Dependence in Statistical Data Analysis:

From Structural Insights Through Decomposition to the Combination of Forecasts and the Evaluation of Tracking Changes

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Jonas Benjamin Rieger

Tag der mündlichen Prüfung: 23. Juli 2024

Referent: Prof. Dr. Oliver Grothe
Korreferent: Prof. Dr. Maximilian Coblenz

Karlsruhe 2024

Danksagung

Diese Arbeit wäre nicht möglich gewesen ohne den Beitrag vieler Menschen, denen ich an dieser Stelle meinen Dank aussprechen möchte. Der erste Dank gilt meinem Doktorvater Professor Dr. Oliver Grothe, der mir bereits während der Masterarbeit erste Einblicke in die Forschung ermöglichte und mir so den Weg in die Promotion ebnete. Seine Unterstützung und das entgegengebrachte Vertrauen haben mir eine freie und selbstbestimmte Forschung ermöglicht und der fachliche Austausch mit ihm, seine konstruktiven Anmerkungen und seine offene und bestärkende Art haben mich motiviert und inspiriert. Ebenso bedanke ich mich bei meinem Korreferenten Professor Dr. Maximilian Coblenz, meiner Prüferin Professorin Dr. Melanie Schienle, sowie dem Prüfungsvorsitzenden Professor Dr. Martin Klarmann für ihre investierte Zeit und ihr Interesse an meiner Arbeit.

Der Austausch mit meinen Kolleginnen und Kollegen am Lehrstuhl für Analytics and Statistics am Institut für Operations Research des Karlsruher Instituts für Technologie hat mir Freude bereitet, mich motiviert und mir neue Impulse für meine Arbeit gegeben. Ich danke insbesondere Dr. Mira Wälde, Dr. Fabian Kächele, Anika Kaplan, Parzival Borlinghaus, Bolin Liu und Na Luo. Darüber hinaus danke ich Marion Rihm für ihre große Unterstützung in allen administrativen Angelegenheiten und ihre immer offene und fröhliche Art. Ich danke Professor Dr. Johan Segers und Professor Dr. Bernd Saugel für die interessanten und inspirierenden Diskussionen, die meine Arbeit bereichert haben. Außerdem danke ich der bischöflichen Studienförderung Cusanuswerk für die finanzielle Unterstützung während meiner Promotion und die Organisation vielfältiger Veranstaltungen, die mir während der Promotion den Blick über das eigene Fachgebiet hinaus ermöglicht haben.

Abschließend danke ich meiner Familie und meinen Freundinnen und Freunden für ihre Unterstützung und ihr Verständnis in den letzten Jahren. Ich danke meinen Eltern Christine und Thomas Rieger, die mir durch ihr bedingungsloses Vertrauen und ihre Förderung ein sorgenfreies Studieren und Promovieren ermöglicht haben. Außerdem danke ich meinen Geschwistern Judith Hoffmann und Dr. Florian Rieger mit ihren Familien, meinen Schwiegereltern Heidrun und Klaus Steiner und Tobias Schürmann und Benedikt Menrad für ihre Anteilnahme, ihr Interesse und die vielen bestärkenden Worte.

Zuletzt danke ich meiner Frau Anna für ihr Verständnis, ihre Unterstützung gerade in stressigen Zeiten und ihren Glauben an mich und ich danke unserer Tochter Magdalena, die mich mit ihrer fröhlichen Art, die Welt zu entdecken, auch in den vertracktesten Situationen auf andere Gedanken gebracht hat.

Contents

List of Figures

List of Tables

List of Abbreviations

ABP	invasive arterial blood pressure
AD	Anderson-Darling
AIC	Akaike information criterion
ATC	ability to track changes
BCa	bias-corrected and accelerated
BLP	beta-transformed linear pool
BMA	Bayesian model averaging
BS	Brier score
CA	Cuadras-Augé
CCD	copula combined density
CDF	cumulative distribution function
CRPS	continuous ranked probability score
DDNN	distributional deep neural network
EMOS	ensemble model output statistics
EPF	electricity price forecasting
FGM	Farlie-Gumbel-Morgenstern
JSU	Johnson's SU
KDE	kernel density estimation
KS	Kolmogorov-Smirnov
LS	logarithmic score
MAE	mean absolute error
MAR	monotonicity anchored representation
NBP	non-invasive blood pressure
PDF	probability distribution function
PIT	probability integral transform
RMSE	root mean squared error
SLP	spread-adjusted linear pool
SVD	singular value decomposition
TADDA	targeted absolute deviation with direction augmentation
TLP	traditional linear pool

1. Introduction

While independence of features simplifies statistical methods considerably, it is often not fulfilled in larger real-world data sets with multiple features. Falsely assuming independence can lead to incorrect results, conclusions, and decisions. Statistical dependence modeling comprises methods to model and analyze multiple-featured data and multidimensional random variables beyond the independence assumption. This thesis delves into three facets of dependence analysis and dependence modeling application and, thus, broadens the understanding and utilization of dependence modeling methods. First, a new decomposition of a specific dependence function type, checkerboard copulas, is presented. The decomposition simplifies the assessment and interpretation of the incorporated dependence information. Various methods and insights are derived from the decomposition, such as decompositions of characteristic copula functionals and new graphical representations. Second, copula dependence models are used to combine point forecasts into a joint density forecast. Thus, the combined forecast integrates knowledge of the forecast error dependence into the combination and can increase accuracy and calibration. Third and finally, new measures and visualization techniques evaluate the joint change direction of predicted and observed change for nowcasts, forecasts, and measurements. Assessing a method's ability to predict the direction of changes provides valuable and easily interpretable information on the model performance for decision-makers.

In the following, a detailed outline of the contributions and the structure of the dissertation is provided. Chapter 3 analyzes a decomposition for a particular class of copulas, the bivariate checkerboard copulas. Copulas are a powerful tool for modeling dependence, as they separate the dependence model from the marginals. While the fundamental work of Sklar (1959) paved the way and coined the term *copula*, copulas are nowadays a widespread and versatile tool for modeling dependence. Their application improves statistical modeling in various applications, ranging from well-known applications in, for example, financial risk management (Embrechts, Lindskog, and Mcneil, 2003) and economics (Patton, 2012; for introductions, see Nelsen, 2006; Joe, 2014; Durante and Sempi, 2015) to specialized and new applications, such as emergency call center call bursts (L'Ecuyer, Gustavsson, and Olsson, 2018), drought and flood risk assessment (Yang, Zhang, and Gao, 2023), and earthquake warning systems (J. Wang et al., 2018). Checkerboard copulas are a particular class of copulas with constant density on an even-spaced unit square grid. Hence, the density has a checkerboard-like structure, and a doubly stochastic matrix can represent the dependence structure.

A checkerboard copula can approximate any copula and is, thus, a theoretically important concept. However, the incorporated information is difficult to assess as the corresponding matrix is large, even for moderate grid sizes. We propose using singular value decomposition to decompose the matrix, providing an easier-to-assess and interpretable representation. Based on the decomposition, we derive insights into the structure of various statistical functionals, such as decompositions of Kendall's τ and Spearman's ρ, and formulate copula similarity measures based on the decomposition. In the spirit of correspondence analysis, we derive new copula profile plots and remark on their interpretation. Two extensions of pure singular value decomposition improve the representation of comonotonicity-like copulas and ensure the truncation's validity. Consequently, the approximations are valid and sparse, in particular, compared to the number of matrix elements, that is, the square of the grid size.

The decompositions are provided for various copulas, symmetric and asymmetric. These examples illustrate the effect of grid size on decomposition and proposed extensions. Using data on fuel injector spray characteristics in jet engines, we illustrate the new copula profile plots on real-world data. We provide the ready-to-use code for the decomposition to facilitate generating further insights with the decomposition. The chapter is based on the joint work with Oliver Grothe (Publ. I).

Chapter 4 improves the combination of point forecasts to a joint density forecast by incorporating dependence information. Combining point forecasts from different models for the same quantity usually improves the forecasts. The combination unites the strengths of different models and reduces the influence of adverse effects, such as the effect of random initialization in model fitting. Thereby, the errors of the different forecasts are usually mutually dependent. The dependence can stem from factors similarly influencing the forecasts, such as a common model structure, common input data, or hard-to-account-for external effects. Existing point forecast combination methods do not respect complex dependence of forecast errors such as asymmetries, leading to a suboptimal combination.

The proposed new combination scheme accounts for the forecasts' serial dependence and the forecast errors' dependence structure by a copula time series model and computes a combined density forecast based on the point predictions. By separating margin and copula in the copula time series model, the combination is more flexible than current approaches; the components can be fitted independently using a wide range of well-established, specialized methods. We prove two theoretical properties of the algorithm, that is, the calibration of the combination and the embedding of the seminal variance-minimizing combination by Bates and Granger (1969) in the case of multivariate Gaussian forecast errors. We show the combination's superior performance compared to other combination algorithms in a simulation study. In an empirical data example of electricity price spot market forecasts, the combined forecast is improved considerably concerning its calibration. Ready-to-use published code facilitates the application of the proposed combination scheme to further data sets. The chapter is based on joint work with Oliver Grothe (Publ. II).

Chapter 5 continues the work on forecasting and focuses on assessing the ability of models to track changes. Being similarly applicable to measurements and nowcasts, the ability to track changes describes a forecasting, nowcasting, or measurement method's ability to correctly predict the direction of changes in values. Whether a quantity increases or decreases is intuitive and easily interpretable information on a quantity measured over time and might, thus, be of high relevance for decision-making. However, it is only implicitly and indistinctly assessed in traditional accuracy measures or probabilistic scoring rules.

We derive and present new measures and visualization techniques based on the joint development of predicted and observed changes, including their dependence, to assess the ability to track changes solely. The measures and visualization techniques are clearly understandable for decision-makers and amend the current evaluation practice. We reduce noise and minor deviations' effect on the evaluation and provide methods to quantify estimation uncertainty using bootstrap confidence intervals. We briefly review rolling estimates to reveal patterns over time and provide a new visualization technique based on multivariate kernel density estimation to inspect local effects in the ability to track changes. Additionally, we provide advice on evaluating the ability to track changes for probabilistic forecasts and nowcasts. We apply the methods to three large data applications: COVID-19-nowcasting, forecasting patient admission to an emergency department, and invasive and non-invasive blood pressure measurements, illustrating the usage of the proposed evaluation technique and providing further insights on the methods. Again, the code is published so that further applications in the medical and healthcare sector and other fields are straightforward. The chapter is based on joint work with Bolin Liu, Bernd Saugel, and Oliver Grothe (Publ. III).

As a basis for the following work, Chapter 2 reviews literature and concepts that spread through several chapters, such as copula models, relevant for Chapters 3 and 4, and forecasting and its evaluation, for Chapters 4 and 5. Chapters 3, 4, and 5 briefly recapitulate the most important concepts and notation to keep them largely self-contained. Similarly, the individual chapters motivate the research question within the respective fields and review relevant literature. Chapter 6 concludes the thesis. The appendix contains supplementary material for the different chapters, including longer proofs, further results, and additional topics. It is organized along the chapters to which the topics are related, and links to the respective sections are given within the text. The author's publications and conference presentations are listed on page 144 and are referenced within the thesis by "Publ." and "Conf.".

2. Copulas and Forecast Evaluation

Copulas, forecasting, and forecast evaluation are fundamental matters of this thesis. While the following sections provide an introduction, the chapters repeat the basic notation and concepts where necessary. Section 2.1 introduces copulas, including methods for high-dimensional modeling and copula properties. Those are the basis for Chapters 3 and 4: A specific type of copula is decomposed, and copula properties are analyzed in Chapter 3 and Chapter 4 applies copula dependence modeling to the combination of point forecasts. Section 2.2 introduces forecasting and forecast evaluation measures, topics of Chapter 4 and Chapter 5. The chapter is based on and, in large parts, identical to Publ. II (Section 2); excerpts are also taken from Publ. III (Section 2).

2.1. Copulas

Copulas are a tool to model multivariate dependencies by modeling the dependence of the variables separately from the marginals. Through Sklar's theorem (Sklar, 1959) every multivariate cumulative distribution function (CDF) H of random variables $X_1, \dots, X_n$ can be expressed in terms of the individual marginal distributions $F_1, \dots, F_n$ and their dependence encoded in the copula C

$$H(\mathbf{x}) = C\left(F_1(x_1), F_2(x_2), \dots, F_n(x_n)\right), \; \forall \mathbf{x} = (x_1, \dots, x_n) \in \bar{\mathbb{R}}^n, \tag{2.1}$$

where $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. For continuous $F_1, \dots, F_n$, C is unique. If C is a copula of dimension n and $F_1, \dots, F_n$ are distribution functions, then the function H defined by (2.1) is a n-dimensional distribution function with margins $F_1, \dots, F_n$ (Nelsen, 2006, p. 46). As $U_i = F_i(X_i) \sim U(0,1)$ for continuous F_i, C is the n-dimensional CDF of $U_1, \dots, U_n \sim U(0,1)$ and any n-dimensional distribution function with uniform margins is a copula. The inverse of Equation (2.1),

$$C(\mathbf{u}) = H(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)), \mathbf{u} = (u_1, \dots, u_n) \in (0,1)^n, \tag{2.2}$$

gives a method of defining copulas, where F_i^{-1} denotes the inverse of F_i ($i = 1, \dots, n$). For H being a multivariate normal distribution with correlation matrix Σ and $F_1, \dots, F_n$ being standard normal distributions, Equation (2.2) defines the Gaussian copula and encodes the dependence structure of a multivariate Gaussian distribution. Analogously, the t-copula reflects the dependence structure of the multivariate t-distribution and can be defined with a multivariate t-distribution instead of the normal distribution. Because of their similar structure and the use of an elliptical distribution in their formulation, Gaussian and t-copula are called elliptical copulas. The independence copula

$$\Pi(u_1, \dots, u_n) = \prod_{i=1}^{n} u_i,$$

models independent variables. It is obtained by various other parametric copulas with specific parameters, such as $\Sigma = \operatorname{diag}(1, \dots, 1)$ for the Gaussian copula.

While elliptical copulas are radially symmetric, the class of Archimedean copulas provides a method of constructing copulas with asymmetries. The generator function φ and its pseudo-inverse φ^{-1} defines the copula. Nelsen (2006, Sections 4.1 and 4.6) formulates requirements and rigorous definitions of these functions. Well-known examples of Archimedean copulas include the Clayton, Gumbel, and Frank copulas. Table 2.1 shows a summary of their functional forms. Archimedean copulas share essential features through their functional forms, such as the straightforward ability to generate random numbers and extendability to higher dimensions.

Elliptical and Archimedean copulas impose strong conditions on the copulas contained; thus, not all copulas belong to one of those classes. The Farlie-Gumbel-Morgenstern (FGM) copula and the Cuadras-Augé (CA) copula used later within this work, for example, belong to neither of them. They have the functional form of

$$C^{\theta,\mathrm{FGM}}(u,v) = uv + \theta uv(1-u)(1-v) \quad (\theta \in [-1,1]) \text{ and}$$
$$C^{\theta,\mathrm{CA}}(u,v) = \begin{cases} uv^{1-\theta} & ,u \le v \\ u^{1-\theta}v & ,u \ge v \end{cases} \quad (\theta \in [0,1])$$

with a parameter θ controlling the strength of dependence. Figure 2.1 shows scatter plots of the above copulas. While Gaussian, Student t, Frank, and FGM copula are symmetric for 180-degree rotations, the Clayton, Gumbel, and CA copulas are not. Note that copulas can also be modeled in a semiparametric or nonparametric way [see, among others, Gijbels and Mielniczuk, 1990; Chen and Huang, 2007; Nagler, Schellhase, and Czado, 2017 and the checkerboard copula in Chapter 3].

Family	$C_\theta(u_1, u_2)$	$\varphi_\theta(t)$	$\theta \in$
Clayton	$\left[\max\left(u_1^{-\theta} + u_2^{-\theta} - 1, 0\right)\right]^{-1/\theta}$	$\frac{1}{\theta}\left(t^{-\theta} - 1\right)$	$[-1, \infty) \setminus \{0\}$
Frank	$-\frac{1}{\theta}\left(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right)$	$-\ln \frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(-\infty, \infty) \setminus \{0\}$
Gumbel	$\exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{\frac{1}{\theta}}\right)$	$(-\ln t)^\theta$	$[1, \infty)$

Table 2.1.: Summary of exemplary Archimedean copulas in two dimensions (Nelsen, 2006, p. 116 et seq.).

Family	λ_l	λ_u
Gaussian	$\mathbb{1}\{\rho = 1\}$	$\mathbb{1}\{\rho = 1\}$
Clayton (for $\theta > 0$)	$2^{-\frac{1}{\theta}}$	0
Frank	0	0
Gumbel	0	$2 - 2^{\frac{1}{\theta}}$
FGM	0	0
CA	$\mathbb{1}\{\theta = 1\}$	θ

Table 2.2.: Upper and lower tail dependencies for some important bivariate copulas with parameters θ (Nelsen, 2006, p. 215). For the Gaussian copula, ρ refers to the correlation between the two variables, Σ_{12}. The tail dependence for Student's t copula is omitted due to its complexity and is computed, for example, in Embrechts, Lindskog, and Mcneil (2003, Sec. 5.3).

Well-known dependence measures can be computed based solely on the copula; for example, Kendall's concordance measure

$$\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u, v)\, \mathrm{d}C(u, v) - 1$$

and Spearman's

$$\rho(X_1, X_2) = 12 \int_0^1 \int_0^1 C(u, v)\, \mathrm{d}u\, \mathrm{d}v - 3.$$

In addition to the overall dependence measures, there are measures for specific areas of interest; for example, the dependence for U_1 and U_2 approaching one or zero. This is called upper or lower tail dependence. The upper tail dependence λ_u of continuous random variables X and Y with marginal distribution functions F_1 and F_2 and copula C is defined as

$$\lambda_u = \lim_{t \uparrow 1} \mathbb{P}\left[X_2 > F_2^{(-1)}(t) \mid X_1 > F_1^{(-1)}(t)\right] = 2 - \lim_{t \uparrow 1} \frac{1 - C(t,t)}{1-t}$$

and for lower tail dependence analogously. Table 2.2 depicts the tail dependencies for the copulas above. Tail dependencies can be distinguished in the scatter plots of the copulas in Figure 2.1 by the sharp spikes near $(0,0)$ or $(1,1)$.

A typical method for the selection of one of the copulas involves the Akaike information criterion (AIC; Akaike, 1974). It is defined as

$$\mathrm{AIC} = 2(l_{n,\max} - p), \tag{2.3}$$

where $l_{n,\max}$ is the maximized likelihood and p is the total number of marginal and copula parameters. Using the AIC is only fully justified when all marginal and copula parameters are estimated simultaneously. As argued by Hofert, Kojadinovic, Machler, et al. (2018, p. 191 et seq.), the use of AIC is theoretically problematic in staged estimation contexts. Still, alternatives with much higher computation costs do not perform considerably better. Therefore, we use AIC to select models in the following. Note that for copulas with the same number of parameters, for example, Gaussian, Gumbel, Clayton, and Frank, maximizing AIC is identical to choosing the copula with the highest likelihood $l_{n,\max}$.

Parametric copulas, such as Gaussian, t, or Archimedean copulas, usually do not provide enough flexibility for data in high dimensions. They have a closed functional form in higher dimensions with only a few parameters to fit. Pair copula constructions are a more flexible approach, with vine copulas being a prominent representative (Joe, 1996; Bedford and Cooke, 2002; Cooke, Joe, and Aas, 2010). For a brief introduction, we refer to Coblenz et al. (2020), for a textbook introduction to Czado (2019). Vine copulas model the high-dimensional copula as a multilayered tree of bivariate, conditional copulas. An optimal structure of the tree is generally challenging to find, and usually, within the tree, the simplifying assumption is imposed (for definition, discussion, and extensions, see, e.g., Hobæk Haff, Aas, and Frigessi, 2010; Coblenz, 2018, Section 6.1.2; Czado and Nagler, 2022, Section 6). There are various implementations of vine copulas for different programming languages, such as MATLAB (Coblenz, 2021) and Python (Nagler and Vatter, 2021).

2.2. Forecast evaluation

Forecasting is intended to predict the future value of a quantity. There has been an ongoing shift from point forecasts to issuing more information on future development in recent years. Adding uncertainty measures such as the expected root mean squared error (RMSE) gives a general remark on a likely spread of the value, whereas more advanced methods such as prediction interval forecasting or quantile regression give values of prescribed quantiles (Nowotarski and Weron, 2018). To include even more information, forecasters can issue distributional forecasts, that is, the distribution of the unknown quantity given the information at hand (Gneiting and Ranjan, 2013). The forecast distribution includes the information mentioned above on the location and uncertainty of the forecast, such as quantiles and moments. However, assessing distributional forecasts is more challenging than point forecasts, as a distribution is forecasted, but a value is observed. We give a brief overview of point and distributional forecast evaluation in the following.

Various measures are available to assess the accuracy of point forecasts, Gneiting (2011) calls them scoring functions. The RMSE and mean absolute error (MAE) are common scale-dependent scoring functions for point forecasts. For a series of forecasts x_t and observations y_t $(t = 1, \dots, T)$, the RMSE and MAE are defined as

$$\mathrm{RMSE}(x, y) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - y_t)^2} \text{ and}$$
$$\mathrm{MAE}(x, y) = \frac{1}{T}\sum_{t=1}^{T}|x_t - y_t|.$$

RMSE and MAE penalize deviations depending on the scale of the forecast (for a concise analysis, see, Hyndman and Koehler, 2006). There are various extensions and other measures to make the assessment less scale-dependent, for example, by dividing by the observation or the error of a benchmark model. However, due to their simplicity and interpretability, for example, having the same scale as the data, the RMSE and MAE are still widely used. Gneiting (2011) analyze point forecast evaluation measures for the statistical quantity minimizing the function. The mean or median are the scoring-rule-minimizing functionals for the RMSE and MAE, respectively.

The concepts of calibration and sharpness are essential to compare different distributional forecasts. Calibration is the statistical consistency between predicted and actual distribution (Thorarinsdottir, 2013; Gneiting and Ranjan, 2013). Thus, observed realizations are indistinguishable from random draws of the forecast distribution for a calibrated forecast. Sharpness refers to the concentration of the predictive distribution in absolute terms (Gneiting, Balabdaoui, and Raftery, 2007). Murphy and Winkler (1987) and Gneiting, Balabdaoui, and Raftery (2007) conclude that statistical forecasting aims at maximizing the sharpness of the forecast subject to calibration.

In practice, calibration is not verifiable, as the actual distribution remains unknown. Therefore, practically applicable calibration concepts focus on specific aspects of statistical consistency. For real-valued random variables, evaluating the calibration of a forecast typically involves investigating the probability integral transform (PIT) of the forecast (Dawid, 1984; Gneiting and Ranjan, 2013). In the case of a continuous forecast F, the PIT reduces to

$$Z_F = F(X),$$

that is the value of the forecast CDF at the observation. For a continuous random variable X and its CDF F_X, the PIT satisfies $Z_{F_X} \sim U(0,1)$ (Rüschendorf, 2009). Due to this, any CDF forecast F for a random variable X is called a probabilistically calibrated forecast for X if $Z_F \sim U(0,1)$ (Gneiting and Ranjan, 2013).

Practical tools to assess the uniformity of a sequence of PIT values include empirical moments, uniformity tests, and graphical analysis of PIT values. A $U(0,1)$ random variable has support on the whole unit interval and variance $1/12$. The PIT of a forecast is called regular and neutrally dispersed if it fits the support and variance, respectively. If the variance is smaller (larger) than $1/12$, the forecast is overdispersed (underdispersed). Often, it can be beneficial to inspect the values graphically. The kind of deviation from uniformity can lead to the identification of differences between the predicted and actual distributions. Skewed PIT histograms correspond to an incorrect location of the forecast distributions. U-shaped (respectively hump-shaped) histograms suggest a too small (respectively too large) variance of the PIT values (Gneiting and Ranjan, 2013). Figure 2.2 visualizes sample PIT histograms for different forecast distributions for the same underlying distribution.

Noceti, Smith, and Hodges (2003) analyze different statistics to assess the uniformity of the PIT values. They identify the test statistic following Anderson-Darling (AD) and Kolmogorov-Smirnov (KS) (Anderson and Darling, 1954; Neave and Worthington, 1988) as powerful and widely used. The AD test statistic, A^2, and the KS test statistic, D, evaluate the uniformity of PIT values by

$$\begin{aligned} D &= \max\left\{\max_{i\in[T]}\{(i/T) - z_{(i)}\}, \max_{i\in[T]}\{z_{(i)} - (i-1)/T\}\right\} \text{ and} \\ A^2 &= -T - \sum_{i=1}^{T}(2i-1)\left[\ln z_{(i)} + \ln(1 - z_{(T+1-i)})\right]/T \end{aligned}$$

for (sorted) PIT values $z_(1), \ldots, z_{(T)}$. Note that other notions of calibration exist, for example, marginal calibration (Gneiting, Balabdaoui, and Raftery, 2007) and complete calibration (Dawid, 1984).

While the above scoring functions are for point forecasts, one can use scoring rules to simultaneously evaluate the two potentially contrary objectives of sharpness and calibration for probabilistic forecasts. They impose penalties based on the issued forecast and the observed realization and are negatively oriented; lower scores indicate better forecasts. A good scoring rule should encourage issuing the actual distribution, reflected by the notion of propriety. A scoring rule is proper if, in expectation, no forecast yields a lower score than the true distribution. For a mathematical formulation, see, for example, Gneiting and Katzfuss (2014, p. 133).

The logarithmic score (LS) for a density forecast f and realising observation y is defined as

$$S_{LS}(f, y) = -\log f(y). \tag{2.4}$$

It dates back to at least Good (1952). In regression evaluation, the logarithmic score is equivalent to the log-likelihood score. It can be computed for continuous and discrete forecast distributions. However, it can be infinite if the forecast support does not include the observation, and the probability distribution function (PDF) has to exist and be computed.

The continuous ranked probability score (CRPS), another proper scoring rule, is a generalized version of the MAE and is, for a forecast CDF F, computed by

$$S_{CRPS}(F, y) = \int_{\mathbb{R}} (F(x) - \mathbb{1}\{x > y\})^2 dx.$$

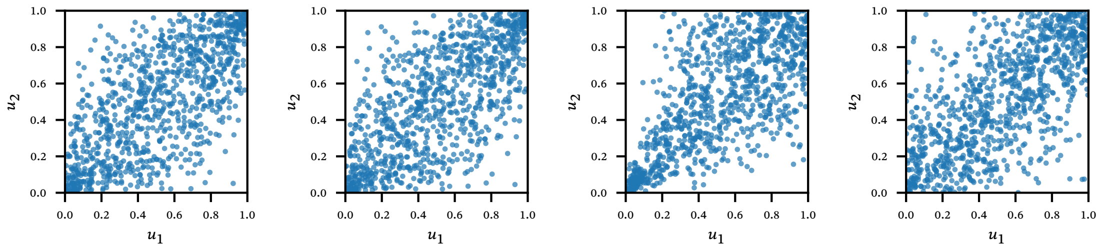
It can also be expressed as an integral over the quantile loss and is approximated by the scaled average pinball loss over different quantiles (Gneiting and Katzfuss, 2014; Hong et al., 2016). For a series of forecasts F_t and observations y_t $(t = 1, \dots, T)$, the average score S_T^F assesses the overall performance of the forecast, that is, for a proper scoring rule S,

$$S_T^F = \frac{1}{T} \sum_{t=1}^{T} S(F_t, y_t).$$

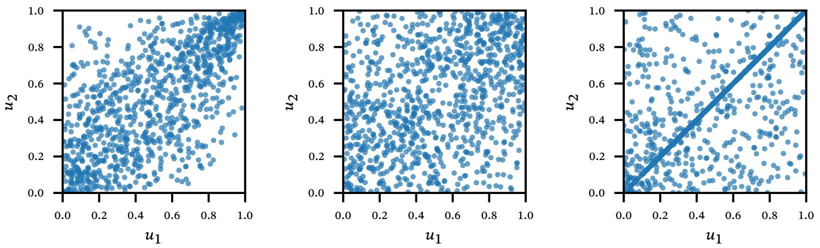
Dichotomous forecasts are an important facet of forecasting, focused on predicting binary events. There, the notation and methods are more straightforward, as only two outcomes are possible, $y = 0$ and $y = 1$, and only the probability p for $y = 1$ is forecasted. The Brier score (BS) is a widely used scoring rule for dichotomous probabilistic forecasts (Brier, 1950) and is defined as

$$S_{\text{BS}}(p, y) = (p - y)^2.$$

Again, lower values indicate a better forecast. Additional graphical calibration assessments are available for dichotomous forecasts, such as reliability diagrams (see Section 5.2.6).

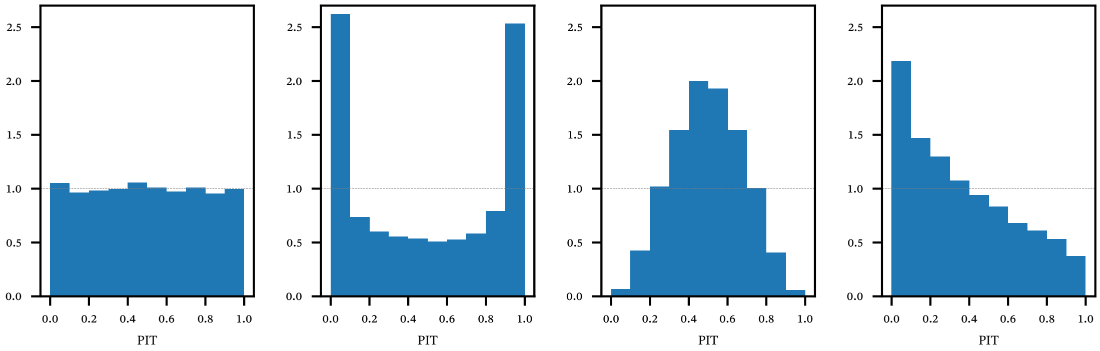


(a) Gaussian **(b)** Student t **(c)** Clayton **(d)** Frank



(e) Gumbel **(f)** FGM, $\tau = 2/9$ **(g)** CA

Figure 2.1.: Scatter plots with 1 000 points of the copulas of Section 2.1 using the R package `copula` (Hofert, Kojadinovic, Maechler, et al., 2023, version 1.1-3) and Mai and Scherer (2012, Example 1.8) for the CA copula. Apart from the FGM copula, the copulas have parameters such that $\tau = 0.5$.



(a) Calibrated forecast.

(b) Underdispersed forecast.

(c) Overdispersed forecast.

(d) Incorrectly-located forecast.

Figure 2.2.: PIT histograms visualizing deviations from calibration. The left histogram is uniformly distributed, indicating a calibrated forecast. The second and third histograms are U-shaped and hump-shaped, revealing an incorrect forecast variance. The right histogram is skewed, suggesting a wrong forecast location. The data is generated using a $N(0,1)$ distribution and forecasts $N(0,1)$, $N(0,0.25)$, $N(0,4)$, and $N(0.5,1)$, respectively.

3. Checkerboard Copula Decomposition

This chapter analyzes, visualizes, and applies singular value decomposition to doubly stochastic matrices representing checkerboard copulas and, thus, provides structural insights into the dependence structure of checkerboard copulas. The chapter is based on joint work with Oliver Grothe (Publ. I) and was presented at the *CMStatistics 2022* (Conf. IV). Dimensionality reduction techniques were also applied to wind speed data to identify clusters of similarity in Publ. IV by the author of this thesis. However, this manuscript is not part of the thesis.

The chapter is structured as follows. Section 3.1 motivates the decomposition of checkerboard copulas and reviews relevant literature. Section 3.2 describes the approach, including the extensions for comonotonicity-like copulas, non-copula truncations, and the computation of statistical functionals. We analyze the difference between decomposed copulas and draw the connection between discrete (checkerboard) and continuous decompositions. Section 3.3 provides the resulting decompositions for well-known copulas of different complexities and symmetric and asymmetric dependencies. We use the graphical tools of correspondence analysis to interpret the two-dimensional graphs of copulas and apply the graphical tools to an empirical checkerboard from data on the fuel injection spray characteristics of jet engines in Section 3.4. Section 3.5 concludes the chapter. The code for this chapter is available at `https://github.com/jo-rie/cca`.

3.1. Introduction

A copula contains information on the likelihood of joint occurrence of random variables on their intrinsic quantile scale (for preliminaries on copulas, see Section 2.1). For two-dimensional vectors, the copula thus encodes a possibly large or infinite two-dimensional frequency table specifying the joint likelihood of the transformed random vector. If finite, square, and scaled appropriately, this table can be interpreted as a checkerboard copula (X. Li et al., 1997; Durante and Sempi, 2015). The tables are generally large and may contain redundant information, and assessing the incorporated dependence information is not straightforward. We apply well-known decomposition and dimensionality reduction techniques of high-dimensional data analysis to this table, thereby decomposing the copula. The decomposition opens a wide range of further analyses, for example, to compute and analyze copula characteristics, plot meaningful two-dimensional plots of the copula, or build simpler, reasonable approximations of complicated dependency structures. Through the low-rank approximation, one can drastically decrease the number of items to be stored compared to the full checkerboard matrix, that is, the square of the lattice size.

Checkerboard copulas can be obtained from empirical data or, for example, by discretizing continuous copulas (see, e.g., Kolesárová et al., 2006; Durrleman, Nikeghbali, and Roncalli, 2000). In either way, the copula frequency table is a doubly stochastic matrix. Taking the doubly stochastic matrix, we apply correspondence analysis methods that are mainly based on singular value decomposition (SVD).

Additive decompositions of copulas using variable-specific functions already exist in the literature, but only for continuous representations. Continuous decompositions are considered, for example, in Mesiar and Najjari (2014) or Rodríguez-Lallena (2004) for the generation of new copulas and in C. M. Cuadras (2015) for the decomposition of copulas. The checkerboard case differs from existing approaches and yields different decompositions, as discussed in Section 3.3.3. Durrleman, Nikeghbali, and Roncalli (2000) mention SVD of checkerboard copulas but do not go into detail, and C. M. Cuadras (2002) considers discrete and continuous decompositions of general bivariate distributions. In contrast to these studies, we concentrate on the decomposition of doubly stochastic matrices that represent checkerboard copulas, allowing us to focus on the features of copulas. We provide formulas for important statistical functionals, including Spearman's ρ, Kendall's τ, and Pearson's ϕ^2. Through the Frobenius distance between the matrices, we express the similarity of two checkerboard copulas in terms of their ϕ^2.

Using the standard kit of correspondence analysis has obstacles for some copulas. Copulas such as the comonotonicity copula are costly to represent in standard SVD, as the corresponding frequency matrix is similar to an identity matrix, having full rank and many equally large singular vectors. Thus, approximations by truncating the SVD series have slowly decaying errors. Therefore, we propose to use a monotonicity anchored representation (adapted from Greenacre, 1984 and Kazmierczak, 1978), taking into account the independence and comonotonicity-like characteristics. This representation does not change the singular vectors for symmetric copulas but can considerably reduce the approximation error. Also, the obtained truncations are not necessarily valid checkerboard copulas, as negative values can occur. We provide an algorithm that yields the nearest valid copula for the Frobenius norm by generalizing an algorithm by Zass and Shashua (2007) and thus maps the obtained truncated (not doubly stochastic) matrix to the nearest doubly stochastic matrix. While this chapter is focused on the Frobenius error norm, we remark on using the Hellinger distance in Appendix A.2.

The frequency table decomposition corresponds to a decomposition of the discretized copula PDF. Section 3.2.6 links our analysis to continuous decompositions, as in the literature on copula generation and continuous copula decomposition, and to cumulative distribution function (CDF) decompositions. Through the decomposition, we motivate a decomposition of the Gaussian copula into transformed Hermite polynomials.

Thus, this chapter makes several contributions. We define the decomposition of checkerboard copulas and give extensions of the approach for comonotonicity-like copulas and non-copula truncations. We link the approach to important existing copula concepts such as dependence measures, similarities of copulas, and continuous decompositions of copulas. We derive characteristics of the graphs obtained by the approach and thus provide a new method of graphical copula representations. Lastly, we apply the approach to theoretical copula families of various complexities and an empirical data example from the engineering context.

3.2. Checkerboard copula decomposition and its characteristics

This section examines the singular value decomposition (SVD) and its truncation for checkerboard copulas, that is, doubly stochastic matrices. We introduce some notation in Section 3.2.1 and then define the truncated decomposition, including a monotonicity anchored representation (MAR) that accounts for dependencies similar to comonotonicity in Section 3.2.2. To correct negative matrix elements in the truncated representation, Section 3.2.3 formulates an algorithm to approximate the truncation by a doubly stochastic matrix. Sections 3.2.4 and 3.2.5 derive statistical functionals and similarity measures using the decomposition. Section 3.2.6 links the decompositions of continuous copulas and their discretized counterparts.

3.2.1. Doubly stochastic matrices from bivariate copulas

Let X and Y be random variables with CDF $F_{X,Y}$ and marginal CDFs F_X and F_Y, respectively. Through the well-known theorem of Sklar (1959) the multivariate CDF $F_{X,Y}(x,y)$ can be decomposed as

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y))$$

whereby the copula C encodes the dependence structure of X and Y. The copula C can also be seen as CDF of $F_X(X)$ and $F_Y(Y)$ and, thus, has the properties of a multivariate CDF with uniform margins, provided that X and Y are continuous. While the copula is unique for continuous random variables, it is only uniquely identified on the image of F_X and F_Y, respectively, in the discrete case. Further information on copulas is available in Section 2.1.

A checkerboard copula (X. Li et al., 1997) is a special type of copula that assumes a uniform mass within the squares of an evenly spaced lattice $I^n \times I^n$ ($I^n = \{0, 1/n, \dots, 1\}$). Checkerboard copulas can be computed from empirical data or by the discretization of continuous copulas. The discretization facilitates the (asymptotic) comparison of discrete and continuous characteristics.

Any continuous copula C defines a doubly stochastic matrix $\mathbf{C}^n$ on the grid $I^n \times I^n$ by evaluating C on $I^n \times I^n$, that is, $\mathcal{C} := C(u, v)$ $(u, v \in I^n)$ (Mayor, Suner, and Torrens, 2005; Kolesárová et al., 2006) and computing

$$\mathbf{C}^n_{i,j} = n\Big[\mathcal{C}\Big(\frac{i}{n}, \frac{j}{n}\Big) - \mathcal{C}\Big(\frac{i-1}{n}, \frac{j}{n}\Big) - \mathcal{C}\Big(\frac{i}{n}, \frac{j-1}{n}\Big) + \mathcal{C}\Big(\frac{i-1}{n}, \frac{j-1}{n}\Big)\Big] \quad \text{for } i, j = 1, \dots, n. \tag{3.1}$$

The properties of $\mathbf{C}^n$ follow from the copula properties of C:

1. $\mathbf{C}^n$ has nonnegative entries as the defining Equation (3.1) coincides with a scaled version of the 2-volume of the copula, which is nonnegative.

2. From $C(u, 1) = C(1, u) = u$ $(u \in [0, 1])$ and $C(u, 0) = C(0, u) = 0$ $(u \in [0, 1])$ follows for $j \in [n]$

   $$\sum_{i=1}^{n} \mathbf{C}^n_{i,j} = \sum_{i=1}^{n} n\Big[\mathcal{C}\Big(\frac{i}{n}, \frac{j}{n}\Big) - \mathcal{C}\Big(\frac{i-1}{n}, \frac{j}{n}\Big) - \mathcal{C}\Big(\frac{i}{n}, \frac{j-1}{n}\Big) + \mathcal{C}\Big(\frac{i-1}{n}, \frac{j-1}{n}\Big)\Big] = 1.$$

   An analogous computation shows $\sum_{j=1}^{n} \mathbf{C}^n_{i,j} = 1$ $(i \in [n])$ and, thus, the row and column sums of $\mathbf{C}^n$ are 1.

The matrix $\mathbf{C}^n$ is by construction square, together with 1. and 2., a doubly stochastic matrix. The element $\mathbf{C}^n_{i,j}$ ($i, j \in [n]$, where $[n] := \{1, 2, \dots, n\}$) corresponds to the density of the checkerboard copula,

$$\hat{c}(u, v) = n\mathbf{C}^n_{\lceil u \cdot n \rceil, \lceil v \cdot n \rceil} = n \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{C}^n_{i,j} \mathbb{1}\Big\{u \in \Big[\frac{i-1}{n}, \frac{i}{n}\Big)\Big\} \mathbb{1}\Big\{v \in \Big[\frac{j-1}{n}, \frac{j}{n}\Big)\Big\} \quad (u, v \in \mathbb{R}), \tag{3.2}$$

in the rectangle

$$R_{i,j} := \Big[\frac{i-1}{n}, \frac{i}{n}\Big] \times \Big[\frac{j-1}{n}, \frac{j}{n}\Big].$$

We denote by $\mathbb{1}\{\cdot\}$ the indicator function. Thus, $\mathbf{C}^n_{i,j}$ can be interpreted naturally as a table of the likelihood of occurrence in the copula domain. Integration over Equation (3.2) yields a checkerboard approximation of the copula CDF $\hat{C}$ and the conditional CDF $\hat{C}_{u|V=v}$. The discretizations approximate the copula C with increasing n, and every copula C is the limit of its discretizations for $n \to \infty$ (see Kolesárová et al., 2006, Theorem 1).

3.2.2. Singular value decomposition and monotonicity anchored representation

Having the table of likelihood of occurrence, $\mathbf{C}^n$, correspondence analysis can be applied to the matrix $\mathbf{C}^n$ to analyze the structural properties of the matrix. Correspondence analysis uses the SVD to compute low-dimensional approximations of the matrix. To this end, the SVD is truncated, yielding the nearest matrix of the specified rank according to the Frobenius or the spectral norm (Mirsky, 1960). In correspondence analysis, the matrix $\mathbf{C}^n$ is usually centered, and some scaling is applied to rows and columns to account for the sum differences of the rows or columns (Greenacre, 1984). In the case of $\mathbf{C}^n$, the centering step is implemented by subtracting the matrix $\mathbf{\Pi}^n \coloneqq n^{-1}\mathbb{1}\mathbb{1}^\top$ from $\mathbf{C}^n$, where $\mathbb{1}$ is the vector of ones of suitable dimension. We denote this by

$$\mathbf{A}^n = G(\mathbf{C}^n) \coloneqq \mathbf{C}^n - \mathbf{\Pi}^n.$$

Note that $\frac{1}{\sqrt{n}}\mathbb{1}$ is a (left and right) singular vector of $\mathbf{C}^n$ with singular value 1, whereby 1 is the largest singular value for doubly stochastic matrices (Perfect and Mirsky, 1965), and, thus, the rank of $\mathbf{A}^n$ is at most $n-1$. We denote the SVD of $\mathbf{A}^n$ by

$$\mathbf{A}^n = \mathbf{U}\mathbf{S}\mathbf{V}^\top, \text{with } \mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_n), \mathbf{S} = \operatorname{diag}(s_1, \dots, s_n), \mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n), \tag{3.3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and the singular values s_k are in $[0,1]$ and are sorted in descending order as usual.

The decomposition in Equation (3.3) may be truncated by using only the $n^\star \leq n$ largest singular values of s, and the corresponding first $n^\star$ columns of $\mathbf{U}$ and $\mathbf{V}$:

$$T_{n^\star}(\mathbf{A}^n) \coloneqq \mathbf{U}_{:,1:n^\star}\mathbf{S}_{1:n^\star,1:n^\star}\left(\mathbf{V}_{:,1:n^\star}\right)^\top,$$

where we will use $T_{n^\star}(\cdot)$ as a truncation operator of the argument's SVD in the following. The truncated $T_{n^\star}(\mathbf{A}^n)$ yields an approximation of $\mathbf{C}^n$ by applying the inverse function of G, that is,

$$G^{-1}(T_{n^\star}(\mathbf{A}^n)) = T_{n^\star}(\mathbf{A}^n) + \mathbf{\Pi}^n.$$

The truncated SVD yields low-rank approximations with small errors for matrices with a few large and many small (or zero) singular values. We will show examples in Section 3.3. However, in the copula context, many copulas share characteristics with the comonotonicity copula, an identity matrix with singular value 1 with multiplicity n, and, thus, high approximation errors for small-rank representations. To “remove” the comonotonicity copula characteristics before applying the SVD, we suggest transforming the matrix $\mathbf{C}^n$ so that we account for high frequencies on the diagonal of the matrix and thus the monotone dependence structures. We denote this transformation by $G_{\mathrm{MAR}}(\cdot)$ and call it monotonicity anchored representation (MAR). As we argue in Lemma 1, through this representation, the singular vectors do not change for symmetric copulas, but the series of singular values s_k decreases faster, leading to better low-rank approximations. The MAR is given by

$$\tilde{\mathbf{A}}^n = G_{\mathrm{MAR}}(\mathbf{C}^n, \eta) := \mathbf{C}^n + \eta I_n - (1+\eta)\frac{1}{n}\mathbb{1}\mathbb{1}^\top, \tag{3.4}$$

with $\eta \in \mathbb{R}$ and I_n denoting the n-by-n identity matrix. The centering step is implemented by the last summand $-(1+\eta)\frac{1}{n}\mathbb{1}\mathbb{1}^\top$, that is, for $\eta \in \mathbb{R}$

$$\begin{aligned}(\tilde{\mathbf{A}}^n)\mathbb{1} &= \mathbf{C}^n\mathbb{1} + \eta I_n\mathbb{1} - \frac{1+\eta}{n}\mathbb{1}\mathbb{1}^\top\mathbb{1} \\ &= \mathbb{1} + \eta\mathbb{1} - \frac{1+\eta}{n}\mathbb{1}\cdot n = 0\cdot\mathbb{1}\end{aligned}$$

and analogously for $(\tilde{\mathbf{A}}^n)^\top\mathbb{1}$. The approach also suits strong negative dependence by rotating the copula first. A similar transformation to G_{MAR} can be found in Kazmierczak (1978), in Greenacre (1984, Section 8.6) formulated in the context of frequency tables. Unlike Greenacre (1984) who use two parameters and demand them to be chosen such that $(\tilde{\mathbf{A}}^n)_{i,j} \geq 0\ \forall (i,j)$ for merely illustrative purposes, we do not require this additional restriction here. We scale the last summand by $1/n$ to preserve the margins shown above.

Note that $G(\mathbf{C}^n)$ is nested within the MAR by setting $\eta = 0$. Later, the parameter η is calculated such that the Frobenius distance between the inverse transformed version of $\tilde{\mathbf{A}}^n$, denoted by $G^{-1}_{\mathrm{MAR}}(T_{n^\star}(\tilde{\mathbf{A}}^n), \eta)$, and (the original) $\mathbf{A}^n$ is minimized.

Analogously to the above notation, we denote the SVD of $\tilde{\mathbf{A}}^n$ by

$$\tilde{\mathbf{A}}^n = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^\top, \text{with } \tilde{\mathbf{U}} = (\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_n), \tilde{\mathbf{S}} = \mathrm{diag}(\tilde{s}_1, \dots, \tilde{s}_n), \tilde{\mathbf{V}} = (\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_n).$$

The following lemma shows that singular values and vectors of $\tilde{\mathbf{A}}^n$ and $\mathbf{A}^n$ are closely connected, provided that $\mathbf{C}^n$ is symmetric.

Lemma 1. *For the SVDs of* $\tilde{\mathbf{A}}^n = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^\top$ *and* $\mathbf{A}^n = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ *of symmetric* $\mathbf{C}^n$*, that is, for* $\mathbf{U} = \mathbf{V}$*, it holds that*

$$\begin{aligned} &\mathbf{u}_k = \tilde{\mathbf{u}}_k = \mathbf{v}_k = \tilde{\mathbf{v}}_k \quad k = 1, 2, \dots \textit{ and} \\ &\tilde{s}_k = (s_k + \eta) \quad k = 1, 2, \dots, n. \end{aligned}$$

Proof. From $\mathbf{v}_1, \dots, \mathbf{v}_{n-1} \perp \mathbb{1}$ follows for $k = 1, \dots, n-1$

$$\begin{aligned} \tilde{\mathbf{A}}^n \cdot \mathbf{v}_k &= \left[\mathbf{C}^n + \eta I_n + \frac{1+\eta}{n}\mathbb{1}\mathbb{1}^\top\right] \cdot \mathbf{v}_k \\ &= \mathbf{C}^n\mathbf{v}_k + \eta I_n \mathbf{v}_k + \frac{1+\eta}{n}\mathbb{1}\, \underbrace{\mathbb{1}^\top\mathbf{v}_k}_{=0} \\ &= s_k\mathbf{u}_k + \eta\mathbf{v}_k. \end{aligned}$$

For symmetric matrices, thus,

$$\tilde{\mathbf{A}}^n \cdot \mathbf{v}_k = (s_k + \eta)\mathbf{v}_k.$$

□

For asymmetric $\mathbf{C}^n$, the singular values and vectors of $\mathbf{A}^n$ and $\tilde{\mathbf{A}}^n$ differ. Lemma 1 yields the $n^\star$-truncated representation of $\tilde{\mathbf{A}}^n$,

$$\begin{aligned} T_{n^\star}(\tilde{\mathbf{A}}^n) &= \sum_{k=1}^{n^\star} \tilde{\mathbf{u}}_k \tilde{s}_k \tilde{\mathbf{v}}_k^\top \\ &\overset{(*)}{=} \sum_{k=1}^{n^\star} (s_k + \eta)\mathbf{u}_k\mathbf{u}_k^\top \qquad |(*):\ \text{for symmetric } \mathbf{C}^n \end{aligned}$$

and thus after back transformation of Equation (3.4)

$$\begin{aligned} G_{\mathrm{MAR}}^{-1}\big(T_{n^\star}(\tilde{\mathbf{A}}^n), \eta\big) &= \left(\sum_{k=1}^{n^\star} \tilde{\mathbf{u}}_k \tilde{s}_k \tilde{\mathbf{v}}_k^\top - \eta I_n + (1+\eta)\frac{1}{n}\mathbb{1}\mathbb{1}^\top\right) && (3.5) \\ &\overset{(*)}{=} \left(\sum_{k=1}^{n^\star} \mathbf{u}_k\mathbf{u}_k^\top (s_k + \eta) - \eta I_n + (1+\eta)\frac{1}{n}\mathbb{1}\mathbb{1}^\top\right) && (3.6) \\ &= \sum_{k=1}^{n^\star} \mathbf{u}_k\mathbf{u}_k^\top (s_k + \eta) - \eta I_n + (1+\eta)\mathbf{\Pi}^n && (3.7) \end{aligned}$$

and Equations (3.6) and (3.7) are, again, only valid for symmetric copulas. PDF and CDF can be computed using $G_{\mathrm{MAR}}^{-1}\big(T_{n^\star}(\tilde{\mathbf{A}}^n), \eta\big)$ analogously to Equation (3.2). The parameter η of Equation (3.4) can be determined by minimizing some error norm of interest. For example, we calculate the fraction η that minimizes the residual inertia (thus, Frobenius error) for a specified approximation of rank $n^\star$ by

$$\eta^*(\mathbf{C}^n, n^\star) = \operatorname*{arg\,min}_{\eta\in\mathbb{R}} \left\| \sum_{k=1}^{n^\star} \tilde{\mathbf{u}}_k \tilde{\mathbf{v}}_k^\top(\tilde{s}_k) - \eta I_n + (1+\eta)\boldsymbol{\Pi}^n - \mathbf{C}^n \right\|_F^2 . \tag{3.8}$$

For a symmetric matrix $\mathbf{C}^n$ and an approximation of dimension $n^\star$, this yields

$$\eta^*(\mathbf{C}^n, n^\star) = \operatorname*{arg\,min}_{\eta\in\mathbb{R}} \sum_{k=n^\star+1}^{n} (s_k + \eta)^2 = - \sum_{k=n^\star+1}^{n} \frac{s_k}{n - n^\star}. \tag{3.9}$$

For asymmetric matrices $\mathbf{C}^n$, the problem in Equation (3.8) can be solved numerically. The simulations in Section 3.3 examine the choices of η and the resulting matrices $\tilde{\mathbf{A}}^n$.

3.2.3. Ensuring double stochasticity of truncations

As noted above, truncations of the SVD can yield low error approximations with considerably lower rank matrices. In general, truncations of the SVD are not necessarily doubly stochastic matrices. Truncations keep the property of having row and column sums of one as the singular vectors $\mathbf{u}_k$ and $\mathbf{v}_k$, or $\tilde{\mathbf{u}}_k$ and $\tilde{\mathbf{v}}_k$, respectively, are perpendicular to $\mathbb{1}$ for $k \in [n-1]$, but the truncations do not necessarily have nonnegative elements. One can approximate the truncation by the nearest, doubly stochastic matrix to ensure nonnegativity. This step does not increase the complexity of the representation, as it does not include any information other than the truncated matrix. We give a general idea of the algorithms for symmetric and asymmetric matrices here; they are more specifically described in Appendix A.1.

