the random vector (W_t, W_{t-1}, W_{t-2}) , which has joint density given by c_{321} .

$$\begin{aligned} F_{U_t, U_{t-1}}(x_1, x_2) &= \int_0^1 \int_0^1 \int_0^1 \left[K_{2|1}(x_1|w_1) c_{321}(w_1, w_2, w_3) \mathbb{1}\{w_1 \leq C_{3|21}^{-1}(K_{2|1}(x_2|w_2)|w_2, w_3)\} \right. \\ &\quad \left. \times c_{321}(w_1, w_2, w_3) \right] dw_1 dw_2 dw_3. \end{aligned}$$

For $p = q = 2$ we have that C_{321} is the copula of (W_t, W_{t-1}, W_{t-2}) and K_{321} the copula of $(U_t, \varepsilon_{t-1}, W_{t-2})$. The joint distribution of consecutive observations can be calculated as follows.

$$\begin{aligned} F_{U_t, U_{t-1}}(x_1, x_2) &= \mathbb{P}(U_t \leq x_1, U_{t-1} \leq x_2) \\ &= \mathbb{P}(\varepsilon_t \leq K_{3|21}(x_1|\varepsilon_{t-1}, W_{t-2}), \varepsilon_{t-1} \leq K_{3|21}(x_2|\varepsilon_{t-2}, W_{t-3})) \end{aligned}$$

We substitute $\varepsilon_{t-2} = C_{3|21}(W_{t-2}|W_{t-3}, W_{t-4})$ and condition on $(\varepsilon_{t-1}, W_{t-2}, W_{t-3}, W_{t-4})$ with joint density c_{321} .

$$F_{U_t, U_{t-1}}(x_1, x_2) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 K_{3|21}(x_1|e_1, w_2) c_{321}(w_2, w_3, w_4) \mathbb{1}\{e_1 \leq K_{3|21}(x_2|e_2, w_3)\} dw_2 dw_3 dw_4 de_1.$$

□

The Field Equations of Penalized non-Parametric Regression

Sven Pappert^{a,*}

^a*Chair of Econometrics, Department of Statistics, TU Dortmund University, Germany*

Abstract

We view penalized risks through the lens of the calculus of variations. We consider risks comprised of a fitness-term (e.g. MSE) and a gradient-based penalty. After establishing the Euler-Lagrange field equations as a systematic approach to finding minimizers of risks involving only first derivatives, we proceed to exemplify this approach to the MSE penalized by the integral over the squared ℓ^2 -norm of the gradient of the regression function. The minimizer of this risk is given as the solution to a second order inhomogeneous PDE, where the inhomogeneity is given as the conditional expectation of the target variable conditioned on the features. We discuss properties of the field equations and practical implications thereof, which also apply to the classical Ridge penalty for linear models, and embed our findings into the existing literature. In particular, we find that we can recover the Rudin-Osher-Fatemi model for image-denoising, if we consider the features as deterministic and evenly distributed. Last, we outline several directions for future research.

Keywords: Artificial Intelligence, Calculus of Variations, LASSO, Non-Parametric Regression, Penalized Regression, Regularization, Ridge Regression, Gradient-Based Regularization

1. Introduction

When training a non-parametric model w.r.t. a penalized risk, it is clear that the model is not estimating the conditional expectation. But what is the model estimating? In this work we want to introduce a perspective on penalized risks in non-parametric regression problems that takes into account this question and offers a way to answer it systematically. Answering this question is crucial for understanding the nature of penalization. To motivate the beforementioned perspective, we begin by considering a standard example, namely that the conditional expectation minimizes the MSE risk.

1.1. Motivation

Assume we observe pairs of *iid* random variables (Y_i, X_i) with $i \in \{1, \dots, n\}$, $X_i \in \Omega \subset \mathbb{R}^p$ with Ω being compact and $Y_i \in \mathbb{R}$. We assume that there is a relation between the Y_i and X_i , i.e. there is a mapping $g_0 : \Omega \rightarrow \mathbb{R}$

*Corresponding author
Email address: pappert@statistik.tu-dortmund.de (Sven Pappert)

such that $Y_i = g_0(X_i) + \varepsilon_i$, where $\varepsilon_i \sim (0, \sigma^2)$ is an idiosyncratic, elliptic and *iid* error term. The goal of non-parametric regression is to estimate the function g_0 . By $\hat{g}_n$ we denote the estimator for g_0 based on n observations. The parameter space, Θ_n , containing the parameters which parametrize g_n can vary with n in accordance to the discussion in [Stone \(1982\)](#). A sensible $\hat{g}_n$ ought to be chosen in such a way that the distance between g and $\hat{g}$ is small. Specifying this distance amounts to specifying an (empirical) risk $\hat{Q} : \mathbb{R}^n \times \Omega^n \times H(\Omega) \rightarrow \mathbb{R}$, where $H(\Omega)$ is a suitable function-space on Ω . Choosing $\hat{Q}$ as the (scaled) empirical MSE risk and plugging in the target observations, $\mathbf{Y} := (Y_1, \dots, Y_n) \in \mathbb{R}^n$ and a function $g \in \mathcal{L}^2(\Omega, \mathbb{P}^X)$ evaluated at the feature observations, $\mathbf{X} = (X_1, \dots, X_n) \in \Omega^n$, i.e. $g(\mathbf{X}) := (g(X_1), \dots, g(X_n))$, we may write

$$\hat{Q}(\mathbf{Y}, \mathbf{X}; g) = \frac{1}{n} \sum_{i=1}^n (Y_i - g(X_i))^2.$$

Note that under suitable regulatory conditions, $\hat{Q}(\mathbf{Y}, \mathbf{X}; g)$ is a consistent and unbiased estimator for the MSE risk:

$$Q[g] = \mathbb{E}[(Y_1 - g(X_1))^2] = \int (y - g(x))^2 f_{Y,X}(y, x) dy dx. \quad (1)$$

The aim of this work is to find the function that minimizes risks as the one in Eq. 1 and risks which are penalized by a functional of the gradient of g . In order to do so, we view the risk Q as a functional of the regression function g and search for the $\tilde{g}$ that minimizes the risk functional. A systematic approach for finding extrema of functionals is offered by the calculus of variations. In Sect. 2.1 we introduce some basic concepts, and refer to [Rindler \(2018\)](#) for a comprehensive exposition. To give a simple illustration, we reiterate the result from Chapter 1.5.5 of [Bishop \(2006\)](#): The function minimizing the MSE-risk functional, Q from Eq. 1, is the mapping $x \mapsto \mathbb{E}(Y|X = x)$. The proof idea goes as follows: The first variation of Q w.r.t. g is derived and set to zero, $\delta Q[g] = 0$.¹ This yields an equation for g , whose solution is an extremum. The extremum actually being a minimum follows from the convexity of the MSE-risk. The complete derivation, which we adopt from [Bishop](#)

¹We suggest the readers who are not familiar with the calculus of variations to think of δQ as the total differential of Q w.r.t. a function instead of a variable for now. The differentiation w.r.t. a function is indicated by the ' δ ' (instead of using a ' d '). In Sect. 2.1, we give a formal definition of the first variation.

(2006) is as follows;

$$\begin{aligned}
\delta Q[g] &= \delta \int \int (y - g(x))^2 f_{Y,X}(y, x) dy dx \\
&= \int \int 2(y - g(x)) \delta g(x) f_{Y,X}(y, x) dy dx \\
&= 2 \int \left[\int (y - g(x)) f_{Y,X}(y, x) dy \right] \delta g(x) dx \\
&= 2 \int \left[\int y \underbrace{\frac{f_{Y,X}(y, x)}{f_X(x)}}_{=\mathbb{E}(Y|X=x)-g(x)} dy - g(x) \right] f_X(x) \delta g(x) dx.
\end{aligned}$$

Setting $\delta Q[g] = 0$, the solution, denoted by $\tilde{g}$, is found as $\tilde{g}(x) = \mathbb{E}(Y|X = x)$ for all $x \in \Omega$. Of course, this is not the only way to show that the conditional expectation minimizes the MSE-risk. However, this way offers a generalizable approach to systematically derive equations whose solutions are the minimizers of risk functionals, i.e. *field equations*. This approach amounts to analyzing risks by understanding the risks as functionals of fields, and to find the minimum of the risk using the calculus of variations. The equations obtained by requiring $\delta Q[g] = 0$ are the field equations (the naming is in analogy to physics). A similar derivation as for the MSE-risk can be conducted for the MAE Risk. The results are given in the first two rows of Table 1.

Notice that the unpenalized risk functionals in the first two rows of Table 1 depend only on the field g itself. Given a certain penalty, e.g. a Ridge-penalty, the functional form of the penalty term on the underlying regression function is not obvious, i.e. how would you write $\sum_i^p \beta_i^2$ in terms of the regression function g ? It is also not obvious what the corresponding field equation and its respective solution is. To emphasize these points, the corresponding cells in the table are just question marks, (?1)-(?3), and the penalty is just written as $\widehat{\text{Pen}}[g]$. The goal of this work is to fill these cells for the case where the penalty is gradient-based.² In the related literature different choices for the penalty are available. Taking the integral of the ℓ^1 or ℓ^2 norm of ∇g (or versions thereof) are popular choices, see e.g. Drucker and Le Cun (1991, 1992), Rosasco et al. (2013), Follain and Bach (2024), Ding et al. (2024) or Shen and Lin (2022). In Sect. 2.2.2 we give an overview over the different approaches and distinguish between the respective risk and its empirical counterpart. The (?1)-cell will then be filled by what we will define as *risk functional*, see Def. 1. In order to fill the (?2)-cell, we take the first variation of the risk functional. In particular, if the penalty only depends on the first derivative, the field equations are given by the Euler-Lagrange equations, cf. proposition 2 of this paper or Rindler (2018), Theorem 3.1. Filling the last cell of the table, (?3), amounts to solving partial differential equations (PDEs). For general cases no closed form solutions are available, however there are special cases where the resulting differential equation can be solved.

²This already answers the question 'how would you write $\sum_i^p \beta_i^2$ in terms of the regression function g ?' for linear g . The Ridge-penalty can be obtained by taking the integral of the squared ℓ^2 -norm of the gradient of g , $\int \|\nabla g(x)\|_2^2 d\mathbb{P}^X(x) = \sum_{i=1}^p \beta_i^2$. For non-linear g . We will further discuss this in Sect. 2.2.2.

Table 1: Empirical risks, risks, their field equations and solutions

Name	empirical risk	risk	Field Eqs.	Solution
MSE	$\frac{1}{n} \sum_{i=1}^n (y_i - g(\mathbf{x}_i))^2$	$\mathbb{E}((Y - g(\mathbf{X}))^2)$	$0 = \mathbb{E}(Y \mathbf{X} = \mathbf{x}) - g(\mathbf{x})$	$g = \mathbb{E}(Y \mathbf{X})$
MAE	$\frac{1}{n} \sum_{i=1}^n y_i - g(\mathbf{x}_i) $	$\mathbb{E}(Y - g(\mathbf{X}))$	$0 = g(\mathbf{x}) - F_{Y \mathbf{X}=\mathbf{x}}^{-1}(0.5)$	$g(\mathbf{x}) = F_{Y \mathbf{X}=\mathbf{x}}^{-1}(0.5)$
Penalized-MSE	$\frac{1}{n} \sum_{i=1}^n (y_i - g(\mathbf{x}_i))^2 + \lambda \widehat{\text{Pen}}[g]$	(?1)	(?2)	(?3)

Some of these are outlined in Sect. 3.1.

1.2. Contribution

The contribution of this paper can be structured in three parts. First, we define the risk functional in the non-parametric regression context and integrate gradient-based penalization from the related literature into this framework. Second, we derive the minimum of a first-order risk functional as the solution to the Euler-Lagrange equations associated to the risk functional. We investigate the special case where the gradient-based penalty is the integral of the squared ℓ^2 -norm of the gradient of the regression function. Third, we relate our framework to existing literature on image processing. In particular, we see that we obtain the Rudin-Osher-Fatemi (ROF) framework, cf. [Rudin et al. \(1992\)](#), under the condition that the features deterministic and evenly distributed. The ROF framework was introduced by the authors in order to remove noise from images, there the ‘features’ are just the places of a pixel. In our framework, we allow for random features following any (continuous) distribution. Consequently the joint distribution of the features influences the risk functional.

1.3. Overview

In Sect. 2, we introduce the (basic) concepts, which are necessary to understand this work. These include the calculus of variations, Sect. 2.1, as well as linear and non-parametric penalized regression, Sect. 2.2.1 and Sect. 2.2.2, respectively. For the penalized non-parametric regression, we focus on gradient-based penalization. We give examples for gradient-based penalization from the related literature, including deep learning, non-parametric regression and image denoising, and embed them in our definition of a risk functional, Def. 1. In Sect. 2 we derive the field equations for risk functionals, represented as the sum of a fitness-term and a gradient-based penalty only involving first derivatives. In Sect. 3.1, we proceed to analyze the special case when the risk functional comprises the MSE-risk and the squared- ℓ^2 -penalty. This is the non-parametric equivalent to the Ridge penalty. In Sect. 4, we discuss our findings, their restrictions, possible applications and further research questions.

2. Basics

2.1. Calculus of Variations

The goal of the calculus of variations is to search for minimizers of a functional in a predetermined space of functions. The space of functions considered in this paper is the Sobolev space on $\Omega \subset \mathbb{R}^p$. As is common, we denote by $W^{k,p}(\Omega)$ the Sobolev space, where L^p -integrability of up to the k -th derivative is required, cf. [Rindler \(2018\)](#), Appendix A5. We note that integrability w.r.t. a non-singular measure μ is implied.