Zass and Shashua (2007) propose an algorithm to find the nearest doubly stochastic matrix for any symmetric matrix $\mathbf{A}^{\mathrm{sym}} \in \{\mathbf{A} \in \mathbb{R}^{n\times n} : \mathbf{A} = \mathbf{A}^\top\}$ according to the Frobenius norm, that is, a solution to the problem $P(\mathbf{A})$, with

$$\begin{aligned} P(\mathbf{A}) = \operatorname*{arg\,min}_{\mathbf{B}\,\in\,\mathbb{R}^{n\times n}} \quad & ||\mathbf{A} - \mathbf{B}||_F^2 \\ \text{s.t.} \quad & \mathbf{B}\mathbb{1} = \mathbb{1}, \\ & \mathbf{B}^\top\mathbb{1} = \mathbb{1}, \\ & \mathbf{B}_{i,j} \geq 0, \quad \forall i, j \in [n]. \end{aligned}$$

According to Zass and Shashua (2007), $P(\mathbf{A})$ can be solved for symmetric $\mathbf{A}$ iteratively by solving two problems, that is, $P(\mathbf{A}) = P_2 P_1 P_2 \ldots P_1(\mathbf{A})$, with

$$\begin{aligned} P_1(\mathbf{A}_1) = \underset{\mathbf{B} \in \mathbb{R}^{n \times n}}{\arg\min} \quad & \|\mathbf{A}_1 - \mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{B}\mathbb{1} = \mathbb{1}, \\ & \mathbf{B}^\top \mathbb{1} = \mathbb{1} \end{aligned}$$

and

$$\begin{aligned} P_2(\mathbf{A}_2) = \underset{\mathbf{B} \in \mathbb{R}^{n \times n}}{\arg\min} \quad & \|\mathbf{A}_2 - \mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{B}_{i,j} \geq 0, \quad \forall i, j \in [n]. \end{aligned}$$

$\mathbf{A}_1$ and $\mathbf{A}_2$ refer to iterative solutions of P_2 and P_1, respectively. Algorithm 1 formulates the algorithm explicitly. P_1 and P_2 have closed-form solutions and calculations for the solution of P_1 and P_2 are carried out in A.1. For asymmetric matrices, Algorithm 1 maintains its general form but must be complemented by a deflection component (Dykstra, 1983). The resulting Algorithm 2 is shown in A.1. Note that there are algorithms for approximations with a particular interest in keeping the sparsity structure of $\mathbf{A}$. Rontsis and Goulart (2020) formulates an algorithm for a slightly modified problem that accounts for the sparsity of the matrix $\mathbf{A}$ based on the alternate direction method of multipliers and applies to symmetric and asymmetric matrices $\mathbf{A}$. Sparsity thereby refers to zero entries in the matrix $\mathbf{A}$. In general, the SVD approximations typically contain many small, nonzero values, and thus the approximation does not benefit from exploiting the sparsity structure.

3.2.4. Statistical functionals of decompositions and truncations

Various statistical properties can be computed using the decomposition, including dependence measures such as Kendall's τ, Spearman's ρ_S, and Pearson's ϕ^2. We start by expressing the well-known dependency measures Kendall's τ and Spearman's ρ_S through the decomposition (for a definition, see Section 2.1). The structure of both measures inherits the SVD structure of the checkerboard copula, and thus, copulas with many high singular values tend to have a measure representation with many terms, subject to the direction of the singular vectors. The empirical computation of dependence measures using the checkerboard copula itself might be inefficient, and the use of a low-rank approximation might be more robust. We leave the asymptotics of the decomposed measures for further research.

Durrleman, Nikeghbali, and Roncalli (2000) show that for checkerboard copulas Kendall's τ and Spearman's ρ_S can be computed by

$$\begin{aligned}\tau(\mathbf{C}^n) &:= 1 - \frac{1}{n^2}\operatorname{trace}\left(\mathbf{E}\mathbf{C}^n\mathbf{E}(\mathbf{C}^n)^\top\right), \text{ and}\\ \rho_S(\mathbf{C}^n) &:= \frac{3}{n}\operatorname{trace}\left(\mathbf{\Omega}\mathbf{C}^n\right) - 3\end{aligned}$$

with $\mathbf{E} \in \mathbb{R}^{n\times n}$ and

$$\mathbf{E}_{i,j} = \begin{cases}1, & \text{if } i = j,\\ 2, & \text{if } i > j,\\ 0, & \text{if } i < j\end{cases}$$

and $\mathbf{\Omega} \in \mathbb{R}^{n\times n}$ where

$$\mathbf{\Omega} := \breve{\omega}\breve{\omega}^\top, \text{ with } \breve{\omega} = \frac{1}{n}\big(2n - 2\cdot 1 + 1, 2n - 2\cdot 2 + 1, \dots\big)^\top.$$

Let, as in Section 3.2.2, the SVD of the centered $\mathbf{C}^n$ be denoted by $\mathbf{A}^n = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, and additionally $\mathbf{u}_0 = \mathbf{v}_0 = \frac{1}{\sqrt{n}}\mathbb{1}^\top$ and $s_0 = 1$, such that $\mathbf{C}^n = \sum_{k=0}^{n-1}\mathbf{u}_k s_k \mathbf{v}_k$. Then follows Spearman's ρ_S with $\omega := ||\breve{\omega}||^{-1}\breve{\omega}$ as

$$\begin{aligned}\rho_S(\mathbf{C}^n) &= \frac{3}{n}\operatorname{trace}\left(\mathbf{\Omega}\sum_{k=0}^{n-1}\mathbf{u}_k s_k \mathbf{v}_k^\top\right) - 3\\ &= (4 - 1/n^2)\sum_{k=1}^{n-1} s_k\langle\omega, \mathbf{u}_k\rangle\langle\mathbf{v}_k, \omega\rangle - 3\\ &= (4 - 1/n^2)\sum_{k=1}^{n-1} s_k\langle\omega, \mathbf{u}_k\rangle\langle\mathbf{v}_k, \omega\rangle \qquad (3.10)\end{aligned}$$

and for Kendall's τ

$$\begin{aligned}\tau(\mathbf{C}^n) &:= 1 - \frac{1}{n^2}\operatorname{trace}\left(\mathbf{E}\mathbf{C}^n\mathbf{E}(\mathbf{C}^n)^\top\right)\\ &= 1 - \frac{1}{n^2}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1} s_k s_l\langle\mathbf{u}_l, \mathbf{E}\mathbf{u}_k\rangle\langle\mathbf{v}_k, \mathbf{E}\mathbf{v}_l\rangle. \qquad (3.11)\end{aligned}$$

Details of the calculations are provided in Appendix A.3. Both dependence measures can also be put in terms of the MAR, for example,

$$\rho_S(\mathbf{C}^n) = (4 - \tfrac{1}{n^2})\left[\sum_{k=1}^{n-1} \tilde{s}_k \langle \tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle\langle \tilde{\mathbf{v}}_k, \boldsymbol{\omega}\rangle - (4 - \tfrac{1}{n^2})\eta + \tfrac{3}{n}(1+\eta)\right] - 3 \tag{3.12}$$

$$= (4 - \tfrac{1}{n^2})\left[\sum_{k=1}^{n-1} (s_k + \eta)\langle \tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle^2 - (4 - \tfrac{1}{n^2})\eta + \tfrac{3}{n}(1+\eta)\right] - 3 \quad |(\text{for symmetric } \mathbf{C}^n). \tag{3.13}$$

Note that η refers to the MAR coefficient of Equation (3.4). The calculations are performed in Appendix A.3. The decompositions of ρ_S and τ are both based on the singular-value-weighted sum of scalar products containing the singular vectors. Thus, they account for the importance and the direction of the component. For τ, the projection vector contains transformations of the other singular vectors, reflecting the integral's measure being the copula CDF. The representations in Equations (3.10) and (3.11) yield approximations for Spearman's ρ_S and Kendall's τ in terms of the truncated representations, that is,

$$\hat{\rho}_{S_{n^\star}}(\mathbf{C}^n) := (4 - 1/n^2) \sum_{k=1}^{n^\star} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle \text{ and} \tag{3.14}$$

$$\hat{\tau}_{n^\star}(\mathbf{C}^n) := 1 - \frac{1}{n^2} \sum_{k=0}^{n^\star}\sum_{l=0}^{n^\star} s_k s_l \langle \mathbf{u}_l, E\mathbf{u}_k\rangle\langle \mathbf{v}_k, E\mathbf{v}_l\rangle. \tag{3.15}$$

In the SVD representation, Pearson's ϕ^2 boils down to the total inertia of the copula from independence (Schmid et al., 2010, p. 223)

$$\begin{aligned}
\phi^2(\mathbf{C}^n) &= \int_0^1 \int_0^1 \hat{c}^2(u, v)\, \mathrm{d}u\, \mathrm{d}v - 1 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} n^2 \left(\mathbf{C}^n_{ij}\right)^2 \frac{1}{n^2} - 1 \\
&= ||\mathbf{A}^n||_F^2 \\
&= \sum_{k=1}^{n} s_k^2,
\end{aligned}$$

where $s_1, s_2, \ldots$ are the singular values of the centered $\mathbf{A}^n$. Note that this is proportional to Pearson's χ^2 statistic for testing independence in an empirical contingency table (for Pearson's χ^2 statistic in the copula context see, e.g., Savu and Trede, 2008). The truncated representations ϕ^2 is

$$\hat{\phi^2}_{n^\star}(\mathbf{C}^n) = \sum_{k=1}^{n^\star} s_k^2 = \phi^2(G^{-1}(T_{n^\star}(\mathbf{A}^n))).$$

In correspondence analysis, the ratio of the total inertia of approximation and the original matrix is a standard measure for the approximation's goodness of fit, that is,

$$\frac{\sum_{k=1}^{n^\star} s_k^2}{\sum_{k=1}^{n} s_k^2} = \frac{\hat{\phi^2}_{n^\star}(\mathbf{C}^n)}{\phi^2(\mathbf{C}^n)}.$$

Counting the number of nonzero singular values yields an estimate of the dimensionality of the representation, that is,

$$\phi_g(\mathbf{C}^n) = |\{s_k : s_k > 0, k \in [n]\}| = \operatorname{rank}(\mathbf{A}^n).$$

It counts the dimensions needed to model all information in $\mathbf{C}^n$ and does not consider the strength of the information, in contrast to, for example, Pearson's ϕ^2. C. M. Cuadras and Díaz (2012) calls this the geometric dimension of a copula. For discretizations of a continuous copula, the values of ϕ^2 and ϕ_g depend on the grid resolution n and are, therefore, the discretized copula's properties and not of the continuous counterpart. The following lemma formulates this explicitly, and the example in Section 3.2.6 shows that the geometric dimension can decrease with increasing grid size when the grid sizes are not nested.

Lemma 2. *Let $n_1 > n_2$, with $n_1 = n_2 m$ ($m \in \mathbb{N}$), be grid resolutions of the discretizations $\mathbf{C}^{n_1}$ and $\mathbf{C}^{n_2}$ of a copula C. Then,*

$$\phi_g(\mathbf{C}^{n_1}) \geq \phi_g(\mathbf{C}^{n_2}).$$

Proof. Let M_1 and M_2 be the discretized copula C with grid sizes n_1 and n_2, respectively. Then,

$$M_2 = AM_1A^\top$$

with $A \in \mathbb{R}^{n_2 \times n_1}$ and $A_{ij} = \frac{1}{\sqrt{m}} \mathbb{1}\{j \in ((i-1)\cdot m, i \cdot m]\}$ and

$$\begin{aligned}\operatorname{rank}(M_2) &= \operatorname{rank}(AM_1A^\top) \leq \min(\operatorname{rank}(A), \operatorname{rank}(M_1), \operatorname{rank}(A^\top)) \\ &= \min(\operatorname{rank}(A), \operatorname{rank}(M_1)) \leq \operatorname{rank}(M_1).\end{aligned}$$

□

3.2.5. Similarity of copulas

Using the decomposition makes it easy to compute the similarity of copulas if they have a shared grid size. We show that this similarity in terms of the Frobenius distance is mainly driven by Pearson's ϕ^2 of the product of the two copulas. The Frobenius distance is highly dependent on the grid size n; thus, we propose two normalizations. Let $\mathbf{A}^A = \mathbf{U}^A\mathbf{S}^A(\mathbf{V}^A)^\top$ and $\mathbf{A}^B = \mathbf{U}^B\mathbf{S}^B\left(\mathbf{V}^B\right)^\top$ be the two matrices after centering the bistochastic matrices $\mathbf{C}^A$ and $\mathbf{C}^B$, respectively. For ease of notation, we omit the common grid size n. Then,

$$\begin{aligned}
||\mathbf{A}^A - \mathbf{A}^B||_F^2 &= ||\mathbf{C}^A - \mathbf{C}^B||_F^2 \\
&= \operatorname{trace}\mathbf{C}^A(\mathbf{C}^A)^\top - 2\operatorname{trace}\mathbf{C}^B(\mathbf{C}^A)^\top + \operatorname{trace}\mathbf{C}^B(\mathbf{C}^B)^\top \\
&= \phi^2(\mathbf{C}^A) + \phi^2(\mathbf{C}^B) - 2\phi^2(\mathbf{C}^P) \\
&= \sum_{k=1}^{n-1}(s_k^A)^2 + \sum_{k=1}^{n-1}(s_k^B)^2 - 2\operatorname{trace}\mathbf{A}^B(\mathbf{A}^A)^\top + 2 \qquad (3.16)
\end{aligned}$$

with the product copula $\mathbf{C}^P = \mathbf{C}^B(\mathbf{C}^A)^\top$ (see Kolesárová et al., 2006, p. 700). Whereas the terms $\phi^2(\mathbf{C}^A)$ and $\phi^2(\mathbf{C}^B)$ depend on the individual copulas solely, $\operatorname{trace}\mathbf{A}^B(\mathbf{A}^A)^\top$ depends on the relative orientation of the singular vectors, that is,

$$\begin{aligned}
\operatorname{trace}\mathbf{A}^B(\mathbf{A}^A)^\top &= \sum_{k=1}^{n-1}\sum_{l=1}^{n-1} s_k^B s_l^A \langle \mathbf{v}_k^B, \mathbf{v}_l^A\rangle\langle \mathbf{u}_l^A, \mathbf{u}_k^B\rangle \\
&= \sum_{k=1}^{n-1}\sum_{l=1}^{n-1} s_k^B s_l^A \cos\alpha(\mathbf{v}_k^B, \mathbf{v}_l^A)\cos\alpha(\mathbf{u}_l^A, \mathbf{u}_k^B),
\end{aligned}$$

where $\alpha(\cdot,\cdot)$ is the angle between the two vectors. Thus, the copula similarity is driven by the similarity of the singular vectors weighted by the singular values.

Although the distance (squared) in Equation (3.16) is straightforward to compute, it depends on the grid size n, as the range of values increases with n. Clearly, $||\mathbf{A}^A - \mathbf{A}^B||_F^2 \geq 0$ and $||\mathbf{A}^A - \mathbf{A}^B||_F^2 = 0$ for $\mathbf{A}^A = \mathbf{A}^B$. The maximum

$$\begin{aligned}
||\mathbf{A}^A - \mathbf{A}^B||_F^2 &= \phi^2(\mathbf{C}^A) + \phi^2(\mathbf{C}^B) - 2\phi^2(C^P) \\
&\leq n + n - 2\cdot 0 = 2n \qquad (3.17)
\end{aligned}$$

is attained, for example, for $\mathbf{C}^A = I_n$ and $\mathbf{C}^B$ any doubly stochastic matrix with ones on off-diagonal elements, for example,

$$\mathbf{C}^B = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & & & & \\ 1 & 0 & \dots & & \end{pmatrix}.$$

Thus, the use of the Frobenius distance suffers from a high dependence on the grid size n, and we propose two simple rescalings of the distance that account for the increase in n. The first one uses the maximal distance from Equation (3.17), yielding

$$\delta_1(\mathbf{C}^A, \mathbf{C}^B) = \frac{\|\mathbf{C}^A - \mathbf{C}^B\|_F}{\sqrt{2n}},$$

so that the values lie within $[0, 1]$. The examples in Section 3.3.4 indicate that this normalization overcorrects, resulting in decreasing δ_1 for checkerboard approximations of the same copulas with increasing n.

Another approach is to standardize the distance by the sum of Pearson's ϕ^2 of the copulas $\mathbf{C}^A$ and $\mathbf{C}^B$, that is,

$$\delta_2(\mathbf{C}^A, \mathbf{C}^B) = \frac{\|\mathbf{C}^A - \mathbf{C}^B\|_F}{\sqrt{\phi^2(\mathbf{C}^A) + \phi^2(\mathbf{C}^B)}},$$

As $\|\mathbf{C}^A - \mathbf{C}^B\|_F \geq 0$, $\delta_2(\cdot, \cdot) \geq 0$ and from $\frac{\|\mathbf{C}^A - \mathbf{C}^B\|_F^2}{\phi^2(\mathbf{A}^A) + \phi^2(\mathbf{A}^B) - 2\phi^2(C^P)} = 1$ follows $\delta_2(\cdot, \cdot) \leq 1$. This standardization yields values that exhibit less variation with n. The similarity measures are applied to copulas in Section 3.3.4.

3.2.6. Some considerations on the link to continuous decompositions

C. M. Cuadras and Díaz (2012) and C. M. Cuadras (2015) define continuous PDF decompositions for continuous copulas. In the following, we briefly expand on the connection between the continuous decomposition and the decomposition of the corresponding checkerboard copulas. Let again C denote the copula CDF, c the copula PDF and

$$c(u, v) = 1 + \sum_{k \geq 1} \lambda_k a_k(u) b_k(v), \tag{3.18}$$

with complete orthonormal sets $\{a_k\}$ and $\{b_k\}$. C. M. Cuadras and Díaz (2012) calls the cardinality of the set $\{\lambda_k : \lambda_k \geq 0\}$ geometric dimensionality, γ, of the copula, provided that γ is finite, analogously to the discretized case. The decomposition exists if the copula's ϕ^2, that is, $\phi^2(C) = \sum_k \lambda_k^2$, is finite and induces a decomposition of the copula CDF

$$\begin{aligned}\mathcal{C}(u,v) &= \int_0^u \int_0^v \left(1 + \sum_{k=1}^{\gamma} \lambda_k a_k(\bar{u}) b_k(\bar{v})\right) \mathrm{d}\bar{u}\; \mathrm{d}\bar{v} \\ &= uv + \sum_{k=1}^{\gamma} \int_0^u \int_0^v \lambda_k a_k(\bar{u}) b_k(\bar{v})\, \mathrm{d}\bar{u}\; \mathrm{d}\bar{v}.\end{aligned}$$

The discretized copula of grid size n yields

$$\begin{aligned}C_{ij}^n &= C\left(\tfrac{i}{n}, \tfrac{j}{n}\right) - C\left(\tfrac{i-1}{n}, \tfrac{j}{n}\right) - C\left(\tfrac{i}{n}, \tfrac{j-1}{n}\right) + C\left(\tfrac{i-1}{n}, \tfrac{j-1}{n}\right) \\ &= \sum_{k=1}^{\gamma} \lambda_k \int_{\frac{i-1}{n}}^{\frac{i}{n}} a_k(u)\, \mathrm{d}u \int_{\frac{j-1}{n}}^{\frac{j}{n}} b_k(v)\, \mathrm{d}v\end{aligned}$$

and with the additional vectors

$$\mathfrak{a}_k = \left(\int_0^{\frac{1}{n}} a_k(u)\, \mathrm{d}u, \int_{\frac{1}{n}}^{\frac{2}{n}} a_k(u)\, \mathrm{d}u, \dots\right)^\top \quad (k \in [\gamma]) \tag{3.19}$$

$$\mathfrak{b}_k = \left(\int_0^{\frac{1}{n}} b_k(v)\, \mathrm{d}v, \int_{\frac{1}{n}}^{\frac{2}{n}} b_k(v)\, \mathrm{d}v, \dots\right)^\top \quad (k \in [\gamma]), \tag{3.20}$$

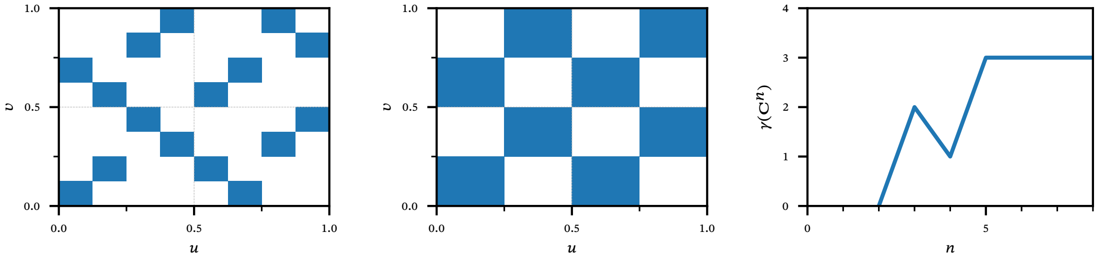
$$\mathbf{C}^n = \sum_{k=1}^{\gamma} \lambda_k \mathfrak{a}_k \mathfrak{b}_k^\top. \tag{3.21}$$

Note that Equation (3.21) denotes an exact decomposition of $\mathbf{C}^n$, but not necessarily the SVD-decomposition. The difference becomes particularly evident if $n < \gamma$ and the summation in (3.21) has more summands than the dimensionality of the $n \times n$ matrix $\mathbf{C}^n$. To be the SVD, the vectors $\mathfrak{a}_k$ and $\mathfrak{b}_k$ must be left and right singular vectors. Take $\mathfrak{a}_l$ with $l \in [\gamma]$,

$$(\mathbf{C}^n)^\top \mathfrak{a}_l = \left(\sum_{k=1}^{\gamma} \lambda_k \mathfrak{a}_k \mathfrak{b}_k^\top\right)^\top \mathfrak{a}_l \tag{3.22}$$

$$= \left(\sum_{k=1}^{\gamma} \lambda_k \mathfrak{b}_k \mathfrak{a}_k^\top\right) \mathfrak{a}_l \tag{3.23}$$

$$= \sum_{k=1}^{\gamma} \lambda_k \mathfrak{b}_k (\mathfrak{a}_k^\top \mathfrak{a}_l). \tag{3.24}$$



(a) The continuous copula's mass is uniformly distributed over the blue rectangles.

(b) The checkerboard copula's mass is uniformly distributed over the blue rectangles for $n = 4$.

(c) The geometric dimension $\gamma(\mathbf{C}^n)$ for increasing n.

Figure 3.1.: Example for a copula C_1 with corresponding doubly stochastic matrix $\mathbf{C}^n$ that has strictly smaller geometric dimension $\gamma(\mathbf{C}^n)$ than $\min(\gamma(C), n-1)$ for $n = 4$.

Thus, $\mathfrak{a}_l$ is the left singular vector if it is orthogonal to the other $\{\mathfrak{a}_k : k \in [\gamma], k \neq l\}$ and if $\mathfrak{b}_l$ is a right singular vector. The corresponding singular value is λ_k. The orthogonality condition is trivially fulfilled for $\gamma = 1$ and generally depends on the grid size n. The orthogonal $\{a_k\}$ and $\{b_k\}$ do not induce the orthogonality of the vectors $\{\mathfrak{a}_k : k \in [n]\}$ and $\{\mathfrak{b}_k : k \in [n]\}$.

In addition, the decomposition in Equation (3.21) bounds the geometric dimension of the discretized decomposition by the geometric dimension of the continuous decomposition. The trivial matrix-order bound is $n-1$. Example 3.2.1 shows that this is indeed an upper bound and not an equality. A representation with fewer summands could be possible with fewer orthogonal vectors.

Example 3.2.1. *Let C_1 be a continuous copula with uniform support on the rectangles shown in Figure 3.1a. The continuous copula's decomposition has geometric dimension $\gamma(C_1) = 3$. Figure 3.1c shows the geometric dimensions of discretizations of C_1 with various grid sizes. For $n = 4$, the geometric dimension of the discretization is 1, and thus strictly smaller than the continuous geometric dimension and $n-1$ (see Figure 3.1b).*

Similarly to the decompositions of the continuous copula, the decompositions of the copula CDF do not directly yield decompositions of the PDF. A continuous decomposition of the CDF with d summands is in general form

$$C(u, v) = uv + \sum_{k=1}^{d} \lambda_k F_k(u) G_k(v), \tag{3.25}$$

with orthogonal $F_1, \dots, F_d$ and $G_1, \dots, G_d$. It implies a decomposition of the PDF for differentiable F_k and G_k ($k = 1, \dots, d$),

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v) = 1 + \sum_{k=1}^{d} \lambda_k \frac{\partial F_k}{\partial u}(u) \frac{\partial G_k}{\partial v}(v),$$

that generally lacks the orthogonality of the function $\partial F_1(u)/\partial u, \partial F_2(u)/\partial u, \dots$. However, the above calculation shows that the number of summands for a representation of PDF is, at most, the number of summands of CDF, such that d is an upper bound for the geometric dimension, γ, of a PDF decomposition.

Equations (3.25) and (3.18) enable constructing copulas from appropriate $\{\lambda_k\}$, $\{f_k\}$, and $\{g_k\}$. Rodríguez-Lallena (2004) formulates conditions on the components to ensure the validity of the resulting copula. Mesiar and Najjari (2014) extend this construction to higher (finite) dimensions. The construction only yields copulas without tail dependence and thus excludes, for example, the CA or Gumbel copula. Instead of estimating all components, some parts in Equation (3.25) can be fixed. Bakam and Pommeret (2022), for example, uses a Legendre polynomial basis and only fits the remaining coefficients. In Section 3.3.3, we show that for (transformed) Hermite polynomials and certain $\{\lambda_k\}$, the Gaussian copula arises. Allowing not only pairs in Equation (3.18) but also the cross products for $a_k(u)b_l(v)$ ($k \neq l$) in the summation leads to further copula decomposition methods; see, for example, called Generalized Partition of Unity Copulas (Pfeifer et al., 2016; Masuhr and Trede, 2020).

We give further examples of the difference between continuous and discretized decomposition for the FGM copula in Section 3.3.1 and for the Gaussian copula in Section 3.3.3.

3.3. Illustrative singular value decompositions of copulas

This section provides the resulting decompositions for some checkerboard approximations of parametric copula families (for an overview of the considered families, see Section 2.1). Section 3.3.1 focuses on symmetric copulas, whereas Section 3.3.2 analyzes asymmetric copulas. These sections give examples of the resulting singular values and singular vectors, and we expand on the Frobenius norm-minimizing choice of η in the MAR. At the end of Section 3.3.1, we provide examples of invalid, that is, non-copula, truncations, and the use of Algorithm 1. Section 3.3.3 compares the checkerboard and continuous decomposition, as introduced in Section 3.2.6, for the Gaussian copula. Section 3.3.4 applies the similarity measures of Section 3.2.5 to various checkerboard copulas.

In this section, we will denote the rank of the truncation by $n^* \in \mathbb{N}_0$ and refer to the non-MAR model by *raw* model.

3.3.1. Decompositions of symmetric copulas

We start with simple copulas with low geometric dimensions and get up to high geometric-dimensional copulas with tail dependence in the later examples in this section. The independence copula

$$C^{\Pi}(u, v) = uv$$

yields the checkerboard copula $\mathbf{C}^n = \Pi$ of geometric dimension 0. The comonotonicity copula

$$C^{M}(u, v) = \min(u, v)$$

yields the checkerboard copula $\mathbf{C}^n = I_n$ with geometric dimension $n-1$. The MAR with $\eta = -1$ fully recovers the matrix for $n^* \geq 0$. Thus, the geometric dimensionality is significantly reduced in the MAR for the comonotonicity copula.

The FGM copula family with CDF

$$C^{\theta,FGM}(u, v) = uv + \theta uv(1-u)(1-v)$$

for $\theta \in [-1, 1]$ is of geometric dimensionality 1. Figure 3.2a depicts the first singular vector with respect to n and Figure 3.2b the first singular value, $s_1 = |\theta|/3$ for the continuous representation with respect to θ. The first singular vector is $u_1 = \alpha(1, 1 - 2/(n-1), \dots, -1)^{\mathsf{T}}$ ($\alpha \in \mathbb{R}$ such that $||u_2|| = 1$), being the checkerboard analog of $a_1(x) = \sqrt{3}(1-2x)$ according to Equation (3.19). The singular vector is the piecewise integrated a_1 since the geometric dimension is one. For $n^* = 1$, the MAR following the optimization in (3.9) has parameters $\eta = s_1/(n-1) = |\theta|/(3(n-1))$. A numerical optimization in MATLAB yields numerically equivalent values, as shown in Figure 3.2c. The approximation is improved with the MAR, but the gain is smaller than for the comonotonicity copula (see Figure 3.2d). The matrix can be fully recovered for any $n^* \geq 1$. The calculation of Spearman's ρ_S according to the representation in Equation (3.10) yields the result for the FGM copula ($n^* \geq 1$)

$$\begin{aligned}\hat{\rho_S}(\hat{C}^{\theta,FGM}) &= (4 - 1/n^2) \sum_{k=1}^{n^*} s_k \langle \omega, u_k \rangle \langle v_k, \omega \rangle \\ &= (4 - 1/n^2) \frac{\theta}{3} \left(\alpha \frac{1}{||\tilde{\omega}||} \langle \tilde{\omega}, (1, 1 - 2/n, \dots, -1)^{\mathsf{T}} \rangle \right)^2 \\ &= \frac{\theta}{3}(1 - 1/n^2).\end{aligned}$$

The approximated $\hat{\rho_S}$ yields the FGM copula's analytical Spearman's ρ_S of $\theta/3$ for $n \to \infty$.

The Cuadras-Augé (CA) family of copulas (C. M. Cuadras and Augé, 1981) with CDF

$$C^{\theta,\mathrm{CA}}(u,v) = \begin{cases} uv^{1-\theta} & ,u \le v \\ u^{1-\theta}v & ,u \ge v \end{cases}$$

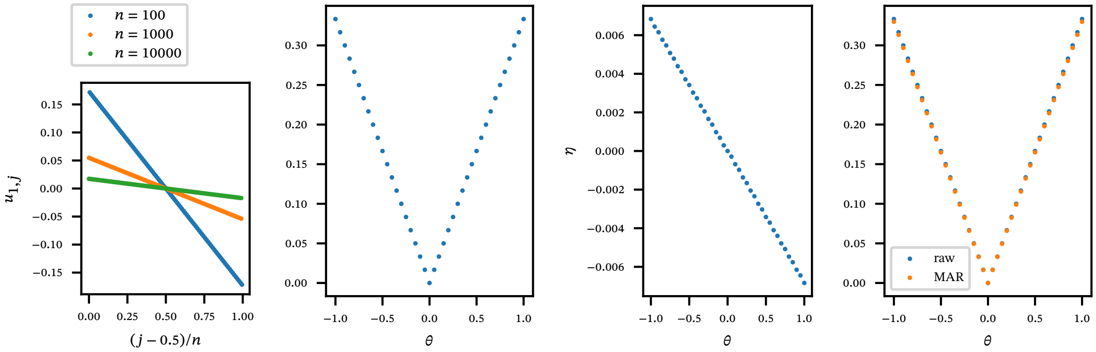
for $\theta \in [0,1]$ has an upper tail dependency of $\lambda_u = \theta$. The corresponding centered, doubly stochastic matrix is of rank $n-1$. For $\theta = 0$, $C^{0,\mathrm{CA}} = C^{\Pi}$, while $C^{1,\mathrm{CA}} = C^{M}$ with the decompositions argued above. Figure 3.3 shows the computed singular vectors and values for $\theta \in (0,1)$. The singular vectors in Figure 3.3a drop near $u = 1$. The decay of singular values starting from s_1 is similar for the different values of θ, but it is shifted upward for higher values of θ, as shown in Figure 3.3b. Figure 3.3d shows that the reconstruction is significantly improved when the MAR is used, especially for large θ. For large absolute values of θ, larger absolute values of η in the MAR are chosen (see Figure 3.3c).

The Gumbel family of copulas with CDF

$$C^{\theta,\mathrm{Gu}}(u,v) = \exp\left(-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{\frac{1}{\theta}}\right)$$

for $\theta \in [1,\infty)$ is an Archimedean copula and exhibits upper tail dependence like the CA copula. The checkerboard copula contains high values in the upper right part (see Figures 3.5a and 3.6a). The Gumbel copula is the independence copula for $\theta = 1$ and approaches the comonotonicity copula for $\theta \to \infty$. The singular vectors in Figure 3.4a contain jumps next to $u = 1$ like the singular vectors for the CA copula. Again, the approximation improves considerably when using the MAR, particularly for higher values of θ, as shown in Figure 3.4c for approximations of rank one or in Figure 3.4d for approximations of rank five. The difference in the MAR and the raw representation approximation reduces when the approximation order increases (see Figure 3.4b).

For higher parameter values θ, the truncated representations of the Gumbel copula contain negative entries. Figures 3.5 and 3.6 show the discretized PDF, its approximation, and an indicator plot for the invalidity of the elements. The negative elements in Figures 3.5b and 3.6b have a waveform. The Gumbel copula contains higher peaks for higher parameters θ, and the approximation tends to have more negative elements. After applying the correction algorithm, all elements are nonnegative, and the Frobenius distance between the (corrected) approximation and the discretized PDF is smaller (see Table 3.1). The corrections are smaller for small values of θ (see Figure 3.5c) than for larger values of θ (see Figure 3.6c).



(a) Elements $u_{1,j}$ ($j \in [n]$) of the first singular vector u_1 for $\theta = 0.8$ and various grid resolutions n. The same plots arise for other values of $\theta \neq 0$. The different slopes result from the normalization of the singular vector.

(b) The first singular value s_1 for various values of θ. The value is, by definition, positive.

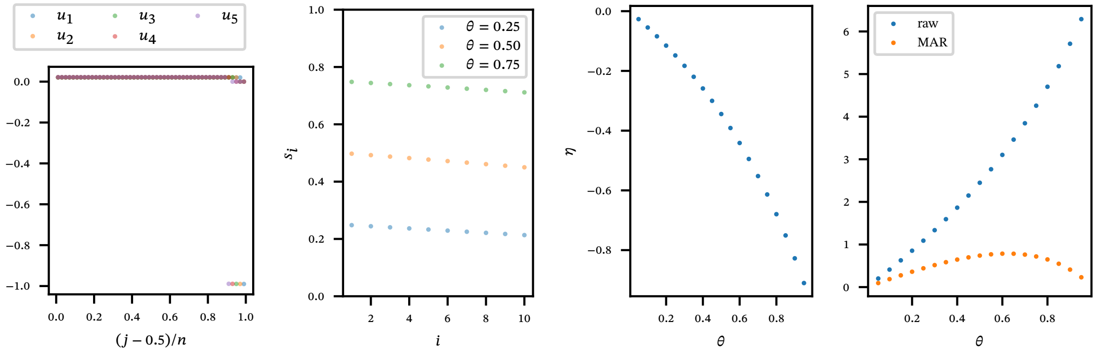
(c) The values of η in the MAR minimizing the Frobenius error for a 0-truncation, that is only the MAR. The values for η are obtained by numerical minimization using MATLAB's `fminsearch`. The resulting values of η coincide with their theoretical counterparts (see Equation (3.9).

(d) Frobenius error of the MAR and the standard representation for 0-truncations. The values of η are in plot c). The MAR reduces the error slightly.

Figure 3.2.: Analysis of the FGM checkerboard copula decompositions using the raw and MAR model.

	$\theta = 2.5$	$\theta = 7.5$
$\lVert\mathbf{C}^{50} - G^{-1}(T_{10}(\mathbf{A}^{50}))\rVert_F$	0.0084	0.6449
$\lVert\mathbf{C}^{50} - P(G^{-1}(T_{10}(\mathbf{A}^{50})))\rVert_F$	0.0084	0.5476
$\lVert G^{-1}(T_{10}(\mathbf{A}^{50})) - P(G^{-1}(T_{10}(\mathbf{A}^{50})))\rVert_F$	0.0008	0.3099
$\lVert G^{-1}(T_{10}(\mathbf{A}^{50})) - P(G^{-1}(T_{10}(\mathbf{A}^{50})))\rVert_F / \lVert\mathbf{C}^{50}\rVert_F$	0.05 %	11.47 %

Table 3.1.: Frobenius distances for the approximation of a Gumbel checkerboard copula with parameter θ and $n = 50$. $G^{-1}(T_{10}(\mathbf{A}^{50}))$ denotes the truncation, and $P(G^{-1}(T_{10}(\mathbf{A}^{50})))$ the result of Algorithm 1. The distance between the original and the approximation decreases with the application of Algorithm 1. The last row displays the relative change through Algorithm 1 with respect to the Frobenius norm of the raw matrix $\mathbf{C}^{50}$.



(a) Elements $u_{i,j}$ ($i \in [5], j \in [n]$) of the first five singular vectors for $\theta = 0.5$. The singular vectors have a similar course for other values of $\theta \in (0,1)$.

(b) The singular values for $\theta \in \{0.25, 0.5, 0.75\}$.

(c) Frobenius-norm minimizing choice of η in the MAR for approximations of rank one.

(d) Frobenius error of the MAR and raw representation for approximations of rank one.

Figure 3.3.: Analysis of the CA checkerboard copula decompositions using the raw and MAR model for various values of θ and $n = 50$.

3.3.2. Decompositions of asymmetric copulas

For asymmetric copulas, the left and right singular vectors do not coincide. We use an asymmetric construction method from Nelsen (2006, p.84), which yields copulas with cubic sections. The copula CDF is

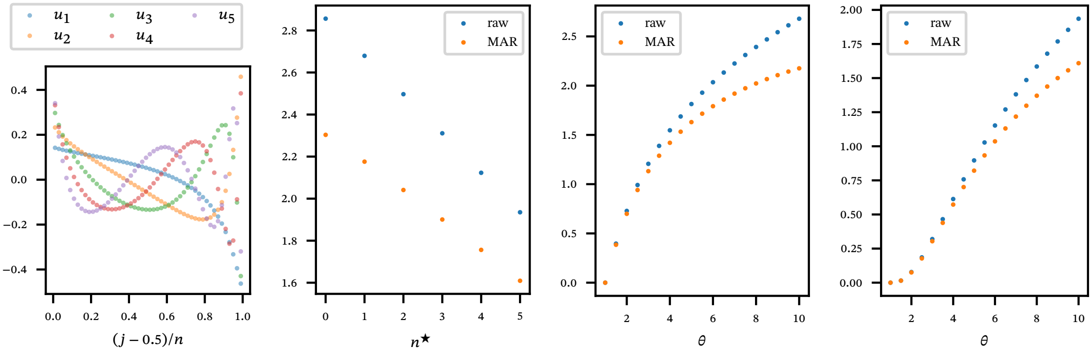
$$C^{a,b,\text{asym}}(u,v) = uv + uv(1-u)(1-v)[(a-b)v(1-u)+b], \tag{3.26}$$

where $|b| \leq 1, [b-3-(9+6b-3b^2)^{1/2}]/2 \leq a \leq 1$, and $a \neq b$. The conditions on a and b ensure the validity of the resulting copula. For $a = b \in [-1,1]$ the FGM copula with parameter b arises. Figure 3.7 shows the resulting SVD for two configurations of a and b. For $a = 0.5$ and $b = -0.5$ the resulting copula CDF is

$$C^{0.5,-0.5,\text{asym}}(u,v) = uv + u(1-u)^2v^2(1-v) - 0.5uv(1-u)(1-v),$$

and for $a = -1.5$ and $b = 0.5$

$$C^{-1.5,0.5,\text{asym}}(u,v) = uv - 2uv^2(1-u)^2(1-v) + 0.5uv(1-u)(1-v).$$



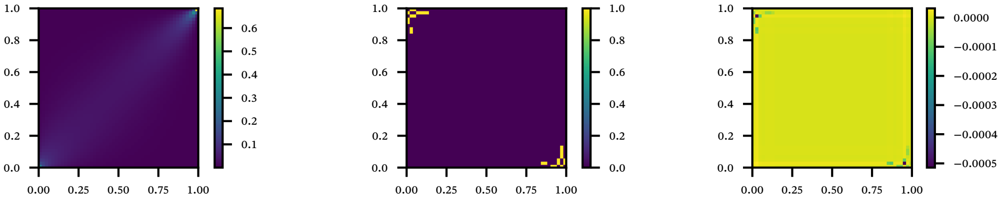
(a) Elements $u_{i,j}$ ($i \in [5], j \in [n]$) of the first singular vectors for $\theta = 10$. The continuous Gumbel copula has an upper tail dependence.

(b) The Frobenius error of the approximation for a Gumbel copula with $\theta = 10$ and increasing approximation order $n^\star$. The MAR reduces the error considerably.

(c) Frobenius error for approximations of rank one.

(d) Frobenius error for approximations of rank five.

Figure 3.4.: Analysis of the Gumbel checkerboard copula decompositions using the raw and MAR model for $\theta = 10$ and $n = 50$.

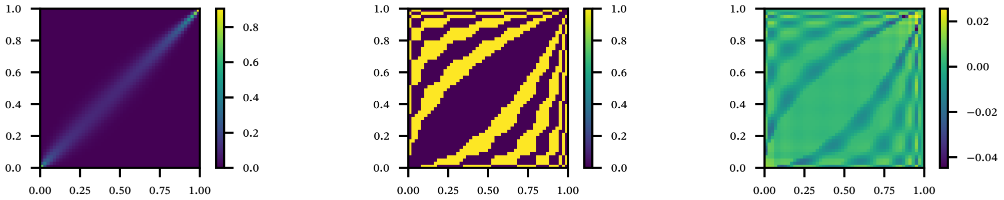


(a) The checkerboard PDF.

(b) The yellow squares indicate negative matrix elements in $T_{10}(\mathbf{A}^{50})$.

(c) Difference of approximation and corrected approximation, $G^{-1}(T_{10}(\mathbf{A}^{50})) - P(G^{-1}(T_{10}(\mathbf{A}^{50})))$, using Algorithm 1. Note the different scaling compared to a).

Figure 3.5.: Analysis of the truncation of order 10 of a Gumbel checkerboard copula with $\theta = 2.5$ and $n = 50$.



(a) The checkerboard PDF.

(b) The yellow squares indicate negative matrix elements in $T_{10}(\mathbf{A}^{50})$. They occur more frequently than for $\theta = 2.5$.

(c) Difference of approximation and corrected approximation, $G^{-1}(T_{10}(\mathbf{A}^{50})) - P(G^{-1}(T_{10}(\mathbf{A}^{50})))$, using Algorithm 1. Note the different scaling compared to a).

Figure 3.6.: Analysis of the truncation of order 10 of a Gumbel checkerboard copula with $\theta = 7.5$ and $n = 50$.

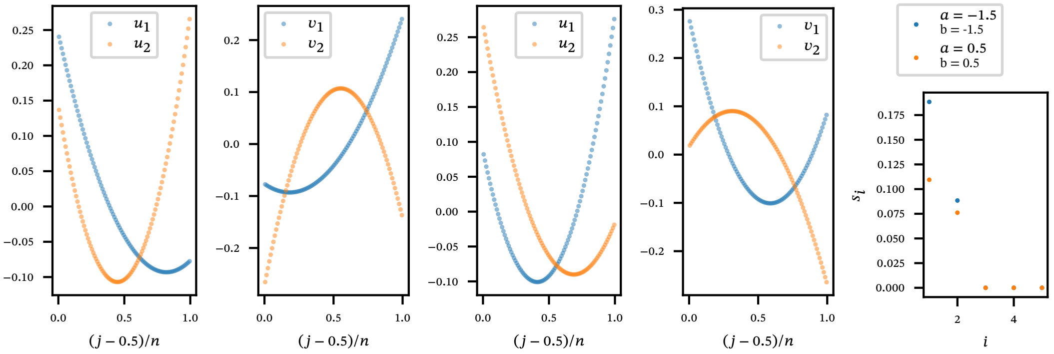
In both cases, the left and right singular vectors are polynomials of degree two. The geometric dimension of the discretized copula is 2. Thus, the singular values in Figure 3.7e drop at 3 to zero. The singular values are larger for the first singular value combination than for the second. The left singular vectors in Figures 3.7a and 3.7c have similar courses but change order. The right singular vectors (Figures 3.7b and 3.7d) exhibit a greater variation between the combinations of parameters than the left singular values. They show y-axis mirroring but also change slope and are shifted.

3.3.3. The Gaussian copula

We end the section with the Gaussian copula and apply the notions of Section 3.2.6. The Gaussian copula models the dependence structure of multivariate Gaussian distributions. Let F_ρ denote a bivariate Gaussian CDF with correlation ρ, variance $(1,1)^\top$ and mean $(0,0)^\top$, the PDF by f_ρ and the standard univariate Gaussian counterparts by Φ and φ, respectively. Then, the CDF of a Gaussian copula with correlation $\rho \in [-1,1]$ is given by

$$C^{\rho,Ga} = F_\rho(\Phi^{-1}(u), \Phi^{-1}(v)).$$

Figures 3.8a and 3.8b show the resulting PDF decompositions for the checkerboard copula. As proven in the following, the singular vectors are identical for different $|\rho| \in (0,1)$ in the continuous decomposition. No noticeable differences can be observed for the singular vectors of the checkerboard approximations for different values of ρ.



(a) Elements $u_{i,j}$ ($i \in [2], j \in [n]$) of the *left* singular vectors u_i with $a = 0.5$ and $b = -0.5$.

(b) Elements $v_{i,j}$ ($i \in [2], j \in [n]$) of the *right* singular vectors v_i with $a = 0.5$ and $b = -0.5$.

(c) Elements $u_{i,j}$ ($i \in [2], j \in [n]$) of the *left* singular vectors u_i with $a = -1.5$ and $b = 0.5$.

(d) Elements $v_{i,j}$ ($i \in [2], j \in [n]$) of the *right* singular vectors v_i with $a = -1.5$ and $b = 0.5$.

(e) The singular values s_i drop to zero after s_2 as the geometric dimension is 2.

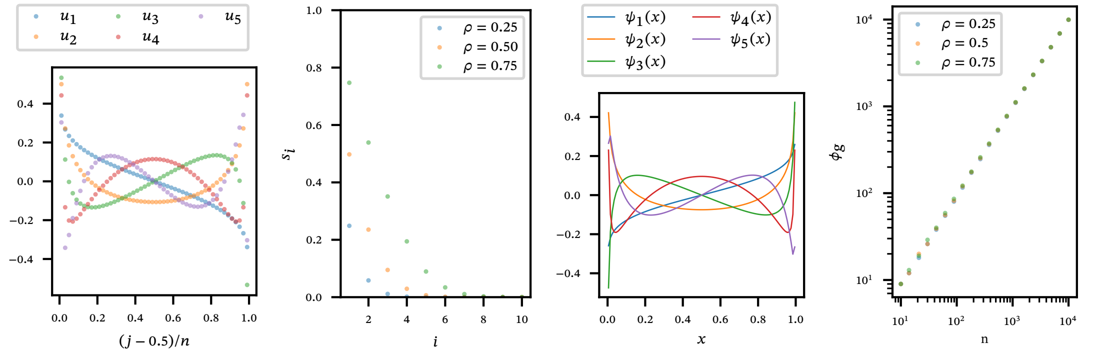
Figure 3.7.: Analysis of the asymmetric checkerboard copula decomposition of the copula following Equation (3.26) with $n = 50$ and two configurations of a and b. The left singular vectors are similar between the two parameter configurations, whereas the right singular values exhibit strong differences.

For a bivariate Gaussian distribution, Hill (1974) shows a PDF decomposition using Hermite polynomials. The following theorem extends its results to the Gaussian copula, yielding a representation in terms of transformed Hermite polynomials. We use the representation of the probabilist's Hermite polynomial ψ_i of order i by

$$\psi_i(x) = (-1)^i \exp\left(x^2/2\right) \frac{d^i}{dx^i} \exp\left(-x^2/2\right).$$

Theorem 3.3.1. *Let c_ρ be a Gaussian copula density with parameter $-1 < \rho < 1$, Φ the standard Gaussian CDF, and ψ_i the probabilist's Hermite polynomial of order i. Then,*

$$c_\rho(u, v) = 1 + \sum_{i=1}^{\infty} \frac{\rho^i}{i!} \psi_i(\Phi^{-1}(u)) \psi_i(\Phi^{-1}(v)), \quad u, v \in (0, 1). \tag{3.27}$$



(a) Elements $u_{i,j}$ ($i \in [5], j \in [n]$) of the singular vectors for $\rho = 0.5$. No discernible difference is evident in the plots for the other ρ.

(b) The singular values s_i increase for larger values of ρ.

(c) The first five (normalized) transformed probabilist's Hermite polynomials ψ_i.

(d) The numerical estimations of the geometric dimensions increase with the grid size and are comparable for the different values of ρ.

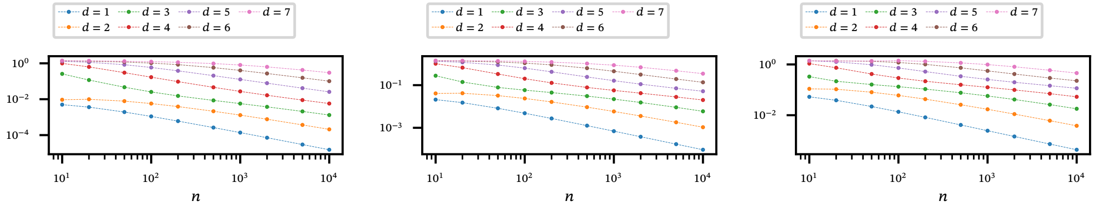
Figure 3.8.: The checkerboard decomposition of the Gaussian family of copulas for $n = 50$, the transformed probabilist's Hermite polynomials, and numerical estimates for the geometric dimension.

Proof.

$$
\begin{aligned}
c_\rho(u,v) &= \frac{f_\rho(\Phi^{-1}(u),\Phi^{-1}(v))}{\varphi(\Phi^{-1}(u))\varphi(\Phi^{-1}(v))} && |x := \Phi^{-1}(u), y := \Phi^{-1}(v) \\
&= \frac{f_\rho(x,y)}{\varphi(x)\cdot\varphi(y)} && |\text{ Hill (1974)} \\
&= \frac{\frac{1}{2\pi}\exp\left[-\frac{1}{2}(x^2+y^2)\right]\left\{1+\sum_{i=1}^\infty \frac{\rho^i}{i!}\psi_i(x)\psi_i(y)\right\}}{\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})} \\
&= \sum_{i=0}^\infty \frac{\rho^i}{i!}\psi_i(\Phi^{-1}(u))\psi_i(\Phi^{-1}(v)).
\end{aligned}
$$

□

Using the well-known maximal correlation property of the Gaussian distribution (Klaassen and Wellner, 1997, Section 6 and the references therein), the representation



(a) $\rho = 0.25$ **(b)** $\rho = 0.50$ **(c)** $\rho = 0.75$

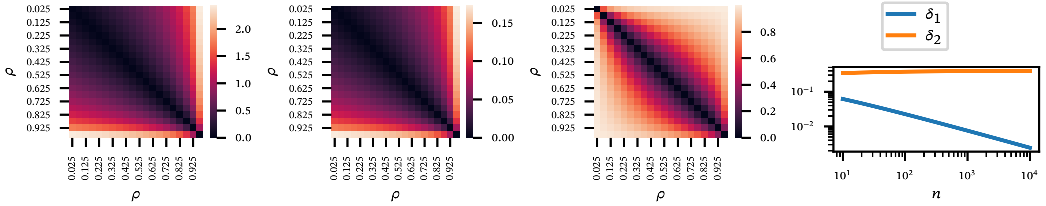
Figure 3.9.: Distance between the d-th piecewise integrated continuous singular vectors and d-th singular vector of the discretized matrix for a Gaussian copula for increasing n and different values of ρ. Distance decreases for all degrees and parameters ρ considered with n.

in Equation (3.27) is the one obtained by canonical correlation and thus a decomposition in the sense of Section 3.2.6 (see Lancaster, 1957).

Figure 3.8c shows the first transformed probabilist's Hermite polynomials ψ_i and Figure 3.8d the geometric dimension of Gaussian checkerboard copulas for various grid sizes n. The geometric dimension is bounded by $n - 1$ and increases with n, whereas the continuous Gaussian copula has an infinite geometric dimension. Figure 3.9 shows the distance between the piecewise integrated transformed Hermite polynomials (see Equations (3.19) and (3.20)) and the singular vectors of the Gaussian checkerboard copula for polynomial degrees 1 to 7. The distance decreases with n for all degrees. The smaller the parameter ρ and the degree, the faster the distance decreases.

3.3.4. Copula similarities

The difference between copulas can be quantified using the calculated measures of Section 3.2.5. Figures 3.10 and 3.11 show examples of the similarity of copulas using the (normalized) Frobenius distance of discretizations of Section 3.2.5. Figure 3.10 shows the distance between Gaussian copulas with different correlations. While Figure 3.10a shows the Frobenius distance, Figures 3.10b and 3.10c show the results using the normalizations. In Figure 3.10b, the distances are scaled by a common factor, which results in pairs of Gaussian copulas with large ρ being considered more dissimilar than pairs with small ρ with identical difference. The second normalization, δ_2, scales based on Pearson's ϕ^2 of the copulas and yields similar differences for pairs of the same ρ difference. With increasing discretization grid size, n, δ_1 tends to shrink the distance, whereas δ_2 maintains the value (see Figure 3.10d). Figure 3.11 shows the two normalizations for various checkerboard approximations of parametric copulas for two values of τ. Using the Frobenius distance or δ_1 in



(a) $\|\cdot\|_F$

(b) $\delta_1(\cdot)$. For small ρ, the copulas are assigned similar differences.

(c) $\delta_2(\cdot)$

(d) Comparison the normalizations for two Gaussian copulas with $\rho = 0.4$ and $\rho = 0.6$, respectively, with increasing grid size n.

Figure 3.10.: Comparison of the normalizations of Section 3.2.5 for a Gaussian copula with various copula correlations ρ and $n = 100$. Normalization δ_1 tends to shrink the distance with increasing n.

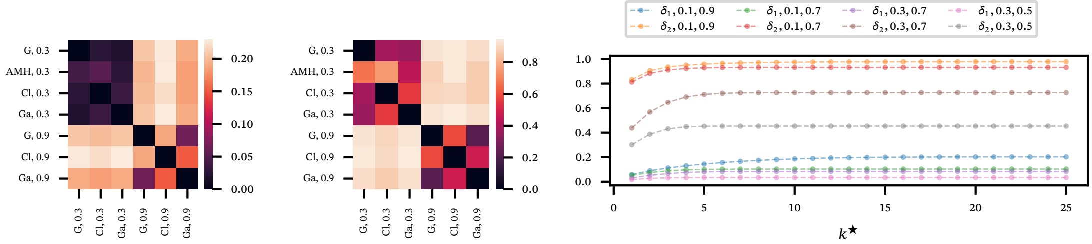
Figure 3.11a generally produces higher distances if at least one copula has a high τ. Normalization δ_2 produces close distance values for similar values of τ as shown in Figure 3.11b and covers a wider range of possible distances between 0 and 1. Figure 3.11c shows results of the truncated approximation of δ_1 and δ_2. Most of the pairs of Gaussian copulas considered already approximate the distance for small k^*.

3.4. Visual exploratory analysis of copulas with profile plots

A primary purpose of correspondence analysis is usually to generate visual representations of high-dimensional data by projecting row and column profiles into a low-dimensional space while maximizing the covered variation of the data (for an introduction, see, e.g., Greenacre, 1984). We start by describing the approach and identifying the characteristics of the copula visible in the graphs and, thus, the characteristics of the graphs to be analyzed in Section 3.4.1. In Section 3.4.2, we use empirical data plots from ranked pseudo-observations to analyze the dependence structure.

3.4.1. Understanding and interpreting profile plots

In profile plots, the similarity of the rows and columns of the checkerboard copula are shown. A row corresponds to the conditional distribution of u given the “row” value of



(a) $\delta_1(\cdot)$. For $\tau = 0.3$, the copulas are assigned similar differences.

(b) $\delta_2(\cdot)$

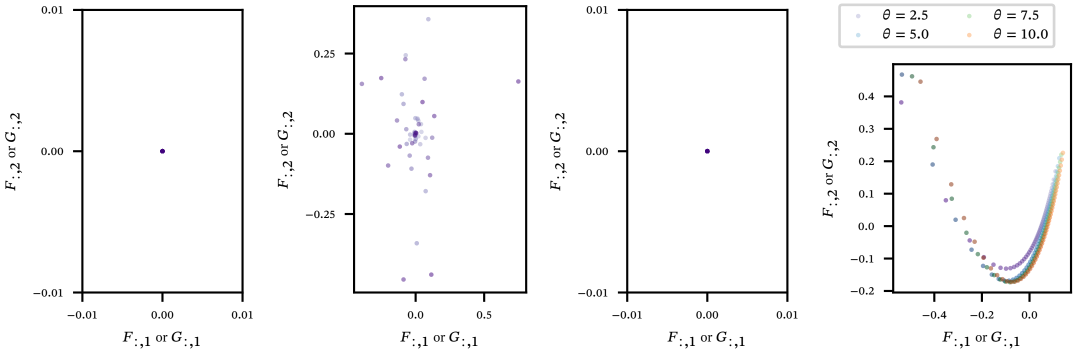
(c) The truncated, normalized distances δ_1 and δ_2 from Section 3.2.5 over truncation order $k^\star$ for pairs of Gaussian copulas with $n = 100$.

Figure 3.11.: Comparison of the normalizations of Section 3.2.5 for a (G)umbel, (C)layton and (Ga)ussian copula for two different values of τ and $n = 100$. Kendall's τ values refer to the copula to be discretized. Figure 3.11c uses the truncated representation of δ_1 and δ_2 and shows the computed values for increasing truncation parameter.

v (where we use the standard notation of u being the horizontal coordinate and v being the vertical coordinate). The row profiles, $\mathbf{F}$, and the column profiles, $\mathbf{G}$, correspond to the singular value weighted coordinates in the space spanned by the opposing singular vectors, that is, in the notation of Section 3.2.2, $\mathbf{F} \coloneqq \mathbf{US}$ and $\mathbf{G} \coloneqq \mathbf{VS}$. All n row and column profiles are shown on the basis of their first two coordinates in the profile plot. Therefore, a row profile shows the two most significant coordinates of the rows with respect to the basis spanned by the columns and vice versa. The proximity of different row profiles reflects the similarity of the corresponding conditional distributions of u given the value of v. For example, for independent variables, the distribution of u given v does not change with the value of v, and all profile points in a profile plot would match. In a case with monotone dependence instead, the distribution of u given v changes with v, and the profile points referring to different values of v would not match, and their distance increases with the dissimilarity of the respective conditional distributions. In a 2-D plot of the first two basis vectors, the v value of the row profile is not visible in the coordinates of the points. Thus, we color the profiles to reflect the position of the profile: the lighter the color, the closer the v value is to zero. Thus, the row profile for the conditional distribution u given $v \approx 0$ is the point with the lightest color, and the profile given $v \approx 1$ is the point with the darkest color.

The profiles of rows and columns, $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{G}}$, using the MAR can be computed analogously. We compare $\mathbf{F}$ and $\tilde{\mathbf{F}}$ for the Clayton copula below. For symmetric copulas, $\mathbf{F}$ and $\tilde{\mathbf{F}}$ differ only in singular values and not in singular vectors that lead to the same shapes but different profile scalings (see Lemma 1). In traditional correspondence analysis, normalizations of the row and column profiles account for the variation of the frequencies of the individual profiles. All profiles have the same frequencies in the copula domain. Therefore, normalizations are not necessary in this setting. Although sometimes plotted in one figure, the distances between a row and a column profile cannot be interpreted directly because the representation is based on a different basis. The profile plots of rows and columns are identical if the underlying copula is symmetric positive definite; that is, $\mathbf{C}^n$ and $(\mathbf{C}^n)^\top$ are equal, and $\mathbf{C}^n$ is positive definite. The matrix $\mathbf{C}^n$ is, in particular, for exchangeable copulas, symmetric but not necessarily positive definite. A copula is exchangeable if $(F_X(X), F_Y(Y))$ has the same distribution as $(F_Y(Y), F_X(X))$. The plot of several copula profiles in one plot displays differences between copulas.

Figure 3.12 shows graphs for some of the copulas of Sections 3.3.1 and 3.3.2, visualizing the general characteristics depicted in the profile plot. Profiles of the raw model lying close to the zero point indicate approximately conditionally independent distributions since the most significant coordinates are close to zero. For an independence copula, all profiles lie close to zero. Significant deviations between the components in the raw model graph and the MAR model graph refer to strong characteristics of the comonotonicity copula. Figure 3.12a shows examples for an independence and in Figures 3.12b and 3.12c for a comonotonicity copula. Through the points' colors, the plots also display how the profiles evolve and how rapidly the profiles change. Points of similar colors lying close together exhibit a smooth evolution of the copula, whereas varying distances show more extensive changes of the copula in certain areas. Increasing changes are evident, for example, in the case of tail dependence, where the profiles change rapidly in the area of the tail. The plot of the comonotonicity copula in the raw model in Figure 3.12b shows unordered profiles. The comonotonicity copulas SVD is ambiguous since any orthonormal set of vectors forms singular vectors of the diagonal matrix. Thus, the calculated basis is merely random, and the profiles are scattered. For the Gumbel copula, the profiles in Figures 3.12d and 3.12e evolve smoothly. Still, the differences become larger for higher values of θ and the profiles closer to one since the copula has an upper tail dependence that increases with θ. Using the MAR affects the profiles only slightly in Figure 3.12e as MAR only changes singular values and not singular vectors for symmetric copulas. Figure 3.12f shows the profiles of a Gaussian copula for different values of ρ in one chart. The similarity of profiles changes most pronounced in the tails of the profiles, whereas the profile differences corresponding to middle columns and rows remain similar. Figure 3.12g depicts the row and column profiles of an asymmetric copula, where the profiles do not coincide but are mirrored with respect to the horizontal axis.

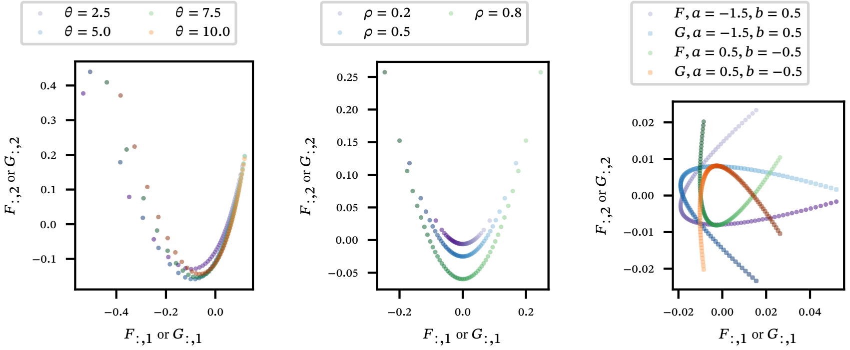


(a) The independence copula in the raw model. All profiles lie close to zero.

(b) The comonotonicity copula in the raw model. The profiles are scattered.

(c) The comonotonicity copula in the MAR. All profiles lie close to zero.

(d) The Gumbel copula in the raw model.



(e) The Gumbel copula in the MAR. The axis limits vary slightly compared to Figure 3.12d.

(f) The Gaussian copula in the raw model.

(g) The asymmetric copula according to Equation (3.26). Row and column profiles differ.

Figure 3.12.: Row and column profiles for four copulas with various parameters, each with grid size $n = 50$. Except for the copula in (3.12g), the displayed copulas are symmetric and have identical row and column profiles. The profiles reflect various copula characteristics, such as the strength of dependence, symmetry for different axes, and areas with high variation.

3.4.2. Profile plots illustrated with a data example

Using data from an engineering context, we apply the graphical dependence assessment to empirical data. Coblenz et al. (2020) model the distribution of fuel drops that are generated by a fuel injector in a jet engine using vine copulas. The droplets are characterized by five variables $x_1, \dots, x_5$, that is, the size of the drop, the position in the x and y directions, and the velocity in the x and y directions. Data are generated using numerical simulations under different operating conditions of jet engines, specified by the air velocity, the air pressure, and the thickness of the atomizing edge. Coblenz et al. (2020) publish statistically simulated data for different operating conditions. We focus on one of the ten operating conditions modeled, that is, an air velocity of $90ms^{-1}$, an air pressure of 5 bar, and a thickness of the atomizing edge of $230\mu m$ since this is the largest of the provided datasets. It consists of 5,252 points in the five dimensions listed above.

The published data of Coblenz et al. (2020) is available in the rank-transformed copula domain, which we denote by $u_1, \dots, u_5$. Note that due to the rank transformation, all values of the u_j are in the discrete set $\{1/5252, 2/2525, \dots, 1\}$. The copula domain's relative frequency table, $\mathbf{C}^n$, is computed by counting the number of points per lattice box in $I^n \times I^n$. Observations lying precisely on a grid boundary are counted for the box below. We use $n = 26$ as a divisor of 5252 for the analysis so that the resulting table has 202 observations in each row and column, and dividing each cell by 202 leads to a doubly stochastic matrix. For each distinct pair of dimensions, we plot the row profiles, the column profiles, and a checkerboard copula plot of the pseudo-observations in Figure 3.13. We focus on five of the pairs here, the graphs for the combinations (u_1, u_3), (u_1, u_5), (u_2, u_5), (u_3, u_5), and (u_4, u_5) are shown in Appendix A.4 in Figure A.1.

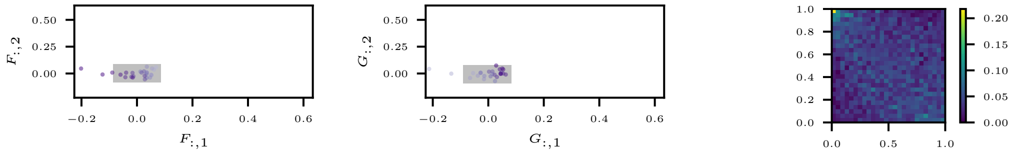
As profile points are obtained from empirical data, they deviate from their theoretical counterparts. To visualize statistical noise in the plots, we show typical minimal and maximal values of profiles for an independence copula by a gray rectangle in the plots. The gray rectangles are obtained by sampling 5252 realizations from an independence copula and computing their row and column profiles. The procedure is repeated 100 times. The rectangles cover 95% of the minimal and maximal point coordinates of the 100 samples in every dimension. Thus, if the profiles are outside the gray box, the underlying copula is unlikely to be the independence copula. This approach aligns with Greenacre (1984), who advocates resampling methods, for example, bootstrapping, over using asymptotic results for profile values. Again, the darker the point's color, the closer the conditional distribution's conditioning variable is to one.

The profile plots for variables u_1 and u_2 in Figures 3.13a and 3.13b show that some profiles deviate from others and that there is a continuous development with the conditioning variable. For row profiles, that is, conditional distributions u_2 given u_1, with u_1 close to one (dark points in the row profile plot) and column profiles u_2 close to zero (light points in the column profile plot), the profiles change and indicate that the variables are not independent. The checkerboard plot in Figure 3.13c shows a peak at $(0, 1)$. Further information is covered by noise. The profiles in Figures 3.13d and 3.13e exhibit a U-shaped pattern and are mirrored with respect to the vertical axis. Thus, the dependence has a countermonotonic characteristic that is reflected loosely in the checkerboard copula plot in Figure 3.13f. While high values are apparent near $(0, 1)$ and $(1, 0)$, the pattern in between is hard to distinguish. The row and column profiles of Figures 3.13g and 3.13h differ clearly. Whereas the row profiles evolve in a similar direction with stronger changes near 0, the column profiles undergo a cyclical transformation. The profiles corresponding to small and large values of u_3 are similar, and the profiles for u_3 near 0.5 are different. This pattern is a sign of U or hump-shaped dependence, which is also reflected in the checkerboard plot. For variables u_2 and u_4, Figures 3.13j to 3.13l show a typical tail-dependence behavior. The profiles change rapidly for small values of u_2 and u_4, whereas they evolve relatively smoothly for larger values. The behavior of the profile plots in Figures 3.13m and 3.13n is similar to variables u_2 and u_3, but is exchanged. The row profiles undergo a cyclical transformation, while the column profiles evolve smoothly. As the U-shaped form is more apparent than for variables u_1 and u_4, the profiles show a stronger pattern for variables u_3 and u_4. The u-shape is distinguishable in Figure 3.13o.

Overall, the row and profile plots provide at least the same amount of information as the checkerboard plots, but they are more transparent and less cluttered than the checkerboard plots.

3.5. Conclusion

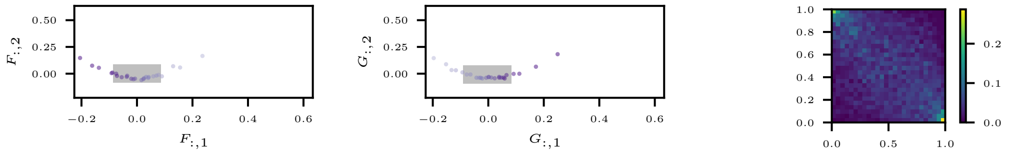
This chapter analyzes truncations of singular value decompositions (SVDs) and correspondence analysis of checkerboard copulas. Checkerboard copulas can be mapped to doubly stochastic matrices, making it straightforward to ensure copula properties for the approximations. We find that some common copulas, for example, comonotonicity-like, have high ranks and thus are poorly represented in the straightforward SVD and that truncations can have negative elements. To account for comonotonicity-like copulas with high ranks, we adapt a representation anchored with the comonotonicity copula and show its performance in examples. We compute the nearest valid doubly stochastic matrix to correct the truncations with negative entries. We analyze representations of statistical characteristics of copulas, like Kendall's τ, Spearman's ρ, or differences between copulas through the decomposition. The truncations can be used to compute discretized versions of continuous decom-



(a) Row profiles for variables u_1 and u_2.

(b) Column profiles for variables u_1 and u_2.

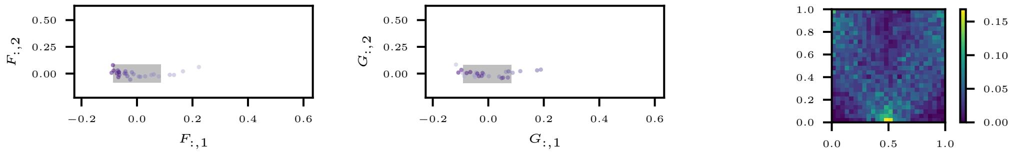
(c) Checkerboard plot for variables u_1 and u_2.



(d) Row profiles for variables u_1 and u_4.

(e) Column profiles for variables u_1 and u_4.

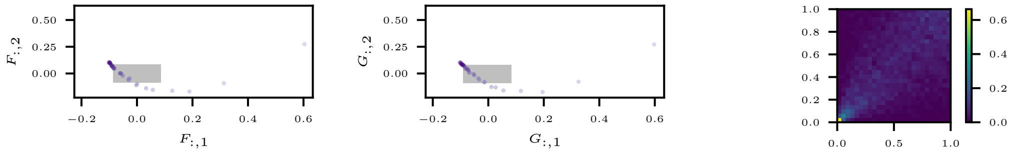
(f) Checkerboard plot for variables u_1 and u_4.



(g) Row profiles for variables u_2 and u_3.

(h) Column profiles for variables u_2 and u_3.

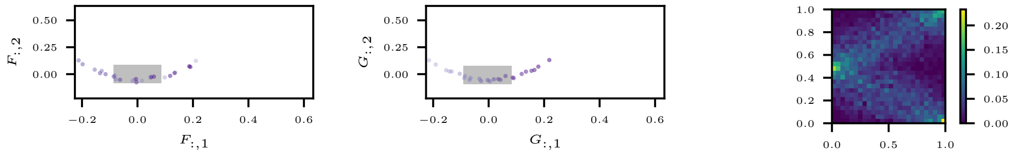
(i) Checkerboard plot for variables u_2 and u_3.



(j) Row profiles for variables u_2 and u_4. A profile at $(0.60, 0.27)$ is out of scope.

(k) Column profiles for variables u_2 and u_4. A profile at $(0.60, 0.27)$ is out of scope.

(l) Checkerboard copula plot for variables u_2 and u_4.



(m) Row profiles for variables u_3 and u_4.

(n) Column profiles for variables u_3 and u_4.

(o) Checkerboard plot for variables u_3 and u_4.

Figure 3.13.: Profile and checkerboard plots of the fuel injector spray characteristics in jet engines from Coblenz et al. (2020). The physical interpretations of the variables are drop size (u_1), x-position (u_2), y-position (u_3), x-velocity (u_4), and y-velocity (u_5).

positions, linking our analysis to continuous decompositions. We derive a decomposition of the Gaussian copula into transformed Hermite polynomials and show that the discretized singular vectors draw closer to the transformed Hermite polynomials with increasing grid size. We analyze correspondence analysis profile plots for copulas and show that they reveal asymmetries and non-monotonic dependence. Profile plots for various copulas are shown, and the graphical analysis is illustrated on a dataset on fuel injector spray characteristics in jet engines.

Other approaches for reducing the comonoticity-like characteristics of the copula are possible, such as using rook copulas (Cottin and Pfeifer, 2014) and, for empirical data, sample-dependent grid sizes (Janssen, Swanepoel, and Veraverbeke, 2012) or anchoring with respect to other copulas while varying the sample size (Cuberos, Masiello, and Maume-Deschamps, 2020). They need more complex fitting of the parameters and components and might use different grid sizes. Thus, we leave the comparison of these methods for further research.

In this chapter, we do not expand on the empirical estimation of the model. It is well known that the empirical checkerboard copula converges to the theoretical checkerboard copula. Perturbation theory analyzes the effect of noise on the results of the SVD (for a concise overview, see, e.g., Stewart, 1991). The singular vectors can suffer from large fluctuations for small noise; the singular values, however, are estimated more robustly. Thus, the visual analysis in Section 3.4 is less prone to noise than plotting the singular vectors directly.

Although the approach can be extended to larger dimensions, it is not straightforward. The concept of the checkerboard copula is viewed in a higher dimension, for example, in Carley and M. D. Taylor (2002). There is no direct analog of SVD in three and higher dimensions, but various approaches exist. See, for example, Kolda and Bader (2009) for an introduction. Copula-specific methods for modeling high-dimensional data include vine copulas (Joe, 1996; Bedford and Cooke, 2002; Panagiotelis et al., 2017; Czado, 2019) and nested Archimedean copulas (Savu and Trede, 2010; Hofert and Mächler, 2011), where the copulas involved could be analyzed using the methods presented here.

4. Combining Point Forecasts to Probabilistic Forecasts Using Copulas

This chapter introduces the copula combined density method to combine multiple point forecasts into a single calibrated probabilistic forecast using copulas. The chapter is based on “Combining Point Forecasts to Calibrated Probabilistic Forecasts Using Copulas”, joint work with Oliver Grothe (Publ. II). It was presented at the *DAGStat Conference 2022* and the *42nd International Symposium on Forecasting* (Conf. I; Conf. III).

The chapter is structured as follows. Section 4.1 motivates the new forecast combination algorithm based on relevant literature, and the contributions of this chapter are outlined. The existing algorithms for combining forecasts and their application are described in Appendix B.1. Section 4.2 establishes the copula combined density (CCD) method, which is motivated by copula time series models and a Bayesian derivation. Then, we prove its calibration for a known dependence and marginal structure and show that the forecast is equal to the combination of point forecast by Bates and Granger (1969) for Gaussian forecast errors. We give examples of the effect of model components and their misspecification on the combined forecast. The simulation studies in Section 4.3 compare the CCD method to other forecast aggregation algorithms and show the effect of the training size on estimation. These are followed by a worked-through example for employing the CCD method to energy price forecasts in Section 4.4. Section 4.5 concludes the chapter. The code for further combinations using the CCD method is available at `https://github.com/jo-rie/ccd`.

4.1. Introduction

Combining point forecasts from various sources into a distributional forecast not only gathers the information embedded in the different forecasts but also incorporates information on the uncertainty of the forecast into the result. Unlike point forecasts, distributional forecasts report not only a characteristic value of a forecast, for example, the mean, but also information on the uncertainty of the forecast by spread, prediction intervals, and tail decay (see Dawid, 1984; Tay and Wallis, 2000; Gneiting, Balabdaoui, and Raftery, 2007; Hall and Mitchell, 2007 and Section 2.2).

The point forecasts to be combined into a joint probability forecast might come from different institutions, models, or sensors measuring the same quantity with errors. However, overlapping information sets, model structures, or other influencing factors analogously affect several forecasts. They might introduce complex dependence structures between forecast errors (see, for example, the electricity price forecasts in Section 4.4). We propose a point forecast combination method that explicitly models the forecast dependence structure through a copula model and generates a combined density forecast. The copula model accounting for dependencies between the forecast errors can alter the combined forecast's location, spread, and shape. Generating a density forecast makes all statistical information of a distributional forecast available.

There are various forecast combination methods without dedicated dependence modeling in multiple applications. Forecast combination methods are, among others, used to combine weather predictions from various models or parameter configurations (Gneiting, Raftery, et al., 2005; Berrocal, Raftery, and Gneiting, 2007; Glahn et al., 2009; Kleiber, Raftery, and Gneiting, 2011), wind power generation (J. W. Taylor, McSharry, and Buizza, 2009), economic data (Stock and Watson, 2004) or multi-sensor data (Khaleghi et al., 2013). Early aggregation methods address the combination of airline passenger forecasts and are based solely on the correlation of the forecast errors instead of the complete dependence structure (Bates and Granger, 1969). More recent approaches for combining forecasts can be divided into methods for point forecast combination and methods for distributional forecast combination. The earliest of the latter ones is the traditional linear pool (TLP; Genest and Zidek, 1986), combining the distributional forecasts' CDFs linearly. The TLP tends to issue too wide forecasts (see Section 2.2), thus, the spread-adjusted linear pool (SLP) and beta-transformed linear pool (BLP) were proposed (Gneiting and Ranjan, 2013). They extend the TLP by a spread-adjustment parameter for the individual forecasts and the following transformation by the CDF of the beta distribution, respectively. Established algorithms for point forecast combinations are ensemble model output statistics (EMOS) in Wilks (2011) and Bayesian model averaging (BMA) in Raftery et al. (2005).

In general, the combination of point forecasts can profit from modeling the forecast error dependence, and the accuracy of the combined forecast can increase. Copulas are a well-known approach for modeling dependencies between random variables due to the possibility to differentiate between dependency and marginal modeling and are employed in various applications (Nelsen, 2006; Joe, 2014 and also Cherubini and Luciano, 2004; Grothe and Schnieders, 2011). Copulas are already being used in the context of forecasting as well. Schefzik, Thorarinsdottir, and Gneiting (2013) propose ensemble copula coupling for ensemble forecasts. A model based on the dependence structure of the ensemble members generates new ensemble realizations, that is, new samples of the same dimensionality.

Grothe, Kächele, and Krüger (2023) uses the Shaake shuffle to generate multivariate probabilistic forecasts from univariate point forecasts. Madadgar and Moradkhani (2014) use copulas within BMA for every member of the ensemble. In contrast, we use an ensemble of different forecasts in our setting to compute a single combined forecast. This problem is touched upon in the literature but has not yet received broader attention. Jouini and Clemen (1996) lies the ground by mentioning a copula combination for expert models in their highly influential research paper. Yazdandoost, Zakipour, and Izadi (2021) uses a copula to model the joint distribution of precipitation observations and forecasts and compute (combined) point forecasts. De Oliveira et al. (2017) uses copulas to compute maximum-likelihood (point) forecasts, and Mitchell (2013) proposes a method for combining density forecasts using copulas and, thus, directly employing the method of Jouini and Clemen (1996). Whereas ensemble copula coupling and copulas in BMA are widely accepted, copulas in the combination of point or density forecasts did not receive greater attention.

In this chapter, we formulate the CCD method. In its general notion formulated throughout the chapter, the CCD method combines point forecasts into a single, calibrated density forecast based on copulas and time series techniques. The point forecasts might be generated from uncalibrated density forecasts (see Section 4.4 for an example) so that the method combines uncalibrated densities into calibrated ones. Since the method embeds the approaches by de Oliveira et al. (2017) and Mitchell (2013), as a special case, the method's copula combination part can also directly combine several calibrated density forecasts. We prove the method's theoretical properties and analyze and validate them in simulation studies. Additionally, we give criteria in which the combined forecasts profit from the advanced dependence modeling. We expand on the procedure for fitting the marginals and copula and the influence of incorrect or incomplete characterizations of the marginal distributions. We apply the CCD method to combine the electricity price forecasts of Marcjasz et al. (2023).

The contribution of this chapter is manifold: We formulate a unifying approach to combine point or distributional forecasts into a common density forecast for a quantity of interest. We prove theoretical properties and evaluate the method's performance in simulation studies. The specification of model components is described, and the effects of misspecification are inspected. Through simulations, practical advice on the characteristics of the data, making it beneficial to use the method, is derived, for example, in terms of the necessary training sample size and the dependence structure. The methods are applied in a real-world data study, and the improvement in calibration is quantified.

4.2. The copula combined density (CCD) method

The copula combined density method is a method to combine K point forecasts from various sources into a single density forecast. Thereby, it uses the common density of all K point forecast errors $\boldsymbol{\varepsilon}_{t+h|t}$ issued at time t for time $t+h$:

$$f_{\boldsymbol{\varepsilon}_{t+h|t}}\left(\varepsilon^1_{t+h|t}, \dots, \varepsilon^K_{t+h|t}\right). \tag{4.1}$$

It lies at hand to model the density using a copula and thus split the individual error margin from the dependence structure model. We motivate Equation (4.1) in Section 4.2.1 and then derive the CCD forecast starting from the above equation in Section 4.2.2. Subsequently, we show that the forecasts produced are calibrated when knowing the dependence structure of the forecast errors and that it embeds the point forecast combination by Bates and Granger (1969) in Section 4.2.3. Section 4.2.4 reviews the effects of component misspecification on the CCD forecasts.

4.2.1. Modeling the forecast error density

We start by introducing some notation. Let

$$\hat{\mathbf{x}}_{t+h|t} = \left(\hat{x}^1_{t+h|t}, \dots, \hat{x}^K_{t+h|t}\right) \tag{4.2}$$

be K point forecasts for the value of some quantity of interest x_{t+h} at time $t+h$, made with the information available at time t, for forecasters $1, \dots, K$. We restrict x_{t+h} and $\hat{\mathbf{x}}_{t+h|t}$ to take values in $\mathbb{R}$ (resp. $\mathbb{R}^K$) and the set of time indices to be discrete. Let

$$\varepsilon^k_{t+h|t} = \hat{x}^k_{t+h|t} - x_{t+h} \sim F_{\varepsilon^k_{t+h|t}}(\cdot) \tag{4.3}$$

be the individual error of forecaster $k = 1, \dots, K$, which is unknown until $t+h$. Note that the error can alternatively be defined as the true value minus the forecast (Nowotarski and Weron, 2018). Let $f_{\varepsilon^k_{t+h|t}}$ denote the corresponding marginal PDF to $F_{\varepsilon^k_{t+h|t}}$. The vector of all individual point forecast errors is denoted by

$$\boldsymbol{\varepsilon}_{t+h|t} = \left(\varepsilon^1_{t+h|t}, \dots, \varepsilon^K_{t+h|t}\right) \sim F_{\boldsymbol{\varepsilon}_{t+h|t}}(\cdot, \dots, \cdot)$$

with common density of the forecast errors $f_{\boldsymbol{\varepsilon}_{t+h|t}}(\cdot, \dots, \cdot)$.

In practice, one has to estimate the distributions of $\varepsilon^k_{t+h|t}$ $(k = 1, \dots, K)$ and $\boldsymbol{\varepsilon}_{t+h|t}$ using historical data

$$\left\{\left(\varepsilon^1_{1+h|1}, \dots, \varepsilon^K_{1+h|1}\right),\ \dots,\ \left(\varepsilon^1_{T+h|T}, \dots, \varepsilon^K_{T+h|T}\right)\right\}.$$

In general, one could use a (non)parametric multivariate model for the density $f_{\varepsilon_{t+h|t}}$, like a multivariate Gaussian distribution or a multivariate kernel density estimation. These models are restrictive in the parametric case or suffer from the curse of dimensionality in the nonparametric case. On the contrary, using a copula allows splitting the margin and dependence structure model and provides flexibility and feasibility in higher dimensions (see Section 2.1).

By applying Sklar's theorem (Sklar, 1959), there is some copula $C_{\varepsilon_{t+h|t}}$ such that

$$F_{\varepsilon_{t+h|t}}\left(\varepsilon^1_{t+h|t}, \dots, \varepsilon^K_{t+h|t}\right) = C_{\varepsilon_{t+h|t}}\left(F_{\varepsilon^1_{t+h|t}}(\varepsilon^1_{t+h|t}), \dots, F_{\varepsilon^K_{t+h|t}}(\varepsilon^K_{t+h|t})\right). \tag{4.4}$$

Differentiation of Equation (4.4) yields

$$f_{\varepsilon_{t+h|t}}\left(\varepsilon^K_{t+h|t}\right) = c_{\varepsilon_{t+h|t}}\left(F_{\varepsilon^1_{t+h|t}}(\varepsilon^1_{t+h|t}), \dots, F_{\varepsilon^K_{t+h|t}}(\varepsilon^K_{t+h|t})\right) \cdot \prod_{k=1}^{K} f_{\varepsilon^k_{t+h|t}}(\varepsilon^k_{t+h|t}) \tag{4.5}$$

with $c_{\varepsilon_{t+h|t}} = \frac{\partial^K}{\partial u_1 \dots \partial u_K} C_{\varepsilon_{t+h|t}}(u_1, \dots, u_K)$ being the corresponding copula density to C.

The problem of finding and fitting the appropriate models for the components in Equation (4.5) is called multivariate copula time series forecasting. Models for marginal densities and copula in Equation (4.4) can be chosen independently. In our worked-through example in Section 4.4, we use neural networks and kernel density estimation for the marginal densities and a vine copula for their dependence structure.

For the general setup of multivariate time series forecasting, we refer to Patton (2012). Margins and copula can either be modeled parametrically or nonparametrically, respectively. Entirely (non-)parametric models use a (non-)parametric approach for the margins and the copula. In a semi-parametric approach, the margins are estimated nonparametrically, and the copula model is parametric. Patton (2012) suggests using the model with maximum likelihood on historical error data; however, scoring rules could also be used for model selection and estimation. Section 2.1 summarizes essential parametric copula models and references for nonparametric copula models.

In the next section, we use Equation (4.5) to derive a density forecast for x_{t+h}. However, other quantities of interest can be derived from Equation (4.5), such as a maximum likelihood point forecast for x_{t+h} (de Oliveira et al., 2017).

4.2.2. The CCD forecast

Through the methods of Section 4.2.1 we can model the density

$$f_{\varepsilon_{t+h|t}}\left(\varepsilon^1_{t+h|t}, \dots, \varepsilon^K_{t+h|t}\right) \overset{4.3}{=} f_{\varepsilon_{t+h|t}}\left(\hat{x}^1_{t+h|t} - x_{t+h}, \dots, \hat{x}^K_{t+h|t} - x_{t+h}|x_{t+h}\right) \tag{4.6}$$

for any marginal and dependence structure for continuous forecast errors. This denotes a density in $\boldsymbol{\varepsilon}$ or $\hat{\mathbf{x}}_{t+h|t}$ knowing x_{t+h}, respectively. At time t, the forecasts $\hat{\mathbf{x}}_{t+h|t}$ can be observed, but x_{t+h} remains unknown until $t + h$. Taking a Bayesian perspective, we can reformulate Equation (4.6) as a likelihood.

Instead of varying $\hat{\mathbf{x}}_{t+h|t}$ in Equation (4.5) and holding x_{t+h} fixed, x_{t+h} can be varied

$$\mathcal{L}_{t+h|t}(x_{t+h}) = f_{\varepsilon_{t+h|t}}\left(\hat{x}^1_{t+h|t} - x_{t+h}, \dots, \hat{x}^K_{t+h|t} - x_{t+h}|x_{t+h}\right). \tag{4.7}$$

Intuitively, Equation (4.5) then calculates the likelihood of getting the observed point forecasts for different true values x_{t+h}. By normalising Equation (4.7), one gets the forecast density $\hat{f}_{t+h|t}$ for x_{t+h}

$$\hat{f}_{t+h|t}(x_{t+h}) = \left(\int_{\mathbb{R}} \mathcal{L}_{t+h|t}(y)dy\right)^{-1} \cdot \mathcal{L}_{t+h|t}(x_{t+h}). \tag{4.8}$$

$\hat{f}_{t+h|t}(x_{t+h})$ is the CCD forecast for the random quantity x_{t+h}.

Remark 4.2.1. *The numerical computation of the normalizing constant $\int_{\mathbb{R}} \mathcal{L}_{t+h|t}(y)\ \mathrm{d}y$ might not be necessary if only samples of the distributions are of interest. Some random number generation methods work with likelihood and do not require a normalized density, for example, the Metropolis-Hasting algorithm (Hastings, 1970; H. Wang et al., 2017).*

If one of the forecasts is missing for a time step, the CCD can easily be adapted by integration over $\mathbb{R}$. Let the j-th forecast be missing. Then, the CCD forecast is

$$\begin{aligned}
\hat{f}_{t+h|t}(x_{t+h}) \propto \mathcal{L}_{t+h|t}(x_{t+h}) \\
&= c^{-j}_{\varepsilon_{t+h|t}}\big(F_{\varepsilon^1_{t+h|t}}(\hat{x}^1_{t+h|t} - x_{t+h}), \dots, F_{\varepsilon^{j-1}_{t+h|t}}(\hat{x}^{j-1}_{t+h|t} - x_{t+h}), \\
&\qquad F_{\varepsilon^{j+1}_{t+h|t}}(\hat{x}^{j+1}_{t+h|t} - x_{t+h}), \dots, F_{\varepsilon^K_{t+h|t}}(\hat{x}^K_{t+h|t} - x_{t+h})\big) \cdot \\
&\prod_{k=1, k\neq j}^{K} f_{\varepsilon^k_{t+h|t}}(\varepsilon^k_{t+h|t})
\end{aligned}$$

where $c^{-j}_{\varepsilon_{t+h|t}}$ denotes the marginal copula density without j. This approach generalizes naturally to more than one missing forecast.

4.2.3. Properties of CCD forecasts

We prove two outstanding properties of the CCD method: combining a vast range of point forecasts to a calibrated forecast and embedding the well-known combination by Bates and Granger (1969).

Theorem 4.2.1 (Calibration of the CCD method). *Let $\hat{\mathbf{x}}$ be continuous random point forecasts for X as motivated in Section 4.2.2 with finite error density. Let $\hat{f}$ be the CCD forecast for X using the true error marginals and copula and $\hat{F}$ the corresponding CDF. Then the CCD forecast is calibrated, that is*

$$Z_{\hat{F}} = F(X) \sim U(0,1).$$

The proof of Theorem 4.2.1 is given in Appendix B.2.1. Since any multivariate distribution that admits a density can be expressed in the copula representation, any point forecast combination with bounded density can be combined into a calibrated forecast.

For two forecasts with bivariate Gaussian errors, that is, they have Gaussian marginals and a Gaussian copula, the CCD density forecast can be computed explicitly and takes the form of a Gaussian PDF. Its parameters are equal to the forecast error variance minimizing combination stated by Bates and Granger (1969).

Theorem 4.2.2. *Let x be the quantity of interest, $\hat{x}_1$ and $\hat{x}_2$ point forecasts for x and $\varepsilon_i = \hat{x}_i - x$ the error of forecast i for $i \in \{1,2\}$. Let ε_1 and ε_2 be jointly Gaussian with mean $\mathbf{0}$, standard deviations σ_1 and σ_2, and correlation coefficient ρ. Then, the CCD forecast $\hat{f}$ according to Equation* (4.8) *is normally distributed with mean*

$$\mu_c = k \cdot \hat{x}_1 + (1-k) \cdot \hat{x}_2, \ \textit{with}\ k = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{4.9}$$

and standard deviation

$$\sigma_c = \frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}}. \tag{4.10}$$

The proof of Theorem 4.2.2 follows directly from Equation (4.8). The necessary calculations are carried out in Appendix B.2.2.

Bates and Granger (1969) establish the density forecast by minimizing the mean squared error. Thus, the CCD method minimizes the mean squared error for two jointly Gaussian forecast errors. Whether this holds for other margins and copulas is an open question.

4.2.4. Effects of model components

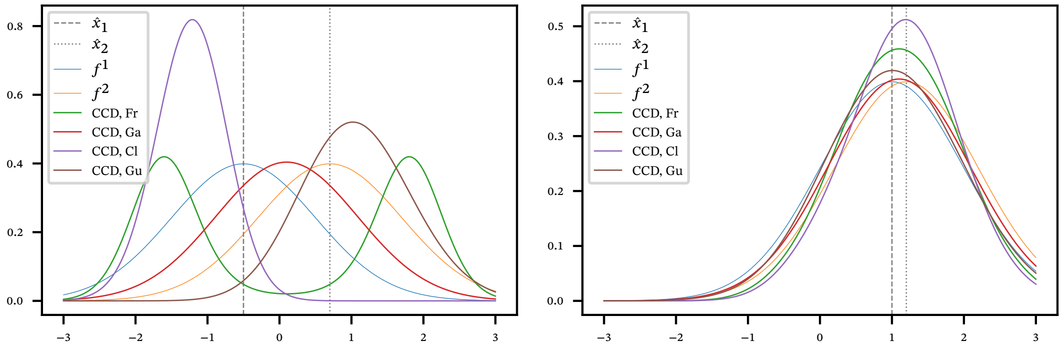
The following section expands on the average effects of the model components on the combined forecast, shows which PIT histograms indicate incorrect specification of which components, and thus lays the foundation for explaining the CCD method's performance in the simulations of Section 4.3. In two simulations, either the copula family or the marginal distribution is specified incorrectly, and the resulting forecasts are displayed through their PIT histograms. Of course, incorrectly specified margins influence the copula fitting process, typically leading to misspecified copulas. Thus, the two effects shown here can be mutually dependent.

Misspecified copula

Figure 4.1 starts by visualizing the effect of the copula on a single CCD forecast. For point forecasts $\hat{x} = (-0.5, 0.7)$ and standard normal margins, the combined forecasts resulting from different copula families with $\tau = 0.8$ are shown in Figure 4.1a. The point forecasts are far from each other in quantile terms (0.31, 0.76). Thus, when assuming a Clayton copula, the two forecast errors are assumed to lie in the upper tail, as in the lower tail, widespread points are unlikely due to the lower tail dependence. Thus, the combined forecast is shifted to the left. For the Frank copula, the point forecasts are assumed to lie in one of the tails; thus, the combination is bimodal, with modes below the lower and above the higher point forecast. The Gauss combination lies between the two point forecasts, and for the Gumbel copula, the reverse of the Clayton copula is the case.

Figure 4.1b shows a similar plot for $\hat{x} = (1, 1.2)$. The point forecast quantiles lie close together (0.84, 0.88), and the CCD combinations differ not as wildly as for Figure 4.1b. The combinations for Clayton and Gumbel are shifted to the converse directions, as argued above. Using the Frank or Gaussian copula, the mean lies within the two point forecasts. In summary, the copula has a major influence on the shape and the width of the combined prediction.

The sample results in Figure 4.2a show the averaged effect of the copula specification on the combined forecasts: For point forecast errors with Frank, Gaussian, Clayton, and Gumbel copulas, 10 000 two-dimensional point forecasts are generated and evaluated using all those copulas. Standard normal distributions are used in the generation of point forecast errors and the evaluation in the CCD method. Figure 4.2a shows results for $\tau = 0.4$, while Figure 4.2b shows results for $\tau = 0.8$. For the small value of $\tau = 0.4$, the influence of the copula family on the calibration is relatively small. If Clayton and Gumbel are mistaken, the histograms contain discernible deviations from equal height bars. For $\tau = 0.8$, the influence of the copula on the calibration is larger. The histograms exhibit skewness



(a) $\hat{x} = (-0.5, 0.7)$

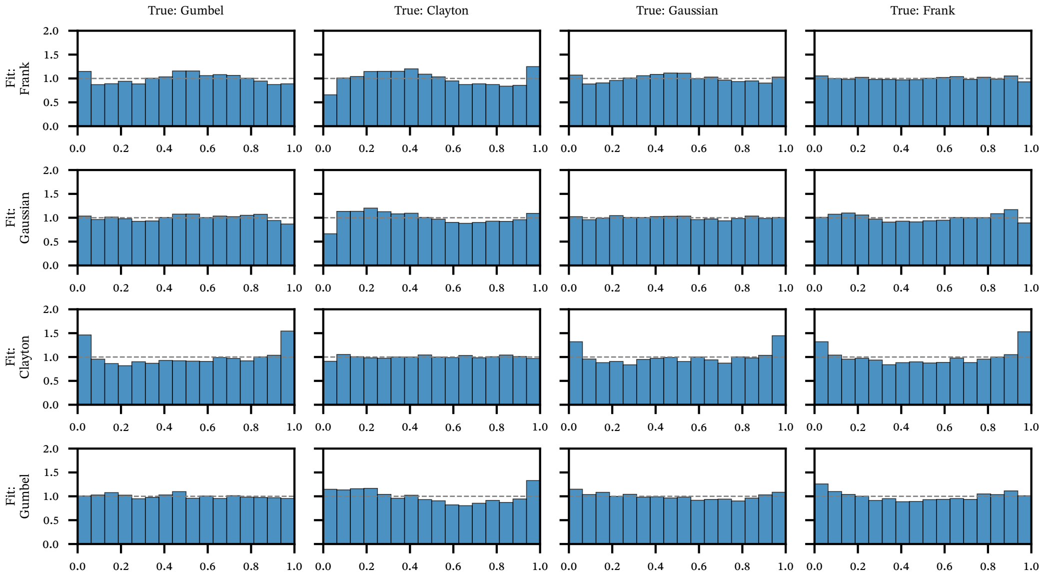
(b) $\hat{x} = (1, 1.2)$

Figure 4.1.: Illustrative CCD combinations for two point forecasters with standard normally distributed forecast errors and different parametric copula families. All copulas have parameters such that Kendall's $\tau = 0.8$. Depending on the copula family, the CCD forecast mean is shifted, and the spread is decreased compared to the two raw forecasts as different regions of the copula are likely to have quantiles being far apart or close together.

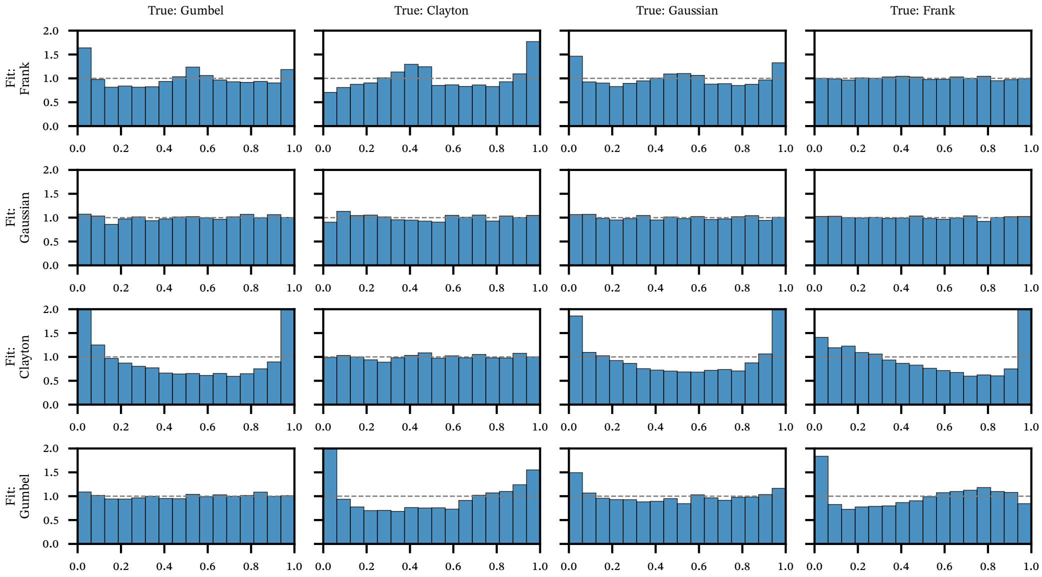
and underdispersion for incorrect copulas, for example, if a Clayton instead of a Gumbel copula is used in the CCD method. As argued above and in Figure 4.1, the copula in the CCD combination impacts the shift and width of the combined forecast. Again, misspecifications, including Clayton and Gumbel copula, influence the resulting histograms most because they have inverse tail dependence properties. The results for using a Gaussian copula seem well-calibrated for $\tau = 0.8$ and normal margins in Figure 4.2b. However, this is not the case for $\tau = 0.4$ or non-normal margins (see Figure B.1). For normal margins, Figure 4.1 illustrates how the Gaussian copula leads to less sharp densities than the true model. Thus, fitting the true copula in the CCD method is also in the case of normal margins and strong dependencies between the forecasts favorable.

Misspecified margins

Figure 4.3 shows the effect of incorrect margins on the combined forecast. The point forecast errors are generated using a t distribution with mean 0, variance 1, and 4 degrees of freedom. In the CCD method, a standard normal distribution is used. Thus, the mean and variance are specified correctly, but the tails decay too fast. 10 000 samples are generated for different copula families, each with $\tau = 0.4$. The true copula but the wrong marginal distributions are used in the evaluation. The combined forecasts are overdispersed like the marginal forecasts. Thus, the too-light tails carry through the CCD method. For Clayton and

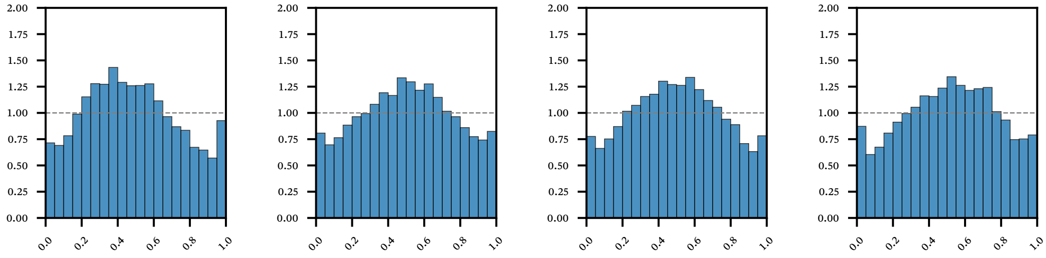


(a) $\tau = 0.4$. The effect of an incorrect copula estimation is rather small for small rank correlations.



(b) $\tau = 0.8$. For higher rank correlations, the effects are significant and unsymmetric.

Figure 4.2.: Effect of a misspecified copula on the calibration of the CCD forecast. 10 000 points are evaluated using the “fit” copula family for data stemming from the “true” copula. Both copulas use the same, “true” τ. The margins are standard normal.



(a) Clayton **(b)** Frank **(c)** Gaussian **(d)** Gumbel

Figure 4.3.: Effect of a misspecified margin on the calibration of the CCD forecast. For all copulas, the combined forecast is overdispersed; for Gumbel and Clayton copula, the histogram is, in addition, unsymmetric.

Gumbel copula, this is particularly evident in the regions of tail dependence, where more values than predicted lie in the tail.

4.3. Simulations

In this section, simulation studies evaluate the overall performance of the CCD method and an appropriate training sample size. In Section 4.3.1, we compare the CCD method to other forecast combination algorithms, that is, TLP, SLP, BLP and EMOS (see Appendix B.1). Section 4.3.2 reviews the effect of the training sample on the resulting CCD forecast if an automatic fitting algorithm is used.

4.3.1. Comparison to other forecast combination algorithms

We conduct a simulation study that underpins the ability of the CCD method to produce calibrated forecasts for individual forecasts with a complicated dependence structure compared to the forecast combination algorithms (see Appendix B.1). We identify situations where CCD is particularly powerful by varying the marginal error distribution, the copula family, and the rank correlation of the errors. Table 4.1 summarizes the considered marginal distributions, copulas, and rank correlations. For $\tau = 0$, all copulas are equal to the independence copula. Therefore, distinguishing between different copulas is redundant in this case. As Section 4.2.2 points out, CCD can combine any number of point forecasts. We restrict the simulation to the case of two forecasts to maintain the number of combinations feasible.

All scenarios are evaluated through the same scheme. Based on the copula and margins, 6 000 two-dimensional error data points are sampled and divided into a training data set

Component	Values	
Copula	Clayton, Frank, Gaussian, Gumbel	
Margins	$\mathcal{M}_1$	Uniform on $\left(-\sqrt{12}/2, \sqrt{12}/2\right)$
	$\mathcal{M}_2$	Gaussian $(0, 1)$
	$\mathcal{M}_3$	Student's t with $\left(\mu = 0, \sigma = \sqrt{1/2}, \nu = 4\right)$
τ	0, 0.4, 0.8	

Table 4.1.: Overview of the components in the simulation study in Section 4.3.1. The marginal distribution parameters are chosen to have the same mean and variance.

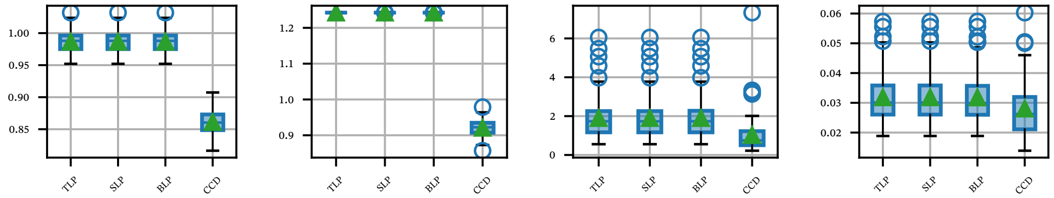
of 5 000 points and an evaluation set containing 1 000 points. The algorithms TLP, SLP, BLP, and EMOS are fit numerically using the logarithmic score minimizing parameters for the training data set. The CCD method's copula family and parameter are fit based on the AIC using the `pyvinecopulib` python library's `select` function (Nagler and Vatter, 2021), whereby the set of copulas contains all parametric copulas of the package without the BB7 copula due to numerical instabilities in the tail. The TLP, SLP, BLP, and the CCD method use the true marginal distributions. We report the EMOS results only for normal margins, as we use an EMOS formulation resulting in a normal distribution. To our knowledge, no generalizations of the EMOS approach to uniform or t distributions are available, and the optimization for those distributions is not straightforward. BMA with access to the true marginal distribution results in the TLP; thus we do not include BMA separately. The overall approach of CCD is compared to EMOS and BMA. Through the comparison with TLP, SLP, BLP, we analyze the copula-combination-step in the CCD method.

Then, the different methods are evaluated based on the test set data points. KS and AD test statistics assess the probabilistic calibration of the resulting forecast combination. Furthermore, the RMSE of the forecast combination means is calculated to evaluate the forecast location and the LS as a simultaneous measure of calibration and sharpness.