The functionals, whose minima we are interested in, are mappings from a Sobolev space to the real numbers, $\Gamma : W^{k,p}(\Omega) \rightarrow \mathbb{R}$. A major result of the calculus of variations are the well-known Euler-Lagrange equations, which we present in Thm. 1 and make use of in Sect. 3. Minimizers of functionals of the form $\Gamma[\phi] = \int f(x, \phi(x), \nabla \phi(x))$ are subject to the Euler-Lagrange equation. We state this well-known result in the following theorem and refer to Theorem 3.1 and Proposition 3.3 from [Rindler \(2018\)](#) for a formulation with predetermined boundary conditions.³ Alternatively, the first theorem in Sect. 2.2 from [Giaquinta and Hildebrandt \(1996\)](#) may be referenced for a formulation without predetermined boundary conditions. Furthermore, we require regulatory conditions, which are given in [Rindler \(2018\)](#). For completeness, we restate them in [Appendix A](#).

Theorem 1 (Euler-Lagrange Equations). *Let $\Gamma : W^{2,p}(\Omega) \rightarrow \mathbb{R}$ be given as $\Gamma[\phi] = \int f(x, \phi(x), \nabla \phi(x)) dx$. Furthermore let Assumptions A1)-A3) from [Appendix A](#) be fulfilled. If $\phi \in W^{1,p}(\Omega)$ minimizes Γ , then ϕ is subject to the following boundary problem*

$$\begin{cases} 0 = -\nabla \frac{\partial f(x, \phi(x), \nabla \phi(x))}{\partial \nabla \phi(x)} + \frac{\partial f}{\partial \phi(x)}(x, \phi(x), \nabla \phi(x)), & x \in \Omega, \\ 0 = \frac{\partial f}{\partial \nabla \phi} \cdot \hat{n}, & x \in \partial\Omega, \end{cases} \quad (2)$$

where $\hat{n}$ denotes the normal vector of Ω .

Remarks:

- a) In Theorem 3.1 from [Rindler \(2018\)](#), the boundary condition is different from the one we give here. The reason for this is that we do not have requirements on ϕ on the boundary ab initio. Instead, we adopt the *natural boundary conditions*, cf. [Giaquinta and Hildebrandt \(1996\)](#), Sect. 2.4.
- b) Theorem 1 does not assure that the solution to the Euler-Lagrange equation is a minimum. The conditions under which a solution to the Euler-Lagrange equation is a minimum are part of Prop. 3.3 from [Rindler \(2018\)](#). For completeness, we restate this proposition as Prop. 1 below.

³Theorem 3.1 assures that minimizers of Γ are subject to the Euler-Lagrange equations while proposition 3.3 assures that solutions to the Euler-Lagrange equations are also minimizers of the functional.

Proposition 1 (Conditions for solution to the E.-L.-Eqs. to be a minimum, from [Rindler \(2018\)](#), Prop. 3.3). Let $\Gamma : W^{2,p}(\Omega) \rightarrow \mathbb{R}$ be given as $\Gamma[\phi] = \int f(x, \phi(x), \nabla \phi(x)) dx$. If the following conditions hold, the solution to the Euler-Lagrange equation of Γ is a minimum of Γ : Let $x \in \Omega$, $s \in \mathbb{R}$ and $u \in \mathbb{R}^p$.

- i) $|\frac{\partial}{\partial_s} f(x, s, u)| \leq C(1 + |s|^{p-1} + \|u\|^{p-1})$ and $\|\frac{\partial}{\partial_u} f(x, s, u)\| \leq C(1 + |s|^{p-1} + \|u\|^{p-1})$ for $C > 0$ and $p > 1$.
- ii) For fixed x , $f(x, s, u)$ is convex in (s, u) .

We note that the proposition in [Rindler \(2018\)](#) also holds for vector-valued fields. Since it is sufficient for our case, we only gave the one-dimensional version of the proposition. The x, s and u are placeholders for $x, \phi(x), \nabla \phi(x)$. [Rindler \(2018\)](#) named these placeholders x, v and A , respectively.

The Euler-Lagrange equations can be derived by taking the first variation of the functional of interest. This is also what we do in the motivation of this paper, Sect. 1.1. The first variation of a functional is defined as

$$\delta\Gamma[\phi] = \lim_{\varepsilon \rightarrow 0} \frac{\Gamma[\phi + \varepsilon\eta] - \Gamma[\phi]}{\varepsilon}. \quad (3)$$

where η is a test function on which we enforce the same assumptions as on ϕ (cf. [Rindler, 2018](#), Eq. 3.1). This expression is similar to the usual differential quotient. The difference is that it is not a function whose value changes by changing the argument by a small amount, rather it is a functional which is changed by changing its argument by adding a test function with a small coefficient. The Euler-Lagrange equations are the result of setting

$$\delta\Gamma[\phi] = 0$$

for Γ as in Theorem. 1. This procedure is analogous to setting the derivative (or the gradient) of a function to zero to find its extrema. Now, the other main concept of this paper – Penalized Regression – is introduced.

2.2. Penalized Regression

Let $(Y_i, \mathbf{X}_i), i \in \{1, \dots, n\}$ be pairs of *iid* and integrable random variables/vectors. $Y_i \in \mathbb{R}$ is the (continuous) target variable, while $\mathbf{X}_i \in \Omega \subset \mathbb{R}^p$ are (continuous) features. It is assumed that there exist square-integrable, *iid* and elliptic random variables $\varepsilon_i \sim (0, \sigma^2)$ and a function $g_0 : \Omega \rightarrow \mathbb{R}$ s.t. we may write $Y_i = g_0(\mathbf{X}_i) + \varepsilon_i$. Now we give a brief to penalized linear regression, followed by the non-parametric case afterwards.

2.2.1. Penalized Regression for Linear Models

Assuming that g_0 is linear, i.e. we may write $g_0(x) = \gamma_0 + \sum_{i=1}^p \gamma_i x_i$, $\forall x \in \Omega$, we take the regression function g to be linear as well, $g(x) = \beta_0 + \sum_{i=1}^p \beta_i x_i$. Then we can write the penalized empirical scaled MSE risk,

evaluated at observations $\mathbf{Y} = (Y_1, \dots, Y_n)$ and $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$, respectively, and parameters β as

$$\hat{Q}(\mathbf{Y}, \mathbf{X}; \beta) = \frac{1}{n} \sum_{i=1}^n \left(Y_i - (\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \right)^2 + \lambda \widehat{\text{Pen}}(\beta) = \frac{1}{n} (\mathbf{Y} - \mathbf{X}\beta)^2 + \lambda \widehat{\text{Pen}}(\beta). \quad (4)$$

where $\lambda > 0$ is the penalization parameter and $\widehat{\text{Pen}}(\beta)$ is a suitable penalty, e.g. the Ridge-penalty, $\widehat{\text{Pen}}(\beta) = \sum_{i=1}^p \beta_i^2$. The idea is to penalize deviations from constancy s.t. simplicity is favored. Notice that the penalty depends on a vector of parameters, β , and not on a field as in the considerations for non-parametric regression in the introduction (and in the next subsection, Sect. 2.2.2). The theoretical counterpart to Eq. 4, i.e. the penalized MSE-risk for linear functions, evaluated at parameters $\beta = (\beta_0, \dots, \beta_p)$, is given as

$$Q(\beta) = \mathbb{E} \left(Y_1 - (\beta_0 + \sum_{j=1}^p \beta_j X_{1j})^2 \right) + \lambda \widehat{\text{Pen}}(\beta).$$

Since $\widehat{\text{Pen}}(\beta)$ only depends on deterministic β , the penalty takes the same form in the empirical and the non-empirical risk. This is a major difference to non-parametric penalization. Depending on the choice of the penalty, the empirical risk and the respective estimators have different properties. We discuss these properties in the following and thereby introduce different penalties.