Each combination of the simulation scheme is run 100 times. Boxplots summarize the results for each scenario and each evaluation measure separately. We show some boxplots here to underpin the most significant results. The remaining figures are in the appendix. Table 4.2 links each simulation scenario to the associated figure.

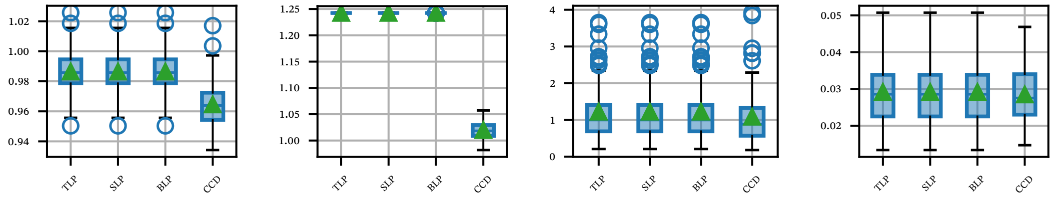
Results

The simulations show that modeling the dependence structure of the forecast through its copula in CCD increases the power of the combined forecast in all the aspects considered.



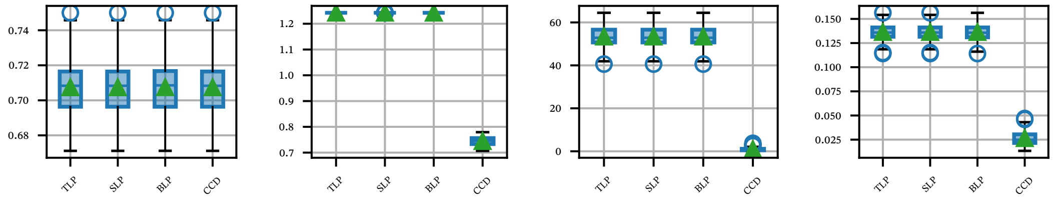
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure 4.4.: CCD simulation results for the scenario with margins $\mathcal{M}_1$, $\tau = 0.8$, and a Clayton copula.



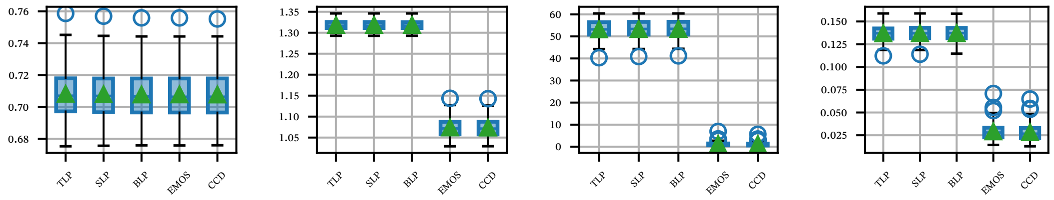
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure 4.5.: CCD simulation results for the scenario with margins $\mathcal{M}_1$, $\tau = 0.8$, and a Gumbel copula.



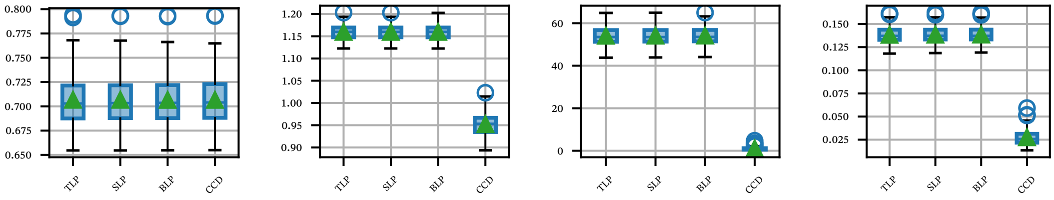
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure 4.6.: CCD simulation results for independent forecasts with margins $\mathcal{M}_1$.



(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure 4.7.: CCD simulation results for independent forecasts with margins $\mathcal{M}_2$.



(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure 4.8.: CCD simulation results for independent forecasts with margins $\mathcal{M}_3$.

The CCD forecasts always perform at least comparably well as any other forecast method in terms of probabilistic calibration, forecast means, and combined calibration and sharpness.

The means of the combined forecasts differ slightly between the different methods. For dependence structures with tail dependence, for example, Clayton and Gumbel copula with $\tau = 0.8$, the CCD means are closer to the realization. With an upper tail dependence, CDF-transformed forecasts lying close together in the upper tail cause the CCD-combined forecast to shift to the left. The other algorithms cannot deal with this situation separately. Figure 4.5 shows this as an example for standard normal margins, $\tau = 0.8$, and the Gumbel copula. The forecast mean is considerably closer in terms of the RMSE than for the other forecast combination algorithms. For a lower tail dependence, the same applies the other way around (see Figure 4.4).

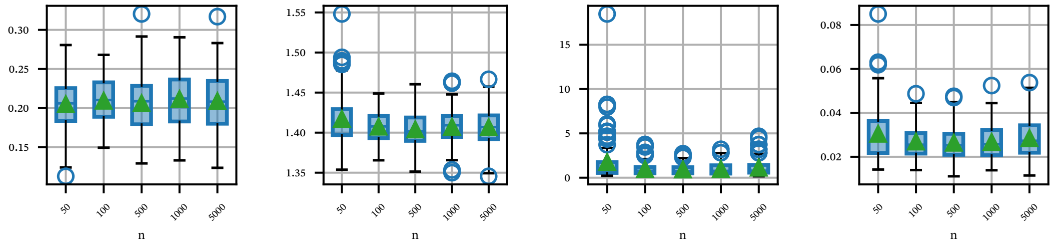
The calibration differs widely for the different algorithms in the scenarios. The TLP is particularly poorly calibrated for $\mathcal{M}_1$ and $\mathcal{M}_3$ with low rank correlations (Figures 4.6 and 4.8). As proven in Gneiting and Ranjan (2013), the TLP combined forecasts are uncalibrated in general for calibrated individual forecasts, while BLP can combine any marginal forecast to a calibrated forecast. Nevertheless, for high-rank correlations, TLP, SLP, and BLP perform comparably well. Low-rank correlations cause TLP and SLP to produce uncalibrated forecasts.

EMOS is only evaluated for margins $\mathcal{M}_2$ and the only algorithm in which calibration and score evolve incoherently. The forecasts are for scenarios with large τ among the worst calibrated (for example, Figure 4.4 and B.6), but EMOS performs better in terms of the logarithmic score.

Overall, the CCD method remarkably outperforms the other algorithms in the scenarios with nonlinear dependence while being systematically at least as good as the other method in all cases.

τ		0.4			0.8	
	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
Clayton	Fig. B.2	Fig. B.8	Fig. B.16	Fig. 4.4	Fig. B.12	Fig. B.20
Frank	Fig. B.3	Fig. B.9	Fig. B.17	Fig. B.6	Fig. B.13	Fig. B.21
Gaussian	Fig. B.4	Fig. B.10	Fig. B.18	Fig. B.7	Fig. B.14	Fig. B.22
Gumbel	Fig. B.5	Fig. B.11	Fig. B.19	Fig. 4.5	Fig. B.15	Fig. B.23

Table 4.2.: List of evaluation figures for the simulations in Section 4.3.1. Some figures are displayed in Appendix B.3.

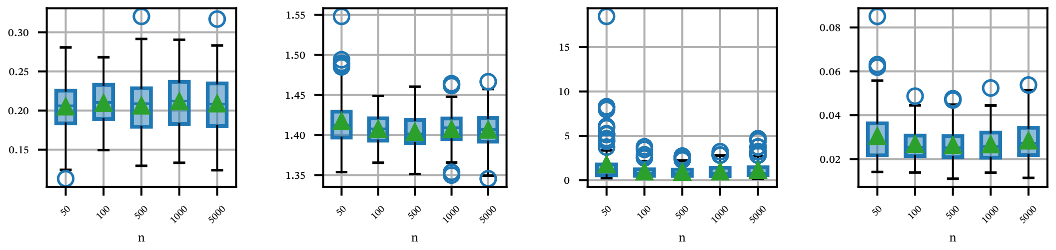


(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure 4.9.: CCD simulation results for different training sizes n and an independence copula.

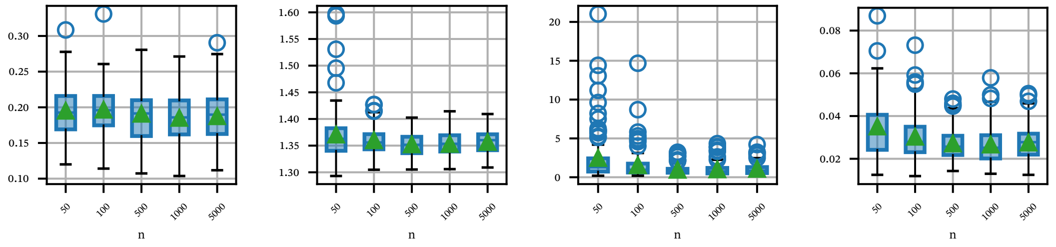
4.3.2. Effect of the training sample size

The following simulation investigates the likelihood of misspecification of the copula, family, and parameter with respect to the training sample size. Thus, it shows how much training data should be available to apply the method reasonably. The marginal distributions are standard normal for two forecasters and are known to the CCD method. The copula is automatically adjusted using the `select`-function of the package `pyvinecopula` (Nagler and Vatter, 2021) using a sample of n points. The CCD method is evaluated on 1 000 data points with the estimated copula. The procedure is repeated 50 times for each configuration of copula and training sample size n. Figures 4.9, 4.10 and 4.11 show the results for the independence, the Gaussian and Gumbel copula, the latter two with $\tau = 0.8$. For all copulas, the fit is reasonably good for $n = 500$ in terms of all measures considered and for all true copulas. For larger n, the fluctuation, possibly due to the influence of outliers, increases slightly. A smaller data size is sufficient for the simple dependence structure as for the independence copula in Figure 4.9.



(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure 4.10.: CCD simulation results for different training sizes n and a Gaussian copula with rank correlation $\tau = 0.8$.



(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure 4.11.: CCD simulation results for different training sizes n and a Gumbel copula with rank correlation $\tau = 0.8$.

4.4. Electricity price forecast combination

Electricity price forecasting (EPF) has received much attention in recent years as the increasing number of volatile electricity production plants, such as solar or wind power farms, raises large fluctuations in electricity prices. Although the literature on electricity price point forecasting is much broader (see, for example, the review in Lago et al., 2021), there are works on probabilistic electricity price forecasting (Nowotarski and Weron, 2018; Petropoulos et al., 2022; Grothe, Kächele, and Krüger, 2023). The distributional approach of Marcjasz et al. (2023) outperforms not only other probabilistic algorithms but also point prediction algorithms when using the forecasted mean as a point forecast in forecasting the day-ahead electricity prices for Germany. It uses distributional deep neural networks (DDNNs) for probabilistic forecasts of electricity prices with a wide range of input features. Two different parametric output distributions are considered in different setups: the normal and Johnson's SU (JSU) distributions. To account for random effects on hyperparameter tuning, the hyperparameters for the DDNN are fit four times, leading to four different DDNN for each of the distributions. Forecasts of the same distribution family are combined using a traditional linear pool with equal weights and horizontal averaging. We show in the following that the combined forecasts generated in Marcjasz et al. (2023) – although outperforming competing forecasting procedures – can be considerably improved in terms of calibration by using the CCD method to combine the results of the different models. The remainder of the section is structured as follows. We start by briefly describing the setup of the considered data. Then, we expand on the models considered in Section 4.4.1, including the model parameters and the CCD components. Section 4.4.2 evaluates the results and compares them to those stated in Marcjasz et al. (2023).

The model's input data are aligned with the availability of the data for the day-ahead auction. The spot market day-ahead auction takes place at noon, and all hours of the following day are traded in a uniform price auction. The models aim to predict Germany's hourly electricity price for the next day at 11:30 am. The data available to estimate the model are historical day-ahead electricity prices, day-ahead load forecasts, day-ahead combined renewable electricity sources forecasts, EU emission allowance prices, and coal, gas, and oil prices, as well as weekday dummies. The data comprises the years 2015 to 2020. The data is divided into several subsequent training periods for the different model components as the models build upon each other (see Figure 4.15). The initial training period of neural networks and hyperparameter tuning runs from January 1, 2015, to December 25, 2018. The marginal models and copula are trained with the data from December 26, 2018, to June 26, 2019, and June 27, 2019 to December 31, 2019, respectively. After the initial training, the models are retrained daily to perform the 24 hourly distributional forecasts and evaluated from January 1, 2020, to December 31, 2020.

4.4.1. Models

The models proposed in Marcjasz et al. (2023) are distributional deep neural network (DDNN) models, being neural networks where the last layer represents one or several distributions, for example, for subsequent time steps. Marcjasz et al. (2023) identify the normal and the JSU distribution as particularly powerful in EPF. The models have a complex structure with various hyperparameters to fit. Thus, the fitting is usually initialized at random. The hyperparameter tuning is repeated four times to account for random effects, yielding four different model setups for each distribution. For more information on the models, hyperparameters, and tuning, we refer to Marcjasz et al. (2023). We denote these models by DDNN-Normal-i and DDNN-JSU-i ($i \in [4]$). To combine the results, Marcjasz et al. (2023)

Model abbr.	Description
DDNN-Normal-i	Distributional deep neural network (DDNN) model by Marcjasz et al. (2023) forecasting normal distribution parameters.
DDNN-JSU-i	DDNN model forecasting JSU distribution parameters.
KDE-d-i	Kernel density estimation (KDE) estimation using the DDNN-d-i mean forecast as point forecast and past errors.
Normal-TLP	TLP with equal weights of the DDNN-Normal models.
JSU-TLP	TLP with equal weights of the DDNN-JSU models.
CCD-All	CCD model using the KDE-Normal-1 ... 4 as marginal models and a vine copula as dependence model. If *-Hourly* is appended, a separate copula per hour of the day is fit. All data is used to fit the vine copula.
CCD-1Y	Like CCD-All, but at most, the last year is used when fitting the copula.
CCD-13/14	Only KDE-Normal-1 and KDE-Normal-3 or KDE-Normal-1 and KDE-Normal-4 are used in the CCD method. The appendices are analogous to the ones for the CCD model with all marginals.

Table 4.3.: Model abbreviation overview for electricity price forecasting (EPF). The model number $i \in [4]$ refers to a specific result of the hyperparameter tuning, and d denotes the normal or JSU distribution.

suggests simple linear aggregation. We denote by Normal-TLP and JSU-TLP the TLP with equal weights of the DDNN-Normal and DDNN-JSU models.

The pair plots in Figures 4.14 and B.24 show that the point forecast errors of the different models exhibit high dependencies with Kendall's τ ranging from 0.726 to 0.776 for the DDNN-normal models. Thus, the simulations in Section 4.3 suggest that the combination benefits from using the CCD method. For the CCD method, we do not use the DDNN model output directly as the marginal distributions are not calibrated with skewed histograms and show slightly too light tails (see Figure 4.12). As the point forecasts by the DDNN perform well, we use them as point forecast input for the CCD method. We focus on normal forecasts as they produce better point forecasts in terms of MAE and RMSE and use a kernel density estimation (KDE) for the marginal distributions. KDE is a flexible and straightforward approach for data that do not follow a parametric distribution, as is the case here. We do not apply a further time series model, as the DDNN models already incorporate sequential time series information. The KDE uses the last half-year of data and a Gaussian kernel. The resulting PIT histograms for the KDE models are shown in Figure B.26 and provide reasonably calibrated, although not perfectly calibrated, marginals.

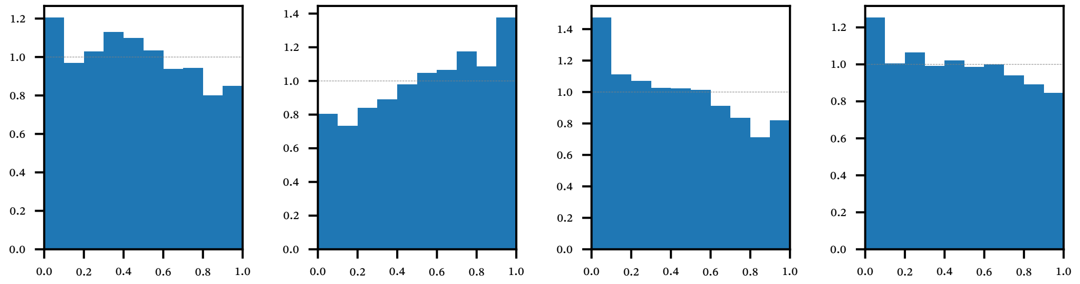
We employ a vine copula model for the dependence model in the CCD method (see Section 2.1), as it is a flexible and straightforward applicable model for high-dimensional dependence structures. We use the standard settings of Nagler and Vatter (2021) but exclude the BB7 copula as it suffers from numerical instabilities in the tails in our calculations. The copula is refit every 24 hours and then used to compute the 24 hourly probabilistic forecasts. Figure 4.15 visualizes the training periods for the different model components. Further information and executable code are available on GitHub (Link to be inserted).

We compute four different CCD models combining all four marginal models. We differentiate between the size of the copula training data and the number of fitted copulas per day. The size of the training data is all available data ("All") or data from at most the last year ("1Y"). The number of copulas is either one for all hours of the day or 24 if a separate copula is fitted for each hour ("Hourly"). Taking this together yields CCD-All, CCD-1Y for the one-copula-per-day models, and CCD-All-Hourly and CCD-1Y-Hourly. As noted in Hong et al. (2016), restricting the set of combined models to the best-performing models is often beneficial. Thus, we also compute the combination using only DDNN-Normal-1 and DDNN-Normal-3 and DDNN-Normal-1 and DDNN-Normal-4. We report only the non-hourly fitting for those combinations, as it performs considerably better on the combination of all models.

Figure 4.16 shows the copula fit for the combinations of DDNN-Normal-1 and DDNN-Normal-3 and DDNN-Normal-1 and DDNN-Normal-4. The fit of the copula remains stable over time, with a student's t copula having a rank correlation of approximately 0.65 and 0.72, respectively. Only in one of the fits is a BB1 copula used (Joe and Hu, 1996; Nikoloulopoulos, Joe, and H. Li, 2012). There is no simple way of visualizing the dependence structure over time for the higher-dimensional copula models.

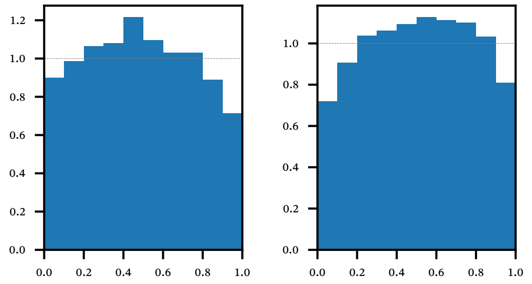
4.4.2. Results

We analyze the results of the different models through their PIT histograms and the numerical scores listed in Table B.1. The evaluation period runs from 2020-01-01 to 2020-12-31 for all models. Note that Marcjasz et al. (2023) uses a longer evaluation period and reports the average pinball loss, which produces approximately half as high as the CRPS values listed here. Figure 4.12 shows the resulting PIT histograms for the four individual DDNN-Normal models. All of them are skewed and show slightly too light tails. As typical for TLPs, the resulting combination is overdispersed as Figure 4.13 displays, regardless of distribution. Table B.1 also shows the KS test statistic of the PIT values assessing the calibration numerically. The TLP tends to be calibrated slightly worse than the best of the individual models for both distributions, but MAE, RMSE, and CRPS are improved considerably through the



(a) DDNN-Normal-1 **(b)** DDNN-Normal-2 **(c)** DDNN-Normal-3 **(d)** DDNN-Normal-4

Figure 4.12.: PIT histograms for the forecasts by the DDNN-Normal models with different hyperparameter setups for EPF. The analog results for the DDNN-JSU models are shown in Figure B.25.



(a) Normal-TLP **(b)** JSU-TLP

Figure 4.13.: PIT histograms for the TLP of normal or JSU models in EPF. The forecasts are overdispersed and skewed for the normal distribution.

combination. As the PIT histograms indicate, the TLP forecasts are too broad, producing a too large variance.

Figure 4.17 shows the resulting PIT histograms for the non-hourly fit of all models and the combination of DDNN-Normal-1 and DDNN-Normal-4. The histograms are remarkably better calibrated than any non-CCD method. The KS test statistics are approximately half the value of the TLP counterparts. For CCD-All, it is 0.0174, while it is 0.0399 for Normal-TLP. The forecast location is comparable to the TLP approach with a slightly better MAE and a slightly higher RMSE for all CCD models.

Combining the four models in CCD performs better than combining only two models. While the KS test statistics are better than for any of the non-CCD models, RMSE, MAE and CRPS are lower for CCD-All and CCD-1Y than for CCD-All-13 or CCD-1Y-13. The dif-

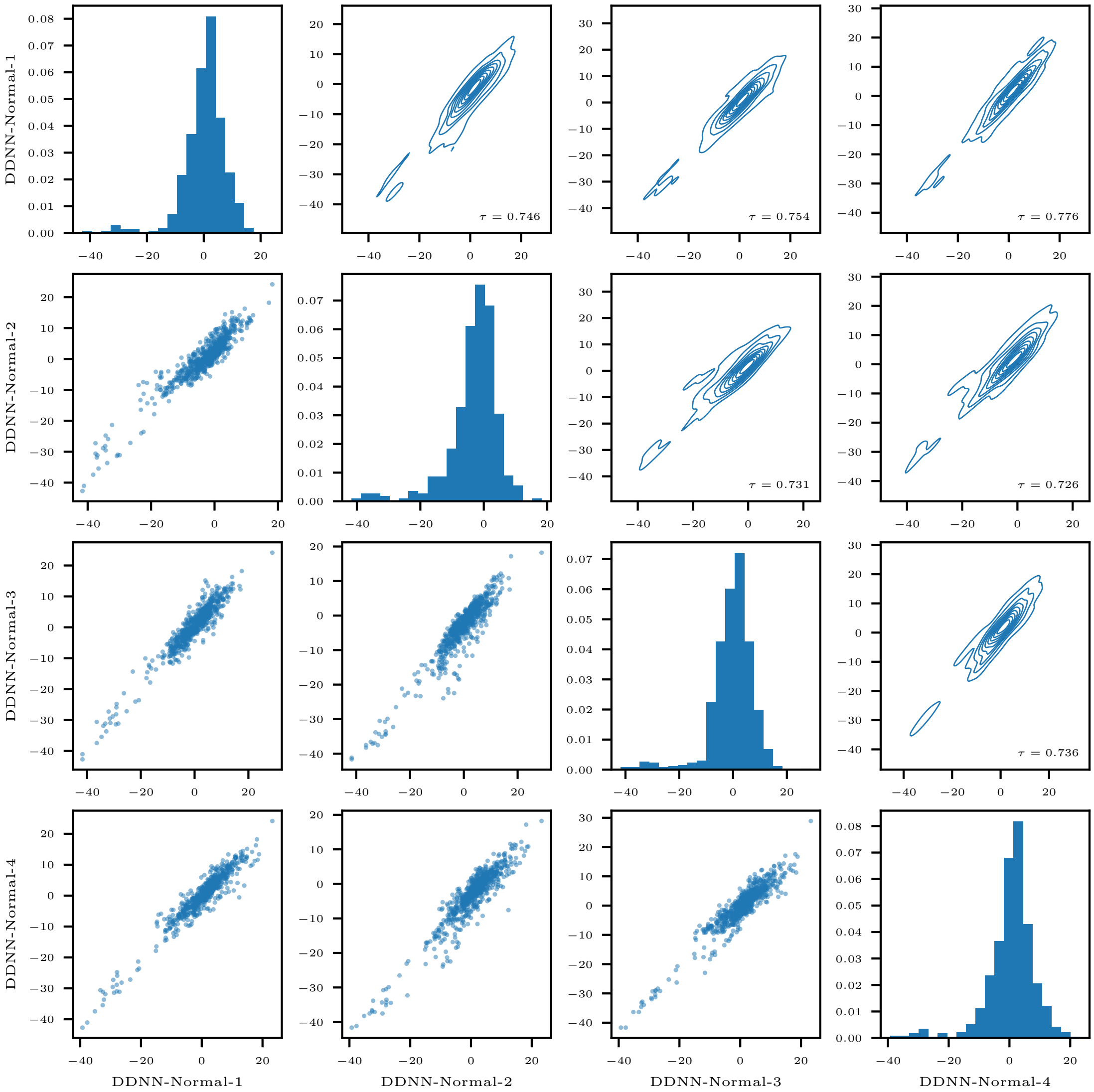


Figure 4.14.: Forecast error dependence plot for the DDNN-Normal models in EPF. The lower left plots display scatter plots of the errors, whereas the upper right plots visualize the contours of a two-dimensional kernel density estimation. On the diagonal, histograms of the errors for the individual models are shown. The resulting Kendall's τ is displayed in the lower right corner of the contour plots. The models have a strong pairwise dependence with many large negative errors.

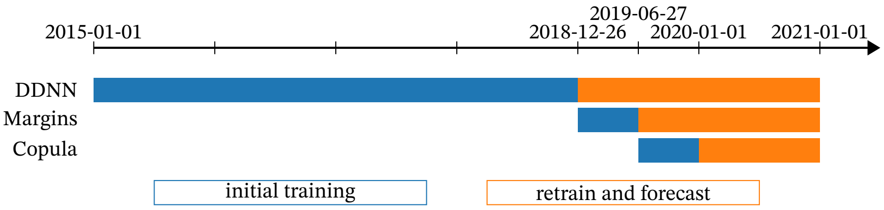


Figure 4.15.: Visualization of training and evaluation data sets for the different CCD model components in EPF.

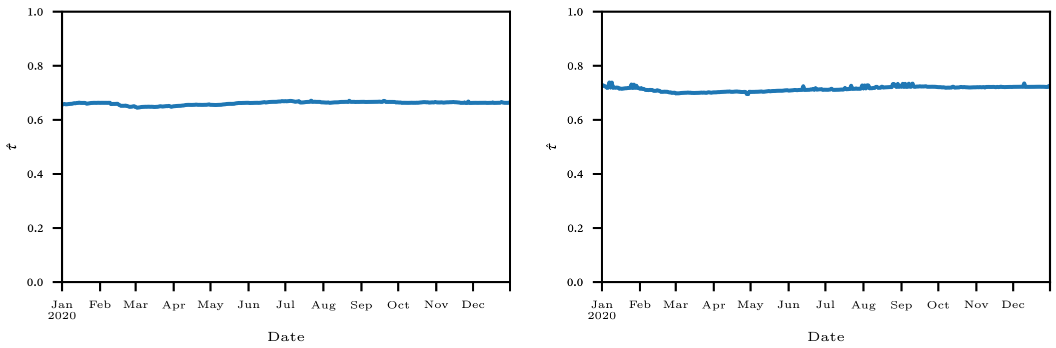
ference between CCD-All and CCD-1Y is slight, with CCD-1Y yielding better MAE, RMSE, and CRPS while being slightly less calibrated.

An overall copula for all hours yields a better fit than the hourly copula fitting (see Figures B.27 (a) and (b) in the appendix). The copula estimation for the hourly models has only 1/24th of the training data available. As argued in Section 4.3.2, this is insufficient for a reasonable copula fit, particularly in four dimensions. However, the results are better calibrated than for the non-CCD models. The combination of DDNN-Normal-1 and DDNN-Normal-3 produces better-located results with smaller CRPS and is comparably well calibrated compared to the combination of DDNN-Normal-1 and DDNN-Normal-4.

4.5. Conclusion

We formulate a point forecast combination method based upon the likelihood of a copula time series model, the copula combined density (CCD) method. It uses a copula model to combine individually fitted time series for every point forecast error to a joint density forecast. We show that this approach produces calibrated forecasts for correctly estimated time series and copula models. The CCD method embeds the squared error minimizing point forecast combination method by Bates and Granger (1969). We give remarks on missing forecasts and expand on methods for fitting model components, with a particular focus on estimating the copula.

In a simulation study, we identify scenarios in which CCD is superior to other algorithms, including TLP, SLP, BLP, and EMOS in terms of calibration and scores. The scenarios comprise, in particular, those with complicated dependence structures, e.g., high-rank correlation, asymmetric dependencies, and non-normal marginal distributions. In all the scenarios considered, the CCD method is at least as good as any other method. We describe and analyze the effects of incorrectly specified marginal and dependence models and conduct simulations on typical sample sizes to estimate the copula components sufficiently in stan-



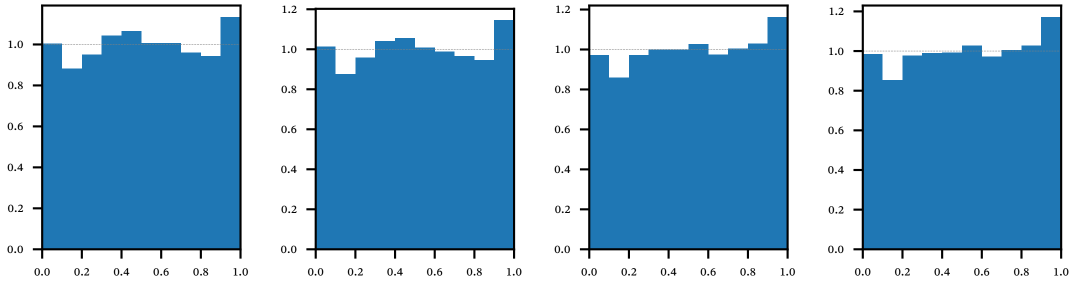
(a) Estimation for the combination of KDE-Normal-1 and KDE-Normal-3. The estimated copula is non-rotated student t.

(b) Estimation for the combination of KDE-Normal-1 and KDE-Normal-4. The estimated copula is non-rotated student t except for 2020-01-25, where a BB1 copula with a 180-degree rotation is fitted.

Figure 4.16.: Estimation of the copula rank correlation for the daily fitting in 2020 in the CCD method for EPF. The copulas are fit using all data from 2019-06-27.

dard settings. As a rule of thumb, 500 data points are necessary to estimate the components. This number might be strongly reduced if parts of the model components are known due to a-priori-knowledge or parametric assumptions.

We apply the method to hourly electricity price forecasting in Germany. We combine state-of-the-art distributional deep neural networks (DDNNs) outputs using a kernel density estimation of the margins and a vine copula for their dependence structure. This approach yields better-calibrated forecasts than the competing methods.



(a) CCD-All **(b)** CCD-1Y **(c)** CCD-All-13 **(d)** CCD-1Y-13

Figure 4.17.: PIT histograms for the CCD models using all data or at most one year of data to fit and using all DDNN-Normal models or only models 1 and 3.

5. Assessing the Ability to Track Changes for Forecasts, Nowcasts, and Measurements

This chapter proposes a toolbox for evaluating the ability to track changes for forecasts, nowcasts, and measurements. It is based on joint work with Bolin Liu, Bernd Saugel, and Oliver Grothe (Publ. III).

Section 5.1 begins with introducing the ability to track changes and the fields of nowcasting, forecasting, and measurement and reviewing existing evaluation schemes in the respective fields. Section 5.2 formalizes the ability to track changes and investigates several extensions, such as noise-aware methods, confidence intervals, and probabilistic forecasts and nowcasts. Additionally, we introduce a new graphical method, which can simultaneously assess different intervals and asymmetries. In Section 5.3, we show the results of applying our method to several practical examples from measurement, forecasting, and nowcasting. Hence, we provide blueprints for applying the proposed evaluation in practice and make remarks on interpreting the measures. We conclude the chapter in Section 5.4. Ready-to-use code for further assessment of the ability to track changes (ATC) is available at `https://github.com/jo-rie/aatc`.

5.1. Introduction

Measurements, nowcasts, or forecasts ideally should correctly reflect changes in the values of interest. It is thus important to meticulously assess the ability of measurements, nowcasts, or forecasts to correctly predict the direction of changes in values – which we refer to as the ability to track changes (ATC). Although measurements, nowcasts, and forecasts fundamentally differ as they either measure or predict a value, similar methods can be used to assess their ATC.

Forecasting methods predict the future based on historical data, patterns, or exogenous factors. A forecast is computed based on the current value of interest and an estimate of its future development. In medicine and healthcare, forecasting – for example – is used to predict patient volumes in emergency departments (Jones et al., 2008; Rostami-Tabar, Browell, and Svetunkov, 2023) or the demand of emergency medical services (Haugsbo Hermansen and Mengshoel, 2021).

Methodologically evolved from forecasting (Browning and Collier, 1989), nowcasting methods focus on predictions for the present, the immediate future, and the recent past (Bańbura et al., 2013; World Meteorological Organization (WMO), 2017). Nowcasting methods use high-frequency indicators or preliminary measurements related to the value of interest and focus on updating predictions using currently available information (Castle, Hendry, and Kitov, 2017). Nowcasting, for example, can assess the current situation during an ongoing epidemic, considering the main pathogenic, epidemiological, clinical, and socio-behavioral factors (Wu et al., 2021) or provide daily numbers of COVID-19 cases for events that have occurred but have not yet been reported (Günther et al., 2021; Wolffram et al., 2023).

Measurements aim to obtain accurate and precise values of a measurable quantity (measurand; Squara et al., 2021a; Squara et al., 2021b). Repeated measurements can be used to track changes in a value over time. When introducing new measurement methods, they are evaluated against current reference methods, often called the "gold standard", by simultaneously measuring the same quantity with the new method and the reference method – often in various individuals or different clinical settings.

In forecasting and nowcasting, the evaluation of the measurement performance is based on statistical methods quantifying the accuracy such as the RMSE, probabilistic scoring rules, and calibration measures (Gneiting, Balabdaoui, and Raftery, 2007; Günther et al., 2021; Wolffram et al., 2023). None of them assesses the method's ATC.

In a prediction competition on armed conflicts, the assessment of the ATC recently gained attention as Vesco et al. (2022) proposed the novel targeted absolute deviation with direction augmentation (TADDA) score with an additive tracking-changes-component for evaluation. However, the score poses an unintuitive incentive to forecasters and is thus theoretically problematic (Bracher et al., 2023). When evaluating the measurement performance of a measurement method, comparative statistics such as Bland-Altman analysis (Bland and Altman, 1986) and the percentage error (L. A. H. Critchley and J. A. J. H. Critchley, 1999) are commonly used. How to best assess the ATC of measurement methods is a field of ongoing research (Saugel, Grothe, and Wagner, 2015; Saugel, Grothe, and Nicklas, 2018; Hiraishi, Tanioka, and Shimokawa, 2021).

In this chapter, we focus on how to assess the ability of measurements, nowcasts, or forecasts to track changes. We formalize the concept of ATC and present visual techniques and quantitative measures to assess it – considering both noiseless data and data with noise and small non-informative changes. We introduce the conditional ATC plot, a new graphical

Application	Predicted change computation
Measurement	$(x_t - x_{t-l})_{t=l}^T$
Nowcasting	$\mathbf{x}^{\Delta,l} = \begin{cases} (x_{t\mid t} - x_{t-l\mid t})_{t=l}^T, & \text{if } y_{t-l} \text{ is not known at time } t, \\ (x_{t\mid t} - y_{t-l})_{t=l}^T, & \text{otherwise.} \end{cases}$
Forecasting	$\mathbf{x}^{\Delta,l} = (x_{t\mid t-l} - y_{t-l})_{t=l}^T$

Table 5.1.: Computation of the predicted change in the different applications. For nowcasting and forecasting, $x_{t|\tau}$ refer to values issued at τ with a target time t. For measurement, x_t denotes the test device measurement at time t.

method for assessing the local ability, and review bootstrap methods for calculating confidence intervals. We extend the concept of assessment to probabilistic predictions. We exemplarily illustrate the proposed methods to assess the ATC for nowcasting during the COVID-19 pandemic, patient admissions to an emergency department, and non-invasive blood pressure measurements – and thus provide blueprints for future assessments.

5.2. Assessment of the ability to track changes (ATC)

This section presents methods for assessing the ability to track changes for measurements, nowcasts, and forecasts. Section 5.2.1 introduces the notation and gives notes on the computation of changes in the different fields. Section 5.2.2 reviews the four-quadrant plot, a graphical method, and Section 5.2.3 derives numerical measures. In Sections 5.2.4 and 5.2.5, we extend the measures to account for noise and non-informative changes and introduce the conditional ATC plot. Section 5.2.6 considers measures and graphical tools for probabilistic forecasts and nowcasts.

5.2.1. Computing changes and notation

The assessment of ATC is based on the measured/observed/true and the predicted changes in a value of interest over a time horizon l. The *measured* change is straightforward to compute for all types of measurement, nowcast, or forecast. Let $\mathbf{y} = (y_t)_{t=0}^T$ denote the actual values for nowcasting or forecasting, or gold standard measurements up to time T. The sequence of changes is then given by the differences of values in $\mathbf{y}$ with horizon l, that is,

$$\mathbf{y}_t^{\Delta,l} = (y_t - y_{t-l}) \quad \text{for } t = l, \dots, T. \tag{5.1}$$

The definition of the *predicted* change depends on the context; Table 5.1 summarizes the notation for measurements, nowcasts, or forecasts and the computation of the predicted change. For nowcasting, let $x_{t|\tau}$ denote the nowcast for time t computed with the knowledge of time τ. We call t the *target time* and τ the *issue time*. The predicted change is computed by

$$\mathbf{x}^{\Delta,l} = \begin{cases} (x_{t|t} - x_{t-l|t})_{t=l}^{T} & \text{if } y_{t-l} \text{ is not known at time } t, \\ (x_{t|t} - y_{t-l})_{t=l}^{T} & \text{otherwise.} \end{cases} \tag{5.2}$$

When computing the predicted change of a nowcast for a time t, we use the best knowledge available at that time t, and the true value might not be known yet. If the true value y_{t-l} is known at time t, the predicted change is computed by the difference between the nowcast and the true value, as y_{t-l} is also known by the nowcaster and incorporated into the nowcast. Through the computation in Equation (5.2), the predicted change can be computed with the knowledge of the nowcaster at time t.

The notation is similar for forecasting: Let $x_{t|\tau}$ denote the forecast for target time t and issue time τ. The predicted change is computed by

$$\mathbf{x}^{\Delta,l} = (x_{t|t-l} - y_{t-l})_{t=l}^{T} \tag{5.3}$$

with the same structure as in the nowcasting case and consistent indices with $\mathbf{y}^{\Delta,l}$. If the true value y_{t-l} is not known at time $t-l$, a similar modification can be made as in Equation (5.2).

The distinction between forecast and issue time is unnecessary in measurement analysis, as the measurement is typically available with a very short time lag. Thus, x_t denotes the test method measurement for time t. The computation

$$(x_t - x_{t-l})_{t=l}^{T} \tag{5.4}$$

yields the change by the test method. It is computed purely by the test method without the gold standard y_t to analyze whether the gold standard and test method changes are consistent. Accordingly, y_{t-l} is not used in the computation even if known at time t in contrast to forecasting and nowcasting.

In applications, data are often not available for all time steps, for example, due to technical problems or delays in data transfer (see the examples in Sections 5.3.1 and 5.3.2). We refer to time steps for which either measurement, nowcast, or forecast or true values are unavailable as missing values. Systematical missing values could lead to a biased assessment, and missing data should be inspected for any underlying patterns. If the missing values are not systematic, random, and occur scarcely, data pairs with missing values can be excluded from the data to calculate the measures (see Van Buuren, 2018, Section 1.3). Note that in the case of measurement data, one missing value in the time series leads to two undefined differences in the change series; that is if x_t is missing, x_t^{Δ} and x_{t+l}^{Δ} are undefined; if an observation y_t is missing, y_t^{Δ} and y_{t+l}^{Δ} are undefined. The data pair is excluded even if the corresponding nowcast or forecast is available.

5.2.2. The four-quadrant plot

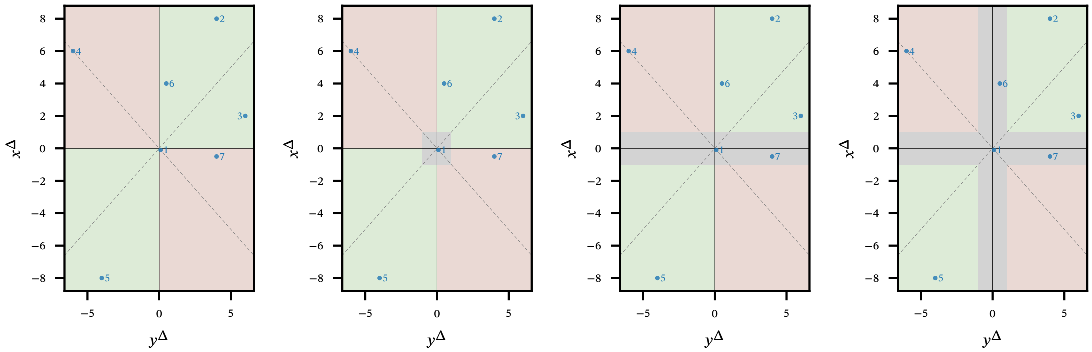
Formally, the assessment of ATC is the same for measurements, nowcasts, and forecasts, given the notation for the respective application of Section 5.2.1. In the following, we omit the horizon l for ease of notation; $\mathbf{x}^{\Delta}$ and $\mathbf{y}^{\Delta}$ refer to $\mathbf{x}^{\Delta,l}$ and $\mathbf{y}^{\Delta,l}$ for a common horizon l. The ATC is maximal if all predicted change directions are correct; that is, the sign of all elements of $\mathbf{x}^{\Delta}$ and $\mathbf{y}^{\Delta}$ coincide. Consequently, when assessing the ATC, we examine the statistical consistency of $\text{sign}(\mathbf{x}^{\Delta})$ and $\text{sign}(\mathbf{y}^{\Delta})$. A simple yet insightful method is the four-quadrant plot (see, e.g., Perrino, Harris, and Luther, 1998; Saugel, Grothe, and Wagner, 2015). In a four-quadrant plot, the occured changes and the predicted changes are plotted together, that is, $(y_t^{\Delta}, x_t^{\Delta})$ for $t = l, \dots, T$. Thus, the x-axis of a four-quadrant plot shows the true value differences, whereas the y-axis displays the prediction data differences. Points in the green upper right and lower left quadrants reflect a correct change direction for the respective time step, whereas points in the remaining red quadrants show incorrectly predicted changes. Figure 5.1a displays a basic four-quadrant plot, and Figure 5.2a shows a four-quadrant graph for simulated data with $T = 1461$, for example, four years of daily data (for the data generation, see Appendix C.1.1).

The four-quadrant plot can be extended by including information on the time index in the point color to reveal effects over time. In Figure 5.2b, the point colors turn from blue to green for higher time indices t, that is, more recent values; $(y_l^{\Delta}, x_l^{\Delta})$ is blue and turns green until $(y_T^{\Delta}, x_T^{\Delta})$. However, four-quadrant plots become crowded for larger datasets, and sequential information on the differences is complex to assess thoroughly.

The four-quadrant plot is intuitive to interpret, and the magnitude and direction of change are shown simultaneously. Other visualization techniques, such as polar plots, lack the four-quadrant plot's clarity and intuition without adding more information on the ATC (Saugel, Grothe, and Wagner, 2015).

5.2.3. The ATC ratio and other measures

Analyzing the number of points in the green versus red quadrants is a standard approach in the ATC assessment of measurement data (L. A. Critchley, Lee, and Ho, 2010; Saugel, Grothe, and Wagner, 2015). With that, we estimate the probability of a correctly predicted change direction, $P(X^{\Delta}Y^{\Delta} > 0)$, where Y^{Δ} and X^{Δ} denote random variables for future incremental changes. Since $z_1 z_2 > 0$ imposes the same condition as $\text{sign}(z_1) = \text{sign}(z_2)$



(a) Basic four-quadrant plot.

(b) Four-quadrant plot with rectangular exclusion area. Point 1 is excluded.

(c) Four-quadrant plot with horizontal exclusion area. Points 1 and 7 are excluded.

(d) Four-quadrant plot with cross-shaped exclusion area. Points 1, 6 and 7 are excluded.

Figure 5.1.: Illustrations of the four-quadrant plot with sample points and with and without exclusion areas. The rectangular exclusion area in Figure 5.1b excludes only points where both components are likely to be noise-driven, while the exclusion areas in Figures 5.1c and 5.1d exclude points where at least one component is noise-driven.

($z_1, z_2 \in \mathbb{R} \setminus \{0\}$), the standard estimator for $P(X^\Delta Y^\Delta > 0)$ is

$$\mu(\mathbf{x}^\Delta, \mathbf{y}^\Delta) := \frac{\sum_{t\in\mathcal{T}} \mathbb{1}\{x_t^\Delta y_t^\Delta > 0\}}{|\mathcal{T}|}. \tag{5.5}$$

Here, the numerator counts the number of same-sign-changes, while the denominator is the number of considered pairs (y_t^Δ, x_t^Δ). Thus, μ is the proportion of concordant changes on all changes. We refer to this estimator as the *ATC ratio* of the prediction and set $\mathcal{T} = \{l, \dots, T\}$. Visually, the measure computes the fraction of points in the upper right or lower left quadrant. Similar evaluations are used in other scientific areas, for example, with contingency tables in dichotomous forecasting or with confusion matrices in classification analysis (see, e.g., the introductions in James et al. 2021, Ch. 4, and Jolliffe and Stephenson, 2012, Ch. 3). Many other measures can be adapted from those fields to deepen the analysis. Two simple measures that focus on a positive or negative predicted change are the positive and negative

ATC ratios μ^+ and μ^-, respectively. They are defined as

$$\mu^+(\mathbf{x}^\Delta, \mathbf{y}^\Delta) := \frac{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{x_t^\Delta y_t^\Delta > 0\right\} \mathbb{1}\left\{x_t^\Delta > 0\right\}}{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{x_t^\Delta > 0\right\}}, \text{ and} \tag{5.6}$$

$$\mu^-(\mathbf{x}^\Delta, \mathbf{y}^\Delta) := \frac{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{x_t^\Delta y_t^\Delta > 0\right\} \mathbb{1}\left\{x_t^\Delta < 0\right\}}{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{x_t^\Delta < 0\right\}}. \tag{5.7}$$

They estimate the probability of a correct prediction of the direction of change, given that the predicted direction is positive or negative, that is, $P(X^\Delta Y^\Delta > 0 | X^\Delta > 0)$ and $P(X^\Delta Y^\Delta > 0 | X^\Delta < 0)$.

Rolling estimates of the above measures detect changes in performance over time and can give a sharper estimate of the current ATC. For the ATC ratio, a rolling estimate with a backward-looking window of length w at time t is given by

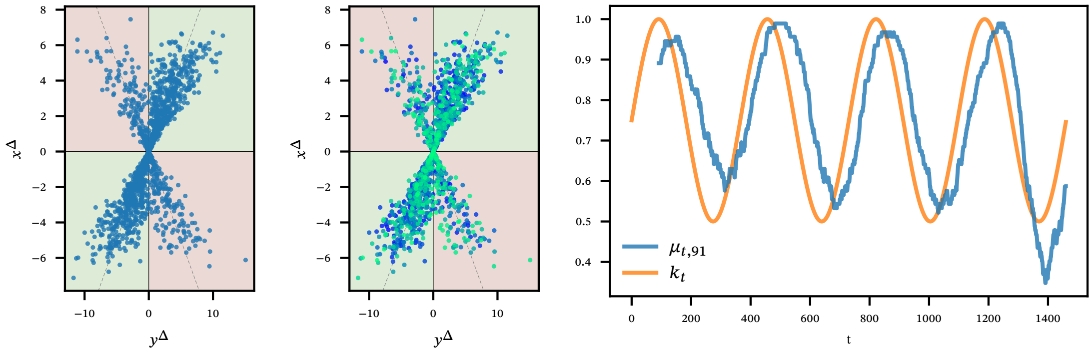
$$\mu_{t;w}(\mathbf{x}^\Delta, \mathbf{y}^\Delta) := \frac{\sum_{t^\star=t-w+1}^{t} \mathbb{1}\left\{x_{t^\star}^\Delta y_{t^\star}^\Delta > 0\right\}}{w}.$$

Backward-looking windows estimate the ATC ratio at a time t considering the w time steps before time t. The window length w controls the smoothing of the estimate; a larger w gives smoother results, while a small w focuses on local variations. Plotting the rolling estimates for $t = w - 1, \dots, T$ yields an estimate of the ATC ratio over time. Figure 5.2c depicts a rolling window estimate of the ATC ratio for the simulated data of Figures 5.2a and 5.2b. While colored four-quadrant plots, as in Figure 5.2b, illustrate ongoing overall drifts in the ATC, seasonal aspects are only revealed in rolling window estimates.

5.2.4. Accounting for noise and non-informative small changes and bootstrapping confidence intervals

The above measures can be extended to account for information on the point's location within the quadrant. For example, points close to the zero point may have less explanatory power or may be less reliable than points far away from zero on one of the diagonals. Suppose noise or non-systematic effects are present in the true values or predictions. In that case, noise can drive a point's assignment to a quadrant instead of a systematic ATC. This is more likely for points with at least one small coordinate.

Using an exclusion area around the zero point, as further defined below, is a straightforward and highly interpretable extension of the measures of Section 5.2.3 accounting for such effects (see, e.g., Saugel, Grothe, and Wagner, 2015; L. A. Critchley, Lee, and Ho, 2010). Points within that area are omitted in the calculation of the measures. In particular, the



(a) Four-quadrant plot with simulated data.

(b) The same data is colored according to the time index t, the greener, the later.

(c) Rolling estimate of the ATC ratio over time with window length 91.

Figure 5.2.: Visualizations for data with a time-varying ATC ratio. We defer information on the data generation process to the appendix (see C.1.1). The ATC ratio for the entire data set is $\mu = 0.7577$. The strong seasonality of the ATC ratio becomes visible in Figure 5.2c. The green curve k_t shows the theoretical probability that x_t^{Δ} has the same sign as y_t^{Δ} for each time step. The rolling estimates are shifted to the right compared to k_t as the windows look backward. Thus, the yearly course of the ATC ratio can be detected. The ATC ratio has a pronounced sinus-shaped seasonality with a peak after a quarter of a year and a low point after three quarters.

measurement, nowcast, or forecast is likely to have a noise component; thus, $\mathbf{x}^{\Delta}$ should be subject to an exclusion area. The measures of Equations (5.5), (5.6) and (5.7) without points in the exclusion area E are

$$\mu_e(\mathbf{x}^{\Delta}, \mathbf{y}^{\Delta}, E) := \frac{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{\mathbf{x}^{\Delta}\mathbf{y}^{\Delta} > 0\right\} \mathbb{1}\left\{(y_t^{\Delta}, x_t^{\Delta}) \notin E\right\}}{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{(y_t^{\Delta}, x_t^{\Delta}) \notin E\right\}}, \tag{5.8}$$

$$\mu_e^{+}(\mathbf{x}^{\Delta}, \mathbf{y}^{\Delta}, E) := \frac{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{x_t^{\Delta}y_t^{\Delta} > 0\right\} \mathbb{1}\left\{x_t^{\Delta} > 0, (y_t^{\Delta}, x_t^{\Delta}) \notin E\right\}}{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{x_t^{\Delta} > 0, (y_t^{\Delta}, x_t^{\Delta}) \notin E\right\}}, \text{ and} \tag{5.9}$$

$$\mu_e^{-}(\mathbf{x}^{\Delta}, \mathbf{y}^{\Delta}, E) := \frac{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{x_t^{\Delta}y_t^{\Delta} > 0\right\} \mathbb{1}\left\{x_t^{\Delta} < 0, (y_t^{\Delta}, x_t^{\Delta}) \notin E\right\}}{\sum_{t\in\mathcal{T}} \mathbb{1}\left\{x_t^{\Delta} < 0, (y_t^{\Delta}, x_t^{\Delta}) \notin E\right\}}. \tag{5.10}$$

The measures are then estimators for the probability of predicting the correct direction, given that the point's location is not driven by noise or non-informative changes.

The estimators accept various shapes of the exclusion area (see Figure 5.1). A rectangular exclusion area, $E = \{(y, x) \in \mathbb{R}^2 : (-\varepsilon_x \leq x \leq \varepsilon_x) \wedge (-\varepsilon_y \leq y \leq \varepsilon_y)\}$ for $\varepsilon_x, \varepsilon_y > 0$, leaves out points that are small in both components. An exclusion area along one axis, for example, $E = \{(y, x) \in \mathbb{R}^2 : (-\varepsilon_x \leq x \leq \varepsilon_x)\}$ for $\varepsilon_x > 0$, removes points in which one of the components could change sign by a small amount of noise. A cross-shaped exclusion area, $E = \{(y, x) \in \mathbb{R}^2 : (-\varepsilon_x \leq x \leq \varepsilon_x) \vee (-\varepsilon_y \leq y \leq \varepsilon_y)\}$ for $\varepsilon_x, \varepsilon_y > 0$, along both axes accounts for the sign reversal in both components.

In most applications, the shape and size of the exclusion area can be chosen based on domain knowledge or expert opinions. The size determination can also be based on a proportion of the total variance or the total range of the data; for example, the 10% smallest absolute values in each component determine the exclusion area size. A third approach is to visualize the ATC ratio for different sizes of E and thus inspect the effects of the exclusion area size on the estimates. For examples of such plots, see Section 5.3.1.