We start with the case where no penalization is present, i.e. $\lambda \widehat{\text{Pen}}(\beta) = 0, \forall \beta \in \mathbb{R}^p$. With no penalization, we are in the OLS setting and under regularity conditions, the Gauss-Markov theorem assures that the OLS estimator, $\hat{\beta}^{\text{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$, is the best linear unbiased estimator, called 'BLUE'.⁴ While theoretically being the best estimator, the OLS-estimator has some problems in certain set-ups. First, [Hoerl and Kennard \(1970\)](#) point out that, if the features are highly correlated, the estimator is not stable, i.e. it has a high variance. Second, [Tibshirani \(1996\)](#) also points out that the OLS estimator is subject to high variance and that the estimator suffers from estimates deviating from their true value, especially zero, just by chance. These problems become especially relevant in high-dimensional settings. Attempts to overcome these problems lead us to the Ridge and LASSO regularization methods, discussed next.

[Hoerl and Kennard \(1970\)](#) introduce the Ridge estimator as $\beta^{\text{Ridge}} = (\mathbf{X}^T \mathbf{X} + \lambda I_{p \times p})^{-1} \mathbf{X}^T \mathbf{Y}$. The BLUE property of the OLS-estimator is sacrificed in order to obtain a more stable estimator. The relation of $\hat{\beta}^{\text{Ridge}}$ to the empirical loss in Eq. 4 is that $\hat{\beta}^{\text{Ridge}}$ minimizes $\hat{Q}$ when $\widehat{\text{Pen}}(\beta)$ is chosen as $\widehat{\text{Pen}}(\beta) = \sum_{i=1}^p \beta_i^2$ (see Chapter 3.4.1 of [Hastie et al., 2009](#)). Consequently, the term $\sum_{i=1}^p \beta_i^2$ is called Ridge-penalty. An interesting property of the Ridge estimator is that $\mathbb{P}(\beta_i^{\text{Ridge}} = 0) = 0$, i.e. the probability that parameters are shrunk to zero exactly is zero. To this end, the LASSO (least absolute shrinkage and selection operator) penalty was introduced by

⁴See Chapter 1.3 of [Hayashi \(2000\)](#) for the Gauss-Markov theorem and the BLUE property

[Tibshirani \(1996\)](#). Using the LASSO, shrinking parameters to exactly zero becomes possible. The LASSO is given as $\widehat{\text{Pen}}(\beta) = \sum_{i=1}^p |\beta_i|$. Due to the absolute value function not being differentiable at 0, we will not analyze penalties of this form in the current framework. However, in the discussion, Sect. 4.2, we pick up penalties involving absolute values again and sketch a way to find minima of the corresponding risk functionals.

2.2.2. Gradient-based Penalization in non-Parametric Regression

Before discussing approaches from the gradient-based penalization literature, we want to give a general form of an (empirical) risk which is to be analyzed by the calculus of variations. Later in this section we will embed these approaches into our general framework. The risk functional we are interested in is the sum of two terms; the first term assesses a model's fit, while the second term is the penalty term. We denote the fitness term by L and the penalty by Pen , with their empirical counterparts being $\hat{L}$ and $\widehat{\text{Pen}}$, respectively. Examples for L are the MSE, (log-)Likelihood, MAE or the Energy Distance. Examples for Pen are integrals of ℓ^2 or ℓ^1 norms (or versions thereof) of the gradient.

Definition 1 (Risk Functional). Let $\widehat{\text{Pen}} : W^{k,p}(\Omega) \times \Omega \rightarrow \mathbb{R}$ and $\hat{L} : W^{k,p}(\Omega) \times \Omega \times \mathbb{R} \rightarrow \mathbb{R}$. The empirical risk functional evaluated at $\mathbf{Y} \in \mathbb{R}$, $\mathbf{X} \in \Omega$ and $g \in W^{k,p}(\Omega)$, is given as

$$\hat{Q}(\mathbf{Y}, \mathbf{X}, g) = \frac{1}{n} \sum_{i=1}^n \hat{L}(Y_i, g(X_i)) + \lambda \widehat{\text{Pen}}(g, \mathbf{X}). \quad (5)$$

The penalty may include partial derivatives of g up to order $(k-1)$. The fitness term may only include g itself. The theoretical counterpart to the empirical risk functional, evaluated at $g \in W^{k,p}(\Omega)$, is given as $Q : W^{k,p}(\Omega) \rightarrow \mathbb{R}$ such that

$$Q[g] = \int \hat{L}(y, g(x)) d\mathbb{P}^{(X,Y)}(x, y) + \lambda \int \widehat{\text{Pen}}(g, x) d\nu(x) \quad (6)$$

where $\mathbb{P}^{(X,Y)}$ is the image-measure of the random vector (X, Y) and ν is a measure on $(\Omega, \mathcal{B}(\Omega))$. Both measures are assumed to be σ -finite measures which are absolutely continuous w.r.t. the Lebesgue-measure on the respective space. The Radon-Nikodym Lebesgue densities may be called $f_{X,Y}$ and f_ν .

Remarks 1

- a) The measure ν does not need to be a probability measure nor does it need to be the measure associated to the random variable X .
- b) Many gradient-based penalization approaches from the literature are related to the risk functional we define here. Def. 1 is an attempt to 1) generalize risks with gradient-based penalties in such a way that popular approaches from the literature are included and 2) have the risk in a form which allows for analysis by variational calculus.

d) Note that $g \in W^{k,p}$ and that we allow up to $(k-1)$ -order partial derivatives in the penalty term. This is due to the need for integration by parts when taking the first variation of the penalty. Furthermore, $g \in W^{k,p}$ ensures L^p -integrability w.r.t. the Lebesgue-measure of up to the k -th derivative. In the risk functional, we integrate w.r.t. σ -finite measures which are absolutely continuous w.r.t. the Lebesgue-measure. The existence of the integrals is guaranteed by the Lebesgue-integrability of g .