Confidence intervals can account for the estimation uncertainty of the measures above. Bootstrap confidence intervals are based on resampling and not on parametric assumptions as classical confidence intervals are (for introductions see Hesterberg, 2011; Bittmann, 2021). Many new samples are drawn with replacement from the dataset, and the statistic of interest is computed for each sample, yielding an estimate for the distribution of the statistic of interest. Based on the derived “new” samples of the statistic, the confidence intervals can be derived through different bootstrapping methods. We use the bias-corrected and accelerated (BCa) approach for bootstrapping in the following, as it holds the confidence level for small and large samples and has a moderate computation time (see the simulation study in Appendix C.1.2).

5.2.5. The conditional ATC plot

The estimators described above give information on the probabilities $P(X^\Delta Y^\Delta > 0 | X^\Delta Y^\Delta \notin E)$, $P(X^\Delta Y^\Delta > 0 | X^\Delta > 0, X^\Delta Y^\Delta \notin E)$ and $P(X^\Delta Y^\Delta > 0 | X^\Delta < 0, X^\Delta Y^\Delta \notin E)$. A still finer analysis might be gained by considering the conditional distribution $P(X^\Delta Y^\Delta > 0 | X^\Delta = \chi)$ to assess the ATC of a prediction for a specific change $X^\Delta = \chi$ of the measurement, nowcast, or forecast. Thereby, $P(X^\Delta Y^\Delta > 0 | X^\Delta = \chi)$ denotes the probability of a correct direction given a predicted change of χ. Thus, if a change of χ is observed in practice, one can directly assess its credibility. A multivariate kernel density estimation (KDE) facilitates a continuous estimation of $P(X^\Delta Y^\Delta > 0|X^\Delta = \chi)$ by estimating the components f_{X^Δ,Y^Δ} and f_{X^Δ} of

$$P(X^\Delta Y^\Delta > 0|X^\Delta = \chi) = \begin{cases} \int_{-\infty}^{0} \frac{f_{X^\Delta,Y^\Delta}(\chi,y)}{f_{X^\Delta}(\chi)} \,\mathrm{d}\,y & \text{if } \chi < 0, \\ \int_{0}^{\infty} \frac{f_{X^\Delta,Y^\Delta}(\chi,y)}{f_{X^\Delta}(\chi)} \,\mathrm{d}\,y & \text{if } \chi > 0, \end{cases}$$

for $\chi \neq 0$ through a KDE. Gramacki (2018) provides a comprehensive introduction to multivariate KDE, and implementations are available in many programming languages, for example, in the `statsmodels` in Python (Seabold and Perktold, 2010). The KDE yields estimates for $P(X^\Delta Y^\Delta > 0|X^\Delta = \chi)$ for all values of $\chi \in \mathbb{R}$. Multivariate KDE takes a kernel and bandwidth selector as modeling parameters. We advise using a Gaussian kernel and the cross-validation maximum likelihood as bandwidth selector (see Appendix C.1.3).

Assessing $P(X^\Delta Y^\Delta > 0|X^\Delta = \chi)$ graphically by drawing $P(X^\Delta Y^\Delta > 0|X^\Delta = \chi)$ against χ eases the simultaneous evaluation of various χ. Furthermore, the graph facilitates the comparison of various methods in a single graph, and asymmetries of $P(X^\Delta Y^\Delta > 0|\mathbf{x}^\Delta = \chi)$ with respect to χ in the ATC can be detected. We refer to the plot as a *conditional ATC plot*.

5.2.6. Probabilistic evaluation

In nowcasting and forecasting, probabilistic predictions have become more prevalent in recent years (see Section 2.2 and the applications in Sections 5.3.1 and 5.3.2). In this section, we develop ATC assessments for probabilistic measurements and nowcasts. Probabilistic predictions issue a probability distribution for the quantity of interest based on their available information and, thus, include a point estimate and information on the prediction uncertainty and quantiles simultaneously. Probabilistic predictions thus also contain a probability of a positive or negative change. For ATC assessment, we compare the predicted probability of positive change, denoted by p_t, with the occurrence of positive changes.

Probabilistic predictions can be a CDF, PDF, or quantiles. The CDF is the most general and can be used to derive the others, given that they exist. Let us first assume that the prediction is a CDF, and that y_{t-l} is known at time t (see Table 5.1). Appendix C.1.4 extends the analysis to quantile predictions or unknown true values.

Let for a forecast $F_{t|t-l}(x)$ denote the predicted CDF for target time t and issue time $t-l$, where the index is analogous to the point notation of Section 5.2.1. The CDF $F_{t|t-l}(x)$ specifies the forecasted probability that the quantity of interest is at most x. A positive change occurs for any value at t larger than the true value y_{t-l} and the CDF $F_{t|t-l}(y_{t-l})$ yields the predicted probability of any value at most y_{t-l}, and, thus, a negative change. Accordingly, the forecasted probability of a positive change is

$$p_t = 1 - F_{t|t-l}(y_{t-l}) \quad t = l, \dots, T.$$

The computation differs slightly for nowcasts, that is,

$$p_t = 1 - F_{t|t}(y_{t-l}) \quad t = l, \dots, T,$$

with analogous derivations as above. Let z_t denote the indicator that the observed change at time t is positive, that is,

$$z_t = \mathbb{1}\left\{y_t^{\Delta} > 0\right\} \quad t = l, \dots, T.$$

The predictive power of $\mathbf{p} = (p_t)_{t=l}^{T}$ for $\mathbf{z} = (z_t)_{t=l}^{T}$ can be assessed using probabilistic dichotomous forecast evaluation methods. Dichotomous forecasts predict a binary outcome, such as a positive or negative change, and are evaluated numerically using scoring rules or visually through reliability diagrams. For general notes on assessing probabilistic forecasts, see Section 2.2.

The BS is a widely used scoring rule for dichotomous probabilistic forecasts (Brier, 1950). In our context, it is

$$BS(\mathbf{p}, \mathbf{z}) = \frac{1}{T - l + 1} \sum_{t=l}^{T} (p_t - z_t)^2.$$

Lower values indicate the considered method's higher probabilistic ATC. The BS assesses the calibration and sharpness of the forecast and the observation simultaneously (Ranjan and Gneiting, 2010; Mitchell and Wallis, 2011). Calibration refers to the statistical consistency of forecasts and observations; that is, the event occurs with the issued probability and is considered the more fundamental quality (Gneiting, Balabdaoui, and Raftery, 2007). Sharpness refers to the spread of the forecast; probabilities close to zero and one are preferable as they convey a higher certainty.

Graphical methods are a standard tool for evaluating the calibration of probabilistic forecasts in detail. In dichotomous forecasting, the reliability diagram is frequently used (Ranjan and Gneiting, 2010). The reliability diagram plots the observed frequency of the positive outcome against the (binned) predicted probability. For example, it shows the proportion of observed increases, given that the predicted probability of increase was approximately 0.7. Ideally, the predicted probability equals the observed frequency, and the reliability diagram is a 45-degree line. Local deviations from the 45-degree line indicate a miscalibration for specific forecast probabilities. Thus, the reliability visualizes the local and overall calibration simultaneously. For an example of a reliability diagram, see Section 5.3.2.

5.3. Application to medical/healthcare nowcasting, forecasting, and measurement data

This section applies the ATC assessment to three different datasets from the medical and healthcare sector to illustrate the usage. The respective code is supplied on Github to ease further applications. Section 5.3.1 analyzes nowcasts for the seven-day hospitalization rate during the COVID-19 pandemic in Germany. In Section 5.3.2, the ATC of forecasts for the number of patient admissions to a hospital's emergency department is assessed. Section 5.3.3 examines the assessment of blood pressure measurements.

5.3.1. Nowcasting during the COVID-19 pandemic

In Germany, the seven-day hospitalization rate was established as a central steering measure in November 2021 during the COVID-19 pandemic, and the imposition of severe public restrictions was based on it (Robert Koch Institute, 2021). However, the publication of the definite hospitalization rate was substantially delayed and partially flawed for two main reasons. First, the reporting process was delayed because – among other reasons – different authorities were involved in passing the data to the RKI (Robert Koch Institute, 2024). Second, the seven-day hospitalization rate allocated all COVID-19-related hospitalizations to the date of the first positive test (for a detailed description, see Wolffram et al., 2023). The COVID-19-Nowcasting-Hub (C19-Hub) collected various nowcasts in a predefined setup, including the mean, median and other quantiles of the predicted seven-day hospitalization rate (for further information see Wolffram et al., 2023 and Table C.2 for the abbreviations used). In addition to those nowcasts, Wolffram et al. (2023) construct two ensemble methods using the ensembles' mean or median. We denote them by ENS-MEAN and ENS-MED. In line with the initial study design, we consider the period from November 22, 2021, to April 29, 2022, as the evaluation period. We use the data from February 8, 2024, for the true values and focus on nowcasts for all inhabitants of Germany. Figure C.4 in Appendix C.2 displays the true and nowcast data for the evaluation period. The time comprises the fourth wave's end in December 2021 and nearly the entire fifth wave of the pandemic in Germany, lasting until May 28, 2022 (Tolksdorf, Loenenbach, and Buda, 2022).

To assess the impact of taken measures and the direction of the curve, the knowledge of whether hospitalization rates rise or fall is essential. Thus, the ATC assessment is particularly relevant for the nowcasts of the seven-day hospitalization rate. If hospitalization rates rise, measures should be tightened, while falling rates might allow for loosening measures. Asymmetries are especially relevant for assessing whether some models are better at recognizing a fall than a rise or vice versa.

Model	RMSE	MAE	Count
ILM	648	504	153
RKI	810	670	156
RIVM	820	674	159
ENS-MED	832	675	158
ENS-MEAN	841	666	158
LMU	979	810	159
SZ	1,048	834	159
SU	1,127	899	159
KIT	1,161	912	159
EPI	1,513	1,006	159

Table 5.2.: Point evaluation measures for the issued mean of the different models in COVID-19 nowcasting. “RMSE” and “MAE” are accuracy measures, while “Count” lists the number of non-missing values. The RMSE orders the models. The evaluation period comprises 159 days, and only a few nowcasts are missing (for explanations of the missing values, see Wolffram et al., 2023, Tables A2, A3, and A4). Note that the high values for the EPI model could be driven by an exceptionally far-off value at the end of the evaluation period (see Figure C.4).

Results

Table 5.2 summarizes the non-ATC-aware point evaluation measures for the issued mean of the different models. The best-performing models in terms of RMSE and MAE are the ILM and RKI models. The ensemble methods ENS-MED and ENS-MEAN perform worse than the best models regarding the mean location. The performance of the models is diverse, with more than twice as high RMSE values for the worst models compared to the best models.

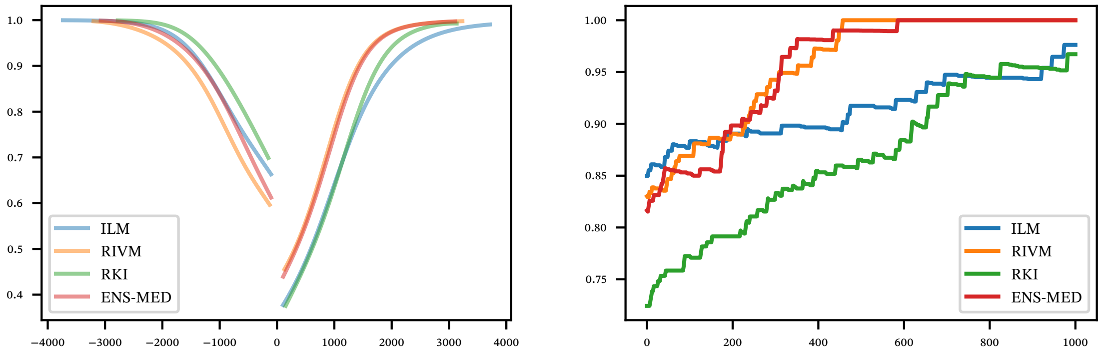
In the following, we apply the ATC assessment for the short-term horizons one and medium-term horizons seven and 14 days. The horizons seven and 14 reflect a typical period until new policy changes are taken. We start by providing background information on the marginal distributions of the actual value and nowcast changes for the different horizons in Table C.3 in Appendix C.2 such as standard deviation and quantiles of the nowcasts and true values. The variability and general level of changes grow with the horizon: The standard deviation increases from roughly 300 for horizon one to 1,200 for horizon seven and 2,000 for horizon 14 days. Similarly, the 10%-quantile of changes, the basis for the exclusion area size, increases. The exclusion area is rectangular; a point falls within it if both $\mathbf{y}^{\Delta}$ and $\mathbf{x}^{\Delta}$ are below the respective 10%-quantile of the absolute changes. Thus, points are still included in the ATC assessment if they are large in one dimension but not in the other, thus ensuring that substantial changes in, for example, $\mathbf{y}^{\Delta}$ are to be recognized by the nowcast and vice versa.

	μ^7	$\mu^{+,7}$	$\mu^{-,7}$	$\mu^7_{q_{0.1}}$	$\mu^{+,7}_{q_{0.1}}$	$\mu^{-,7}_{q_{0.1}}$
EPI	0.77 (0.71, 0.82)	0.67 (0.58, 0.75)	0.87 (0.79, 0.92)	0.78 (0.72, 0.83)	0.68 (0.59, 0.77)	0.88 (0.81, 0.93)
ILM	0.85 (0.80, 0.89)	0.73 (0.64, 0.80)	0.99 (0.94, 1.00)	0.85 (0.80, 0.90)	0.74 (0.65, 0.81)	0.99 (0.94, 1.00)
KIT	0.74 (0.69, 0.79)	0.64 (0.55, 0.72)	0.87 (0.80, 0.93)	0.75 (0.69, 0.80)	0.64 (0.55, 0.72)	0.88 (0.81, 0.94)
LMU	0.80 (0.74, 0.85)	0.70 (0.62, 0.79)	0.91 (0.84, 0.95)	0.81 (0.75, 0.86)	0.72 (0.63, 0.79)	0.92 (0.85, 0.96)
ENS-MEAN	0.82 (0.76, 0.86)	0.71 (0.63, 0.79)	0.94 (0.89, 0.99)	0.82 (0.76, 0.87)	0.71 (0.63, 0.78)	0.96 (0.90, 0.99)
ENS-MED	0.82 (0.76, 0.87)	0.70 (0.62, 0.78)	0.96 (0.90, 0.99)	0.83 (0.77, 0.87)	0.72 (0.63, 0.79)	0.96 (0.90, 0.99)
RIVM	0.83 (0.77, 0.87)	0.74 (0.65, 0.81)	0.92 (0.86, 0.96)	0.83 (0.78, 0.88)	0.74 (0.65, 0.81)	0.93 (0.87, 0.97)
RKI	0.72 (0.65, 0.77)	0.60 (0.51, 0.67)	0.98 (0.92, 1.00)	0.73 (0.67, 0.78)	0.61 (0.52, 0.68)	0.98 (0.92, 1.00)
SU	0.81 (0.75, 0.86)	0.71 (0.62, 0.78)	0.92 (0.85, 0.96)	0.81 (0.75, 0.85)	0.71 (0.63, 0.79)	0.92 (0.85, 0.96)
SZ	0.78 (0.72, 0.83)	0.67 (0.58, 0.75)	0.91 (0.84, 0.96)	0.78 (0.72, 0.83)	0.67 (0.58, 0.75)	0.92 (0.85, 0.97)

Table 5.3.: The ATC ratio μ^7, positive ATC ratio $\mu^{+,7}$, and negative ATC ratio $\mu^{-,7}$ for the models without and with exclusion areas for the horizon seven days in COVID-19 nowcasting. The exclusion areas are rectangles centered on the zero points with a width and height of twice the 10%-quantile of the absolute values of nowcast and true values. The subscript $q_{0.1}$ denotes the measures with exclusion area.

Table 5.3 lists the ATC ratios for all models without and with exclusion areas for the horizon of seven days. The ATC ratios without exclusion area range from 0.72 to 0.85 for the horizon of seven days. The negative ATC ratios are higher than the positive ATC ratios for all models. The confidence intervals for the positive and negative ATC ratios do not overlap for all models, indicating that the ATC ratios are indeed different. The 10%-quantile exclusion areas have, at most, an influence of 0.03 on the ratios. The model with the highest ATC ratio is the ILM model, and the model with the lowest is the RKI model. The confidence intervals between all models without and with exclusion areas overlap. The positive ATC ratio implies a similar ranking of the models than the overall ATC, while the negative ratio provides a different ranking, for example, for the RKI model. For the horizons of one and 14 days, we refer to Table C.4 in Appendix C.2.

Figure 5.3 shows the conditional ATC plots and the ATC ratio over the exclusion area for the horizon seven days; the respective plots for the horizons one day and 14 days are shown in Figure C.6. Here, only the best models in point evaluation measures, ILM, RKI, RIVM, and ENS-MED, are shown to keep the plots easily readable. If RKI or ILM issues a fall in the hospitalization rate, the probability of a fall is higher than if RIVM or ENS-MED issues a fall. The opposite is the case for a nowcasted hospitalization rate increase, and the difference between the models' performance is more prominent than for a fall. Similar observations can be made for the horizon of 14 days in Figure C.6b. For a horizon of one day, the models'



(a) Conditional ATC plot.

(b) ATC ratio over exclusion area size in $\mathbf{x}^{\Delta}$.

Figure 5.3.: Conditional ATC plot and ATC ratio over exclusion area for the nowcasts of the seven-day hospitalization rate ILM, RKI, RIVM, and ENS-MED for the horizon seven days in COVID-19 nowcasting.

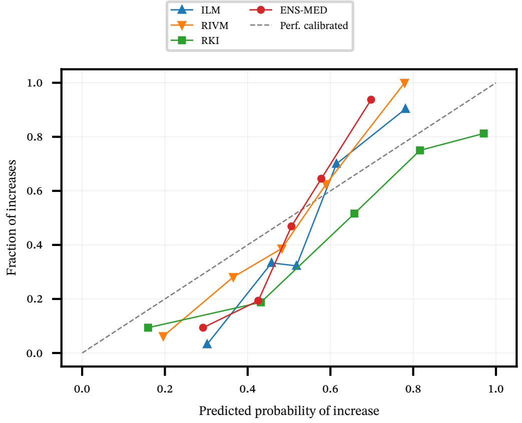
conditional ATC difference is less pronounced (see Figure C.6a). The RKI model is still less conclusive when issuing an increase in the hospitalization rate, while RIVM is most informative in that case. The curves cross for an issued fall, with ENS-MED being on top for issued falls above 250.

The ATC ratios for various exclusion areas are shown in Figure 5.3b. The ATC ratio generally increases with larger exclusion areas. While the RIVM and ENS-MED ATC ratios evolve similarly, the RKI and ILM ATC ratios get closer. For the horizon of one day, the RKI ATC ratio decreases with increasing exclusion area size while the other models rise (see Figure C.6c). For the horizon of 14 days, all ATC ratio curves increase with the exclusion area size (see Figure C.6d).

Figure 5.4 shows the Brier score (BS) and reliability diagrams for the same subset of models, the ILM, RKI, RIVM, and ENS-MED. The probabilities of increase for the different models are computed using the nowcast quantiles. For each horizon l, $10,000$ samples of the forecast date t and the forecast date $t-l$ based on the nowcasts of issue date t are generated, and the proportion of positive changes is computed (see Appendix C.1.4). Remember that a low BS and a reliability diagram along the diagonal are signs of a high ATC. The BS is the lowest for the RIVM model for a one-day-horizon, while the ENS-MED model has the best BS for the horizon of seven and 14 days. The RKI model yields the highest BS for all horizons. Note that the BS is 0.25 for random guessing; thus, all models perform better than random guessing. The reliability diagrams show that the models are not well calibrated for the horizon of one day. While for predicted probabilities below 0.5, the observed ratio of increases is smaller, it is higher than predicted for probabilities above 0.5. Figure C.7c shows

	1 d	7 d	14 d
ILM	0.1783	0.1269	0.1119
RIVM	0.1606	0.1136	0.1274
RKI	0.1893	0.2200	0.1712
ENS-MED	0.1812	0.1096	0.1066

(a) BS for the different models and horizons.



(b) Reliability diagram for horizon one day.

Figure 5.4.: Brier scores and reliability diagrams for the COVID-19 nowcasting models ILM, RKI, RIVM, and ENS-MED. The reliability diagram bins are chosen according to the empirical quantiles of the predicted probabilities. In the computation of BS and reliability diagram, missing values are excluded. The reliability diagram for the horizons seven and 14 days is in the appendix (see Figure C.7).

the observed predicted increase probabilities for all horizons. For the horizons of seven and 14 days, the nowcasters issue only a few moderate probabilities, and most probabilities are near zero and one.

Discussion

For all horizons, the influence of the exclusion area on the 10%-quantile level is negligible. For example, the ATC ratio changes at most by 0.03 for the EPI model with $\mu^{-,14}$. The exclusion areas are thus not crucial for the ATC assessment in the case of the nowcasts of the seven-day hospitalization rate. The lower bound of confidence intervals is at least 0.68 for all models, indicating that they perform better than random guessing the trend.

ATC assessment evaluates the models differently from point evaluation measures. RKI is among the best in point evaluation measures but performs worse in ATC assessment. The assessment of asymmetry in the conditional ATC plots is crucial for interpreting the ATC ratios, with the RKI model being the most prominent example.

Figure 5.3b shows that larger exclusion areas increase the ATC ratio, indicating that the predicted direction is more accurate for large predicted changes.

The probabilistic ATC assessment shows that the models are better than random guessing. The reliability diagram cannot provide information if specific probabilities are issued scarcely. Thus, the reliability diagrams for the horizons of seven and 14 days do not contain information on moderate probabilities. The BS values, however, work well for those examples and provide a good measure for the ATC of the models.

A more extensive data size would be beneficial for assessing the models' performance. For the evaluation period of 159 days, the ATC ratio confidence intervals overlap; thus, no conclusions can be drawn from the ATC assessment comparing the models.

5.3.2. Forecasting patient admissions to an emergency department

In a second example, we consider forecasting patient admissions to an emergency department per hour with data and models by Rostami-Tabar, Browell, and Svetunkov (2023). Every 12 hours, the models issue hourly forecasts for the next 48 hours.

Rostami-Tabar, Browell, and Svetunkov (2023) publish means and probabilistic quantile forecasts for various models and input data. We use the published mean as a point forecast for the ATC assessment and evaluate the probabilistic ATC based on the quantile forecasts subsequently. Considering only the forecasts of at least 36 hours ahead, we restrict the evaluation period to March 2, 2018, at noon, to February 28, 2019, at 23:00, comprising 8,724 hours.

In this setup, ATC assessment is a simple and intuitive way to assess the models' performance. The ATC perspective is easy for the management to understand and implement, as simple comparisons of the expected workload to a recent shift can be made. If, for example, the staff was near capacity during the last shift and an increase in the number of patient admissions is expected, management can take measures to adjust the workload.

The number of patient admissions has a strong weekly and daily pattern. Thus, we consider the horizons of 72 hours, the last already observed shift of the same hour of day, and seven days, the previous shift of the same hour and day, in ATC assessment.

Results

Table 5.4 lists the point evaluation measures and the count of available forecasts. The best-performing models regarding RMSE and MAE are the NBI-2 and Poisson-2 models. More than 8,600 forecasts are available for all models, with changes in the number due to missing values on four afternoons in 2018.

We start by analyzing the marginal distributions for the predicted and observed changes for the three- and seven-day horizons in Table 5.5, again. The computed difference aligns with Section 5.2.1, that is, the difference between the forecasted mean and true value of three and seven days before, as the actual value is available when issuing the forecast. The positive change fraction varies between 0.39 and 0.63 for the horizon of three days and between 0.37

Model	RMSE	MAE	Count
NBI-2	8.883	3.200	8688
Poisson-2	8.884	3.200	8688
Poisson-1	9.164	3.238	8688
Benchmark-2	9.246	3.236	8688
Ttr-2	9.394	3.266	8688
NOtr-1	9.413	3.276	8688
NOtr-2	9.413	3.276	8688
Poisson-2-I	9.458	3.276	8688
Benchmark-1	10.065	3.331	8688
GBM-2	11.663	3.542	8688
tbats	12.905	3.912	8724
Prophet	13.078	3.877	8724
qreg-1	13.337	3.758	8688
Regression-Poisson	21.162	4.818	8724
ADAM-iETSX	28.000	5.561	8724
ETS	29.358	5.742	8724

Table 5.4.: Point evaluation measures for the forecasting models for patient admissions to an emergency department. The smaller count for some models stems from missing forecasts scattered throughout the evaluation period. Note that the reported values for the RMSE differ from those in Rostami-Tabar, Browell, and Svetunkov (2023) due to differences in the evaluation period.

and 0.63 for the horizon of seven days. The variability of changes decreases for the larger horizon for most models; only for the ETS model does it increase. The 10%-quantile of the changes is between zero and one for all models and horizons. Thus, we use an exclusion area of size 1. The resulting fraction of included values in the computation is also listed in Table 5.5 and is at least 79% of the values.

Table 5.6 lists the ATC ratios for all models for three and seven-day horizons. The ATC ratios range from 0.68 to 0.84 for a horizon of three days and from 0.68 to 0.82 for seven days. The negative and positive ATC ratios differ for all models and horizons. For some models, for example, the GBM-2 model, the positive ATC ratio is higher than the negative ATC ratio, and for some models, for example, the tbats model, vice versa. The confidence interval width is at most 0.02 for the ATC ratios and at most 0.03 for the positive and negative ATC ratios. The models GBM-2, qreg-1, and Benchmark-1 have the highest positive ATC ratio for the three and seven-day horizons, while Poisson-2 and NBI-2 have the highest negative ATC ratio.

Figure 5.5 shows the conditional ATC plots for the models Benchmark-1, GBM-2, NBI-2, Poisson-2, and qreg-1 for the horizons three and seven days and thus inspects the local ATC of the models with highest positive and negative ATC ratio. The conditional ATC plots show similar courses for the two horizons, though the curves are shifted downwards for the horizon of seven days. The model's relative ATC evolves consistently for the two horizons, with

	(1), l=3	$\sigma_{x^{\Delta,3}}$	$q_{0.1}(x^{\Delta,3})$	(2), l=3	(1), l=7	$\sigma_{x^{\Delta,7}}$	$q_{0.1}(x^{\Delta,7})$	(2), l=7
ADAM-iETSX	0.57	7.76	0.83	0.88	0.57	7.49	0.78	0.87
Benchmark-1	0.45	5.05	0.50	0.80	0.44	4.43	0.47	0.78
Benchmark-2	0.51	5.11	0.52	0.80	0.50	4.29	0.45	0.78
ETS	0.58	7.49	0.78	0.87	0.58	7.68	0.84	0.88
GBM-2	0.39	4.93	0.51	0.80	0.37	4.61	0.49	0.79
NBI-2	0.53	5.04	0.52	0.81	0.53	4.41	0.48	0.79
NOtr-1	0.52	5.03	0.51	0.81	0.51	4.41	0.49	0.79
NOtr-2	0.52	5.03	0.51	0.81	0.51	4.41	0.49	0.79
Poisson-1	0.53	5.04	0.51	0.81	0.52	4.38	0.48	0.79
Poisson-2	0.53	5.05	0.52	0.80	0.53	4.42	0.48	0.78
Poisson-2-I	0.51	5.03	0.51	0.81	0.50	4.42	0.49	0.79
Prophet	0.62	5.27	1.00	0.91	0.62	5.15	1.00	0.91
Regression-Poisson	0.51	6.65	0.67	0.85	0.51	6.49	0.67	0.85
Ttr-2	0.51	5.03	0.50	0.81	0.50	4.41	0.49	0.79
qreg-1	0.39	5.01	0.49	0.81	0.39	4.84	0.51	0.80
tbats	0.63	5.35	1.00	0.92	0.63	5.04	1.00	0.92
True	0.54	6.61	1.00	0.93	0.55	5.90	1.00	0.92

Table 5.5.: Marginal analysis of the forecast and true changes in patient admissions to an emergency department. The column (1) shows the fraction of values greater than zero for horizon l, $\sigma_{x^{\Delta,l}}$ the standard deviation, and $q_{0.1}(x^{\Delta,l})$ the 10% quantile of the changes' absolute values. Column (2) shows the fraction of values not in the exclusion area of size one.

the NBI-2 and Poisson-2 models being indistinguishable. The GBM-2 model outperforms the qreg-1 model for all predicted changes. The models NBI-2 and Poisson-2 have the highest ATC for all negative predicted changes and the lowest for all positive predicted changes. Benchmark-1 lies between the other models for all predicted changes.

Figure 5.6 visualizes the probabilistic ATC assessment for the same subset of models. The Brier scores (BSs) are shown in Figure 5.6a, and the reliability diagrams for the horizons three and seven days in Figures 5.6b and 5.6c. The BSs are smallest for NBI-2 and Poisson-2 for both horizons, while the BSs for the other models are larger and differ more. The qreg-1 model has both horizons' highest BS. The reliability diagrams of GBM-2 and NBI-2 are also close and show a too-small fraction of increases for the predicted probability overall. For the other models, the reliability diagrams show a fraction of increases that are too large for the corresponding predicted probability.

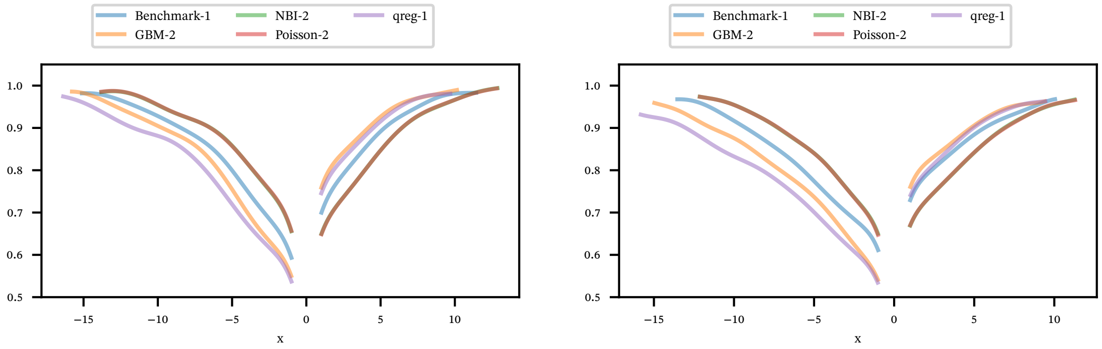
Discussion

The ATC is consistent for the two horizons, with the models' relative ATC evolving similarly for the two horizons. The models' ATC is generally higher for the smaller horizon, but the changes are minor, and confidence intervals overlap.

The positive and negative ATC ratios differ for all models. While some models, such as GBM-2 and qreg-1, have the highest positive ATC ratio, others, such as Poisson-2 and NBI-2,

	μ^3	$\mu^{+,3}$	$\mu^{-,3}$	μ^7	$\mu^{+,7}$	$\mu^{-,7}$
ADAM-iETSX	0.70 (0.69, 0.71)	0.68 (0.67, 0.69)	0.72 (0.71, 0.73)	0.68 (0.67, 0.69)	0.67 (0.66, 0.69)	0.69 (0.67, 0.70)
Benchmark-1	0.83 (0.82, 0.84)	0.86 (0.85, 0.87)	0.81 (0.79, 0.82)	0.81 (0.80, 0.82)	0.86 (0.85, 0.87)	0.78 (0.76, 0.79)
Benchmark-2	0.84 (0.83, 0.84)	0.83 (0.82, 0.85)	0.84 (0.83, 0.85)	0.82 (0.81, 0.83)	0.83 (0.82, 0.84)	0.80 (0.79, 0.82)
ETS	0.68 (0.67, 0.69)	0.66 (0.65, 0.67)	0.70 (0.69, 0.72)	0.67 (0.66, 0.68)	0.66 (0.64, 0.67)	0.68 (0.66, 0.69)
GBM-2	0.82 (0.81, 0.82)	0.90 (0.89, 0.91)	0.77 (0.76, 0.78)	0.78 (0.77, 0.79)	0.88 (0.87, 0.90)	0.73 (0.72, 0.74)
NBI-2	0.84 (0.83, 0.85)	0.83 (0.82, 0.84)	0.85 (0.84, 0.86)	0.82 (0.81, 0.83)	0.82 (0.81, 0.83)	0.82 (0.81, 0.83)
NOtr-1	0.83 (0.83, 0.84)	0.83 (0.82, 0.84)	0.84 (0.82, 0.85)	0.81 (0.80, 0.82)	0.82 (0.81, 0.83)	0.80 (0.79, 0.81)
NOtr-2	0.83 (0.83, 0.84)	0.83 (0.82, 0.84)	0.84 (0.82, 0.85)	0.81 (0.80, 0.82)	0.82 (0.81, 0.83)	0.80 (0.79, 0.81)
Poisson-1	0.84 (0.83, 0.84)	0.82 (0.81, 0.83)	0.85 (0.84, 0.86)	0.82 (0.81, 0.82)	0.82 (0.81, 0.83)	0.81 (0.80, 0.82)
Poisson-2	0.84 (0.83, 0.85)	0.83 (0.82, 0.84)	0.85 (0.84, 0.86)	0.82 (0.81, 0.82)	0.82 (0.81, 0.83)	0.82 (0.80, 0.83)
Poisson-2-I	0.83 (0.83, 0.84)	0.84 (0.83, 0.85)	0.83 (0.82, 0.84)	0.81 (0.80, 0.82)	0.83 (0.81, 0.84)	0.80 (0.79, 0.81)
Prophet	0.75 (0.74, 0.76)	0.72 (0.71, 0.73)	0.79 (0.77, 0.80)	0.74 (0.73, 0.74)	0.72 (0.70, 0.73)	0.76 (0.75, 0.77)
Regression-Poisson	0.72 (0.71, 0.73)	0.73 (0.71, 0.74)	0.72 (0.70, 0.73)	0.70 (0.69, 0.71)	0.71 (0.70, 0.73)	0.69 (0.67, 0.70)
Ttr-2	0.84 (0.83, 0.84)	0.84 (0.83, 0.85)	0.83 (0.82, 0.85)	0.81 (0.80, 0.82)	0.83 (0.82, 0.84)	0.80 (0.79, 0.81)
qreg-1	0.80 (0.79, 0.80)	0.88 (0.87, 0.89)	0.75 (0.74, 0.76)	0.77 (0.76, 0.78)	0.86 (0.85, 0.88)	0.71 (0.70, 0.72)
tbats	0.75 (0.74, 0.76)	0.72 (0.71, 0.73)	0.80 (0.78, 0.81)	0.73 (0.72, 0.74)	0.71 (0.69, 0.72)	0.76 (0.74, 0.77)

Table 5.6.: ATC ratio μ, positive ATC ratio μ^+, and negative ATC ratio μ^- for the models with the exclusion of zero-containing points for the horizons 72 hours and seven days in the forecasting of patient admissions to an emergency department.



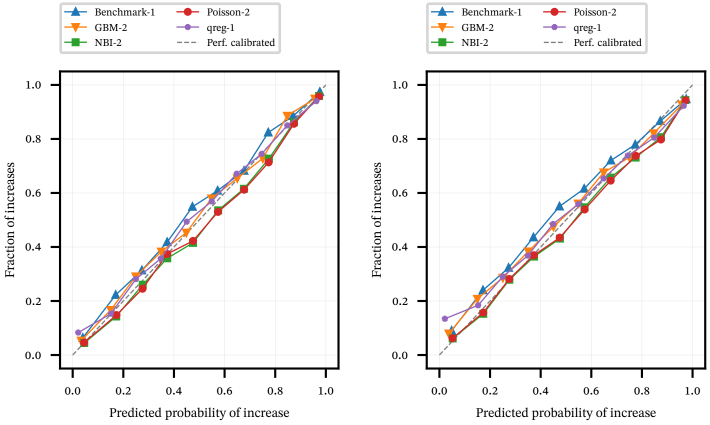
(a) Horizon three days.

(b) Horizon seven days.

Figure 5.5.: Conditional ATC plots for the horizons three and seven days and the models with the best positive or negative ATC in forecasting the patient admissions to an emergency department. The plots of NBI-2 and Poisson-2 are indistinguishable.

	3 d	7 d
Benchmark-1	0.1586	0.1761
GBM-2	0.1590	0.1759
NBI-2	0.1549	0.1717
Poisson-2	0.1549	0.1714
qreg-1	0.1679	0.1843

(a) BS for the different models and horizons.



(b) Reliability diagram for horizon three days.

(c) Reliability diagram for horizon seven days.

Figure 5.6.: Probabilistic ATC assessment for the models Benchmark-1, GBM-2, NBI-2, Poisson-2, and qreg-1 for the horizons three and seven days in forecasting patient admissions to an emergency department. The Brier score in Figure 5.6a evaluates the calibration and sharpness of the probabilistic ATC simultaneously, while the two plots on the right assess solely the calibration, that is, whether the predicted probability of increase occurs empirically. Probabilistic ATC by the BS is best for the models NBI-2 and Poisson-2 for both lags.

have the highest negative ATC ratio. Thus, the uncertainty of the model's predicted change has to be assessed differently based on the direction.

The probabilistic ATC assessment results endorse the point ATC assessment and assign the best scores to NBI-2 and Poisson-2. The reliability diagrams show that they underestimate the fraction of increases slightly.

Overall, the example provides performance assessments that are different from standard point evaluation measures and thus provide further insights into the strengths and weaknesses of the models. While the models with the lowest RMSE, NBI-2, and Poisson-2, also have a high ATC, three models with below-average point evaluation measures, Benchmark-1, GBM-2, and qreg-1, have a high positive ATC.

5.3.3. Non-invasive blood pressure monitoring

We here consider the ATC of non-invasive blood pressure measurements from the MIMIC-III database that includes data of critically ill patients treated in intensive care units of the Beth Israel Deaconess Medical Center in Boston (Massachusetts, USA, A. E. Johnson et al., 2016, Moody et al., 2017; available through Goldberger et al., 2000). We focus on invasive arterial blood pressure (ABP) and non-invasive blood pressure (NBP) measurements and thus limit our analysis to datasets containing simultaneous measurements of ABP and NBP simultaneously. 2,548 datasets include at least one measurement of systolic ABP and NBP, and 1,327 include at least one pair of simultaneously measured systolic ABP and NBP; for the mean ABP and NBP, the numbers are 2,605 and 1,516, respectively. We assess the ATC of non-invasive blood pressure measurements (test method) compared to intraarterial blood pressure measurements (reference method, gold standard). We consider the horizons of one minute, five minutes, and 15 minutes for the ATC assessment, as those are typical intervals of NBP measurements.

Results

Again, we exclude the smallest 10% of absolute changes in ATC assessment. The resulting four-quadrant plots of the mean and systolic blood pressure measurements for the different horizons are shown in Figure 5.7. The number of points in the four-quadrant plot is smaller due to the restriction to data records with measurements of mean or systolic ABP and NBP simultaneously for two consecutive times with the specified horizons. Thus, we use the NBP measurements as the test method and the ABP measurements as the gold standard. For the systolic measurements, 290, 332, and 442 points are available for the horizons of one, five, and 15 minutes; for the mean measurements, 406, 430, and 542.

The ATC ratios, including confidence intervals for the different horizons, are listed in Table 5.7. The confidence intervals have lower bounds of 0.5 or slightly above for the measurements with a horizon of one minute. For larger horizons, the ATC ratio increases. The difference between positive and negative ATC ratios is small for all types and horizons, with overlapping confidence intervals.

Figure 5.8 shows the conditional ATC plots for the different horizons and the systolic and mean blood pressure measurements. It becomes apparent that the systolic measurements have a higher ATC than the mean measurements, except for small negative predicted changes.

Discussion

The four-quadrant plots contain a considerable number of extreme points. Whether these points are due to measurement errors or extreme values is not distinguishable. Some authors argue to exclude the measurements below the 10%-quantile of the absolute changes and the points above the 90%-quantile (see L. A. Critchley, Lee, and Ho, 2010). We do not follow this

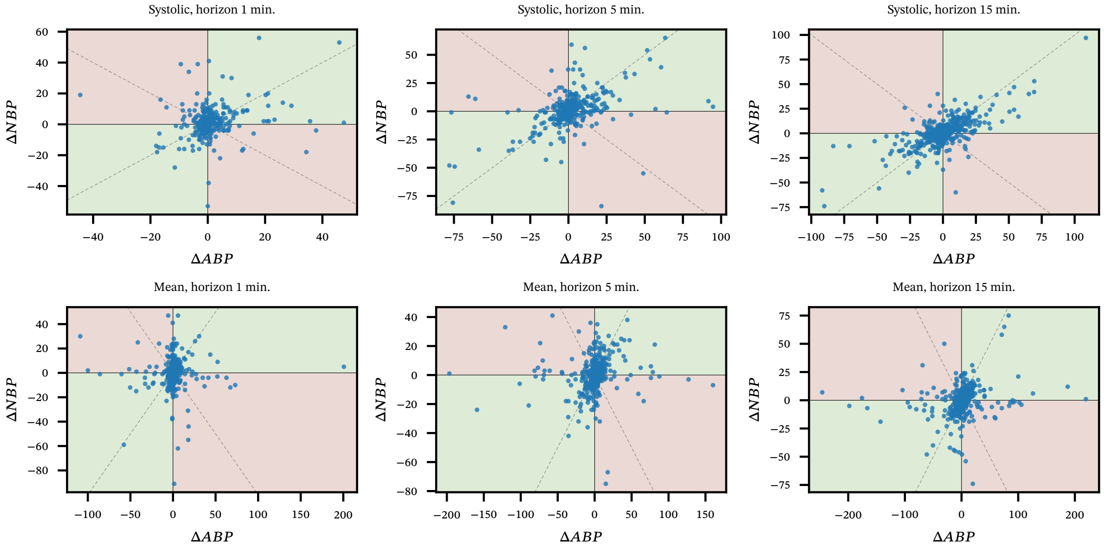


Figure 5.7.: Four-quadrant plots for the different horizons *l* and the systolic and mean blood pressure measurements. The upper row contains systolic measurements, and the lower row contains mean measurements. The columns contain the horizons one, five, and 15 minutes.

approach here, as the extreme values are not necessarily measurement errors and could be particularly relevant.

The difference between positive and negative ATC ratios is small in this example. The positive and negative ATC ratios have overlapping confidence intervals, the conditional ATC plots do not contain prominent deviations in the course, and the four-quadrant plots do not display asymmetries.

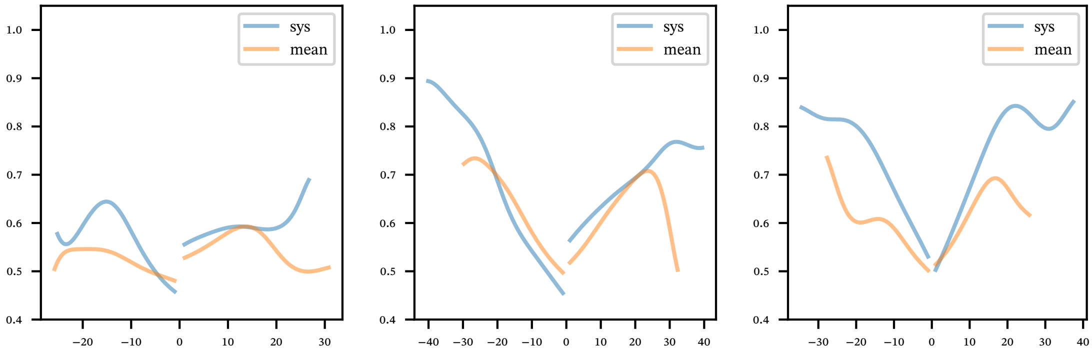
The bootstrap confidence intervals are wide. The width is around 0.1 for the ATC ratio, while it gets up to 0.16 for the negative ATC ratio for systolic measurement and the horizon of one minute. Thus, the confidence intervals cover 0.5 for systolic measurement and the horizon of one minute, and the equality to random guessing cannot be excluded.

5.4. Discussion and conclusion

In this chapter, we examine various methods to assess the ability to track changes (ATC) for measurements, nowcasts, or forecasts, that is, whether they correctly predict the direction of changes in values. While the computation of predicted change varies between the application areas of measurement, nowcasting, and forecasting, the assessment can be based on

Type	l	μ^{l}	$\mu^{+,l}$	$\mu^{-,l}$
Systolic	1	0.55 (0.50, 0.60)	0.59 (0.52, 0.65)	0.58 (0.50, 0.66)
Systolic	5	0.63 (0.59, 0.68)	0.70 (0.64, 0.75)	0.62 (0.56, 0.69)
Systolic	15	0.69 (0.65, 0.73)	0.72 (0.66, 0.76)	0.74 (0.69, 0.79)
Mean	1	0.55 (0.51, 0.59)	0.62 (0.56, 0.68)	0.56 (0.50, 0.62)
Mean	5	0.59 (0.55, 0.64)	0.65 (0.59, 0.71)	0.62 (0.56, 0.68)
Mean	15	0.62 (0.58, 0.65)	0.65 (0.60, 0.70)	0.66 (0.61, 0.71)

Table 5.7.: ATC ratios for the different horizons l and the systolic and mean blood pressure measurements.



(a) Horizon one minute. **(b)** Horizon five minutes. **(c)** Horizon 15 minutes.

Figure 5.8.: Conditional ATC plot for the systolic and mean blood pressure measurements and the horizons one, five, and 15 minutes.

the same methods. The ATC assessment can accompany other evaluation techniques, such as measures of deviation or probabilistic scoring rules.

Four-quadrant plots facilitate the visual inspection of the ATC for a measurement, nowcast, or forecast (see Section 5.2.2). The ATC ratio, the ratio of change directions predicted correctly over the total number of changes, numerically evaluates ATC. Visually, it is the proportion of concordant points in a four-quadrant plot (see Section 5.2.3). The positive and negative ATC ratios analyze the ATC ratio given whether the predicted change is positive or negative, respectively. Thus, they quantify the credibility of the respective predictions. The applications of Section 5.3.3 show that models, in general, indeed have different positive and negative ATC and that they add valuable information to the ATC ratio. In the applications, the bootstrap confidence intervals of Section 5.2.3 are used to quantify the estimation uncertainty of the ATC measures. The width of the confidence intervals is around 0.1 for around 100 samples, while it is around 0.01 for 8000 samples. For models with reasonably high ATC, 100 samples are thus sufficient to differentiate from random guessing or to assess models with high ATC differences.

A conditional ATC plot visualizes the probability of correct change direction prediction over the predicted change of the measurement, nowcast, or forecast (see Section 5.2.5). It is based on a multivariate kernel density estimation (KDE) of predicted and observed change. In the application, the conditional ATC plot gives reasonable insights into the local effects of the ATC. Section 5.2.6 adapts measures of probabilistic forecast evaluation to the ATC assessment of probabilistic forecasts and nowcasts. The Brier score (BS) as numerical assessment of probabilistic ATC is introduced, and reliability diagrams are used to visualize the local ATC of probabilistic forecasts.

The methods of ATC assessment are applied to COVID-19-nowcasting, forecasting the patient admissions to an emergency department, and invasive and non-invasive blood pressure measurements in Section 5.3. While ATC assessment should not be the only aspect, it is a valuable addition to evaluating nowcasts, forecasts, and measurements. Models with highly different accuracies are usually scored similarly in ATC assessment, but ATC assessment can differentiate between models with similar accuracies. As in the application in Section 5.3.1, models with medial point forecast evaluation measures can have the most meaningful positive ATC.

We did not expand on two modeling aspects throughout this chapter, which we leave for further research. In the estimation, we did not consider sequential correlation. The computation of differences is a standard procedure in time series analysis to remove sequential dependence, but, in general, some could remain, and the estimators could account for it. Similarly, the bootstrap confidence intervals could be adapted to consider sequential correlation using time-series bootstrap methods (Härdle, Horowitz, and Jens-Peter Kreiss, 2003; Jens-Peter Kreiss and Lahiri, 2012).

The estimators of Section 5.2.3 do not account for imbalances in the number of observed positive and negative changes (for theoretical analysis, see Jolliffe and Stephenson, 2012, Chapter 3). Significant differences in the number of observed positive and negative changes are unlikely in the ATC setting, as $\mathbf{y}^{\Delta}$ is obtained from differencing time series data and occur, for example, if the true value contains a few high jumps in one direction and many smaller jumps in the other. However, if the number of positive and negative observed changes differs widely, unbalanced-data-aware measures should be considered. There are various adapted measures for unbalanced outcomes, for example, Cohen's κ (Cohen, 1960) or those listed in Jolliffe and Stephenson (2012, Table 3.3).

6. Conclusion

This thesis presents contributions to the decomposition and analysis of checkerboard copulas, the combination of point forecasts using copula time series models, and the evaluation of the ability to track changes for forecasts, nowcasts, and measurements. In all contributions, dependence modeling plays a prominent role. While the first contribution provides insights into a specific class of dependence model structures, the second applies copula dependence modeling to derive a combined density forecast and improve forecast calibration. The third contribution assesses the joint development of predicted and actual changes and infers new measures and visualization techniques to evaluate forecasts, nowcasts, and measurements.

After reviewing the fundamentals of copula models, forecasting, and its evaluation in Chapter 2, Chapter 3 introduces the singular value decomposition of doubly stochastic matrices representing checkerboard copulas. Thus, a better comprehensible representation of the checkerboard copula and various insights into the dependence structure are obtained. The decomposition yields a structural understanding of various checkerboard copula properties, such as Spearman's ρ and Kendall's τ. Several methods extend the decomposition, such as the monotonicity anchored representation for comonotonic copulas and the mapping to the nearest doubly stochastic matrix for truncations with negative entries. The decomposition is illustrated for various checkerboard approximations of parametric copulas and empirical data on fuel injector spray characteristics in jet engines.

Chapter 4 presents a point forecast combination method that accounts for the dependence structure of the forecast errors through a copula time series model. Thus, particular properties of the combination, such as asymmetric dependencies, can be incorporated into the combined density, improving the calibration and location of combined forecasts. The method theoretically yields calibrated combined density forecasts, and simulations show superior results to other combination methods in calibration, sharpness, and forecast location. An empirical data example of electricity price spot market forecasts illustrates the method's ability to improve calibration in practice.

Chapter 5 raises new measures and visualization techniques for assessing the ability to track changes for nowcasts, forecasts, and measurements. The ability to track changes (ATC) refers to the ability to correctly predict the direction of changes in values. In standard evaluation measures, the ATC is not distinctly assessed, although it is usually central information in practice, whether quantities rise or fall. Our derived measures are easily interpretable, and ready-to-use published code makes further applications straightforward. Additionally, we provide extensions to analyze local effects or quantify estimation uncertainty, and we extend the scope to assessing probabilistic nowcasts and forecasts. The methods are illustrated in three large data applications: COVID-19-nowcasting, forecasting patient admissions in an emergency department, and invasive and non-invasive blood pressure measurements.

The contributions raise interesting topics for further research. The approach of Chapter 3 could be extended to Bernstein copulas, which are also based on a doubly stochastic matrix. For Bernstein copulas, the empirical aspects of fitting could be investigated (for example, based on Janssen, Swanepoel, and Veraverbeke, 2012). The copula combined density (CCD) approach could be included in a larger framework for forecasting. Recently, foundation models, pre-trained neural networks for time series analysis, have been proposed. In the approach of Woo et al. (2024), the foundation model results in a linear combination of densities, which could be replaced by the CCD method. As those models are pre-trained on large datasets, the copula fit would be based on a broad data basis and account for within-model dependencies in the results. Besides, the result of the CCD method could be modified to a probability mass function for a discrete quantity of interest using discrete copulas (Genest and Nešlehová, 2007). The combination of discrete forecasts has received less attention but is crucial for count data, such as patient admissions (see Section 5.3.2).

In summary, this thesis provides various improvements in the analysis and application of dependence modeling by simplifying the assessment of incorporated dependence information for checkerboard copulas, enhancing point forecast combination to a density forecast, and evaluating the ability of forecasts, nowcasts, and measurements to predict the correct direction of change.

Appendices

A. Appendix to Chapter 3

The section provides additional material to Chapter 3 and is based on Publ. I. Appendix A.1 contains the proofs for the algorithms of Appendix 3.2.3 and, thereby, highlights the difference for symmetric and asymmetric matrices. Appendix A.2 discusses the decomposition of the checkerboard copula in terms of the Hellinger distance instead of the Frobenius one. In Appendix A.3, we derive the formulas of Kendall's τ and Spearman's ρ for Section 3.2.4. Finally, Appendix A.4 provides additional plots for the empirical example of fuel injector spray characteristics in Section 3.4.2.

A.1. Calculations for the algorithms of Section 3.2.3

We consider the problem

$$P_1(\mathbf{A}) := \underset{\mathbf{B}\in\mathbb{R}^{n\times n}}{\arg\min} \|\mathbf{A}-\mathbf{B}\|_F^2 \quad \text{s.t.} \quad \mathbf{B}\mathbb{1} = \mathbb{1}, \mathbf{B}^\top \mathbb{1} = \mathbb{1},$$

with a symmetric matrix $\mathbf{A}$ in Appendix A.1.1 and an asymmetric $\mathbf{A}$ in Appendix A.1.2.