Now we proceed to give examples gradient-based penalization from the literature and, when possible, embed them into the risk functional framework from Def. 1.

First, we consider ℓ^2 -based penalties and start with the squared- ℓ^2 -penalty:

$$\widehat{\text{Pen}}(\nabla g) = \frac{1}{n} \sum_{i=1}^n \|\nabla g(X_i)\|_2^2, \quad (7)$$

$$\text{Pen}(\nabla g(X_1)) = \int \|\nabla g(x)\|_2^2 d\nu, \quad (8)$$

respectively. This penalty is sometimes called ‘Sobolev-regularization’ (see e.g. Rosasco et al. (2013), remark 1; Ding et al. (2024), Sect. 1). Ding et al. (2024) also consider a penalty of the form $\frac{1}{n} \sum_{i=1}^n \|\nabla g(Z_i)\|_2^2$, where Z_i are *iid* random variables on Ω which do not have to correspond to the measure associated to X (see Eq. 3 of their paper). This of course amounts to specifying the measure ν in Eq. 8 accordingly. When g is linear, i.e. we may write $g(x) = \beta_0 + \sum_{i=1}^p \beta_i x_i$, $\forall x \in \Omega$, then $\widehat{\text{Pen}}(\nabla g) = \text{Pen}(\nabla g(X_1)) = \sum_{i=1}^p \beta_i^2$. This is the classical ridge penalty corresponding to the ridge estimator from Hoerl and Kennard (1970), compare also Sect. 2.2.1. The embedding into the risk functional framework from Def. 1 is obvious: $\text{Pen} : W^{2,2} \rightarrow \mathbb{R}$ is only dependent on first-order partial derivatives. Risks of such form may be called *First-order risk functionals*. The penalty in smoothing spline models is also of squared- ℓ^2 -form. There, the form of the penalty is $\int_{\Omega} (g^{(m)}(x))^2 dx$, where $g^{(m)}$ denotes the m -th order derivative, cf. Wahba (1990), Chapter 1. Choosing $\widehat{\text{Pen}}(g, X) = g^{(m)}$ and $\nu = \lambda$, we may embed the smoothing spline risk into the risk functional framework too. Risks with m -th order derivatives may be called *m-th-order risk functionals*. Additionally, we recognize that the empirical fitness term for a smoothing spline is $\frac{1}{n} \sum_{i=1}^n (Y_i - g(x_i))^2$. Only $\{Y_i\}_{i=1}^n$ are considered random variables, while $\{x_i\}_{i=1}^n$ are considered to be deterministic. Accordingly, the image-measure $\mathbb{P}^{(X,Y)}$ in Eq. 6 can be written as $\mathbb{P}^{(X,Y)} = \frac{\#p}{n} \otimes \mathbb{P}_Y$ to recover the smoothing spline risk.

A related penalty is introduced in Drucker and Le Cun (1991, 1992). There, the ‘backward energy function’, $\hat{E}_b = \sum_{i=1}^n \|\nabla (\frac{1}{2}(Y_i - g(X_i))^2)\|_2^2$, is used as penalty in the context of artificial neural networks (ANNs). The authors introduce double-backpropagation to calculate $\hat{E}_b$ efficiently when g is an ANN. The backward energy function can be embedded into the risk functional framework by writing $E_b = \int \|\frac{1}{2}\nabla(y - g(x))^2\|_2^2 d\mathbb{P}^{(X,Y)}(x, y)$ and $\text{Pen} : W^{2,2} \rightarrow \mathbb{R}$ with $\text{Pen} = E_b$. Then, the penalty depends g and its first partial derivatives.⁵ All before

⁵In the original paper, the notation is different and the risk and penalty are given for multiple output networks. We adjusted the

mentioned penalties are squared ℓ^2 -based penalties. Without squaring, the penalty involves the euclidean norm. Such penalties are popular in image denoising. In the seminal paper by [Rudin et al. \(1992\)](#), the authors use the penalty $\int \|g(\mathbf{x})\|_2 d\mathbf{x}$.⁶ In the context of signal-denoising this penalty is often called 'total variation'; amongst others it is used in [Torres et al. \(2014\)](#) for gravitational wave denoising. Similar to the penalization in the spline risk, the penalty does not involve random variables.

Next, we discuss the ℓ^1 -penalty:

$$\widehat{\text{Pen}}(\nabla g, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \|\nabla g(X_i)\|_1, \quad (9)$$

$$\text{Pen}(\nabla g) = \int \|\nabla g\|_1 d\nu. \quad (10)$$

A penalty of this form is used in [Shen and Lin \(2022\)](#) for penalization of ANNs with the rectified linear unit (ReLU) activation function. It is clear that for a linear function g , the classical LASSO penalty from [Tibshirani \(1996\)](#) is recovered. For the case where g is parametrized as a GAMLSS, the penalty from [Groll et al. \(2019\)](#) is recovered.

Last, we want to introduce the RKHS & gradient-based penalization by [Rosasco et al. \(2013\)](#), which is introduced for general purpose regularization. In the paper, the authors analyze an empirical risk of the following form:

$$\frac{1}{n} \sum_{i=1}^n (Y_i - g(X_i))^2 + \lambda \left(2 \sum_{a=1}^p \left\| \frac{\partial g}{\partial x_a} \right\|_n + v \|g\|_{\mathcal{H}}^2 \right).$$

The first part of the empirical risk is just the usual MSE. The second part, the penalty, consists of a derivative based term, $\sum_{a=1}^p \left\| \frac{\partial g}{\partial x_a} \right\|_n = \sum_{a=1}^p \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial g(X_i)}{\partial x_a} \right)^2}$, and a RKHS-penalty, $v \|f\|_{\mathcal{H}}^2$. The former is based on squared derivatives and shares resemblance with the group LASSO, see [Yuan and Lin \(2006\)](#) for the group LASSO. The term $v \|g\|_{\mathcal{H}}^2$ is a norm on the Hilbert space consisting of functions which fulfill the reproducing property w.r.t. a pre-determined kernel K (see [Rosasco et al. \(2013\)](#), Sect. 4.) The norm depends on the explicit choice of the function space. The non-empirical risk can be obtained by replacing the MSE by the expected MSE and $\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial g(X_i)}{\partial x_a} \right)^2$ by $\mathbb{E}[(\frac{\partial g}{\partial x_a}(X))^2]$, given that the corresponding conditions for convergence are fulfilled. The embedding into the risk functional framework depends on the explicit choice of the kernel K . In its general form, the risk can not be embedded into the risk functional.

Another gradient-based penalization approach, which can not be represented as a risk functional is the one by [Follain and Bach \(2024\)](#). There, an ℓ^2 -type penalty is used but the authors focus on Hermite polynomials and introduce coefficients for each summand of the Hermite polynomial series. The introduction of the coefficient

notation and dimensionality to fit in the context of this paper.

⁶In signal denoising the fidelity term has a different structure and does not include a density. In the paper, the authors

hinders the analysis with our approach. Of course, there are much more regularization approaches available in the literature and this is not an exhaustive discussion. Here, we merely introduced what we think are the basic building blocks of gradient-based penalization. In the next section, Sect. 3, we proceed to derive the field equations of the a first-order risk functional.

3. Field equations of Penalized Non-Parametric Regression

Assume the set-up from above, i.e. a non-parametric regression setting with a risk penalized by a gradient-based penalty. In this section we analyze penalized risks by means of the calculus of variations. We start by proposing the general field equations for first-order risk functionals, i.e. penalties only involving differentiable mappings of the first gradient. Then we proceed by explicitly stating the field equations of the squared- ℓ^2 penalized MSE-risk.

Proposition 2. *Let $(Y_i, X_i)_{i=1}^n$ be iid pairs of random variables and $Y_i \in \mathbb{R}$, $X_i \in \Omega \subset \mathbb{R}^p$. Let $\mathbb{P}^X$ be the image-measure associated to the random variables X_i , $\mathbb{P}^Y$ the image measure associated to the random variables Y_i and $\mathbb{P}^{(X,Y)}$ their joint image-measure. Furthermore, let ν be a measure on $(\Omega, \mathcal{B}(\Omega))$ s.t. $\text{Pen}(\nabla g) \in \mathcal{L}(\Omega, \nu)$. All measures are assumed to be σ -finite and absolutely continuous w.r.t. the respective Lebesgue-measure on the space and their Lebesgue-densities are denoted as f_X , f_Y , $f_{X,Y}$ and f_ν , respectively. Let $g \in W^{2,p}(\Omega)$ and let Q be a risk functional of the form*

$$Q[g] = \int L[y; g(x)] d\mathbb{P}^{(X,Y)} + \lambda \int \text{Pen}[\nabla g] d\nu. \quad (11)$$

Furthermore, let assumptions (A1)-(A5) from [Appendix A](#) hold. Then Q is minimized by the solution to the following differential equation (Euler-Lagrange equation):

$$0 = \int \frac{\partial L(x, y, g)}{\partial g} f_{X,Y}(x, y) dy - \lambda \nabla \left[\frac{\partial \text{Pen}}{\partial \nabla g} [\nabla g(x)] f_\nu(x) \right], \quad (12)$$

subject to boundary conditions: $\frac{\partial \text{Pen}}{\partial \mathbf{n}}[g] = 0$ on $\partial\Omega$, where $\frac{\partial}{\partial \mathbf{n}}$ denotes the normal derivative.

Proof. We write Q as

$$Q[g] = \int f(x, g(x), \nabla g(x)) dx, \text{ where} \quad (13)$$

$$f(x, s, u) = \int L(y, s) f_{X,Y}(x, y) dy + \lambda \text{Pen}(u) f_\nu(x). \quad (14)$$

Then the minimizer of Q is given as the solution to the Euler-Lagrange equations of the system, see Theorem. 1. Since no boundary conditions are imposed a priori, the natural boundary conditions are adopted. $\square$

Remarks:

- a) We note that the field equations in Eq. 12 are the Euler-Lagrange equations of the first-order risk functional. The difference to usual use-cases of the Euler-Lagrange equations, e.g. in physics, is that the integrals here are measure integrals and, as a consequence, the respective Lebesgue-densities enter the equation.
- b) We take $g(0) = \mathbb{E}(Y|X=0)$ as auxiliary boundary condition when necessary. The necessity occurs e.g. in the limiting case $\lambda \rightarrow \infty$. This particular choice will be motivated in the following example of the squared- ℓ^2 penalized MSE-risk. It would also be possible to impose boundary conditions such as $g = \mathbb{E}(Y|X = \cdot)$ on $\partial\Omega$ a priori. Imposing no boundary conditions a priori is an explicit choice that we make.

Now we explicitly give the field equations for the squared- ℓ^2 penalized MSE-risk.

3.1. Field Equations of the squared- ℓ^2 Penalized MSE-Risk

Let $\Omega = [0, 1]^p$, $L(x, y) = (x - y)^2$ and $\text{Pen}(w) = \|w\|_2^2$ and $\nu = \mathbb{P}^X$. Then the risk functional is given as

$$\begin{aligned} Q[g] &= \mathbb{E}[(Y - g(X))^2] + \lambda \mathbb{E}(\|\nabla g(X)\|_2^2) \\ &= \int (y - g(x))^2 d\mathbb{P}^{(Y,X)}(x, y) + \lambda \int \|\nabla g(x)\|_2^2 d\mathbb{P}^X(x) \\ &= \int (y - g(x))^2 f_{Y,X}(y, x) dx dy + \lambda \int \|\nabla g(x)\|_2^2 f_X(x) dx. \end{aligned}$$

This is a first-order risk functional. There is a solution to the minimization problem $\min_{g \in W^{2,2}} Q[g]$ and the function $g : \Omega \rightarrow \mathbb{R}$ that minimizes the risk functional on Ω is given as the solution to the following Neumann boundary problem,

$$\begin{cases} -\lambda \Delta g(x) - \lambda [\nabla \log(f_X(x)) \cdot \nabla g(x)] + g(x) = \mathbb{E}[Y|X=x] & x \in \Omega, \\ \frac{\partial g(x)}{\partial x_i} \Big|_{x_i=0} = \frac{\partial g(x)}{\partial x_i} \Big|_{x_i=1} = 0. \end{cases}$$

Again, we additionally impose $g(0) = \mathbb{E}(Y|X=0)$, when necessary as in e.g. Remark c) below.

Remarks:

- a) The conditions from [Appendix A](#) have to be checked. We recognize that

$$f(x, s, u) = \int (y - s)^2 f_{X,Y}(x, y) dy + \lambda \|u\|_2^2 f_X(x).$$

Then: 1) $f(x, s, u) = \int (s - y)^2 f_{X,Y}(x, y) dy + \lambda \|u\|_2^2$ is a Carathéodory integrand since f is measurable and continuous in all arguments; 2) f is continuously differentiable in s and u since f is quadratic in both arguments; 3) $|\partial_v f(x, s, u)| = |2 \int (s - y) f_{X,Y}(x, y) dy| \leq c_1 |s|$ and $|\nabla_u f(x, s, u)| = \lambda 2|u| \leq c_2 |u|$; 4) f is convex in (s, u) . So the assumptions are fulfilled and the solution to the field equation is a minimum of the risk functional.