The solution of $P_1(\mathbf{A}_1)$ has a closed form, if the matrix $\mathbf{A}_1$ is symmetric, that is,

$$P_1^{\text{sym}}(\mathbf{A}_1) = \mathbf{A}_1 + \left(\frac{1}{n}I_n - \frac{1}{n}\mathbf{A}_1 + \frac{1}{n^2}\mathbb{1}^\top\mathbf{A}_1\mathbb{1}I_n\right)\mathbb{1}\mathbb{1}^\top - \frac{1}{n}\mathbb{1}\mathbb{1}^\top\mathbf{A}_1.$$

In the case of an asymmetric matrix $\mathbf{A}_1$, the problem $P_1(\mathbf{A}_1)$ boils down to a linear system (see Appendix A.1.2).

For $P_2(\mathbf{A}_2)$,

$$\begin{aligned} P_2(\mathbf{A}_2) = \underset{\mathbf{B}\,\in\,\mathbb{R}^{n\times n}}{\arg\min} \quad & \|\mathbf{A}_2-\mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{B}_{i,j} \geq 0, \quad \forall i,j \in [n], \end{aligned}$$

there exists a closed-form solution independent of the symmetry of $\mathbf{A}_2$. As the Frobenius norm can be minimized elementwise,

$$\|\mathbf{B}-\mathbf{A}_2\|_F^2 = \sum_{i,j\,:\,(\mathbf{A}_2)_{ij}<0} (\mathbf{B}-\mathbf{A}_2)_{ij}^2 + \sum_{i,j\,:\,(\mathbf{A}_2)_{ij}\geq 0} (\mathbf{B}-\mathbf{A}_2)_{ij}^2 \quad B \in \mathbb{R}^{n\times n},$$

the solution of $P_2(\mathbf{A}_2)$ is the elementwise positive part of A_2. Algorithm 1 combines P_1 and P_2.

A.1.1. Symmetric copula

The proof is analogous to Zass and Shashua (2007) for symmetric $\mathbf{A}$. We provide it here for completeness and to emphasize its inapplicability to asymmetric matrices. Let $\mathbf{A} = \mathbf{A}^\top$ and $P_1^*(\mathbf{A})$ be the relaxation

$$P_1^*(\mathbf{A}) \coloneqq \underset{\mathbf{B}\in\mathbb{R}^{n\times n}}{\arg\min} \|\mathbf{A}-\mathbf{B}\|_F^2 \quad \text{s.t.} \quad \mathbf{B}\mathbb{1} - \mathbb{1} + \mathbf{B}^\top\mathbb{1} - \mathbb{1} = 0.$$

If $\mathbf{B} = \mathbf{B}^\top$, $\mathbf{B}\mathbb{1} - \mathbb{1} + \mathbf{B}^\top\mathbb{1} - \mathbb{1} = 0 \Rightarrow \mathbf{B}\mathbb{1} = \mathbb{1} = \mathbf{B}^\top\mathbb{1}$. Thus, if $\mathbf{B} \coloneqq P_1^*(X) = \mathbf{B}^\top$, $\mathbf{B}$ is also the solution of $P_1(X)$. Let $L(\mathbf{A},\mu)$ be the Lagrangian of $P_1^*(\mathbf{A})$ with

$$L(\mathbf{B},\mu) = \text{trace}(\mathbf{B}^\top\mathbf{B}) - \text{trace}(2\mathbf{A}^\top\mathbf{B}) - 2\mu^\top(\mathbf{B}\mathbb{1} + \mathbf{B}^\top\mathbb{1} - 2\mathbb{1}).$$

Then,

$$\begin{aligned} \frac{\partial L(\mathbf{B},\mu)}{\partial \mathbf{B}} &= 2\mathbf{B} - 2\mathbf{A} - 2\mu\mathbb{1}^\top - 2\mathbb{1}\mu^\top \overset{!}{=} 0 && \Leftrightarrow \\ \mathbf{B} &= \mathbf{A} + \mu\mathbb{1}^\top + \mathbb{1}\mu^\top && \Leftrightarrow |(\cdot)^\top \qquad \text{(A.1)} \\ \mathbf{B}^\top &= \mathbf{A}^\top + \mathbb{1}\mu^\top + \mu\mathbb{1}^\top \end{aligned}$$

and

$$\begin{aligned} \mathbf{B} + \mathbf{B}^\top &= (\mathbf{A}+\mathbf{A}^\top) + 2\cdot\mu\mathbb{1}^\top + 2\cdot\mathbb{1}\mu^\top && \Leftrightarrow \qquad \text{(A.2)} \\ \underbrace{\mathbf{B}\mathbb{1} + \mathbf{B}^\top\mathbb{1} = 2\cdot\mathbb{1}}_{\text{follows from } \frac{\partial L(\mathbf{B},\mu)}{\partial\mu}=0} &= (\mathbf{A}+\mathbf{A}^\top)\mathbb{1} + 2\cdot\mu\underbrace{\mathbb{1}^\top\mathbb{1}}_{=n} + 2\cdot\mathbb{1}\mu^\top\mathbb{1} && \Leftrightarrow \ldots \\ \mu &= \frac{1}{n}(I_n - \frac{1}{2n}\mathbb{1}\mathbb{1}^\top)(I_n - \frac{1}{2}(\mathbf{A}+\mathbf{A}^\top))\mathbb{1}, && \qquad \text{(A.3)} \end{aligned}$$

using $(nI_n + \mathbb{1}\mathbb{1}^\top)^{-1} = \frac{1}{n}(I_n + \frac{1}{2n}\mathbb{1}\mathbb{1}^\top)$. Plugging μ from Equation (A.3) into (A.1) yields

$$\begin{aligned} \mathbf{B} &= \mathbf{A} + \Big(\frac{1}{n}(I_n - \frac{1}{2n}\mathbb{1}\mathbb{1}^\top)(I_n - \frac{1}{2}(\mathbf{A}+\mathbf{A}^\top))\mathbb{1}\Big)\mathbb{1}^\top + \mathbb{1}\Big(\frac{1}{n}(I_n - \frac{1}{2n}\mathbb{1}\mathbb{1}^\top)(I_n - \frac{1}{2}(\mathbf{A}+\mathbf{A}^\top))\mathbb{1}\Big)^\top \\ &= \mathbf{A} + \Big(\frac{1}{n}I_n - \frac{1}{2n}(\mathbf{A}+\mathbf{A}^\top) + \frac{1}{2n^2}\mathbb{1}^\top(\mathbf{A}+\mathbf{A}^\top)\mathbb{1}I_n\Big)\mathbb{1}\mathbb{1}^\top - \frac{1}{2n}\mathbb{1}\mathbb{1}^\top(\mathbf{A}+\mathbf{A}^\top) \\ &\overset{\mathbf{A}=\mathbf{A}^\top}{=} \mathbf{A} + \Big(\frac{1}{n}I_n - \frac{1}{n}\mathbf{A} + \frac{1}{n^2}\mathbb{1}^\top\mathbf{A}\mathbb{1}I_n\Big)\mathbb{1}\mathbb{1}^\top - \frac{1}{n}\mathbb{1}\mathbb{1}^\top\mathbf{A} \qquad \text{(A.4)} \end{aligned}$$

```
input  : Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\epsilon > 0$, number of maximal iterations $c_{\max}$
output: Nearest doubly stochastic matrix $\mathbf{B}$
Set $\mathbf{B} = \mathbf{A}$ and $c = 1$;
Update $\mathbf{B} = P_1(\mathbf{B})$;
while $\exists i, j : \mathbf{B}_{i,j} < -\epsilon \wedge c \leq c_{max}$ do
    Update $\mathbf{B} = P_2(\mathbf{B})$;
    Update $\mathbf{B} = P_1(\mathbf{B})$;
    Update $c = c + 1$;
end
```

Algorithm 1: Algorithm to compute the nearest doubly stochastic matrix in terms of the Frobenius error following Zass and Shashua (2007) for symmetric matrices $\mathbf{A}$ and $\epsilon > 0$. The stopping criterion $c \leq c_{\max}$ ensures the termination of the algorithm. The solutions of P_1 and P_2 can be found in Appendix A.1.

The result for $\mathbf{B}$ in Equation (A.4) is symmetric and, thus, also solution for $P_1(\mathbf{A})$:

$$\begin{aligned}
\mathbf{B}^\top &= \left(\mathbf{A} + \left(\frac{1}{n}I_n - \frac{1}{n}\mathbf{A} + \frac{1}{n^2}\mathbb{1}^\top\mathbf{A}\mathbb{1}I_n\right)\mathbb{1}\mathbb{1}^\top - \frac{1}{n}\mathbb{1}\mathbb{1}^\top\mathbf{A}\right)^\top \\
&= \mathbf{A} + \left(\frac{1}{n}I_n - \frac{1}{n}\mathbf{A} + \frac{1}{n^2}\mathbb{1}^\top\mathbf{A}\mathbb{1}I_n\right)\mathbb{1}\mathbb{1}^\top - \frac{1}{n}\mathbb{1}\mathbb{1}^\top\mathbf{A} \\
&= \mathbf{B},
\end{aligned}$$

using

$$(\mathbf{A}\mathbb{1}\mathbb{1}^\top)_{ij} = (\sum_{k=1}^{n}\mathbf{A}_{ik}) = (\sum_{k=1}^{n}\mathbf{A}_{kj}) = (\mathbb{1}\mathbb{1}^\top\mathbf{A})_{ij}, \quad i, j \in [n].$$

A.1.2. Asymmetric copula

For asymmetric $\mathbf{A}$, the result of $P_1^*(\mathbf{A})$ is not symmetric and thus is not a solution to the original problem $P_1(\mathbf{A})$.

Instead, the solution of a Karush-Kuhn-Tucker equation system yields the solution for P_1. The problem

$$P_1(\mathbf{A}) \coloneqq \underset{\mathbf{B}\in\mathbb{R}^{n\times n}}{\arg\min} ||\mathbf{A} - \mathbf{B}||_F^2 \quad \text{s.t.} \quad \mathbf{B}\mathbb{1} = \mathbb{1}, \mathbf{B}^\top\mathbb{1} = \mathbb{1}$$

with the Lagrange function and its derivative

$$\begin{aligned} L(\mathbf{B},\lambda,\mu) &= \operatorname{trace}(\mathbf{B}^\top\mathbf{B}) - \operatorname{trace}(2\mathbf{A}^\top\mathbf{B}) - \lambda^\top(\mathbf{B}\mathbb{1} - \mathbb{1}) - \mu^\top(\mathbf{B}^\top\mathbb{1} - \mathbb{1}) \\ \frac{\partial L}{\partial \mathbf{B}_{ij}}(\mathbf{B},\lambda,\mu) &= 2\mathbf{B}_{ij} - 2\mathbf{A}_{ij} - \lambda_j - \mu_i \text{ for } i,j \in [n] \end{aligned}$$

yields the system

$$2\mathbf{B}_{ij} - \lambda_j - \mu_i = 2\mathbf{A}_{ij} \qquad \forall (i,j) \in [n]\times[n] \tag{A.5}$$

$$\sum_{i=1}^{n} \mathbf{B}_{ij} = 1 \qquad \forall j \in [n] \tag{A.6}$$

$$\sum_{j=1}^{n} \mathbf{B}_{ij} = 1 \qquad \forall i \in [n]. \tag{A.7}$$

The solution of the Karush-Kuhn-Tucker equation system is the solution of the linear equation system $Kb = a$ with

$$\tilde{K} = \begin{pmatrix} 2I_{n^2\times n^2} & [\mathbb{1}\otimes I_{n\times n}] & [I_{n\times n}\otimes \mathbb{1}] \\ [\mathbb{1}^\top \otimes I_{n\times n}] & 0 & 0 \\ [I_{n\times n}\otimes \mathbb{1}^\top] & 0 & 0 \end{pmatrix}, \tilde{b} = \begin{pmatrix} \operatorname{vec}(\mathbf{B}) \\ \lambda \\ \mu \end{pmatrix}, \tilde{a} = \begin{pmatrix} \operatorname{vec}(2A) \\ 0^{n\times 1} \\ 0^{n\times 1} \end{pmatrix}$$

and $K = \tilde{K}_{1:n^2+2n-1,1:n^2+2n-1}, b = \tilde{b}_{1:n^2+2n-1}, a = \tilde{a}_{1:n^2+2n-1}$. Thereby, $\otimes$ denotes the Kronecker product of the matrices, and $\operatorname{vec}(\cdot)$ denotes the column-wise stacking of a matrix into a vector. The last row and column are excluded, as the matrix K is singular and the constraint $\sum_{j=1}^{n} \mathbf{B}_{nj} = 1$ is guaranteed by the remaining Constraints (A.6) and (A.7). Then, the first n^2 elements of the solution b rearranged as matrix $\mathbf{B}$ are the solution of P_1.

Additionally, Algorithm 1 has to be expanded by a deflection component (see Zass and Shashua, 2007; Dykstra, 1983), as summarized in Algorithm 2.

A.2. Decomposition in terms of the Hellinger distance

The singular value decomposition and the algorithms of Section 3.2.3 yield minimal errors in terms of the Frobenius norm. The SVD is also the best low-rank approximation considering the spectral norm (Mirsky, 1960). In statistics, the Hellinger distance is often used to assess the proximity of densities (see Aya-Moreno, Geenens, and Penev, 2018; Meier, Kirch, and Meyer, 2020, for two recent contributions). In this section, we analyze Hellinger distance-based decompositions for two different versions of the Hellinger distance for matrices, as, to

```
input : Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\epsilon > 0$, number of maximal iterations $c_{\max}$
output: Nearest doubly stochastic matrix $\mathbf{B}$
Set $\mathbf{B} = \mathbf{A}$ and $c = 1$;
Set $\mathbf{B}_{1,1} = P_1(\mathbf{B})$ and $I_{1,1} = \mathbf{B}_{1,1} - \mathbf{B}$ ;
Set $\mathbf{B}_{1,2} = P_2(\mathbf{B} + I_{1,1})$ and $I_{1,2} = \mathbf{B}_{c,2} - (\mathbf{B} + I_{c,1})$ ;
repeat
    Update $c = c + 1$;
    Set $\mathbf{B}_{c,1} = P_1(\mathbf{B} + I_{c-1,2})$ and $I_{c,1} = \mathbf{B}_{c,1} - (\mathbf{B} + I_{c-1,2})$ ;
    Set $\mathbf{B}_{c,2} = P_2(\mathbf{B} + I_{c,1})$ and $I_{c,2} = \mathbf{B}_{c,2} - (\mathbf{B} + I_{c,1})$ ;
until
    $(\mathbf{B}+I_{c,1}+I_{c,2}) \geq -\epsilon \wedge |(\mathbf{B}+I_{c,1}+I_{c,2})\mathbb{1} - \mathbb{1}| < \epsilon \wedge |(\mathbf{B}+I_{c,1}+I_{c,2})^{\top}\mathbb{1} - \mathbb{1}| < \epsilon \wedge c \leq c_{max}$;
```

Algorithm 2: Algorithm to compute the nearest doubly stochastic matrix in terms of the Frobenius error following Zass and Shashua (2007) for asymmetric matrices **A** and $\epsilon > 0$. The stopping criterion $c \leq c_{\max}$ ensures the termination of the algorithm. The solutions of P_1 and P_2 can be found in A.1.2. $\mathbf{B}_{c,d}$ denotes the solution in iteration c in the subproblem d and $I_{c,d}$ the corresponding incremental change.

our knowledge, there is no agreed definition in the matrix case yet. For a matrix square root-based Hellinger distance, the decomposition generalizes from the Frobenius case, while it is of a different and more complicated structure for an elementwise square root Hellinger distance.

For discrete probability distributions $p = (p_1, \dots, p_n)$ and $q = (q_1, \dots, q_n)$, the Hellinger norm d_H is computed by

$$
\begin{aligned}
d_H(p,q) &= \frac{1}{\sqrt{2}}\left[\sum_{i=1}^{n}\left(\sqrt{p_i} - \sqrt{q_i}\right)^2\right]^{1/2} \\
&= \frac{1}{\sqrt{2}}\left[\sum_{i=1}^{n}(p_i + q_i) - 2\sum_{i=1}\sqrt{p_i q_i}\right]^{1/2}.
\end{aligned}
$$

For matrices, there are different notions of the Hellinger distance in literature. We consider a formulation based on the matrix square root first (Bhatia, Gaubert, and Jain, 2019) and then turn to an elementwise square root method (Rao, 1995).

Bhatia, Gaubert, and Jain (2019) starts from the decomposition of the Hellinger distance for densities into an arithmetic and geometric mean. As the geometric mean for matrices can be interpreted in various ways, different notions of the distance can be obtained. We use their most straightforward generalization yielding the Hellinger distance for positive semidefinite, and, thus symmetric, matrices A and B

$$d_H(A,B) = \|A^{1/2} - B^{1/2}\|_F = \left[\operatorname{trace}(A+B) - 2\operatorname{trace}(A^{1/2}B^{1/2})\right]^{1/2}.$$

Thereby, $A^{1/2}$ denotes the matrix square root with $(A^{1/2})^\top A^{1/2} = A^{1/2}A^{1/2} = A$.

Lemma 3. *The low-rank approximation problem of a positive definite matrix A yields the same eigenvectors and eigenvalues for the Frobenius and the Hellinger distance.*

Proof. Let $P_F(A,k)$ the low-rank approximation problem in terms of the Frobenius distance

$$P_F(A,k) := \min_{B\in\mathbb{R}^{n\times n}} \|A-B\|_F \quad \text{s.t.} \quad \operatorname{rank}(B) \le k$$

and $P_H(A,k)$ in terms of the Hellinger distance

$$P_H(A,k) = \min_{B\in\mathbb{R}^{n\times n}} \|A^{1/2} - B^{1/2}\|_F \quad \text{s.t.} \quad \operatorname{rank}(B) \le k, B \text{ positive semidefinite,} \tag{A.8}$$

as the matrix square root is unique and $\operatorname{rank}(B) = \operatorname{rank}(B^{1/2})$ (Horn and C. R. Johnson, 2012, Theorem 7.2.6).

Thus, the minimizing argument, $B^{1/2}$, of Problem $P_H(A,k)$ is the k-truncated SVD of $A^{1/2}$. Due to the positive definiteness of A, left and right singular values are identical, and the eigenvalue decomposition

$$A^{1/2} := U\Lambda(U)^\top \tag{A.9}$$

exists with eigenvector matrix U and diagonal eigenvalue matrix Λ. This yields for the matrix square root of the minimizing argument, $B^{1/2}$, of $P_H(A,k)$ and the minimizing argument, B,

$$\begin{aligned} B^{1/2} &= U_{[k]}\Lambda_{[k]}(U_{[k]})^\top \text{ and} \\ B = B^{1/2}B^{1/2} &= U_{[k]}\Lambda^2_{[k]}(U_{[k]})^\top. \end{aligned}$$

The eigenvectors of $A^{1/2}$ and A are identical and the singular vectors are squared, as

$$A = A^{1/2}A^{1/2} = (U\Lambda U^\top)(U\Lambda U^\top)^\top = U\Lambda^2 U^\top.$$

Thus, the minimizing argument of $P_F(A,k)$ is

$$U_{[k]}\Lambda^2_{[k]}(U_{[k]})^\top$$

and equal to the minimizing argument of $P_H(A,k)$.

□

This definition of the Hellinger distance obtains the same decomposition as with the Frobenius distance. The coefficient η in the MAR can also be computed with the Hellinger distance instead of the Frobenius distance in Equation (3.8). However, the definition of the Hellinger distance only for positive semidefinite copulas is restrictive, as valid bistochastic matrices do not need to be positive definite, for example, unsymmetric.

Rao (1995) and C. M. Cuadras and D. Cuadras (2006) define the Hellinger distance in terms of an elementwise square root, thus considering only matrices with non-negative elements. Let $\sqrt{A}$ denote the elementwise square root of a matrix A. Then, the decomposition based on the elementwise Hellinger decomposition is for a symmetric checkerboard copula $\mathbf{C}^n$

$$\sqrt{\mathbf{C}^n} = U^H \Lambda^H (U^H)^\top. \tag{A.10}$$

Truncations $T_{n^\star}(\sqrt{\mathbf{C}^n})$ have to be squared elementwise to get a low-rank approximation of the Hellinger decomposition. Note that the squared decomposition does not keep the rank of $T_{n^\star}(\sqrt{\mathbf{A}^n})$. The MAR could be used in the elementwise Hellinger decomposition, and the optimization in Equation (3.8) could be adapted and solved by a general optimization problem solver. To our knowledge, no optimizations similar to those in the Frobenius case are available for the elementwise Hellinger scenario, as either the objective function contains square roots or the constraints are non-linear. While the elements of the squared decomposition are non-negative, the row and column sums are not one, in general, and, thus, the squared decomposition is not doubly stochastic.

All in all, the elementwise Hellinger decomposition is not as straightforward as the Frobenius decomposition, as the squared decomposition does not keep the rank of the truncation, and the attached optimization problems get more complex. Through the elementwise square root, the influence of peaks in the checkerboard copula on the objective function is reduced compared to the Frobenius case. It is a modeling choice, whether this is desired or not. Rao (1995) and C. M. Cuadras and D. Cuadras (2006) point out that elementwise Hellinger-based decomposition's main advantage is the independence from the row and column marginals. However, the marginals are constant in the checkerboard copula setting; thus, the correspondence analysis does not depend on them. Thus, we do not expand on the Hellinger decompositions in the main part of the thesis.

A.3. Computations for Spearman's ρ and Kendall's τ in Section 3.2.4

As in Section 3.2.4, $\mathbf{A}^n = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ denotes the centered copulas SVD. Let additionally be $\mathbf{u}_0 = \mathbf{v}_0 = 1/\sqrt{n} \cdot \mathbb{1}$ and $s_0 = 1$ to ease the notation. The equations for Spearman's ρ in

Section 3.2.4 follow from

$$
\begin{aligned}
\rho(\mathbf{C}^n) &:= \frac{3}{n}\operatorname{trace}\left(\mathbf{\Omega}\mathbf{C}^n\right) - 3 \\
&= \frac{3}{n}\operatorname{trace}\left(\Omega \sum_{k=0}^{n-1} \mathbf{u}_k s_k \mathbf{v}_k^\top\right) - 3 \\
&= \frac{3}{n}\sum_{k=0}^{n-1} s_k \operatorname{trace}\left(\Omega \mathbf{u}_k \mathbf{v}_k^\top\right) - 3 \\
&= \frac{3\|\check{\boldsymbol{\omega}}\|^2}{n}\sum_{k=0}^{n-1} s_k \operatorname{trace}\left(\boldsymbol{\omega}\boldsymbol{\omega}^\top \mathbf{u}_k \mathbf{v}_k^\top\right) - 3 \\
&= \frac{3\cdot(\frac{4}{3}n - \frac{1}{3n})}{n}\sum_{k=0}^{n-1} s_k \operatorname{trace}\left(\boldsymbol{\omega}^\top \mathbf{u}_k \mathbf{v}_k^\top \boldsymbol{\omega}\right) - 3 \\
&= (4 - \tfrac{1}{n^2})\sum_{k=0}^{n-1} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle - 3 \\
&= (4 - \tfrac{1}{n^2})\left(\left\langle \boldsymbol{\omega}, \tfrac{1}{\sqrt{n}}\cdot \mathbb{1}\right\rangle^2 + \sum_{k=1}^{n^*} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle\right) - 3 \\
&= (4 - \tfrac{1}{n^2})\sum_{k=1}^{n-1} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle + 3 - 3 \\
&= (4 - \tfrac{1}{n^2})\sum_{k=1}^{n-1} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle.
\end{aligned}
$$

Similarly follows for the MAR decomposition, $\tilde{\mathbf{A}}^n = \tilde{\mathbf{U}}\tilde{\mathbf{S}}(\tilde{\mathbf{V}})^\top$,

$$
\begin{aligned}
\rho(\mathbf{C}^n) &= \frac{3}{n}\operatorname{trace}\left(\mathbf{\Omega}\left(\sum_{k=1}^{n-1}\tilde{\mathbf{u}}_k \tilde{s}_k \tilde{\mathbf{v}}_k^\top - \eta I_n + (1+\eta)\mathbf{\Pi}^n\right)\right) - 3 \\
&= (4 - \tfrac{1}{n^2})\left[\sum_{k=1}^{n-1}\tilde{s}_k \operatorname{trace}\left(\boldsymbol{\omega}\boldsymbol{\omega}^\top \tilde{\mathbf{u}}_k \tilde{\mathbf{v}}_k^\top\right) - \eta \operatorname{trace}\left(\boldsymbol{\omega}\boldsymbol{\omega}^\top I_n\right) + (1+\eta)\operatorname{trace}\left(\boldsymbol{\omega}\boldsymbol{\omega}^\top \mathbf{\Pi}^n\right)\right] - 3 \\
&= (4 - \tfrac{1}{n^2})\left[\sum_{k=1}^{n-1}\tilde{s}_k \langle \tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle\langle \tilde{\mathbf{v}}_k, \boldsymbol{\omega}\rangle - \eta\|\boldsymbol{\omega}\|^2 + (1+\eta)\left\langle \boldsymbol{\omega}, \tfrac{1}{\sqrt{n}}\cdot \mathbb{1}\right\rangle^2\right] - 3 \\
&= (4 - \tfrac{1}{n^2})\left[\sum_{k=1}^{n-1}\tilde{s}_k \langle \tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle\langle \tilde{\mathbf{v}}_k, \boldsymbol{\omega}\rangle - (4 - \tfrac{1}{n^2})\eta + \tfrac{3}{n}(1+\eta)\right] - 3 \\
&\overset{(*)}{=} (4 - \tfrac{1}{n^2})\left[\sum_{k=1}^{n-1}(s_k + \eta)\langle \tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle^2 - (4 - \tfrac{1}{n^2})\eta + \tfrac{3}{n}(1+\eta)\right] - 3,
\end{aligned}
$$

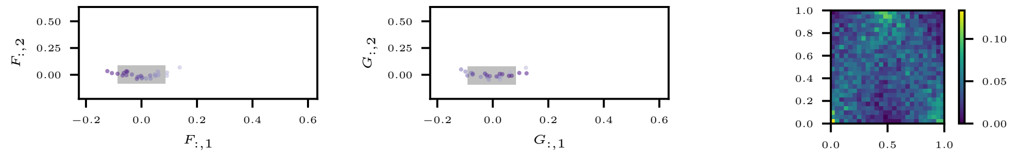
where $(*)$ is only valid for symmetric $\tilde{\mathbf{A}}^n$.

The respective computations for Kendall's τ are

$$\begin{aligned}
\tau(\mathbf{C}^n) &:= 1 - \frac{1}{n^2}\operatorname{trace}\left(\mathbf{E}\mathbf{C}^n\mathbf{E}(\mathbf{C}^n)^\top\right) \\
&= 1 - \frac{1}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} s_{k_1}s_{k_2}\operatorname{trace}\left(\mathbf{E}\mathbf{u}_{k_1}\mathbf{v}_{k_1}^\top\mathbf{E}\mathbf{v}_{k_2}\mathbf{u}_{k_2}^\top\right) \\
&= 1 - \frac{1}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} s_{k_1}s_{k_2}\operatorname{trace}\left(\mathbf{u}_{k_2}^\top\mathbf{E}\mathbf{u}_{k_1}\mathbf{v}_{k_1}^\top\mathbf{E}\mathbf{v}_{k_2}\right) \\
&= 1 - \frac{1}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} s_{k_1}s_{k_2}\langle\mathbf{u}_{k_2},\mathbf{E}\mathbf{u}_{k_1}\rangle\langle\mathbf{v}_{k_1},\mathbf{E}\mathbf{v}_{k_2}\rangle
\end{aligned}$$

and for the MAR analogously.

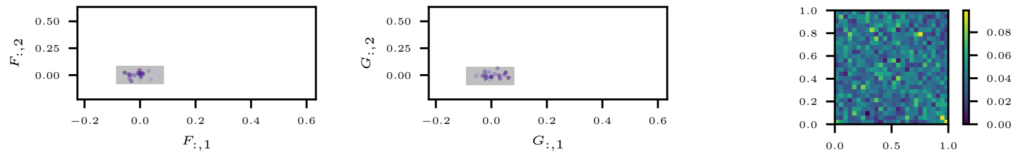
A.4. Further figures for Section 3.4.2



(a) Row profiles for variables u_1 and u_3.

(b) Column profiles for variables u_1 and u_3.

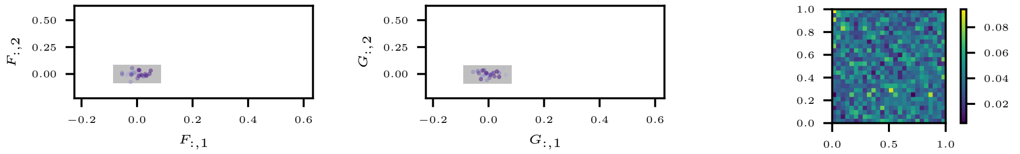
(c) Checkerboard plot for variables u_1 and u_3.



(d) Row profiles for variables u_1 and u_5.

(e) Column profiles for variables u_1 and u_5.

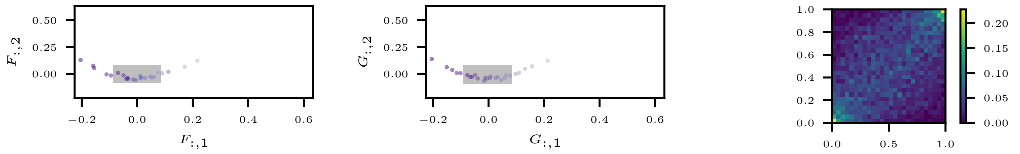
(f) Checkerboard plot for variables u_1 and u_5.



(g) Row profiles for variables u_2 and u_5.

(h) Column profiles for variables u_2 and u_5.

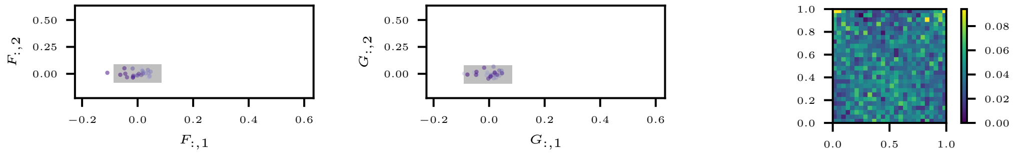
(i) Checkerboard plot for variables u_2 and u_5.



(j) Row profiles for variables u_3 and u_5.

(k) Column profiles for variables u_3 and u_5.

(l) Checkerboard plot for variables u_3 and u_5.



(m) Row profiles for variables u_4 and u_5.

(n) Column profiles for variables u_4 and u_5.

(o) Checkerboard plot for variables u_4 and u_5.

Figure A.1.: The remaining profile and checkerboard plots of the fuel injector spray characteristics in jet engines from Coblenz et al. (2020) from Section 3.4.2. The other dimension combinations are shown in Figure 3.13. The physical interpretations of the variables are drop size (u_1), x-position (u_2), y-position (u_3), x-velocity (u_4), and y-velocity (u_5). For the variable pairs (u_1, u_5) and (u_2, u_5), no deviation from independence is discernible. A weak hump-shape can be observed for variables u_1 and u_3. Again, the course of column profiles is reversed in the middle of the profiles. The plots show a Gaussian-like behavior for variables u_3 and u_5. The profile plots for variables u_4 and u_5 show a weak deviation from independence for the profiles near $u_4 = 1$ and extreme values of u_5.

B. Appendix to Chapter 4

The section presents additional material for Chapter 4 and is based on Publ. II. Appendix B.1 introduces other forecast combination methods used in the simulation study in Section 4.3.1. Appendix B.2 contains the proofs of the theorems presented in Section 4.2.3. Appendices B.3 and B.4 present additional results for the simulations of Section 4.3 and the electricity price forecast combination of Section 4.4.

B.1. Forecast combination methods

In many situations, different forecasts are available in the form of predictive distributions or point forecasts, for example, in the area of weather predictions where forecasts of different weather services or originating from different models are available (see, for example, the reviews in Hall and Mitchell, 2007; Gneiting and Katzfuss, 2014; X. Wang et al., 2023). Combining these different forecasts can form a new predictive distribution that is better than all individual ones since it combines the information of the different forecasters. Depending on the form of the predictions, different algorithms can be applied to combine them. As for individual forecasts, the aggregated predictive distribution should be sharp subject to calibration (see Section 2.2). Thus, scoring rules can be used to fit all the model parameters.

In practice, one has to determine the combination scheme and parameters considering training data from the past. This training data includes the forecasts in the form of CDFs issued by different forecasters $1, \dots, K$ at times $1, \dots, T$ and the corresponding realizing observation y_t

$$\{(F_{1,j}, \dots, F_{K,j}, y_t) : j = 1, \dots, T\}.$$

Training data can similarly be given as densities or point forecasts and realizing observations.

We start by introducing algorithms for the combination of distributional forecasts. Although the CCD method accepts point forecasts, the following algorithms are used in the simulation study in Section 4.3.1 with the true marginal distributions. A simple example of an aggregation method is the traditional linear pool (TLP), where the individual forecasts are combined linearly

$$G_\theta^{TLP}(F_1, \dots, F_K) = \sum_{i=1}^{K} \theta_i F_i$$

with the parameter $\theta = (\theta_1, \dots, \theta_K) \in \Theta = \{\theta : \theta_i > 0, \sum_{i=1}^{K} \theta_i = 1\}$. Ranjan and Gneiting (2010) proves that the TLP tends to increase the dispersion and combinations of calibrated individual forecasts are overdispersed.

Berrocal, Raftery, and Gneiting (2007), Glahn et al. (2009) and Kleiber, Raftery, and Gneiting (2011) propose a nonlinear aggregation method to overcome the shortcomings of the TLP, which Gneiting and Ranjan (2013) generalized to the spread-adjusted linear pool (SLP). With $F_i(y) = F_i^0(y - \mu_i)$ and $f_i(y) = f_i^0(y - \mu_i)$ for μ_i being the unique median of F_i for $i = 1, \dots, K$, the SLP can be written as

$$G_c(y) = \sum_{i=1}^{K} \theta_i F_i^0(\frac{y - \mu_i}{c})$$

with density

$$g_c(y) = \frac{1}{c} \sum_{i=1}^{K} \theta_i f_i^0(\frac{y - \mu_i}{c}).$$

Hereby, $\theta_1, \dots, \theta_K$ are again strictly positive weights summing up to 1, and c is the spread-adjustment parameter taking strictly positive values. For $c = 1$, the traditional linear pool arises.

The beta-transformed linear pool (BLP) is again an extension of the TLP, where the result of the linear combination is transformed with the CDF of a beta distribution $B_{\alpha,\beta}$ with parameters α and β (Ranjan and Gneiting, 2010)

$$G_{\alpha,\beta} = B_{\alpha,\beta}(\sum_{i=1}^{K} \omega_i F_i).$$

The parameter space is then

$$\Theta = \{(\alpha, \beta, \theta_1, \dots, \theta_K) : \alpha, \beta, \theta_1, \dots, \theta_K > 0, \sum_{i=1}^{K} \theta_i = 1\}.$$

α and β are the parameters of the beta distribution and $\theta_1, \dots, \theta_K$ are linear weights. Through the transformation, the BLP can combine a wide class of forecasts to calibrated forecasts (Gneiting and Ranjan, 2013).

In addition to those methods for combining distributional forecasts, there are methods for combining point forecasts from various sources. A set of point forecasts is often called an ensemble prediction system; therefore, these methods are also called ensemble output statistics.

An early point forecast combination algorithm dates back to Bates and Granger (1969). They compute the linear combination of two unbiased point forecasts $\hat{x}_1$ and $\hat{x}_2$ with error variances σ_1^2 and σ_2^2 and correlation coefficient ρ that minimizes the variance of the combined forecast. They conclude that the forecast

$$\hat{x} = k\hat{x}_1 + (1-k)\hat{x}_2$$

with

$$k = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

and resulting standard deviation

$$\sigma = \frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}}$$

is optimal if all observations are equally weighted in the variance computation. Using $\hat{x}$ and σ as parameters of a Gaussian distribution yields a probabilistic forecast for the quantity of interest.

Two more advanced and flexible methods for point forecast combination are Bayesian model averaging (BMA) by Raftery et al. (2005) and ensemble model output statistics (EMOS, Gneiting, Raftery, et al., 2005). In BMA, the combined PDF is a weighted average of the forecast PDFs. The forecast PDF is based on individual bias-corrected point forecasts. The BMA model for y is then

$$p(y|x_1, \dots, x_m) = \sum_{i=1}^{m} w_m g_m(y|x_m).$$

Thereby, $x_1, \dots, x_m$ are the bias-corrected point forecasts, $g_k(y|x_k)$ is the conditional PDF of y dependent on x_k, given that x_k is the best forecast in the ensemble, and w_k are nonnegative weights that sum up to 1.

The EMOS method issues a single parametric distribution, where the distribution parameters are calculated based on the ensemble predictions (see for general instructions, for example, Wilks, 2011, Chapter 7.7). The normal predictive distribution of EMOS for ensemble members $X_1, \dots, X_m$ and ensemble variance S^2 is

$$\mathcal{N}(a + b_1X_1 + \cdots + b_mX_m, c + dS^2). \tag{B.1}$$

For a more straightforward interpretation, $b_1, \dots, b_m$ can be restricted to be nonnegative. Other parametric distributions, such as truncated normal or log-normal distributions, are possible (Baran and Lerch, 2018).

B.2. Proofs for Chapter 4

B.2.1. Proof of Theorem 4.2.1

Proof. Due to Sklar's theorem (Sklar, 1959), the density in Equation (4.5) is the actual density of the forecast errors. In the following, we will omit the copula notation and write in terms of the error density f.

The proof proceeds by a rotation of the $\mathbb{R}^K$ such that the first basis vector of the new basis has coordinates $(1, \dots, 1)$ with respect to the original space. Let T be the corresponding orthonormal transformation matrix, that is, $\det(T) = 1$ and the first column of the inverse matrix $(T^{-1})_{\cdot 1} = (1/\sqrt{n}, \dots, 1/\sqrt{n})'$. Let then $\phi : \mathbb{R}^K \to \mathbb{R}^K, x \mapsto \phi(x) = x \cdot T$ be a linear mapping, and $g(z) := f(z \cdot T)$ the density with points according to the transformed coordinates. Note that x is a row vector. Then the CCD forecast CDF can be expressed as

$$\begin{aligned}
\hat{F}(y) &= \frac{1}{l}\int_{-\infty}^{y} \mathcal{L}(\gamma|\hat{\mathbf{x}})d\gamma \\
&= \frac{1}{l}\int_{-\infty}^{y} f(\hat{x}_1 - \gamma, \dots, \hat{x}_K - \gamma)d\gamma \\
&= \frac{1}{l}\int_{-\infty}^{y} f\big(\hat{\mathbf{x}} - \gamma \cdot (1, \dots, 1)\big)d\gamma \\
&= \frac{1}{l}\int_{-\infty}^{y} f\Big(\big(\hat{\mathbf{x}} - \gamma \cdot (1, \dots, 1)\big) \cdot T^{-1} \cdot T\Big)d\gamma \\
&= \frac{1}{l}\int_{-\infty}^{y} g(\underbrace{\hat{\mathbf{x}} \cdot T^{-1}}_{=:\hat{\mathbf{z}}} - \gamma \cdot (1, \dots, 1) \cdot T^{-1})d\gamma \\
&= \frac{1}{l}\int_{-\infty}^{y} g\big(\hat{\mathbf{z}} - \gamma \cdot \sqrt{n} \cdot (1, 0, \dots 0)\big)d\gamma =: G(\hat{\mathbf{z}}).
\end{aligned}$$

As the random variable Y models the error, its true distribution is a point mass in 0. Instead, the forecasts $\hat{\mathbf{x}}$ resp. $\hat{\mathbf{z}}$ are probabilistic. This is reflected in the following notion G^z where $y = 0$ is fixed and the values of $\hat{\mathbf{z}} = (\hat{z}_1, \hat{\mathbf{z}}_{-1})$ vary

$$G^z(\hat{z}_1, \hat{z}_{-1}) = \frac{1}{l(\hat{z}_1, \hat{z}_{-1})}\int_{-\infty}^{0} g\big((\hat{z}_1, \hat{z}_{-1}) - \gamma \cdot \sqrt{n} \cdot (1, 0, \dots 0)\big)d\gamma. \tag{B.2}$$

The notation $\mathbf{z}_{-i}$ takes the vector $\mathbf{z}$ without the i-th entry. Let now $\hat{\mathbf{z}}_{-1} \in R^{K-1}$ be fixed. Then,

$$\begin{aligned} G^z(\hat{z}_1, \hat{\mathbf{z}}_{-1}) &\to 1, \qquad &&\text{for } \hat{z}_1 \to -\infty \\ G^z(\hat{z}_1, \hat{\mathbf{z}}_{-1}) &\to 0, \qquad &&\text{for } \hat{z}_1 \to \infty. \end{aligned}$$

G^z is as integral of a bounded function, a continuous function in $\hat{z}_1$. Therefore, for all $u \in (0,1)$ the minimum

$$\min\{z_1 \in \mathbb{R} : G^z(z_1, \mathbf{z}_{-1})\} =: z_1^*(\mathbf{z}_{-1}).$$

exists. Let now $u \in (0,1)$ be fixed.

$$\begin{aligned} \mathbb{P}\{G^z(z_1, \mathbf{z}_{-1}) \leq u | \mathbf{z}_{-1}\} &= \frac{1}{l(z_1, \mathbf{z}_{-1})} \cdot && \text{(B.3)} \\ &\int_{-\infty}^{\infty} \mathbb{1}\{G^z(z_1, \mathbf{z}_{-1}) \leq u\} g(z_1, \mathbf{z}_{-1}) dz_1 && \text{(B.4)} \\ &= \frac{1}{l(z_1, \mathbf{z}_{-1})} \int_{z_1^*(\mathbf{z}_{-1})}^{\infty} g(z_1, \mathbf{z}_{-1}) dz_1 && \text{(B.5)} \\ &= G^z(z_1^*, \mathbf{z}_{-1}) && \text{(B.6)} \\ &= u. && \text{(B.7)} \end{aligned}$$

Then, with $g_{\mathbf{z}_{-1}}$ being the marginal density of $\mathbf{z}_{-1}$ and $g_{z_1|\mathbf{z}_{-1}}$ being the density of z_1 conditioning on $\mathbf{z}_{-1}$, for all $z_{-1} \in \mathbb{R}^{K-1}$ and $u \in (0,1)$

$$
\begin{aligned}
\mathbb{P}\{G(0) \leq u\} &= \int_{\mathbb{R}^K} g(z_1, \dots, z_K) \mathbb{1}\{G^z(z_1, \dots, z_K) \leq u\} \,\mathrm{d}\, z && \text{(B.8)} \\
&= \int_{\mathbb{R}^{K-1}} \int_{\mathbb{R}} g_{z_1|\mathbf{z}_{-1}}(z_1|\mathbf{z}_{-1}) g_{\mathbf{z}_{-1}}(\mathbf{z}_{-1}) && \text{(B.9)} \\
&\qquad \mathbb{1}\{G^z(z_1, \dots, z_K) \leq u\} \,\mathrm{d}\, z_1 \,\mathrm{d}\, \mathbf{z}_{-1} && \text{(B.10)} \\
&= \int_{\mathbb{R}^{K-1}} g_{\mathbf{z}_{-1}}(\mathbf{z}_{-1}) \int_{\mathbb{R}} g_{z_1|\mathbf{z}_{-1}}(z_1|\mathbf{z}_{-1}) && \text{(B.11)} \\
&\qquad \mathbb{1}\{G^z(z_1, \dots, z_K) \leq u\} \,\mathrm{d}\, z_1 \,\mathrm{d}\, \mathbf{z}_{-1} && \text{(B.12)} \\
&= \int_{\mathbb{R}^{K-1}} g_{\mathbf{z}_{-1}}(\mathbf{z}_{-1}) \mathbb{P}\{G^z(z_1, \mathbf{z}_{-1}) \leq u | \mathbf{z}_{-1}\} \,\mathrm{d}\, \mathbf{z}_{-1} && \text{(B.13)} \\
&= \int_{\mathbb{R}^{K-1}} g_{\mathbf{z}_{-1}}(\mathbf{z}_{-1}) u \,\mathrm{d}\, \mathbf{z}_{-1} && \text{(B.14)} \\
&= u. && \text{(B.15)}
\end{aligned}
$$

Thus,

$$
Z_{\hat{F}} = F(Y) = G(0) \sim U(0,1).
$$

□

B.2.2. Proof of Theorem 4.2.2

Proof. According to Equation (4.8)

$$
\begin{aligned}
\hat{f}(x) &= \alpha \cdot \mathcal{L}(x) \\
&= \alpha \cdot f_{(\varepsilon_1, \varepsilon_2)}(\hat{x}_1 - x, \hat{x}_2 - x | x)
\end{aligned}
$$

with $\alpha \in \mathbb{R}$ being the normalising constant such that $\int_{\mathbb{R}} \hat{f}(x)dx = 1$. Let C_R^{Ga} be a Gauss-Copula with mean $\mu = \mathbf{0}$ and correlation matrix $R = \operatorname{Corr}((\varepsilon_1, \varepsilon_2)')$. Let $F_{\varepsilon_1} = \Phi_{\varepsilon_1}$ resp. $F_{\varepsilon_2} = \Phi_{\varepsilon_2}$ be the (normally distributed) CDF of ε_1 resp. ε_2 and $F_{\varepsilon_1}^{-1} = \Phi_{\varepsilon_1}^{-1}$ resp. $F_{\varepsilon_2}^{-1} = \Phi_{\varepsilon_2}^{-1}$ the corresponding inverse. Let Φ_R be a bivariate Gaussian CDF with mean $\mu = \mathbf{0}$ and correlation matrix $R = \operatorname{Corr}((\varepsilon_1, \varepsilon_2)')$ and ϕ_R the corresponding density. Then

$$\begin{aligned}
&f_{(\varepsilon_1,\varepsilon_2)}(\hat{x}_1 - x, \hat{x}_2 - x|x) \\
&= \frac{\partial^2}{\partial \hat{x}_1 \partial \hat{x}_2} F_{(\varepsilon_1,\varepsilon_2)}(\hat{x}_1 - x, \hat{x}_2 - x|x) \\
&= \frac{\partial^2}{\partial \hat{x}_1 \partial \hat{x}_2} C_R^{Ga}(F_{\varepsilon_1}(\hat{x}_1 - x|x), F_{\varepsilon_2}(\hat{x}_2 - x|x)) \\
&= \frac{\partial^2}{\partial \hat{x}_1 \partial \hat{x}_2} \Phi_R(\Phi^{-1}(F_{\varepsilon_1}(\hat{x}_1 - x|x)), \Phi^{-1}(F_{\varepsilon_2}(\hat{x}_2 - x|x)) \\
&= \phi_R(\hat{x}_1 - x, \hat{x}_2 - x|x) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \\
&\quad \exp\left[-\frac{1}{2(1-\rho^2)} \underbrace{\left(\frac{(\hat{x}_1 - x)^2}{\sigma_1^2} - 2\frac{\rho(\hat{x}_1 - x)(\hat{x}_2 - x)}{\sigma_x\sigma_y} + \frac{(\hat{x}_2 - x)^2}{\sigma_2^2} \right)}_{=:\beta} \right].
\end{aligned}$$

Thereby, β can be simplified to

$$\begin{aligned}
\beta &= \frac{(\hat{x}_1 - x)^2}{\sigma_1^2} - 2\frac{\rho(\hat{x}_1 - x)(\hat{x}_2 - x)}{\sigma_1\sigma_2} + \frac{(\hat{x}_2 - x)^2}{\sigma_2^2} \\
&= \frac{1}{\sigma_1^2}\left(x^2 - 2\hat{x}_1 x + \hat{x}_1^2\right) + \frac{2\rho}{\sigma_1\sigma_2}\left(-x^2 + x\left(\hat{x}_1 + \hat{x}_2\right) - \hat{x}_1\hat{x}_2\right) + \\
&\quad \frac{1}{\sigma_2^2}\left(x^2 - 2\hat{x}_2 x + \hat{x}_2^2\right) \\
&= x^2\left(\frac{1}{\sigma_1^2} - \frac{2\rho}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}\right) - 2x\left(\frac{\hat{x}_1}{\sigma_1^2} - \frac{\rho\left(\hat{x}_1 + \hat{x}_2\right)}{\sigma_1\sigma_2} + \frac{\hat{x}_2}{\sigma_2^2}\right) + \\
&\quad \left(\frac{\hat{x}_1^2}{\sigma_1^2} - \frac{2\rho\hat{x}_1\hat{x}_2}{\sigma_1\sigma_2} + \frac{\hat{x}_2^2}{\sigma_2^2}\right) \\
&= \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2}\left(x^2 - 2x\mu_c + l\right) \\
&= \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2}\left(x^2 - 2x\mu_c + \mu_c^2 - (\mu_c^2 - l)\right)
\end{aligned}$$

with the constants with regard to x

$$\mu_c = k\hat{x}_1 + (1-k)\hat{x}_2, \tag{B.16}$$

$$k = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}, \text{ and} \tag{B.17}$$

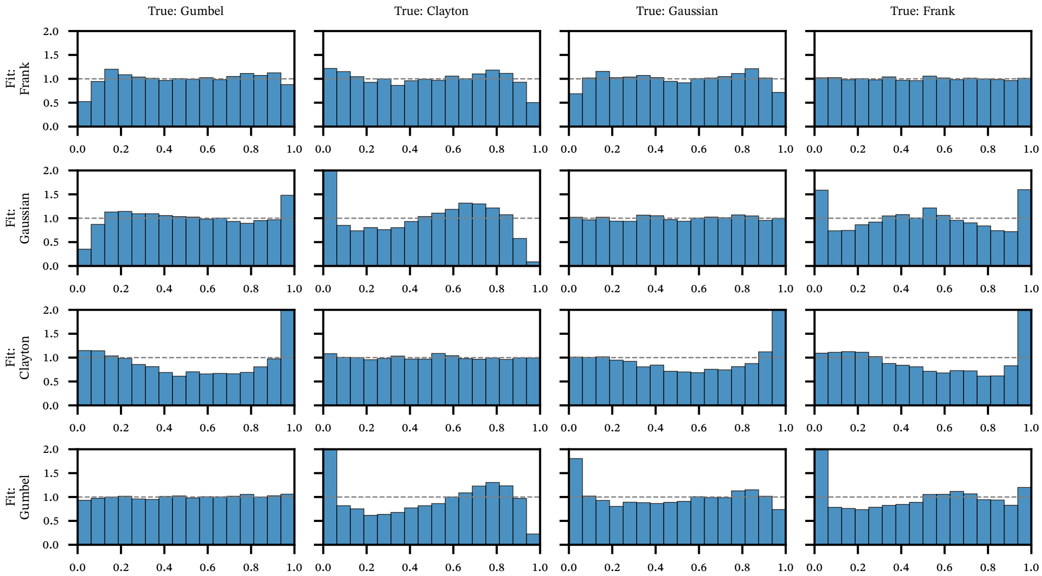
$$l = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}\left(\frac{\hat{x}_1^2}{\sigma_1^2} - \frac{2\rho\hat{x}_1\hat{x}_2}{\sigma_1\sigma_2} + \frac{\hat{x}_2^2}{\sigma_2^2}\right). \tag{B.18}$$

Then it follows for $\hat{f}(x)$

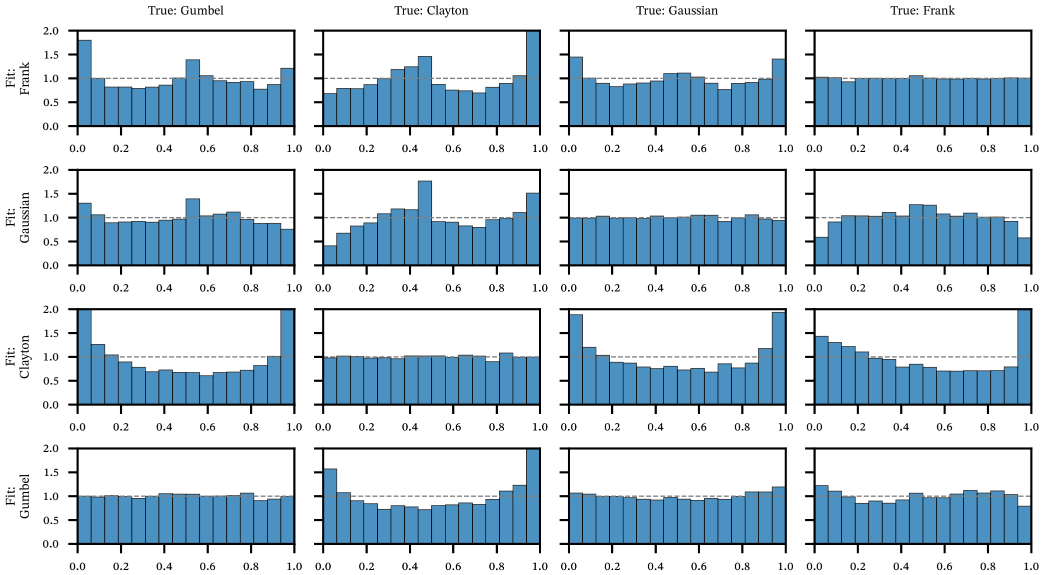
$$\begin{aligned}
\hat{f}(x) &\propto f_{(\varepsilon_1,\varepsilon_2)}(\hat{x}_1 - x, \hat{x}_2 - x|x) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \\
&\quad \exp\left[-\frac{1}{2(1-\rho^2)}\frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2}\left(x^2 - 2x\mu_c + \mu_c^2 - (\mu_c^2 - l)\right)\right] \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(\frac{1}{2\sigma_c^2}\left(\mu_c^2 - l\right)\right) \cdot \exp\left[-\frac{1}{2\sigma_c^2}\left(x^2 - 2x\mu_c + \mu_c^2\right)\right] \\
&\propto \frac{1}{\sqrt{2\pi}\sigma_c}\exp\left[-\frac{1}{2}\frac{(x-\mu_c)^2}{\sigma_c^2}\right],
\end{aligned}$$

which is the density of a normal distributed random variable with mean μ_c and variance σ_c^2. $\hat{f}(x)$ is the density of a normal distribution, as it is proportional to it, and it has to have the same normalizing constant to be a proper density with $\int_{\mathbb{R}} \hat{f}(x)dx = 1$. □

B.3. Further simulation results

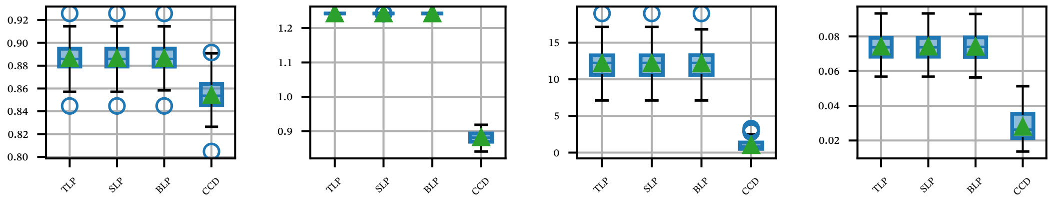


(a) The margins are uniform on $\left(-\sqrt{12}/2, \sqrt{12}/2\right)$.



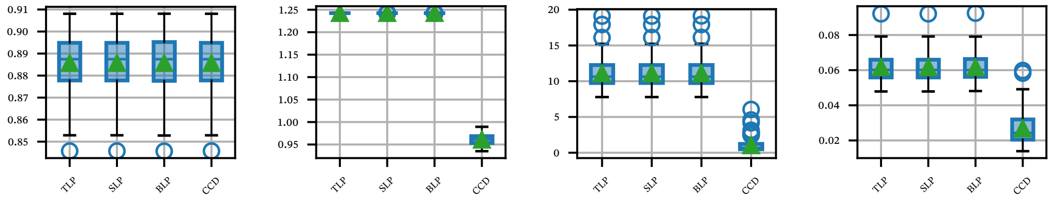
(b) The margins are Student's t with $\left(\mu = 0, \sigma = \sqrt{1/2}, \nu = 4\right)$.

Figure B.1.: Effect of a misspecified copula on the calibration of the forecast. 10 000 points are evaluated using the "fit" copula family for data stemming from the "true" copula. Both copulas use the same, "true" $\tau = 0.8$. For notes on the interpretation, see Section 4.2.4.



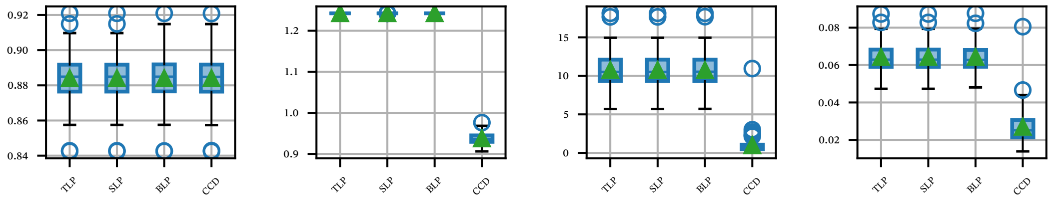
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.2.: CCD simulation results for the scenario with margins $\mathcal{M}_1$, $\tau = 0.4$, and a Clayton copula.



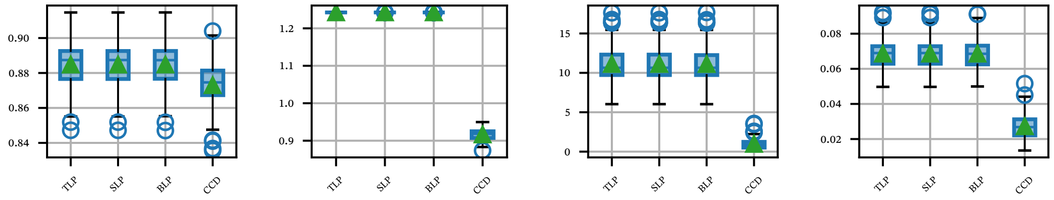
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.3.: CCD simulation results for the scenario with margins $\mathcal{M}_1$, $\tau = 0.4$, and a Frank copula.



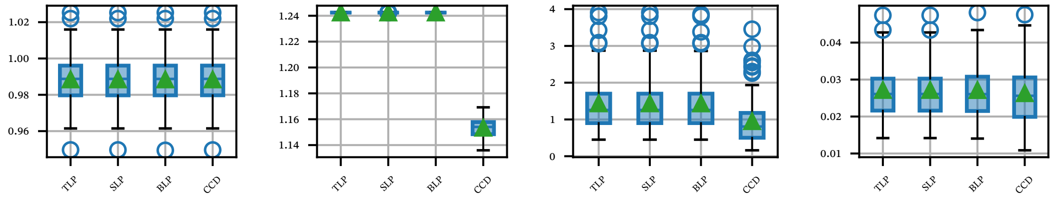
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.4.: CCD simulation results for the scenario with margins $\mathcal{M}_1$, $\tau = 0.4$, and a Gaussian copula.



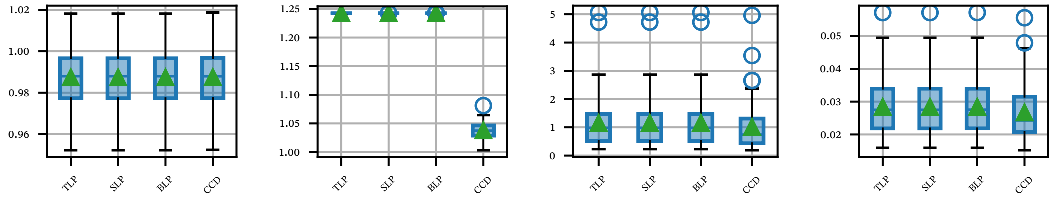
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.5.: CCD simulation results for the scenario with margins $\mathcal{M}_1$, $\tau = 0.4$, and a Gumbel copula.



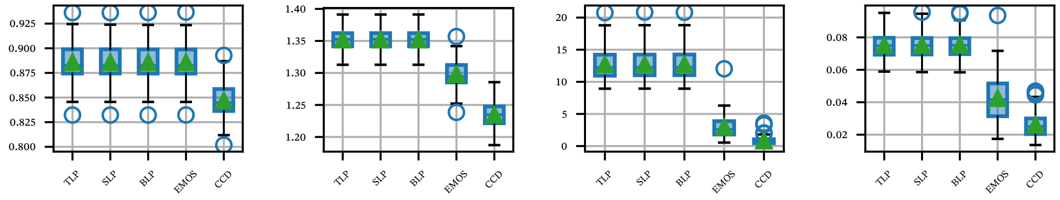
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.6.: CCD simulation results for the scenario with margins $\mathcal{M}_1$, $\tau = 0.8$, and a Frank copula.



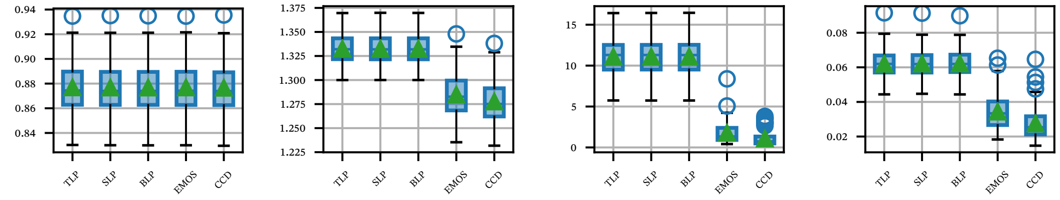
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.7.: CCD simulation results for the scenario with margins $\mathcal{M}_1$, $\tau = 0.8$, and a Gaussian copula.



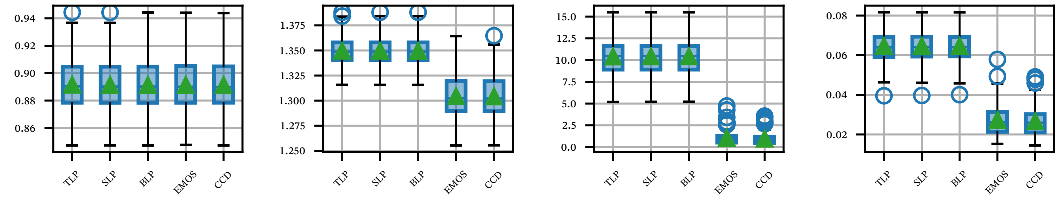
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.8.: CCD simulation results for the scenario with margins $\mathcal{M}_2$, $\tau = 0.4$, and a Clayton copula.



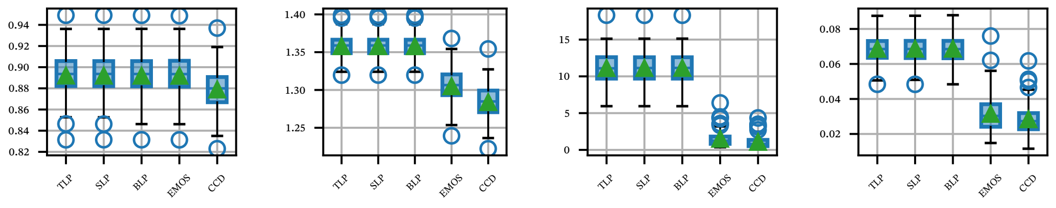
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.9.: CCD simulation results for the scenario with margins $\mathcal{M}_2$, $\tau = 0.4$, and a Frank copula.



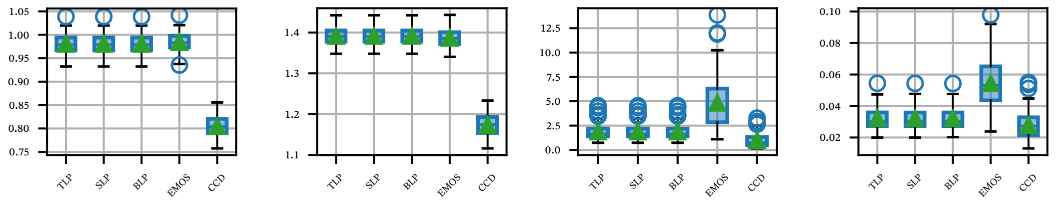
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.10.: CCD simulation results for the scenario with margins $\mathcal{M}_2$, $\tau = 0.4$, and a Gaussian copula.



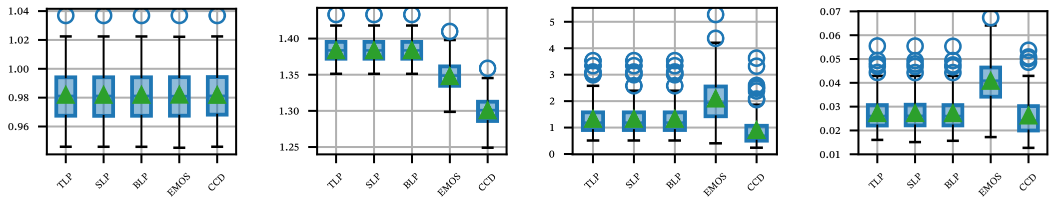
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.11.: CCD simulation results for the scenario with margins $\mathcal{M}_2$, $\tau = 0.4$, and a Gumbel copula.



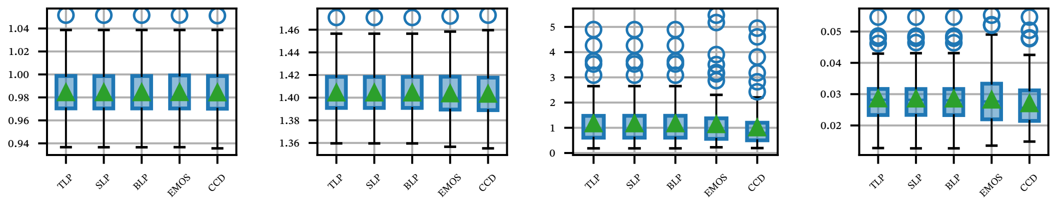
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.12.: CCD simulation results for the scenario with margins $\mathcal{M}_2$, $\tau = 0.8$, and a Clayton copula.



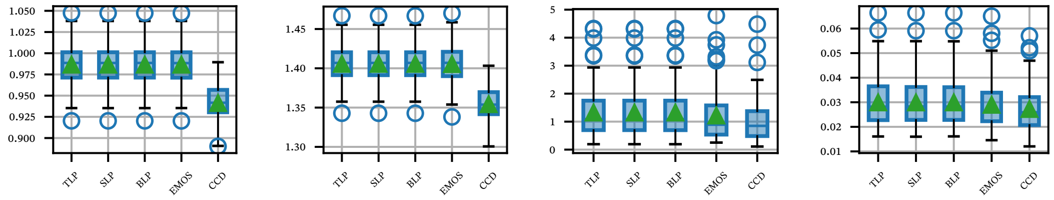
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.13.: CCD simulation results for the scenario with margins $\mathcal{M}_2$, $\tau = 0.8$, and a Frank copula.



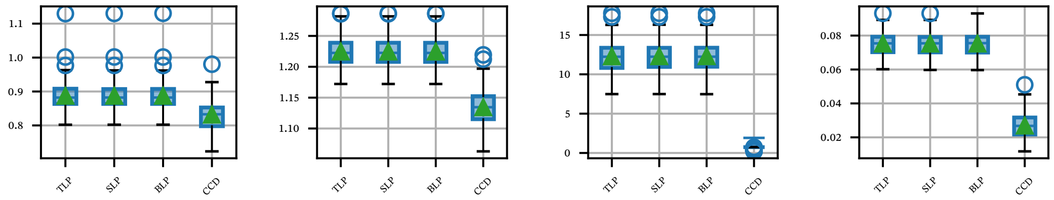
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.14.: CCD simulation results for the scenario with margins $\mathcal{M}_2$, $\tau = 0.8$, and a Gaussian copula.



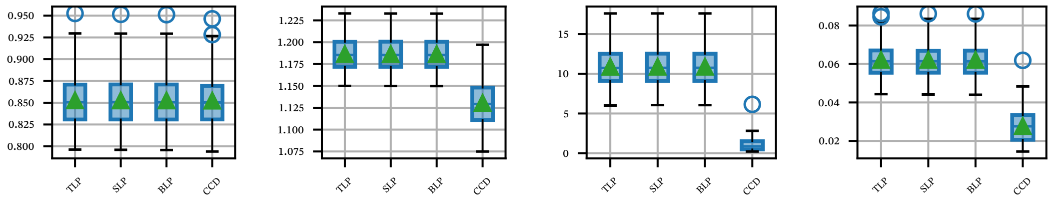
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.15.: CCD simulation results for the scenario with margins $\mathcal{M}_2$, $\tau = 0.8$, and a Gumbel copula.



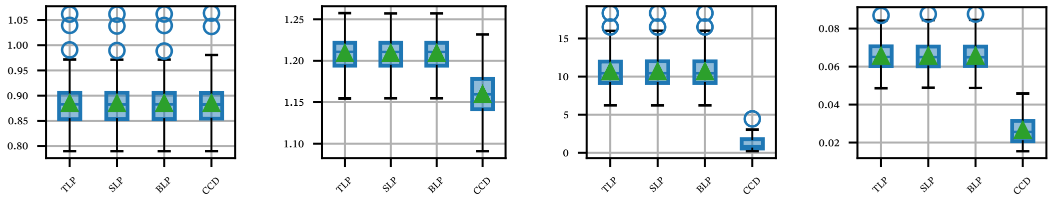
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.16.: CCD simulation results for the scenario with margins $\mathcal{M}_3$, $\tau = 0.4$, and a Clayton copula.



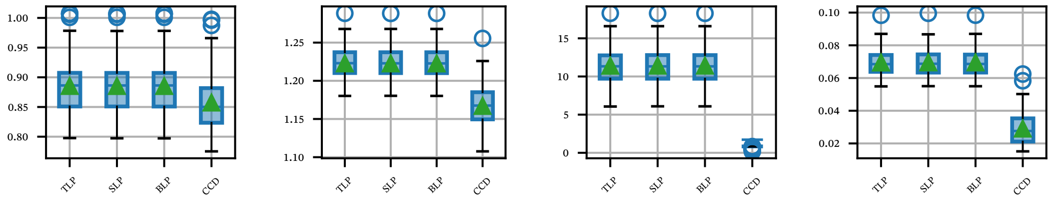
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.17.: CCD simulation results for the scenario with margins $\mathcal{M}_3$, $\tau = 0.4$, and a Frank copula.



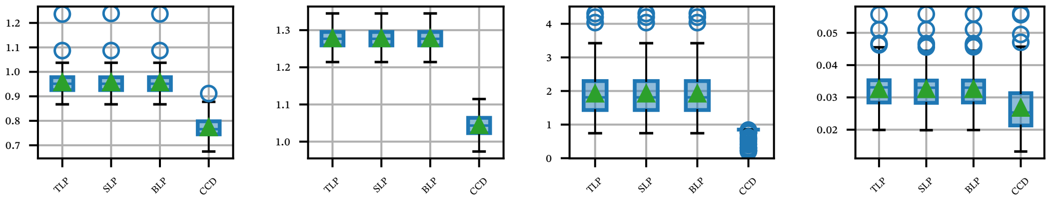
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.18.: CCD simulation results for the scenario with margins $\mathcal{M}_3$, $\tau = 0.4$, and a Gaussian copula.



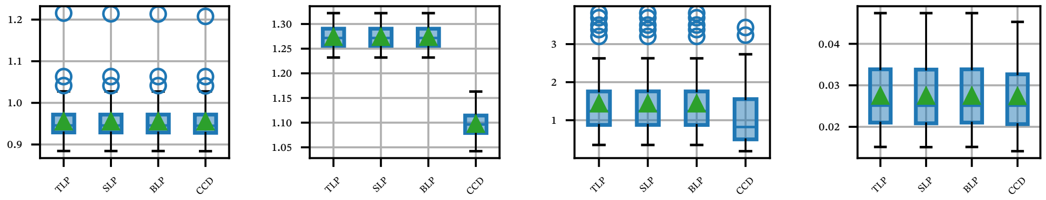
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.19.: CCD simulation results for the scenario with margins $\mathcal{M}_3$, $\tau = 0.4$, and a Gumbel copula.



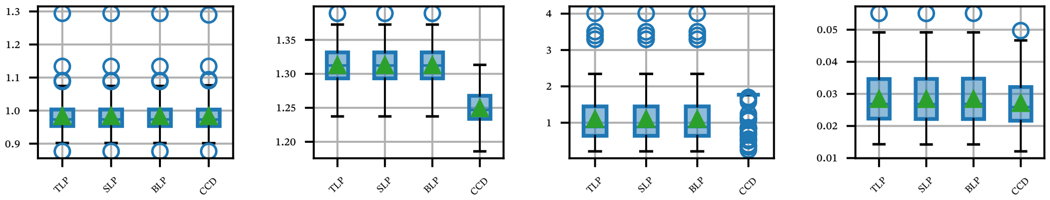
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.20.: CCD simulation results for the scenario with margins $\mathcal{M}_3$, $\tau = 0.8$, and a Clayton copula.



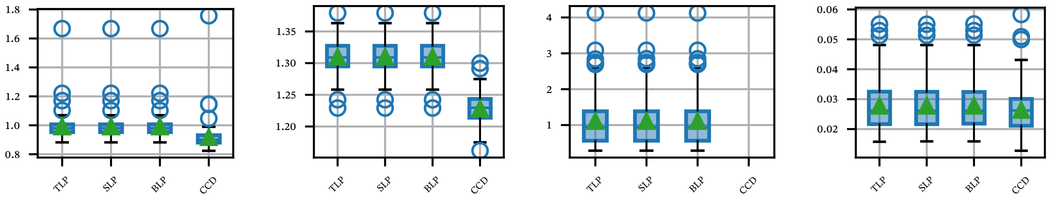
(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.21.: CCD simulation results for the scenario with margins $\mathcal{M}_3$, $\tau = 0.8$, and a Frank copula.



(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.22.: CCD simulation results for the scenario with margins $\mathcal{M}_3$, $\tau = 0.8$, and a Gaussian copula.



(a) RMSE **(b)** LS **(c)** AD **(d)** KS

Figure B.23.: CCD simulation results for the scenario with margins $\mathcal{M}_3$, $\tau = 0.8$, and a Gumbel copula. Note that the AD test statistic cannot be displayed for the CCD method due to numerical instabilities in the test statistic in some simulation runs.

B.4. Further results for Section 4.4

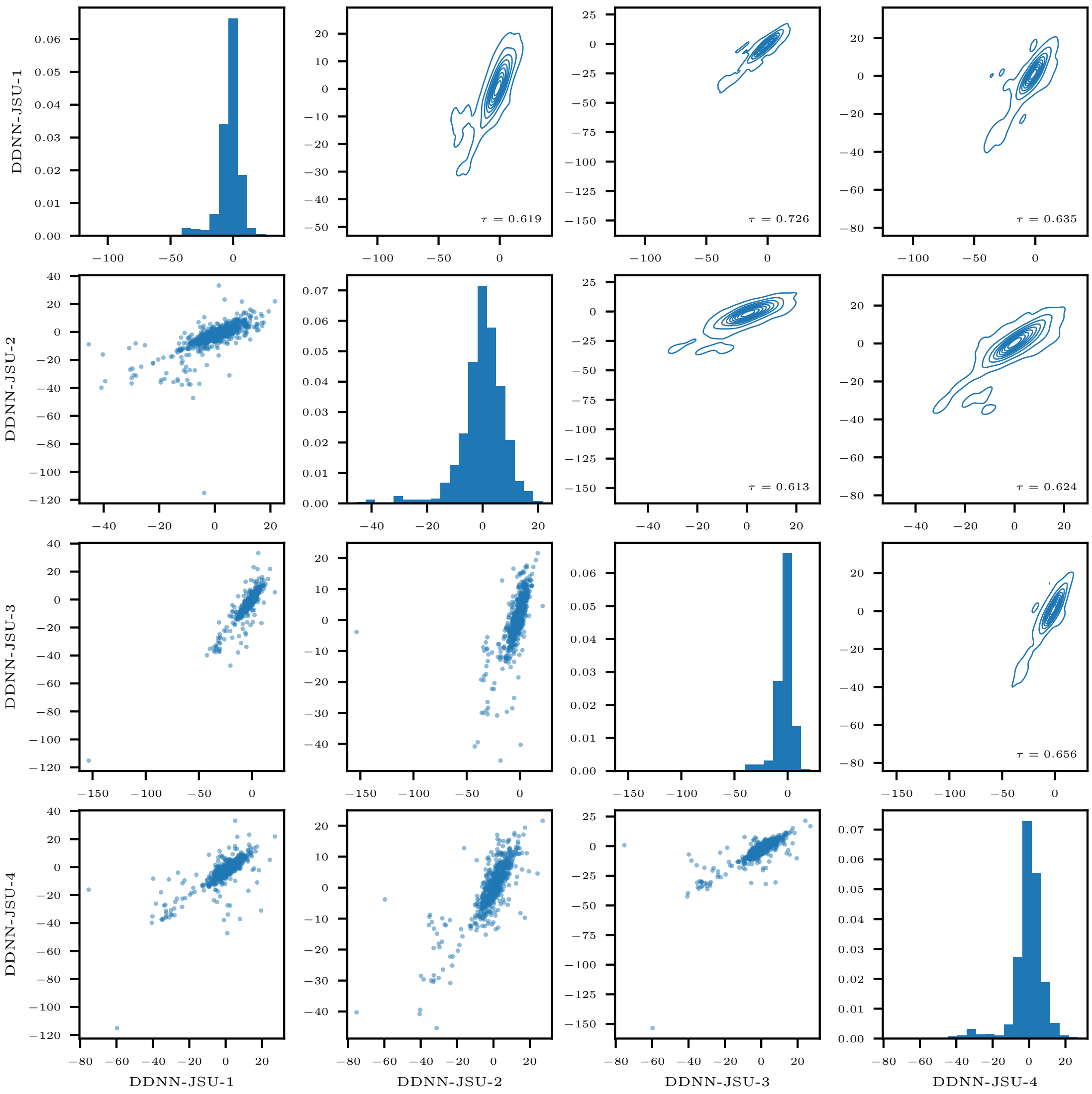
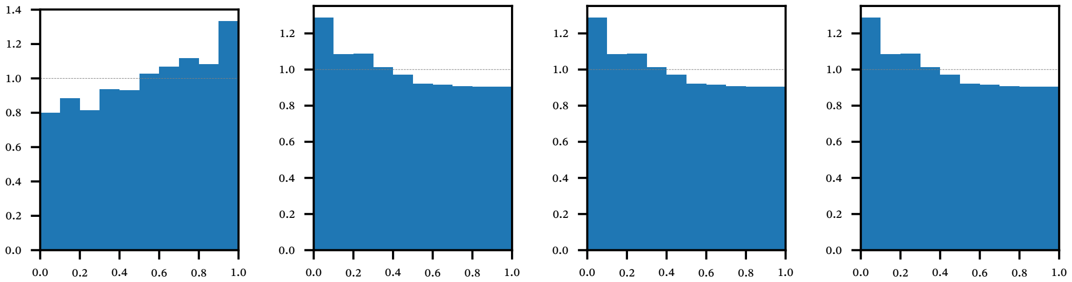


Figure B.24.: Forecast error dependence plot for the DDNN-JSU models in EPF. The lower left plots display scatter plots of the errors, whereas the upper right plots visualize the contours of a two-dimensional kernel density estimation. On the diagonal, histograms of the errors for the individual models are shown. The resulting Kendall's τ is displayed in the lower right corner of the contour plots. The errors are strongly pairwise dependent, and Kendall's τ is smaller than for the DDNN-Normal models. The DDNN-JSU models have more extreme errors than the DDNN-Normal models.

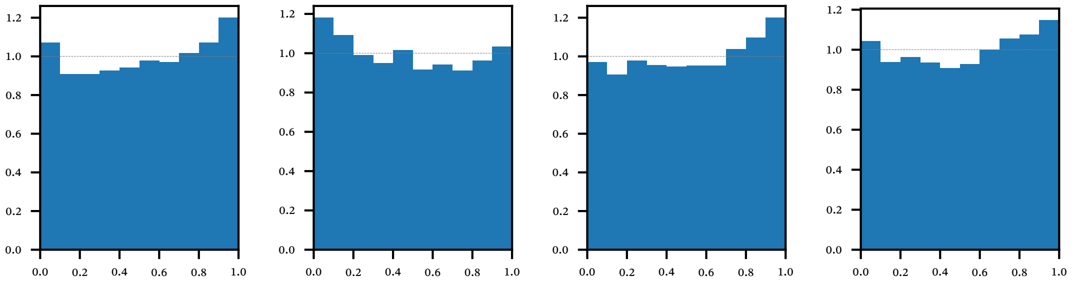
	MAE	RMSE	CRPS	KS
DDNN-JSU-1	4.2641	8.5946	3.0138	0.0650
DDNN-JSU-2	4.2085	6.9521	2.9999	0.0483
DDNN-JSU-3	4.2617	10.7822	3.0108	0.0820
DDNN-JSU-4	4.0343	7.2352	2.8870	0.0323
DDNN-Normal-1	4.0403	6.6909	2.9538	0.0473
DDNN-Normal-2	4.1099	6.6992	3.0098	0.0799
DDNN-Normal-3	4.0714	6.4846	2.9649	0.0733
DDNN-Normal-4	4.0765	6.6180	2.9626	0.0376
KDE-JSU-1	4.2652	8.6027	3.2679	0.0289
KDE-JSU-2	4.2084	6.9546	3.1888	0.0458
KDE-JSU-3	4.2637	10.8575	3.2443	0.0274
KDE-JSU-4	4.0339	7.2532	3.0762	0.0369
KDE-Normal-1	4.0403	6.6909	3.0619	0.0323
KDE-Normal-2	4.1099	6.6992	3.1116	0.0278
KDE-Normal-3	4.0714	6.4846	3.0316	0.0356
KDE-Normal-4	4.0765	6.6180	3.0868	0.0317
JSU-TLP	3.8716	7.6316	2.7436	0.0394
Normal-TLP	3.8626	6.3833	2.8145	0.0399
CCD-All	3.8885	6.2289	2.8838	0.0174
CCD-Hourly	3.9505	6.3056	2.9523	0.0224
CCD-1Y-Hourly	3.9472	6.3002	2.9504	0.0240
CCD-1Y	3.8860	6.2164	2.8816	0.0187
CCD-13-All	3.8751	6.1365	2.8795	0.0240
CCD-13-1Y	3.8764	6.1399	2.8808	0.0240
CCD-14-All	3.9342	6.3185	2.9447	0.0243
CCD-14-1Y	3.9320	6.3094	2.9423	0.0239

Table B.1.: Results of point (MAE, RMSE), probabilistic (CRPS), and calibration (KS) evaluation for the different models on electricity price forecasts. For all columns, lower values indicate better forecasts. The combined forecasts tend to have a lower MAE and RMSE than the individual models and are better calibrated, but do not necessarily have a lower CRPS than the individual models.



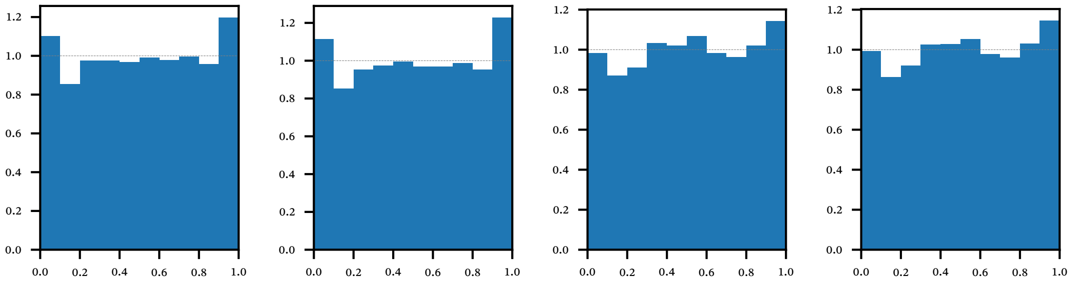
(a) DDNN-JSU-1 **(b)** DDNN-JSU-2 **(c)** DDNN-JSU-3 **(d)** DDNN-JSU-4

Figure B.25.: PIT histograms for the forecasts by the DDNN-JSU models with different hyperparameter setups in EPF. The histograms are skewed, indicating a bad forecast calibration.



(a) KDE-Normal-1 **(b)** KDE-Normal-2 **(c)** KDE-Normal-3 **(d)** KDE-Normal-4

Figure B.26.: PIT histograms for the forecasts by the KDE models with different hyperparameter setups in EPF. The KDE is slightly underdispersed but better calibrated than the corresponding DDNN models. Note the different scales on the y-axis.



(a) CCD-All-Hourly **(b)** CCD-1Y-Hourly **(c)** CCD-All-14 **(d)** CCD-1Y-14

Figure B.27.: PIT histograms for the CCD models using all data or at most one year of data to fit and using all DDNN-Normal models or only models DDNN-Normal-1 and DDNN-Normal-4.

C. Appendix to Chapter 5

The section provides additional material on Chapter 5 and is based on Publ. III. Appendix C.1 contains material on the derivation and the measures of Section 5.2 by presenting the sample data generation schemes and illustrative simulation studies on bootstrapping confidence intervals and bandwidth selection in multivariate KDE for the conditional ATC plot. Appendix C.2 presents additional results for the assessment of ATC for COVID-19 nowcasting in Section 5.3.1.

C.1. Additional material on Section 5.2

C.1.1. Data generation for Section 5.2

The first dataset is generated by sequentially generating $\mathbf{x}^\Delta$ and $\mathbf{y}^\Delta$. First, the x_t^Δ are sampled as a sum of a standard normal random number and a uniform random number on $(-10, 10)$:

$$x_t^\Delta \sim N(0,1) + U(-10,10) \quad t = 1, \dots, T.$$

Subsequently, the $\mathbf{y}^\Delta$ are simulated for a constant ATC ratio k by

$$y_t^\Delta = x_t^\Delta \cdot n_t \cdot b_t,$$

where n_t is a truncated normal distribution with mean one and standard deviation 0.5, truncated at 0, and b_t is a symmetric Bernoulli random variable with parameter k. For a time-varying ATC ratio, the parameter k is modified to have a wave-shape function over time, that is,

$$k_t = 0.75 + \sin(t/365.25 \cdot 2\pi)/4.$$

For the asymmetric ATC ratio, k is a function of x_t^Δ,

$$k(x) = 0.5 + \min\left\{\max\left\{\frac{x+5}{10}, 0\right\}, 1\right\}/2.$$

In the second approach, y_t^Δ and x_t^Δ are modeled to be multivariate normal with mean 0 and covariance matrix

$$\Sigma = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}.$$

Thus, the conditional probability of correct direction prediction can be calculated by a conditional normal distribution to

$$P(Y^\Delta X^\Delta > 0 | X^\Delta = x) = \Phi\left(\frac{3}{2\sqrt{7}} x\right),$$

where Φ is a standard normal CDF.

The four-quadrant plots for the sample realizations of the data generation schemes are shown in Figure C.1.

C.1.2. Simulation study on bootstrapping confidence intervals

We examine three methods for bootstrapping for computing confidence intervals for the ATC ratio: the intuitive percentile and the more sophisticated basic and BCa method. In the *percentile* approach, the confidence interval for the level α is built directly from the empirical distribution of the bootstrap estimators. The *basic* approach computes the confidence interval based on the non-bootstrap estimate using the bootstrapped quantile deviations (Davison and Hinkley, 1997). The BCa method modifies the quantiles of the empirical bootstrap distribution by a bias and an acceleration parameter (Efron, 1987). Typically, the percentile approach needs larger datasets and provides an easy and fast estimate, while the BCa is computationally expensive but requires smaller datasets for reasonable confidence intervals. The basic approach balances these two objectives. We compare the approaches in a small synthetic data study on their small-dataset behavior and computation time.

We vary the number of time steps T to take typical time-series values, such as 30 for daily data in a month, 52 for weekly data, 168, 365, 720, and 1024. The considered datasets are the first dataset with asymmetric dependence and the second dataset outlined in Appendix C.1.1. In the calculations, the `scipy` package's implementation of bootstrap confidence intervals is used (Virtanen et al., 2020). The prescribed confidence level is 90 %, and the number of bootstrap samples is $10,000$. The share of confidence intervals covering the true values per method and T are shown in Table C.1. The true values of the accuracy are computed based on a dataset of size 10^8, yielding 0.7501 and 0.7700 for the two datasets. The computation times per method and dataset are shown in Figure C.2. For the small sample sizes up to $T = 168$, only the BCa method keeps the confidence interval size and yields slightly wider confidence intervals. The method's results are similar for the larger sample

	percentile	basic	BCa
30	0.84 (0.249)	0.86 (0.250)	0.91 (0.257)
52	0.89 (0.194)	0.89 (0.193)	0.89 (0.198)
168	0.91 (0.109)	0.90 (0.109)	0.90 (0.110)
365	0.90 (0.074)	0.90 (0.074)	0.90 (0.074)
720	0.90 (0.053)	0.90 (0.053)	0.90 (0.053)
1024	0.90 (0.044)	0.90 (0.044)	0.89 (0.044)

(a) First dataset with asymmetry.

	percentile	basic	BCa
30	0.87 (0.243)	0.88 (0.242)	0.92 (0.249)
52	0.87 (0.188)	0.89 (0.188)	0.90 (0.192)
168	0.89 (0.106)	0.90 (0.106)	0.90 (0.107)
365	0.90 (0.072)	0.90 (0.072)	0.90 (0.072)
720	0.90 (0.052)	0.90 (0.052)	0.90 (0.052)
1024	0.89 (0.043)	0.90 (0.043)	0.90 (0.043)

(b) Second dataset.

Table C.1.: Proportion of bootstrap confidence intervals covering the true value of ATC ratio per method and sample size T. The average width of the confidence interval is listed in brackets.

sizes. The computation time for the BCa method is slightly larger than for the other methods, but all methods have a moderate computation time. BCa is the only method that maintains the confidence level for small datasets while increasing the computation time only moderately for larger datasets. Therefore, we use the BCa method for confidence intervals in the applications in Section 5.3.

C.1.3. Visualization of different bandwidth selectors in multivariate KDE

We examine the resulting conditional ATC plots for the three well-known KDE bandwidth selectors, rule-of-thumb, cross-validation maximum likelihood, and cross-validation least squares using the `statsmodels` Python package (Seabold and Perktold, 2010) in Figure C.3. While the rule-of-thumb is based only on the covariance matrix, the other two numerically optimize the bandwidth with a hold-one-out least squares or likelihood objective function. The dashed line shows the theoretical $P(Y^{\Delta}X^{\Delta} > 0|X^{\Delta} = \chi)$. The second method, cross-validation least squares, requires long computation times while yielding small or no bandwidth results, even for two relatively small datasets. The rule-of-thumb and cross-validation maximum likelihood methods yield reasonable results at moderate computation times.

C.1.4. Probabilistic ATC evaluation

Section 5.2.6 outlines the assessment of probabilistic ATC for nowcasts and forecasts and specifies the computation for predictions in terms of a CDF and known true values. Here, we outline the computation for quantile forecasts and yet unknown, probabilistic true values.

If forecasts or nowcasts are given as quantiles, p_t can be determined by interpolations among the quantiles. Let q_p denote the quantiles for target time $t + l$ for even-spaced probabilities $p \in \{1/\hat{p}, \dots, (p-1)/\hat{p}\}$ $(\hat{p} \in \mathbb{N} \setminus \{1, 2\})$ and y_t the true value at time t. The quantiles q_p generally differ for each time step, but we omit an index here for ease of notation. The probability p_t^c of a *negative* change is between $p^\star$ and $p^\star + 1/\hat{p}$ for

$$p^\star = \max\{p \in \{1/\hat{p}, \dots, (\hat{p}-1)/\hat{p}\} : q_p \leq y_t\}, \quad \text{if } q_{1/\hat{p}} \leq y_t \leq q_{1-1/\hat{p}}.$$

Quantiles do not determine the location within the interval $[p^\star, p^\star + 1/\hat{p}]$. Under the assumption of a uniform distribution within the quantile interval, the probability of a negative change is

$$p_t^c = \frac{y_t - q_{p^\star}}{\hat{p}(q_{p^\star+1} - q_{p^\star})} + p^\star.$$

The approach does not yet assign probabilities for y_t smaller than the smallest quantile $q_{1/p}$ or greater than the largest quantile. As a simple extension, we assume that the probability mass is uniformly distributed on an interval of the same length as the nearest interval specified by the quantiles. This yields

$$p_t^c = \begin{cases} \max\{\frac{1}{\hat{p}} - \frac{q_{p^\star} - y_t}{\hat{p}(q_{p2/\hat{p}} - q_{1/\hat{p}})}, 0\} & \text{, if } y_t < q_{1/p}, \\ \min\{\frac{1}{\hat{p}} - \frac{y_t - q_{(\hat{p}-1)/\hat{p}}}{\hat{p}(q_{(\hat{p}-1)/\hat{p}} - q_{(\hat{p}-2)/\hat{p}})}, 1\} & \text{, if } y_t > q_{1-1/p}, \\ \frac{y_t - q_{p^\star}}{\hat{p}(q_{p^\star+1} - q_{p^\star})} + p^\star & \text{, otherwise.} \end{cases}$$

The probability of positive change is $p_t = 1 - p_t^c$.

If the true value is given as a distribution because it is still unknown, the probabilities p_t can be computed by integration. Let for two nowcasts the distributions be given by PDFs $f_{t+l|t+l}$ and $f_{t|t+l}$ with CDFs $F_{t+l|t+l}$ and $F_{t|t+l}$. Then, the probability of a negative change can be computed by

$$\begin{aligned} p_t^c &= \int_{\substack{x_1, x_2 \in \mathbb{R}: \\ x_2 < x_1}} f_{t|t+l}(x_1) f_{t+l|t+l}(x_2) \,\mathrm{d}\,(x_1, x_2) \\ &= \int_{x_1 \in \mathbb{R}} \int_{-\infty}^{x_1} f_{t|t+l}(x_1) f_{t+l|t+l}(x_2) \,\mathrm{d}\, x_2 \,\mathrm{d}\, x_1 \\ &= \int_{x_1 \in \mathbb{R}} f_{t|t+l}(x_1) F_{t+l|t+l}(x_1) \,\mathrm{d}\, x_2 \,\mathrm{d}\, x_1. \end{aligned} \tag{C.1}$$

Thereby, the distributions are assumed to be independent. If the nowcasts have the form of a multivariate distribution, including the dependence of the two PDFs, $f_{t+l|t+l}(x_2)$ has to be replaced by the PDF conditional on x_1. As a Monte Carlo approximation of Equation (C.1), the probability can also be calculated by sampling from $f_{t+l|t+l}$ and $f_{t|t+l}$ and calculating the fraction of negative changes. For forecasts, the indexes have to be shifted. If no PDFs are

Abbreviation	Nowcasting hub key
EPI	Epiforecasts-independent
ILM	ILM-prop
KIT	KIT-simple_nowcast
LMU	LMU_StaBLab-GAM_nowcast
RIVM	RIVM-KEW
RKI	RKI-weekly_report
SU	SU-hier_bayes
SZ	SZ-hosp_nowcast
ENS-MEAN	NowcastHub-MeanEnsemble
ENS-MED	NowcastHub-MedianEnsemble

Table C.2.: Matching the abbreviation to the key in the COVID-19 nowcasting hub. Information on the models and references are listed in Wolffram et al. (2023, Table 1).

	(1), l=1	$\sigma_{x^{\Delta,1}}$	$q_{0.1}(x^{\Delta,1})$	(1), l=7	$\sigma_{x^{\Delta,7}}$	$q_{0.1}(x^{\Delta,7})$	(1), l=14	$\sigma_{x^{\Delta,14}}$	$q_{0.1}(x^{\Delta,14})$
EPI	86	520	44	83	1411	78	80	1976	144
ILM	86	281	26	81	1457	102	82	2356	140
KIT	84	354	50	89	1306	171	83	1964	265
LMU	65	285	26	84	1180	124	78	1946	167
ENS-MEAN	85	267	23	86	1213	98	83	1955	235
ENS-MED	88	259	23	88	1206	101	81	1955	186
RIVM	77	241	32	81	1264	123	77	2034	190
RKI	99	362	34	106	1194	145	99	1832	325
SU	91	376	47	85	1390	180	80	2126	263
SZ	92	201	26	89	1154	184	87	1889	241
True	75	262	27	66	1237	126	73	2193	284

Table C.3.: Marginal analysis of the COVID-19 nowcast and true changes for the horizons one, seven, and 14 days. The column (1), $l = l$ shows the number of values greater than zero for horizon l, $\sigma_{x^{\Delta,l}}$ the standard deviation, and $q_{0.1}(x^{\Delta,l})$ the 10% quantile of the changes' absolute values.

available, they can be estimated from the CDF or quantiles, or the CDF or quantiles can be used to generate samples for the Monte Carlo approximation. This approach is applied in Section 5.3.1, as the true values are published with a delay of more than 80 days, and the nowcasts are given as quantiles.

C.2. Additional results for Section 5.3.1

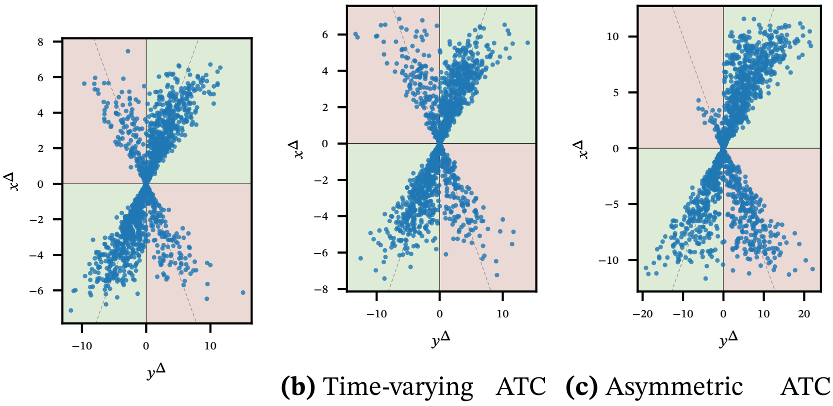
	μ^{1}	$\mu^{+,1}$	$\mu^{-,1}$	$\mu^{1}_{q_{0.1}}$	$\mu^{+,1}_{q_{0.1}}$	$\mu^{-,1}_{q_{0.1}}$
EPI	0.68 (0.62, 0.74)	0.64 (0.55, 0.72)	0.73 (0.63, 0.81)	0.69 (0.63, 0.75)	0.64 (0.56, 0.73)	0.75 (0.65, 0.82)
ILM	0.73 (0.67, 0.79)	0.67 (0.58, 0.76)	0.82 (0.73, 0.89)	0.74 (0.68, 0.79)	0.68 (0.60, 0.76)	0.82 (0.72, 0.89)
KIT	0.62 (0.55, 0.68)	0.58 (0.49, 0.67)	0.65 (0.56, 0.75)	0.62 (0.56, 0.69)	0.59 (0.51, 0.67)	0.66 (0.57, 0.74)
LMU	0.66 (0.60, 0.72)	0.66 (0.57, 0.75)	0.66 (0.57, 0.73)	0.66 (0.59, 0.72)	0.66 (0.55, 0.75)	0.66 (0.57, 0.73)
ENS-MEAN	0.81 (0.75, 0.85)	0.76 (0.68, 0.84)	0.88 (0.81, 0.93)	0.81 (0.75, 0.86)	0.76 (0.68, 0.83)	0.88 (0.81, 0.94)
ENS-MED	0.75 (0.68, 0.80)	0.69 (0.60, 0.77)	0.81 (0.73, 0.89)	0.75 (0.69, 0.80)	0.69 (0.61, 0.77)	0.83 (0.74, 0.90)
RIVM	0.77 (0.72, 0.82)	0.75 (0.66, 0.83)	0.79 (0.71, 0.85)	0.78 (0.72, 0.83)	0.75 (0.66, 0.83)	0.81 (0.73, 0.87)
RKI	0.74 (0.68, 0.80)	0.67 (0.59, 0.75)	0.88 (0.79, 0.93)	0.74 (0.67, 0.79)	0.66 (0.58, 0.73)	0.87 (0.78, 0.93)
SU	0.71 (0.65, 0.77)	0.66 (0.57, 0.74)	0.78 (0.69, 0.85)	0.72 (0.66, 0.78)	0.67 (0.58, 0.75)	0.79 (0.70, 0.87)
SZ	0.74 (0.69, 0.80)	0.68 (0.60, 0.76)	0.82 (0.73, 0.88)	0.74 (0.69, 0.80)	0.68 (0.60, 0.76)	0.82 (0.73, 0.88)

(a) One day.

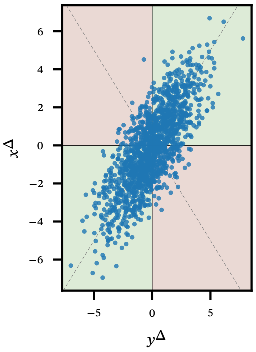
	μ^{14}	$\mu^{+,14}$	$\mu^{-,14}$	$\mu^{14}_{q_{0.1}}$	$\mu^{+,14}_{q_{0.1}}$	$\mu^{-,14}_{q_{0.1}}$
EPI	0.83 (0.77, 0.87)	0.79 (0.70, 0.85)	0.87 (0.80, 0.92)	0.85 (0.80, 0.90)	0.81 (0.73, 0.87)	0.90 (0.83, 0.95)
ILM	0.86 (0.81, 0.90)	0.78 (0.70, 0.85)	0.96 (0.90, 0.99)	0.87 (0.82, 0.91)	0.80 (0.71, 0.86)	0.96 (0.90, 0.99)
KIT	0.81 (0.75, 0.86)	0.76 (0.67, 0.83)	0.87 (0.79, 0.92)	0.82 (0.76, 0.86)	0.76 (0.68, 0.84)	0.88 (0.81, 0.93)
LMU	0.88 (0.83, 0.92)	0.85 (0.77, 0.91)	0.91 (0.85, 0.95)	0.89 (0.85, 0.93)	0.87 (0.79, 0.92)	0.91 (0.85, 0.95)
ENS-MEAN	0.83 (0.77, 0.87)	0.77 (0.69, 0.84)	0.89 (0.83, 0.95)	0.84 (0.78, 0.88)	0.78 (0.70, 0.85)	0.91 (0.84, 0.95)
ENS-MED	0.84 (0.79, 0.89)	0.79 (0.70, 0.85)	0.90 (0.83, 0.95)	0.85 (0.80, 0.90)	0.80 (0.72, 0.86)	0.91 (0.84, 0.96)
RIVM	0.85 (0.80, 0.89)	0.82 (0.74, 0.88)	0.88 (0.80, 0.93)	0.85 (0.80, 0.90)	0.83 (0.75, 0.89)	0.88 (0.80, 0.93)
RKI	0.81 (0.75, 0.86)	0.71 (0.63, 0.77)	0.98 (0.93, 1.00)	0.81 (0.75, 0.86)	0.71 (0.63, 0.78)	1.00 (nan, nan)
SU	0.88 (0.83, 0.92)	0.84 (0.76, 0.90)	0.92 (0.86, 0.96)	0.89 (0.84, 0.93)	0.85 (0.77, 0.91)	0.94 (0.87, 0.97)
SZ	0.82 (0.77, 0.87)	0.76 (0.68, 0.83)	0.90 (0.83, 0.94)	0.83 (0.78, 0.88)	0.78 (0.69, 0.85)	0.90 (0.83, 0.94)

(b) 14 days.

Table C.4.: ATC ratio μ, positive ATC ratio μ^{+}, and negative ATC ratio μ^{-} for the models without and with exclusion areas for the horizon one and 14 days in COVID-19 nowcasting. The exclusion areas are rectangles centered on the zero points with a width and height to exclude the 10%-quantile of the absolute values of nowcast or true values.

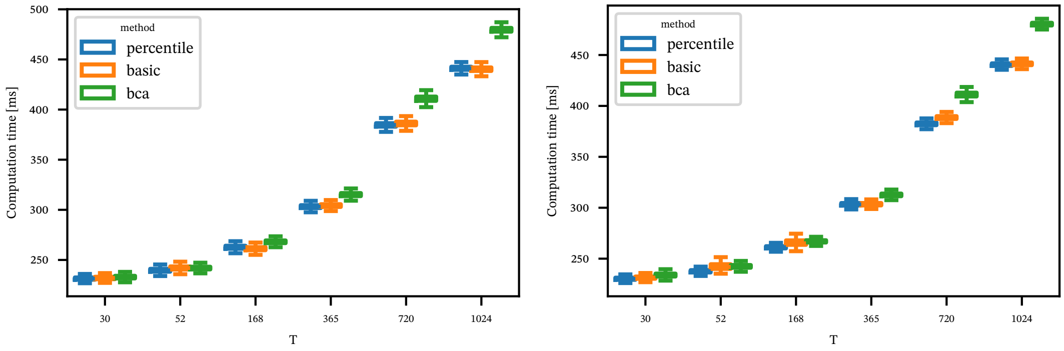


(a) Constant ATC ratio. **(b)** Time-varying ATC ratio **(c)** Asymmetric ATC ratio



(d) Second approach.

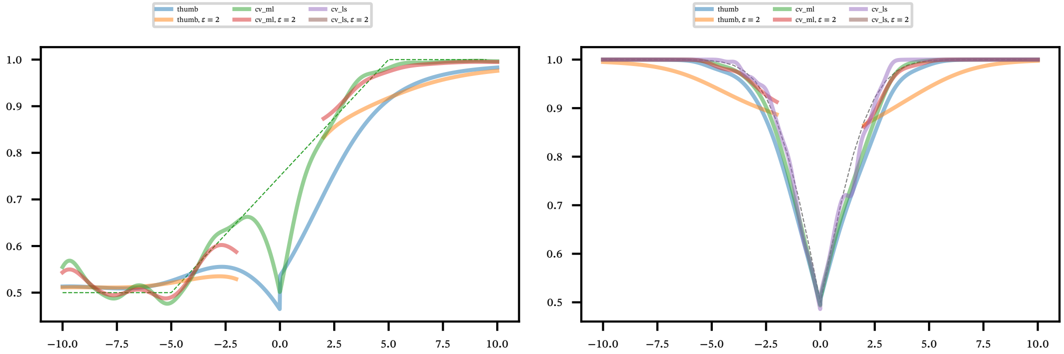
Figure C.1.: Four-quadrant plots for sample realizations of the data generation schemes of Section C.1.1. Although the first and second plots differ over time, their difference is not discernible in the plots. The third data set's asymmetry is visible in the plot, but the decrease in the ATC near 0 is not visible.



(a) First dataset with asymmetry.

(b) Second dataset.