- b) The field equation is a second-order, inhomogeneous, linear PDE. The inhomogeneity is the conditional expectation and the coefficient function for ∇g is the score of the joint distribution of the features, $\nabla \log(f_X(x))$. The PDE may be used for post regularization, i.e. having an unbiased estimator for $\mathbb{E}(Y|X = \cdot)$ which suffers from high variance, it may be used as plug-in estimator in the field equations. Then the solution to the field equation, $\tilde{g}$, will be regularized.
- c) For $\lambda = 0$, we obtain $g(x) = \mathbb{E}(Y|X = x)$, as can be expected. For $\lambda \rightarrow \infty$, the equation simplifies to

$$\Delta g(x) = -\nabla \log(f_X(x)) \cdot \nabla g(x).$$

This equation is trivially solved by $g = C$, $C \in \mathbb{R}$, whereby the Neumann boundary condition is also automatically fulfilled. To determine the constant C , we use the auxiliary condition that $g(0) = \mathbb{E}(Y|X = 0)$. Then $g(x) = \mathbb{E}(Y|X = 0)$, $\forall x \in \Omega$.

For f_X being the independent uniform distribution on Ω , $f_X(x) = \frac{1}{\lambda(\Omega)} = 1$, we obtain

$$\mathbb{E}(Y|X = x) = -\lambda \Delta g(x) + g(x).$$

This reflects the fact that if the distribution of X is unstructured, it does not contribute to the PDE and consequently does not influence the minimum of the risk. Under this assumption and the ℓ^2 -penalty (instead of the squared- ℓ^2 -penalty) this equation is the equivalent to the Rudin-Osher-Fatemi differential equation, cf. [Rudin et al. \(1992\)](#). The equation for the squared- ℓ^2 -penalty and f_X being the uniform distribution is also given in [Ding et al. \(2024\)](#).

- d) The one-dimensional version of the equation is given as

$$\begin{cases} \mathbb{E}(Y|X = x) = -\lambda g''(x) - \lambda (\log(f_X(x)))' g'(x) + g(x) & x \in \Omega \\ g'(0) = g'(1) = 0 \end{cases} \quad (15)$$

This equation is closely related to the Hermite differential equation and may be solved by a power-series ansatz. For the special case where additionally $f_X(x) = 1$, the resulting ODE can be solved explicitly for several forms of $\mathbb{E}(Y|X = x)$. To illustrate the mechanics of the ODE, we set $\mathbb{E}(Y|X = x) = \cos(\omega x)$. Then it is easy to show that the solution, $\tilde{g}$, is given as

$$\tilde{g}(x) = A \sinh\left(\frac{1}{\sqrt{\lambda}}x\right) + \frac{1}{1 + \lambda\omega^2} \cos(\omega x).$$

with $A = \frac{\sqrt{\lambda}\omega \sin(\omega)}{2 \sinh\left(\frac{1}{\sqrt{\lambda}}\right)(1 + \lambda\omega^2)}$. Notice that $\lim_{\lambda \rightarrow 0} \tilde{g}(x) = \cos(\omega x) \forall x \in [0, 1]$ with rate $\sqrt{\lambda}$ and $\lim_{\lambda \rightarrow \infty} \tilde{g}(x) = 0 \forall x \in [0, 1]$ with rate $\sqrt{\lambda}$. Interestingly, for $\lambda \rightarrow \infty$, the first term dominates.

4. Discussion

4.1. Overview

Penalized risks in non-parametric regression settings are analyzed by means of variational calculus. When the risk is comprised of a fitness-term and a gradient-based penalty, under suitable regularity conditions, the minimum of the risk can be found as the solution to the Euler-Lagrange equations associated to the risk. The fitness-term analyzed in this paper is the mean-squared error (MSE) risk, $L[g] = \mathbb{E}((Y - g(X))^2)$. L is a quadratic functional and its unique minimum is $\tilde{g}(x) = \mathbb{E}(Y|X = x)$, cf. Sect. 1.5.5 from [Bishop \(2006\)](#) or the motivation sections of this paper, Sect. 1.1. Adding a penalty term to the MSE-risk complicates finding the minimum of the associated risk functional. The penalty terms considered in this paper are gradient-based penalties. Especially interesting is the case where the penalty is of the form $\text{Pen}[g] = \int \|\nabla g(x)\|_r^s d\nu(x)$. When $s = r = 2$ we may call the penalty squared- ℓ^2 -penalty. From a statistical point of view such penalty decreases the variance of an estimator $\hat{g}$ when it is used in its training. For a linear g , e.g. $g = \beta_0 + \sum_{i=1}^p \beta_i x_i$, the squared- ℓ^2 penalty is $\sum_{i=1}^p \beta_i^2$, which corresponds to the classical ridge penalty, cf. Sect. 3.4.1 from [Hastie et al. \(2009\)](#) or Sect. 2.2.1 of this paper. From a calculus viewpoint, the squared- ℓ^2 -penalty is a Dirichlet functional if ν is the Lebesgue-measure, cf. example 2.8 from [Rindler \(2018\)](#). It is well known that the minimum of the Dirichlet functional is the solution to the Laplace equation. This fact alone motivates the realization that the minimum of a risk comprised of the MSE and a gradient-based penalty is subject to a second order partial differential equation. In fact, the differential equation for the squared- ℓ^2 -penalized MSE is derived as a linear, inhomogeneous second order PDE in Sect. 3.1. In addition to the derivation for the squared- ℓ^2 penalized MSE-risk, we derive the field equations for first-order risk functionals, as a direct consequence of the Euler-Lagrange equations. These field equations may be used for post-processing of estimators. Given a non-parametric estimator for $\mathbb{E}(Y|X = \cdot)$ with a high variance, the variance can be reduced by using the estimator as a plug-in estimator for the conditional expectation in the field equation and (numerically) solving the field equation for $\tilde{g}$ and then taking $\tilde{g}$ as the updated estimator.

4.2. Analysis of other penalization and regularization methods

While we tried to render the results in this paper as general as possible, there are penalization/regularization methods where it is not clear how they may be analyzed using this framework. First and foremost, regularization methods which are not representable by a penalty have to be mentioned. Examples are dropout regularization for MLPs, cf. [Hinton et al. \(2012\)](#) or best subset selection, cf. [Hastie et al. \(2009\)](#), Sect. 3.3.1. For such regularization methods there is almost no hope to formulate the associated effective risk as first (or any order) functional.

Next, there are regularization methods which are based on a penalty, but the penalty can not be formulated in terms of the gradient. E.g., [Follain and Bach \(2024\)](#) use an ℓ^2 -type loss for Hermite polynomials but they add a coefficient to each part of the Hermite polynomial derivative series. These coefficients hinder formulating

Euler-Lagrange equations for this particular risk. Also the penalty from [Rosasco et al. \(2013\)](#) is hard to analyze because of the restriction to a RKHS. In principle, it could be ignored that all functions are from a certain RKHS, but then the minimum of the risk functional could not be included in the RKHS.

Last, we consider gradient-based penalties, which can not be analyzed because of non-differentiability. An obvious example is ℓ^1 -based penalization,

$$\text{Pen}[g] = \int \|\nabla g(x)\|_1 d\nu(x). \quad (16)$$

At the current moment it is not clear how to deal with such set-ups because standard variational calculus methods are not applicable. Nonetheless, there is some hope because at least for the special case where $\nu = \lambda$, the minimum of the functional in Eq. 16 can be seen to be given by monotonic fields.⁷ However, it is not clear yet what the minimum of the functional in conjunction with a fitness term will be.

Next to ℓ^1 -penalization, another interesting research question would be to analyze a generalization of the adaptive LASSO from [Zou \(2006\)](#). A non-parametric generalization of the adaptive LASSO penalty can be conceived as

$$\text{Pen}[g] = \int \sum_{i=1}^p \frac{|\partial_i g(x)|}{\partial_i \mathbb{E}(Y|X=x)} d\nu(x), \quad (17)$$

The interesting question here would be if the oracle properties, cf. [Zou \(2006\)](#) Sect. 3.2, translate to the non-parametric case and if they can be uncovered in the field equations.

4.3. Convergence to the minimum of the risk functional and universal approximation

Notice that throughout this paper, we implicitly assumed that a statistical learner, when trained w.r.t an empirical risk, learns the minimum of the risk. Currently this statement is merely a conjecture. To clarify this, we first would like to make a distinction between learners for which a universal approximation theorem holds and learners for which there is no universal approximation theorem. An example for a universal learner is a feedforward MLP, cf. [Hornik \(1991\)](#). If the solution to the Euler-Lagrange equation lies in the function space which the learner can universally approximate, then we presume that the learner converges to the minimum of the Euler-Lagrange equation for the MSE with gradient-based penalization. However, for now, this is just a presumption, since the universal approximation theorem is a deterministic statement. It does not follow directly that the estimation is consistent.⁸ Now if the function space of the learner is restricted and hence the learner is not subject to a universal approximation theorem on a general function space in which the solution of the

⁷By 'monotonic field' we mean a field g which is subject to either $\frac{\partial g(x)}{\partial x_i} \geq 0$ or $\frac{\partial g(x)}{\partial x_i} \leq 0$ for all $x \in \Omega$ and all $i \in \{1, \dots, p\}$. This result may be of independent interest and will be presented in forthcoming works.

⁸For some risks and learners there are consistency results available in the literature. [Schmidt-Hieber \(2020\)](#) show that ReLU-networks are consistent under certain assumptions (and additionally the author derived the convergence rates) and [Kohler and Langer \(2021\)](#) derive the convergence rates for fully connected MLPs.

Euler-Lagrange equation lies in, then the learner will not converge to the global minimum of the risk. Instead, we presume that the learner converges to the minimum of the risk functional over the restricted function space. The two presumptions are not trivial and have to be proven. This is subject to future research. A possible path would be to first establish the uniform convergence of the risks as functions of parameters and then generalize to uniform convergence of functionals.

Acknowledgement

This research was (partially) funded in the course of TRR 391 *Spatio-temporal Statistics for the Transition of Energy and Transport* (520388526) by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

References

- C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer New York, NY, 2006.
- Z. Ding, C. Duan, Y. Jiao, and J. Z. Yang. Semi-Supervised Deep Sobolev Regression: Estimation, Variable Selection and Beyond. *arXiv preprint arXiv:2401.04535*, 2024.
- H. Drucker and Y. Le Cun. Improving generalization performance in character recognition. In *Neural Networks for Signal Processing: Proceedings of the 1991 IEEE Workshop*, 1991.
- H. Drucker and Y. Le Cun. Improving generalization performance using double backpropagation. *IEEE transactions on Neural Networks*, 3(6):991–997, 1992.
- B. Follain and F. Bach. Nonparametric linear feature learning in regression through regularisation. *Electronic Journal of Statistics*, 18(2):4075–4118, 2024.
- M. Giaquinta and S. Hildebrandt. *Calculus of Variations I*. Springer Berlin, Heidelberg, 1996.
- A. Groll, J. Hambuckers, T. Kneib, and N. Umlauf. LASSO-type penalization in the framework of generalized additive models for location, scale and shape. *Computational Statistics & Data Analysis*, 140:59–73, 2019.
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer New York, NY, 2009.
- F. Hayashi. *Econometrics*. Princeton University Press, 2000.
- G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- A. E. Hoerl and R. W. Kennard. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*, 12(1):69–82, 1970.
- K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.
- M. Kohler and S. Langer. On the rate of convergence of fully connected deep neural network regression estimates. *The Annals of Statistics*, 49(4):2231–2249, 2021.
- F. Rindler. *Calculus of Variations*. Springer Cham, 2018.
- L. A. Rosasco, S. Villa, S. Mosci, M. Santoro, and A. Verri. Nonparametric sparsity and regularization. *Journal of Machine Learning Research*, 14(1):1665–1714, 2013.
- L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

- J. Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. *The Annals of Statistics*, 48(4):1875–1897, 2020.
- X. Shen and J. Lin. Consistency of neural networks with regularization. *arXiv preprint arXiv:2207.01538*, 2022.
- C. J. Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053, 1982.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.
- A. Torres, A. Marquina, J. A. Font, and J. M. Ibáñez. Total-variation-based methods for gravitational wave denoising. *Physical Review D*, 90:084029, 2014.
- G. Wahba. *Spline models for observational data*. SIAM, 1990.
- M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(1):49–67, 2006.
- H. Zou. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.

Appendix A. Assumptions for the Euler-Lagrange Equations

Definition 2 (Assumptions). For a functional $Q : W^{k,q} \rightarrow \mathbb{R}$, with

$$Q[g] = \int L[y; g(x)] d\mathbb{P}^{(X,Y)} + \lambda \int \text{Pen}[\nabla g] d\nu, \quad (\text{A.1})$$

which we may rewrite as

$$Q[g] = \int f(x, g(x), \nabla g(x)) dx, \text{ where} \quad (\text{A.2})$$

$$f(x, s, u) = \int L(y, s) f_{X,Y}(x, y) dy + \lambda \text{Pen}(u) f_\nu(x), \quad (\text{A.3})$$

We define the following assumptions:

- A1) f is a Carathéodory integrand: The mapping $x \mapsto f(x, s, u)$ is Lebesgue measurable $\forall s \in \mathbb{R}, u \in \mathbb{R}^p$ and the mapping $(s, u) \mapsto f(x, s, u)$ is continuous λ almost everywhere.
- A2) f is continuously differentiable in its second and third argument.
- A3) f fulfills the following growth bounds: $\forall x \in \Omega, s \in \mathbb{R}, u \in \mathbb{R}^p$ there is $M > 0$ and $q \geq 1$ s.t. $|\partial_s f(x, s, u)| \leq M(1 + |s|^{q-1} + |u|^{q-1})$ and $|\nabla_u f(x, s, u)| \leq M(1 + |s|^{q-1} + |u|^{q-1})$
- A4) $f(x, s, \cdot)$ is quasiconvex $\forall (x, s) \in \mathbb{R}, \mathbb{R}$

We note that the assumptions are directly taken from Theorem 3.1 and proposition 3.3 from [Rindler \(2018\)](#).