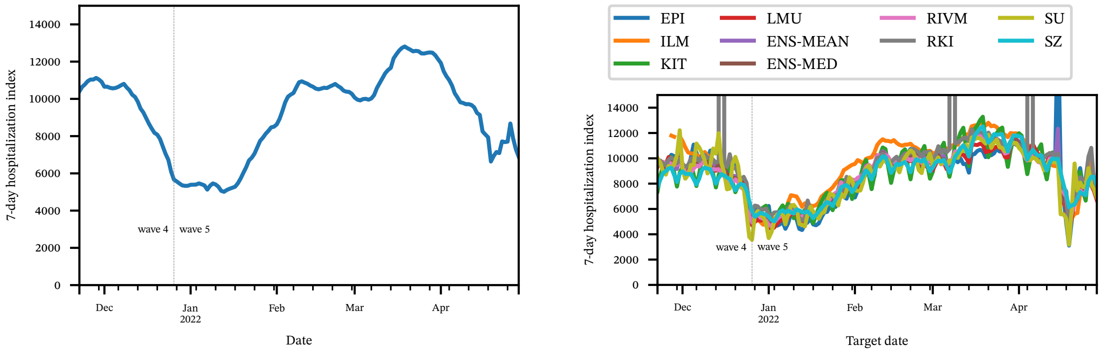
Figure C.2.: Boxplot of the computation time for different bootstrapping methods and data set sizes T. The computation time refers to bootstrapping one confidence interval based upon $10,000$ values. Each boxplot reflects $10,000$ samples. The BCa method takes slightly longer than the other two, but the difference is negligible.



(a) First dataset with asymmetric dependence.

(b) Second dataset.

Figure C.3.: Conditional ATC plot for different bandwidth selection processes. Cross-validation least squares takes a considerably larger computation time. It converges neither for the first nor the second data set with an exclusion area and yields a bandwidth too small for the second data set. The rule of thumb is the fastest method but tends to oversmooth. The cross-validation maximum likelihood method yields a more reasonable bandwidth with moderate computation time. ε specifies an exclusion area $E = \{(x, y) \in \mathbb{R}^2 : (-\varepsilon \leq x \leq \varepsilon)\}$ in $\mathbf{x}^{\Delta}$-direction.



(a) Realisations.

(b) Same-day nowcasts.

Figure C.4.: True and nowcast data of the seven-day-hospitalization in Germany from November 22, 2021, to April 29, 2022 (C19-Hub). The outliers in the RKI model of values above 10^8 are removed before the following analysis.

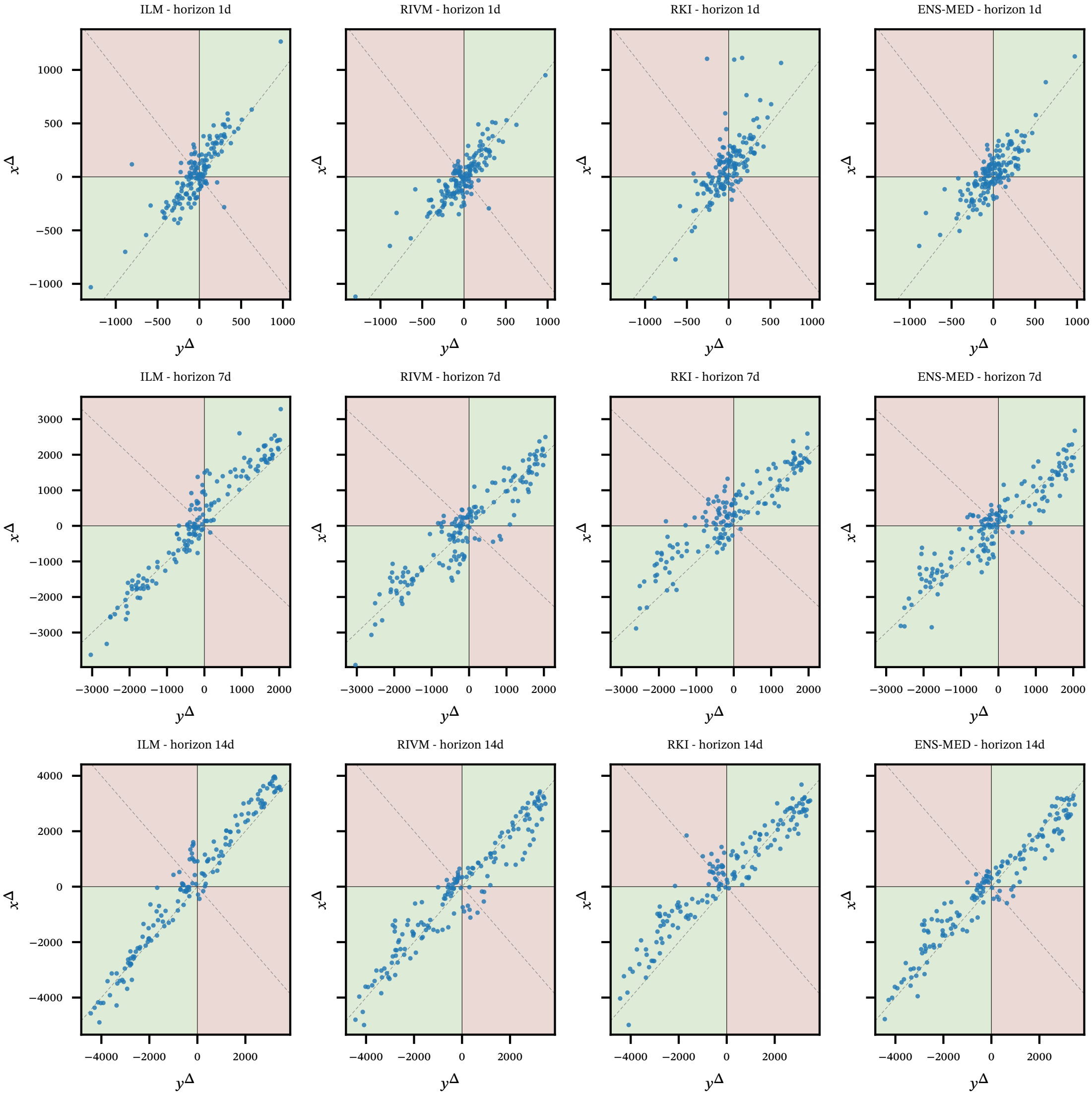
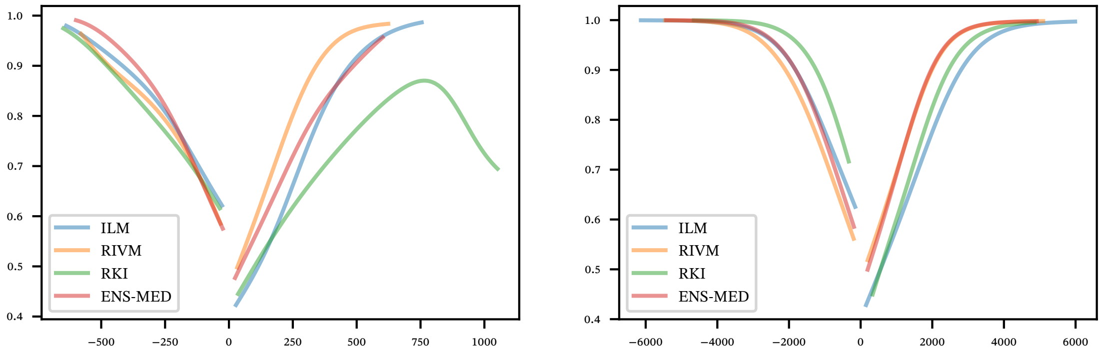
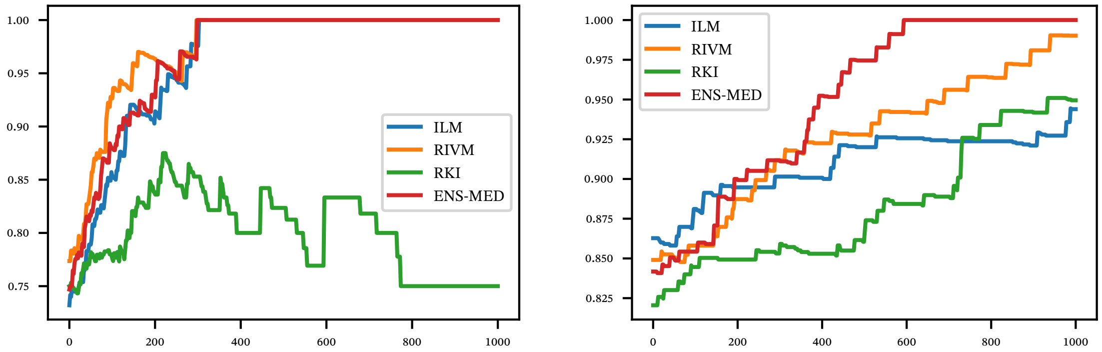


Figure C.5.: Four-quadrant plots for the COVID-19 nowcast models ILM, RIVM, RKI, and ENS-MEAN and the horizons of one, seven, and 14 days. The spread in both directions increases with the horizon.



(a) Conditional ATC plot for horizon one.

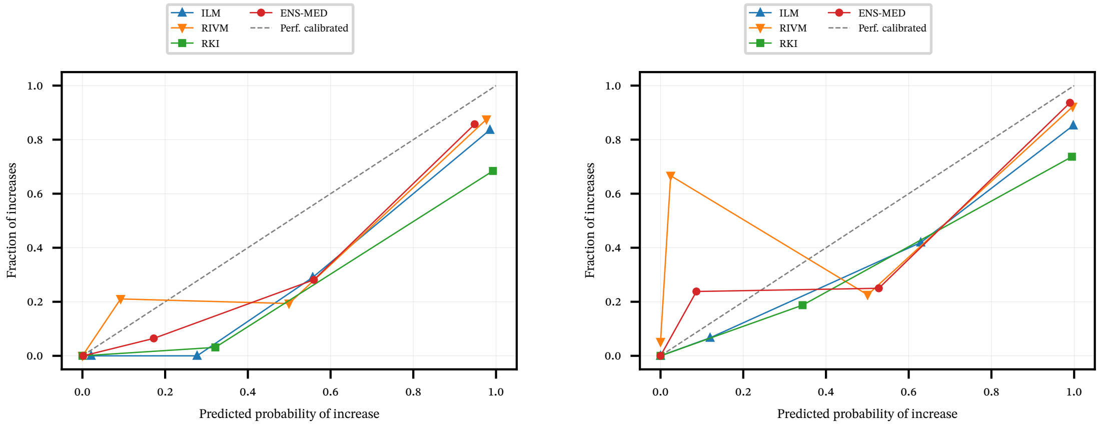
(b) Conditional ATC plot for horizon 14.



(c) ATC ratio over exclusion area size in $\mathbf{x}^{\Delta}$ for horizon one.

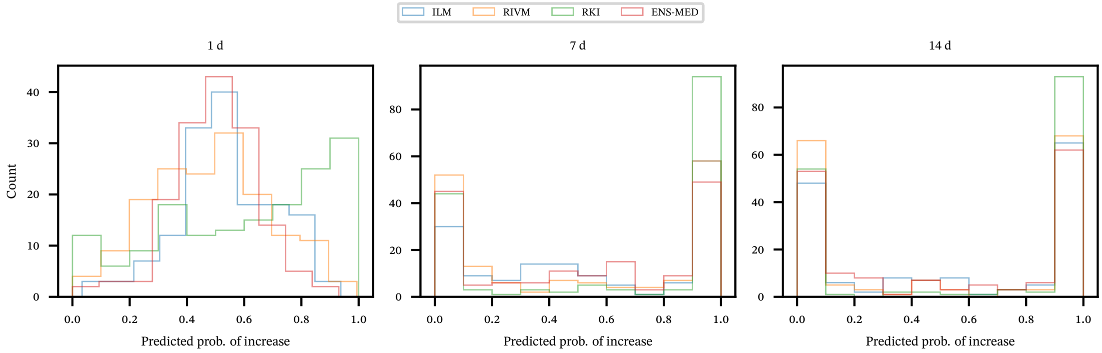
(d) ATC ratio over exclusion area size in $\mathbf{x}^{\Delta}$ for horizon 14.

Figure C.6.: Conditional ATC plot and ATC ratio over exclusion area for the COVID-19 nowcast models ILM, RKI, RIVM, and ENS-MED for the horizon seven days. Note the different axis scalings of the ATC ratios for the two horizons. While the ENS-MED model has a rather low ATC ratio for small exclusion areas, it is increasing fast for larger exclusion areas. The RKI model is among the worst, except for the small negative predicted changes.



(a) Reliability diagram for horizon seven days.

(b) Reliability diagram for horizon 14 days.



(c) Count histogram of the predicted probabilities for the horizon one, seven, and 14 days.

Figure C.7.: The reliability diagram for the COVID-19 nowcasting models ILM, RIVM, RKI, and ENS-MED for the horizon seven and 14 days. Additionally, the count of predicted probabilities for the horizons is shown. The reliability diagram bins are chosen according to the empirical quantiles of the predicted probabilities. As the models issue small or large probabilities of increase for the higher horizons, little information on the accuracy of moderate probability predictions is available.

List of Author’s Publications and Presentations

Publications

Publl. I Grothe, Oliver and Jonas Rieger. “Decomposition and Graphical Correspondence Analysis of Checkerboard Copulas”. Accepted in: *Dependence Modeling*.

Publ. II Grothe, Oliver and Jonas Rieger. “Combining point forecasts to calibrated probabilistic forecasts using copulas”. Submitted to: *International Journal of Forecasting*.

Publ. III Rieger, Jonas, Bolin Liu, Bernd Saugel, and Oliver Grothe. “The assessment of the ability of measurements, nowcasts, and forecasts to track changes”. *Working Paper*.

Publ. IV Grothe, Oliver and Jonas Rieger (June 28, 2022b). “SVD-reduction of high-dimensional German spatio-temporal wind speed data and clusters of similarity”. In: *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*. E-Energy ’22: The Thirteenth ACM International Conference on Future Energy Systems. Virtual Event: ACM, pp. 515–520. doi:10.1145/3538637.3539764.

Conference presentations

Conf. I Grothe, Oliver and Jonas Rieger* (March 30, 2022a). “Combining point forecasts to calibrated probabilistic forecasts using copulas”. DAGStat Conference 2022 (Hamburg, Germany).

Conf. II Grothe, Oliver and Jonas Rieger* (June 28, 2022b). “SVD-reduction of high-dimensional German spatio-temporal wind speed data and clusters of similarity”. E-Energy ’22: The Thirteenth ACM International Conference on Future Energy Systems (Virtual Event).

Conf. III Grothe, Oliver and Jonas Rieger* (July 12, 2022c). “Copula-based combination of point prediction systems to calibrated probabilistic forecasts”. 42nd International Symposium on Forecasting (Oxford, UK).

Conf. IV Grothe*, Oliver and Jonas Rieger (December 17, 2022d). “Singular Value Decomposition Based low-rank Representations of Copulas”. CMStatistics 2022 (London, UK).

The starred name refers to the presenter at the conference.

Bibliography

Akaike, H. (Dec. 1974). “A New Look at the Statistical Model Identification”. In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723. ISSN: 0018-9286. DOI: `10.1109/TAC.1974.1100705`.

Anderson, T. W. and D. A. Darling (1954). “A Test of Goodness of Fit”. In: *Journal of the American Statistical Association* 49.268, pp. 765–769. ISSN: 1537274X. DOI: `10.1080/01621459.1954.10501232`.

Aya-Moreno, Carlos, Gery Geenens, and Spiridon Penev (Nov. 2018). “Shape-Preserving Wavelet-Based Multivariate Density Estimation”. In: *Journal of Multivariate Analysis* 168, pp. 30–47. ISSN: 0047259X. DOI: `10.1016/j.jmva.2018.07.002`.

Bakam, Yves Ismaël Ngounou and Denys Pommeret (Jan. 11, 2022). *K-Sample Test for Equality of Copulas*. arXiv: `2112.05623 [math, stat]`. URL: `http://arxiv.org/abs/2112.05623` (visited on 02/24/2023). preprint.

Bańbura, Marta et al. (2013). “Now-Casting and the Real-Time Data Flow”. In: *Handbook of Economic Forecasting*. Vol. 2. Elsevier, pp. 195–237. ISBN: 978-0-444-53683-9. DOI: `10.1016/B978-0-444-53683-9.00004-9`.

Baran, Sándor and Sebastian Lerch (July 2018). “Combining Predictive Distributions for Statistical Post-Processing of Ensemble Forecasts”. In: *International Journal of Forecasting* 34.3, pp. 477–496. ISSN: 01692070. DOI: `10.1016/j.ijforecast.2018.01.005`. arXiv: `1607.08096`.

Bates, J. M. and C. W. J. Granger (Dec. 23, 1969). “The Combination of Forecasts”. In: *OR* 20.4, p. 451. ISSN: 14732858. DOI: `10.2307/3008764`.

Bedford, Tim and Roger M. Cooke (Aug. 1, 2002). “Vines–a New Graphical Model for Dependent Random Variables”. In: *The Annals of Statistics* 30.4. ISSN: 0090-5364. DOI: `10.1214/aos/1031689016`.

Berrocal, Veronica J., Adrian E. Raftery, and Tilmann Gneiting (2007). “Combining Spatial Statistical and Ensemble Information in Probabilistic Weather Forecasts”. In: *Monthly Weather Review* 135.4, pp. 1386–1402. ISSN: 00270644. DOI: `10.1175/MWR3341.1`.

Bhatia, Rajendra, Stephane Gaubert, and Tanvi Jain (Aug. 2019). “Matrix Versions of the Hellinger Distance”. In: *Letters in Mathematical Physics* 109.8, pp. 1777–1804. ISSN: 0377-9017, 1573-0530. DOI: `10.1007/s11005-019-01156-0`.

Bittmann, Felix (2021). *Bootstrapping: An Integrated Approach with Python and Stata*. Berlin Boston: De Gruyter Oldenbourg. 120 pp. ISBN: 978-3-11-069440-6.

Bland, Martin J. and Douglas G. Altman (Feb. 1986). "STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT". In: *The Lancet* 327.8476, pp. 307–310. ISSN: 01406736. DOI: `10.1016/S0140-6736(86)90837-8`.

Bracher, Johannes et al. (Nov. 2, 2023). "Direction Augmentation in the Evaluation of Armed Conflict Predictions". In: *International Interactions* 49.6, pp. 989–1004. ISSN: 0305-0629, 1547-7444. DOI: `10.1080/03050629.2023.2255923`.

Brier, Glenn W. (Jan. 1950). "VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY". In: *Monthly Weather Review* 78.1, pp. 1–3. ISSN: 0027-0644, 1520-0493. DOI: `10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2`.

Browning, K. A. and C. G. Collier (Aug. 1989). "Nowcasting of Precipitation Systems". In: *Reviews of Geophysics* 27.3, pp. 345–370. ISSN: 8755-1209, 1944-9208. DOI: `10.1029/RG027i003p00345`.

Carley, H. and M. D. Taylor (2002). "A New Proof of Sklar's Theorem". In: *Distributions With Given Marginals and Statistical Modelling*. Dordrecht: Springer Netherlands, pp. 29–34. ISBN: 978-94-017-0061-0. DOI: `10.1007/978-94-017-0061-0_4`.

Castle, Jennifer, David Hendry, and Oleg Kitov (2017). "Forecasting and Nowcasting Macroeconomic Variables: A Methodological Overview". In: *Handbook on Rapid Estimates: 2017 Edition.* Ed. by European Commission. Statistical Office of the European Union. Luxembourg: Publications Office. URL: `https://data.europa.eu/doi/10.2785/488740` (visited on 04/09/2024).

Chen, Song Xi and Tzee Ming Huang (2007). "Nonparametric Estimation of Copula Functions for Dependence Modelling". In: *Canadian Journal of Statistics* 35.2, pp. 265–282. ISSN: 03195724. DOI: `10.1002/cjs.5550350205`.

Cherubini, U and E Luciano (2004). *Copula Methods in Finance*. Chichester: Wiley. ISBN: 0-470-86344-7.

Coblenz, Maximilian (2018). "Advances in Dependence Modeling: Multivariate Quantiles, Copula Level Curve Lengths, and Non-Simplified Vine Copulas". Karlsruhe Institute for Technology (KIT). DOI: `https://dx.doi.org/10.5445/IR/1000088365`.

– (June 2021). "MATVines: A Vine Copula Package for MATLAB". In: *SoftwareX* 14, p. 100700. ISSN: 23527110. DOI: `10.1016/j.softx.2021.100700`.

Coblenz, Maximilian et al. (Aug. 1, 2020). "Modelling Fuel Injector Spray Characteristics in Jet Engines by Using Vine Copulas". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 69.4, pp. 863–886. ISSN: 0035-9254, 1467-9876. DOI: `10.1111/rssc.12421`.

Cohen, Jacob (Apr. 1960). “A Coefficient of Agreement for Nominal Scales”. In: *Educational and Psychological Measurement* 20.1, pp. 37–46. ISSN: 0013-1644, 1552-3888. DOI: `10.1177/001316446002000104`.

Cooke, Roger M., Harry Joe, and Kjersti Aas (Dec. 2010). “Vines Arise”. In: *Dependence Modeling*. World Scientific, pp. 37–71. ISBN: 978-981-4299-88-6. DOI: `10.1142/9789814299886_0003`.

Cottin, Claudia and Dietmar Pfeifer (May 1, 2014). “From Bernstein Polynomials to Bernstein Copulas”. In: *Journal of Applied Functional Analysis* 9, pp. 277–288.

Critchley, Lester A., Anna Lee, and Anthony M.-H. Ho (Nov. 2010). “A Critical Review of the Ability of Continuous Cardiac Output Monitors to Measure Trends in Cardiac Output”. In: *Anesthesia & Analgesia* 111.5, pp. 1180–1192. ISSN: 0003-2999. DOI: `10.1213/ANE.0b013e3181f08a5b`.

Critchley, Lester A. H. and Julian A. J. H. Critchley (Feb. 1, 1999). “A Meta-Analysis of Studies Using Bias and Precision Statistics to Compare Cardiac Output Measurement Techniques”. In: *Journal of Clinical Monitoring and Computing* 15.2, pp. 85–91. ISSN: 1573-2614. DOI: `10.1023/A:1009982611386`.

Cuadras, Carles M. (Apr. 2002). “Correspondence Analysis and Diagonal Expansions in Terms of Distribution Functions”. In: *Journal of Statistical Planning and Inference* 103.1-2, pp. 137–150. ISSN: 03783758. DOI: `10.1016/S0378-3758(01)00216-6`.

– (July 2015). “Contributions to the Diagonal Expansion of a Bivariate Copula with Continuous Extensions”. In: *Journal of Multivariate Analysis* 139, pp. 28–44. ISSN: 0047259X. DOI: `10.1016/j.jmva.2015.02.015`.

Cuadras, Carles M. and Joan Augé (Jan. 1981). “A Continuous General Multivariate Distribution and Its Properties”. In: *Communications in Statistics - Theory and Methods* 10.4, pp. 339–353. ISSN: 0361-0926, 1532-415X. DOI: `10.1080/03610928108828042`.

Cuadras, Carles M. and Daniel Cuadras (Aug. 2006). “A Parametric Approach to Correspondence Analysis”. In: *Linear Algebra and its Applications* 417.1, pp. 64–74. ISSN: 00243795. DOI: `10.1016/j.laa.2005.10.029`.

Cuadras, Carles M. and Walter Díaz (Dec. 31, 2012). “Another Generalization of the Bivariate FGM Distribution with Two-Dimensional Extensions”. In: *Acta et Commentationes Universitatis Tartuensis de Mathematica* 16.1, pp. 3–12. ISSN: 2228-4699, 1406-2283. DOI: `10.12697/ACUTM.2012.16.01`.

Cuberos, A., E. Masiello, and V. Maume-Deschamps (June 17, 2020). “Copulas Checker-Type Approximations: Application to Quantiles Estimation of Sums of Dependent Random Variables”. In: *Communications in Statistics - Theory and Methods* 49.12, pp. 3044–3062. ISSN: 0361-0926, 1532-415X. DOI: `10.1080/03610926.2019.1586936`.

Czado, Claudia (2019). *Analyzing Dependent Data with Vine Copulas*. Vol. 222. Cham: Springer International Publishing. 1-242. ISBN: 978-3-030-13784-7. DOI: `10.1007/978-3-030-13785-4`.

Czado, Claudia and Thomas Nagler (Mar. 7, 2022). "Vine Copula Based Modeling". In: *Annual Review of Statistics and Its Application* 9.1, pp. 453–477. ISSN: 2326-8298, 2326-831X. DOI: `10.1146/annurev-statistics-040220-101153`.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge ; New York, NY, USA: Cambridge University Press. 582 pp. ISBN: 978-0-521-57391-7.

Dawid, A. P. (1984). "Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach". In: *Journal of the Royal Statistical Society. Series A (General)* 147.2, p. 278. ISSN: 00359238. DOI: `10.2307/2981683`.

De Oliveira, Ricardo T.A. et al. (Jan. 2017). "Copulas-Based Time Series Combined Forecasters". In: *Information Sciences* 376, pp. 110–124. ISSN: 00200255. DOI: `10.1016/j.ins.2016.10.022`.

Durante, Fabrizio and Carlo Sempi (2015). *Principles of Copula Theory*. 1-314. ISBN: 978-1-4398-8444-7. DOI: `10.1201/b18674`.

Durrleman, Valdo, Ashkan Nikeghbali, and Thierry Roncalli (2000). "Copulas Approximation and New Families". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: `10.2139/ssrn.1032547`.

Dykstra, Richard L. (Dec. 1983). "An Algorithm for Restricted Least Squares Regression". In: *Journal of the American Statistical Association* 78.384, pp. 837–842. ISSN: 0162-1459, 1537-274X. DOI: `10.1080/01621459.1983.10477029`.

Econometrics and Statistics at Karlsruhe Institute of Technology, Chair of and Computational Statistics Group at Heidelberg Institute for Theoretical Studies (Feb. 23, 2024). *Nowcasts Der COVID-19 Hospitalisierungsinzidenz*. URL: `https://covid19nowcasthub.de/` (visited on 02/23/2024).

Efron, Bradley (Mar. 1987). "Better Bootstrap Confidence Intervals". In: *Journal of the American Statistical Association* 82.397, pp. 171–185. ISSN: 0162-1459, 1537-274X. DOI: `10.1080/01621459.1987.10478410`.

Embrechts, Paul, Filip Lindskog, and Alexander Mcneil (2003). "Modelling Dependence with Copulas and Applications to Risk Management". In: *Handbook of Heavy Tailed Distributions in Finance*. Elsevier, pp. 329–384. ISBN: 978-0-444-50896-6. DOI: `10.1016/b978-044450896-6.50010-8`.

Genest, Christian and Johanna Nešlehová (2007). "A Primer on Copulas for Count Data". In: DOI: `10.3929/ETHZ-B-000422520`.

Genest, Christian and James V. Zidek (1986). “Combining Probability Distributions: A Critique and an Annotated Bibliography”. In: *Statistical Science* 1.1, pp. 114–148. ISSN: 08834237. DOI: `10.1214/ss/1177013825`.

Gijbels, Irene and Jan Mielniczuk (1990). “Estimating the Density of a Copula Function”. In: *Communications in Statistics - Theory and Methods* 19.2, pp. 445–464. ISSN: 1532415X. DOI: `10.1080/03610929008830212`.

Glahn, Bob et al. (2009). “MOS Uncertainty Estimates in an Ensemble Framework”. In: *Monthly Weather Review* 137.1, pp. 246–268. ISSN: 00270644. DOI: `10.1175/2008MWR2569.1`.

Gneiting, Tilmann (June 2011). “Making and Evaluating Point Forecasts”. In: *Journal of the American Statistical Association* 106.494, pp. 746–762. ISSN: 0162-1459, 1537-274X. DOI: `10.1198/jasa.2011.r10138`.

Gneiting, Tilmann, Fadoua Balabdaoui, and Adrian E. Raftery (2007). “Probabilistic Forecasts, Calibration and Sharpness”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 69.2, pp. 243–268. ISSN: 13697412. DOI: `10.1111/j.1467-9868.2007.00587.x`.

Gneiting, Tilmann and Matthias Katzfuss (2014). “Probabilistic Forecasting”. In: *Annual Review of Statistics and Its Application* 1.1, pp. 125–151. ISSN: 2326-8298. DOI: `10.1146/annurev-statistics-062713-085831`.

Gneiting, Tilmann, Adrian E. Raftery, et al. (2005). “Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation”. In: *Monthly Weather Review* 133.5, pp. 1098–1118. ISSN: 00270644. DOI: `10.1175/MWR2904.1`.

Gneiting, Tilmann and Roopesh Ranjan (2013). “Combining Predictive Distributions”. In: *Electronic Journal of Statistics* 7.1, pp. 1747–1782. ISSN: 19357524. DOI: `10.1214/13-EJS823`.

Goldberger, Ary L. et al. (June 13, 2000). “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals”. In: *Circulation* 101.23. ISSN: 0009-7322, 1524-4539. DOI: `10.1161/01.CIR.101.23.e215`.

Good, I. J. (1952). “Rational Decisions”. In: *Journal of the Royal Statistical Society* 14.1, pp. 107–114.

Gramacki, Artur (2018). *Nonparametric Kernel Density Estimation and Its Computational Aspects*. Vol. 37. Studies in Big Data. Cham: Springer International Publishing. ISBN: 978-3-319-71688-6. DOI: `10.1007/978-3-319-71688-6`.

Greenacre, Michael J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press. 364 pp. ISBN: 0-12-299050-1.

Grothe, Oliver, Fabian Kächele, and Fabian Krüger (Apr. 2023). “From Point Forecasts to Multivariate Probabilistic Forecasts: The Schaake Shuffle for Day-Ahead Electricity Price

Forecasting". In: *Energy Economics* 120, p. 106602. ISSN: 01409883. DOI: `10.1016/j.eneco.2023.106602`.

Grothe, Oliver and Julius Schnieders (2011). "Spatial Dependence in Wind and Optimal Wind Power Allocation: A Copula-Based Analysis". In: *Energy Policy* 39.9, pp. 4742–4754. ISSN: 03014215. DOI: `10.1016/j.enpol.2011.06.052`.

Günther, Felix et al. (Mar. 2021). "Nowcasting the COVID-19 Pandemic in Bavaria". In: *Biometrical Journal* 63.3, pp. 490–502. ISSN: 0323-3847, 1521-4036. DOI: `10.1002/bimj.202000112`.

Hall, Stephen G. and James Mitchell (2007). "Combining Density Forecasts". In: *International Journal of Forecasting* 23.1, pp. 1–13. ISSN: 01692070. DOI: `10.1016/j.ijforecast.2006.08.001`.

Härdle, Wolfgang, Joel Horowitz, and Jens-Peter Kreiss (Aug. 2003). "Bootstrap Methods for Time Series". In: *International Statistical Review* 71.2, pp. 435–459. ISSN: 0306-7734, 1751-5823. DOI: `10.1111/j.1751-5823.2003.tb00485.x`.

Hastings, W. K. (Apr. 1, 1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, pp. 97–109. ISSN: 1464-3510, 0006-3444. DOI: `10.1093/biomet/57.1.97`.

Haugsbo Hermansen, Anna and Ole Jakob Mengshoel (Dec. 5, 2021). "Forecasting Ambulance Demand Using Machine Learning: A Case Study from Oslo, Norway". In: *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2021 IEEE Symposium Series on Computational Intelligence (SSCI). Orlando, FL, USA: IEEE, pp. 01–10. ISBN: 978-1-72819-048-8. DOI: `10.1109/SSCI50451.2021.9659837`.

Hesterberg, Tim (Nov. 2011). "Bootstrap". In: *WIREs Computational Statistics* 3.6, pp. 497–526. ISSN: 1939-5108, 1939-0068. DOI: `10.1002/wics.182`.

Hill, M. O (1974). "Correspondence Analysis : A Neglected Multivariate Method". In: *Applied Statistics* 23.3, pp. 340–354.

Hiraishi, Mayu, Kensuke Tanioka, and Toshio Shimokawa (Dec. 2021). "Concordance Rate of a Four-Quadrant Plot for Repeated Measurements". In: *BMC Medical Research Methodology* 21.1, p. 270. ISSN: 1471-2288. DOI: `10.1186/s12874-021-01461-0`.

Hobæk Haff, Ingrid, Kjersti Aas, and Arnoldo Frigessi (May 2010). "On the Simplified Pair-Copula Construction — Simply Useful or Too Simplistic?" In: *Journal of Multivariate Analysis* 101.5, pp. 1296–1310. ISSN: 0047259X. DOI: `10.1016/j.jmva.2009.12.001`.

Hofert, Marius, Ivan Kojadinovic, Martin Machler, et al. (2018). *Elements of Copula Modeling with R*. Cham: Springer International Publishing. 267 pp. ISBN: 978-3-319-89634-2.

Hofert, Marius, Ivan Kojadinovic, Martin Maechler, et al. (2023). *Copula: Multivariate Dependence with Copulas*. manual. URL: `https://CRAN.R-project.org/package=copula`.

Hofert, Marius and Martin Mächler (2011). “Nested Archimedean Copulas Meet *R* : The **Nacopula** Package”. In: *Journal of Statistical Software* 39.9. ISSN: 1548-7660. DOI: `10.18637/jss.v039.i09`.

Hong, Tao et al. (July 2016). “Probabilistic Energy Forecasting: Global Energy Forecasting Competition 2014 and Beyond”. In: *International Journal of Forecasting* 32.3, pp. 896–913. ISSN: 01692070. DOI: `10.1016/j.ijforecast.2016.02.001`.

Horn, Roger A. and Charles R. Johnson (2012). *Matrix Analysis*. 2nd ed. Cambridge ; New York: Cambridge University Press. 643 pp. ISBN: 978-0-521-83940-2.

Hyndman, Rob J. and Anne B. Koehler (Oct. 2006). “Another Look at Measures of Forecast Accuracy”. In: *International Journal of Forecasting* 22.4, pp. 679–688. ISSN: 01692070. DOI: `10.1016/j.ijforecast.2006.03.001`.

James, Gareth et al. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. New York, NY: Springer US. ISBN: 978-1-07-161418-1. DOI: `10.1007/978-1-0716-1418-1`.

Janssen, Paul, Jan Swanepoel, and Noël Veraverbeke (May 2012). “Large Sample Behavior of the Bernstein Copula Estimator”. In: *Journal of Statistical Planning and Inference* 142.5, pp. 1189–1197. ISSN: 03783758. DOI: `10.1016/j.jspi.2011.11.020`.

Joe, Harry (1996). “Families of m-Variate Distributions with given Margins and $m(m-1)/2$ Bivariate Dependence Parameters”. In: *Lecture Notes-Monograph Series* 28, pp. 120–141. ISSN: 0749-2170. DOI: `10.1214/lnms/1215452614`. JSTOR: 4355888.

– (2014). *Dependence Modeling with Copulas*. Boca Raton: CRC Press, Taylor & Francis Group. 1-457. ISBN: 978-1-4665-8323-8. DOI: `10.1201/b17116`.

Joe, Harry and Taizhong Hu (May 1996). “Multivariate Distributions from Mixtures of Max-Infinitely Divisible Distributions”. In: *Journal of Multivariate Analysis* 57.2, pp. 240–265. ISSN: 0047259X. DOI: `10.1006/jmva.1996.0032`.

Johnson, Alistair E.W. et al. (May 24, 2016). “MIMIC-III, a Freely Accessible Critical Care Database”. In: *Scientific Data* 3.1, p. 160035. ISSN: 2052-4463. DOI: `10.1038/sdata.2016.35`.

Jolliffe, Ian T. and David B. Stephenson, eds. (2012). *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*. 2. ed. Oxford: Wiley-Blackwell. 274 pp. ISBN: 978-0-470-66071-3.

Jones, Spencer S. et al. (Feb. 2008). “Forecasting Daily Patient Volumes in the Emergency Department”. In: *Academic Emergency Medicine* 15.2, pp. 159–170. ISSN: 1069-6563, 1553-2712. DOI: `10.1111/j.1553-2712.2007.00032.x`.

Jouini, Mohamed N. and Robert T. Clemen (June 1996). “Copula Models for Aggregating Expert Opinions”. In: *Operations Research* 44.3, pp. 444–457. ISSN: 0030364X. DOI: `10.1287/opre.44.3.444`.

Kazmierczak, J-B (1978). “Migrations Interurbaines Dans La Banlieue Sud de Paris”. In: *Cahiers de l’analyse des données* 3.2, pp. 203–218.

Khaleghi, Bahador et al. (2013). “Multisensor Data Fusion: A Review of the State-of-the-Art”. In: *Information Fusion* 14.1, pp. 28–44. ISSN: 15662535. DOI: `10.1016/j.inffus.2011.08.001`.

Klaassen, Chris A. J. and Jon A. Wellner (Mar. 1997). “Efficient Estimation in the Bivariate Normal Copula Model: Normal Margins Are Least Favourable”. In: *Bernoulli* 3.1, p. 55. ISSN: 13507265. DOI: `10.2307/3318652`. JSTOR: `3318652`.

Kleiber, William, Adrian E. Raftery, and Tilmann Gneiting (2011). “Geostatistical Model Averaging for Locally Calibrated Probabilistic Quantitative Precipitation Forecasting”. In: *Journal of the American Statistical Association* 106.496, pp. 1291–1303. ISSN: 01621459. DOI: `10.1198/jasa.2011.ap10433`.

Kolda, Tamara G. and Brett W. Bader (Aug. 6, 2009). “Tensor Decompositions and Applications”. In: *SIAM Review* 51.3, pp. 455–500. ISSN: 0036-1445, 1095-7200. DOI: `10.1137/07070111X`.

Kolesárová, A. et al. (Oct. 2006). “Discrete Copulas”. In: *IEEE Transactions on Fuzzy Systems* 14.5, pp. 698–705. ISSN: 10636706. DOI: `10.1109/TFUZZ.2006.880003`.

Kreiss, Jens-Peter and Soumendra Nath Lahiri (2012). “Bootstrap Methods for Time Series”. In: *Handbook of Statistics*. Vol. 30. Elsevier, pp. 3–26. ISBN: 978-0-444-53858-1. DOI: `10.1016/B978-0-444-53858-1.00001-6`.

L’Ecuyer, Pierre, Klas Gustavsson, and Leif Olsson (Dec. 2018). “MODELING BURSTS IN THE ARRIVAL PROCESS TO AN EMERGENCY CALL CENTER”. In: *2018 Winter Simulation Conference (WSC)*. 2018 Winter Simulation Conference (WSC). Gothenburg, Sweden: IEEE, pp. 525–536. ISBN: 978-1-5386-6572-5. DOI: `10.1109/WSC.2018.8632536`.

Lago, Jesus et al. (July 2021). “Forecasting Day-Ahead Electricity Prices: A Review of State-of-the-Art Algorithms, Best Practices and an Open-Access Benchmark”. In: *Applied Energy* 293, p. 116983. ISSN: 03062619. DOI: `10.1016/j.apenergy.2021.116983`.

Lancaster, H. O. (1957). “Some Properties of the Bivariate Normal Distribution Considered in the Form of a Contingency Table”. In: *Biometrika* 44.1-2, pp. 289–292. ISSN: 0006-3444, 1464-3510. DOI: `10.1093/biomet/44.1-2.289`.

Li, X. et al. (1997). “On Approximation of Copulas”. In: *Distributions with given Marginals and Moment Problems*. Ed. by Viktor Beneš and Josef Štěpán. Dordrecht: Springer Netherlands, pp. 107–116. ISBN: 978-94-010-6329-6. DOI: `10.1007/978-94-011-5532-8_13`.

Madadgar, Shahrbanou and Hamid Moradkhani (2014). “Improved Bayesian Multimodeling: Integration of Copulas and Bayesian Model Averaging”. In: *Water Resources Research* 50.12, pp. 9586–9603. ISSN: 19447973. DOI: `10.1002/2014WR015965`.

Mai, Jan-Frederik and Matthias Scherer (Aug. 2012). *Simulating Copulas*. Vol. 4. IMPERIAL COLLEGE PRESS. ISBN: 978-1-84816-874-9. DOI: `10.1142/p842`.

Marcjasz, Grzegorz et al. (Sept. 2023). “Distributional Neural Networks for Electricity Price Forecasting”. In: *Energy Economics* 125, p. 106843. ISSN: 01409883. DOI: `10.1016/j.eneco.2023.106843`.

Masuhr, Andreas and Mark Trede (Jan. 1, 2020). “Bayesian Estimation of Generalized Partition of Unity Copulas”. In: *Dependence Modeling* 8.1, pp. 119–131. ISSN: 2300-2298. DOI: `10.1515/demo-2020-0007`.

Mayor, G., J. Suner, and J. Torrens (Aug. 2005). “Copula-like Operations on Finite Settings”. In: *IEEE Transactions on Fuzzy Systems* 13.4, pp. 468–477. ISSN: 1063-6706. DOI: `10.1109/TFUZZ.2004.840129`.

Meier, Alexander, Claudia Kirch, and Renate Meyer (Jan. 2020). “Bayesian Nonparametric Analysis of Multivariate Time Series: A Matrix Gamma Process Approach”. In: *Journal of Multivariate Analysis* 175, p. 104560. ISSN: 0047259X. DOI: `10.1016/j.jmva.2019.104560`.

Mesiar, Radko and Vadoud Najjari (Oct. 2014). “New Families of Symmetric/Asymmetric Copulas”. In: *Fuzzy Sets and Systems* 252, pp. 99–110. ISSN: 01650114. DOI: `10.1016/j.fss.2013.12.015`.

Mirsky, L. (1960). “SYMMETRIC GAUGE FUNCTIONS AND UNITARILY INVARIANT NORMS”. In: *The Quarterly Journal of Mathematics* 11.1, pp. 50–59. ISSN: 0033-5606, 1464-3847. DOI: `10.1093/qmath/11.1.50`.

Mitchell, James (2013). *The Recalibrated and Copula Opinion Pools*. EMF research papers 02. Economic Modelling and Forecasting Group. URL: `https://ideas.repec.org/p/wrk/wrkemf/02.html`.

Mitchell, James and Kenneth F. Wallis (2011). “Evaluating Density Forecasts: Forecast Combinations, Model Mixtures, Calibration and Sharpness”. In: *Journal of Applied Econometrics* 26.6, pp. 1023–1040. ISSN: 08837252. DOI: `10.1002/jae.1192`.

Moody, Benjamin et al. (2017). *MIMIC-III Waveform Database*. Version 1.0. PhysioNet. DOI: `10.13026/C2607M`.

Murphy, Allan H. and Robert L. Winkler (July 1987). “A General Framework for Forecast Verification”. In: *Monthly Weather Review* 115.7, pp. 1330–1338. ISSN: 0027-0644. DOI: `10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2`.

Nagler, Thomas, Christian Schellhase, and Claudia Czado (2017). “Nonparametric Estimation of Simplified Vine Copula Models: Comparison of Methods”. In: *Dependence Modeling* 5.1, pp. 99–120. ISSN: 23002298. DOI: `10.1515/demo-2017-0007`.

Nagler, Thomas and Thibault Vatter (2021). *Pycopulib: High Performance Algorithms for Vine Copula Modeling*. manual. URL: `https://pypi.org/project/pyvinecopulib/`.

Neave, Henry R. and Peter L. Worthington (1988). *Distribution-Free Tests*. 1. publ. London: Unwin Hyman. 430 pp. ISBN: 978-0-04-519019-5.

Nelsen, Roger B. (2006). *An Introduction to Copulas*. Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-0-387-28659-4. DOI: `10.1007/0-387-28678-0`.

Nikoloulopoulos, Aristidis K., Harry Joe, and Haijun Li (Nov. 2012). "Vine Copulas with Asymmetric Tail Dependence and Applications to Financial Return Data". In: *Computational Statistics & Data Analysis* 56.11, pp. 3659–3673. ISSN: 01679473. DOI: `10.1016/j.csda.2010.07.016`.

Noceti, Pablo, Jeremy Smith, and Stewart Hodges (2003). "An Evaluation of Tests of Distributional Forecasts". In: *Journal of Forecasting* 22.6-7, pp. 447–455. ISSN: 02776693. DOI: `10.1002/for.876`.

Nowotarski, Jakub and Rafał Weron (Jan. 2018). "Recent Advances in Electricity Price Forecasting: A Review of Probabilistic Forecasting". In: *Renewable and Sustainable Energy Reviews* 81, pp. 1548–1568. ISSN: 13640321. DOI: `10.1016/j.rser.2017.05.234`.

Panagiotelis, Anastasios et al. (Feb. 2017). "Model Selection for Discrete Regular Vine Copulas". In: *Computational Statistics & Data Analysis* 106, pp. 138–152. ISSN: 01679473. DOI: `10.1016/j.csda.2016.09.007`.

Patton, Andrew J. (Sept. 2012). "A Review of Copula Models for Economic Time Series". In: *Journal of Multivariate Analysis* 110, pp. 4–18. ISSN: 0047259X. DOI: `10.1016/j.jmva.2012.02.021`.

Perfect, Hazel and L. Mirsky (Feb. 1965). "Spectral Properties of Doubly-Stochastic Matrices". In: *Monatshefte für Mathematik* 69.1, pp. 35–57. ISSN: 0026-9255, 1436-5081. DOI: `10.1007/BF01313442`.

Perrino, Albert C., Stephen N. Harris, and Martha A. Luther (Aug. 1, 1998). "Intraoperative Determination of Cardiac Output Using Multiplane Transesophageal Echocardiography". In: *Anesthesiology* 89.2, pp. 350–357. ISSN: 0003-3022. DOI: `10.1097/00000542-199808000-00010`.

Petropoulos, Fotios et al. (July 2022). "Forecasting: Theory and Practice". In: *International Journal of Forecasting* 38.3, pp. 705–871. ISSN: 01692070. DOI: `10.1016/j.ijforecast.2021.11.001`.

Pfeifer, Dietmar et al. (July 21, 2016). "New Copulas Based on General Partitions-of-Unity and Their Applications to Risk Management". In: *Dependence Modeling* 4.1, p. 000010151520160006. ISSN: 2300-2298. DOI: `10.1515/demo-2016-0006`.

Raftery, Adrian E. et al. (2005). "Using Bayesian Model Averaging to Calibrate Forecast Ensembles". In: *Monthly Weather Review* 133.5, pp. 1155–1174. ISSN: 00270644. DOI: `10.1175/MWR2906.1`.

Ranjan, Roopesh and Tilmann Gneiting (2010). “Combining Probability Forecasts”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 72.1, pp. 71–91. ISSN: 13697412. DOI: `10.1111/j.1467-9868.2009.00726.x`.

Rao, C. Radhakrishna (1995). “A Review of Canonical Coordinates and an Alternative to Correspondence Analysis Using Hellinger Distance.” In: *Qüestiió* 19 (1-2-3), pp. 23–63. URL: `http://eudml.org/doc/40238`.

Robert Koch Institute (Sept. 14, 2021). *Aktualisierung Der ControlCOVID-Strategie Zur Vorbereitung Auf Den Herbst/Winter 2021/22*. URL: `https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Downloads/control-covid-2021-09-22.pdf?__blob=publicationFile` (visited on 02/23/2024).

– (Feb. 21, 2024). *RKI - Coronavirus SARS-CoV-2 - Antworten Auf Häufig Gestellte Fragen (FAQ) Zu Akuten Atemwegserkrankungen Und COVID-19*. URL: `https://www.rki.de/SharedDocs/FAQ/COVID-19/FAQ-gesamt.html` (visited on 02/23/2024).

Rodríguez-Lallena, J (Feb. 15, 2004). “A New Class of Bivariate Copulas”. In: *Statistics & Probability Letters* 66.3, pp. 315–325. ISSN: 01677152. DOI: `10.1016/j.spl.2003.09.010`.

Rontsis, Nikitas and Paul Goulart (Aug. 26–28, 2020). “Optimal Approximation of Doubly Stochastic Matrices”. In: *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. Ed. by Silvia Chiappa and Roberto Calandra. Vol. 108. Proceedings of Machine Learning Research. PMLR, pp. 3589–3598. URL: `https://proceedings.mlr.press/v108/rontsis20a.html`.

Rostami-Tabar, Bahman, Jethro Browell, and Ivan Svetunkov (May 1, 2023). “Probabilistic Forecasting of Hourly Emergency Department Arrivals”. In: *Health Systems*, pp. 1–17. ISSN: 2047-6965, 2047-6973. DOI: `10.1080/20476965.2023.2200526`.

Rüschendorf, Ludger (2009). “On the Distributional Transform, Sklar’s Theorem, and the Empirical Copula Process”. In: *Journal of Statistical Planning and Inference* 139.11, pp. 3921–3927. ISSN: 03783758. DOI: `10.1016/j.jspi.2009.05.030`.

Saugel, Bernd, Oliver Grothe, and Julia Y. Nicklas (Apr. 2018). “Error Grid Analysis for Arterial Pressure Method Comparison Studies”. In: *Anesthesia & Analgesia* 126.4, pp. 1177–1185. ISSN: 0003-2999. DOI: `10.1213/ANE.0000000000002585`.

Saugel, Bernd, Oliver Grothe, and Julia Y. Wagner (Aug. 2015). “Tracking Changes in Cardiac Output: Statistical Considerations on the 4-Quadrant Plot and the Polar Plot Methodology”. In: *Anesthesia & Analgesia* 121.2, pp. 514–524. ISSN: 0003-2999. DOI: `10.1213/ANE.0000000000000725`.

Savu, Cornelia and Mark Trede (Mar. 2008). “Goodness-of-Fit Tests for Parametric Families of Archimedean Copulas”. In: *Quantitative Finance* 8.2, pp. 109–116. ISSN: 1469-7688, 1469-7696. DOI: `10.1080/14697680701207639`.

Savu, Cornelia and Mark Trede (2010). “Hierarchies of Archimedean Copulas”. In: *Quantitative Finance* 10.3, pp. 295–304. ISSN: 14697688. DOI: `10.1080/14697680902821733`.

Schefzik, Roman, Thordis L. Thorarinsdottir, and Tilmann Gneiting (Nov. 1, 2013). “Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling”. In: *Statistical Science* 28.4. ISSN: 0883-4237. DOI: `10.1214/13-STS443`.

Schmid, Friedrich et al. (2010). “Copula-Based Measures of Multivariate Association”. In: *Copula Theory and Its Applications*. Ed. by Piotr Jaworski et al. Vol. 198. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 209–236. ISBN: 978-3-642-12465-5. DOI: `10.1007/978-3-642-12465-5_10`.

Seabold, Skipper and Josef Perktold (2010). “Statsmodels: Econometric and Statistical Modeling with Python”. In: *9th Python in Science Conference*.

Sklar, A (1959). “Fonctions de Répartition à n Dimensions et Leurs Marges”. In: *Publications de L’Institut de Statistique de L’Université de Paris* 8, pp. 229–231.

Squara, Pierre et al. (Feb. 2021a). “Metrology Part 1: Definition of Quality Criteria”. In: *Journal of Clinical Monitoring and Computing* 35.1, pp. 17–25. ISSN: 1387-1307, 1573-2614. DOI: `10.1007/s10877-020-00494-y`.

– (Feb. 2021b). “Metrology Part 2: Procedures for the Validation of Major Measurement Quality Criteria and Measuring Instrument Properties”. In: *Journal of Clinical Monitoring and Computing* 35.1, pp. 27–37. ISSN: 1387-1307, 1573-2614. DOI: `10.1007/s10877-020-00495-x`.

Stewart, G. W. (1991). “Perturbation Theory for the Singular Value Decomposition”. In: *SVD and Signal Processing, II: Algorithms, Analysis, and Applications*. Ed. by Richard J. Vaccaro and University of Rhode Island. Amsterdam ; New York : New York, N.Y., U.S.A: Elsevier ; Distributors for the U.S.A. and Canada, Elsevier Science Pub. Co, pp. 99–109. ISBN: 978-0-444-88896-9.

Stock, James H. and Mark W. Watson (2004). “Combination Forecasts of Output Growth in a Seven-Country Data Set”. In: *Journal of Forecasting* 23.6, pp. 405–430. ISSN: 02776693. DOI: `10.1002/for.928`.

Tay, Anthony S. and Kenneth F. Wallis (July 1, 2000). “Density Forecasting: A Survey”. In: *Journal of Forecasting* 19.4, pp. 235–254. ISSN: 0277-6693. DOI: `10.1002/1099-131X(200007)19:4<235::AID-FOR772>3.0.CO;2-L`.

Taylor, James W., Patrick E. McSharry, and Roberto Buizza (2009). “Wind Power Density Forecasting Using Ensemble Predictions and Time Series Models”. In: *IEEE Transactions on Energy Conversion* 24.3, pp. 775–782. ISSN: 08858969. DOI: `10.1109/TEC.2009.2025431`.

Thorarinsdottir, Thordis L. (2013). "Calibration Diagnostics for Point Process Models via the Probability Integral Transform". In: *Stat* 2.1, pp. 150–158. ISSN: 20491573. DOI: `10.1002/sta4.25`.

Tolksdorf, Kristin, Anna Loenenbach, and Silke Buda (Sept. 22, 2022). "Dritte Aktualisierung der „Retrospektiven Phaseneinteilung der COVID-19-Pandemie in Deutschland“". In: *Epidemiologisches Bulletin* 2022.38, pp. 3–6. DOI: `10.25646/10598`.

Van Buuren, Stef (July 17, 2018). *Flexible Imputation of Missing Data, Second Edition*. 2nd ed. Second edition. | Boca Raton, Florida : CRC Press, [2019] |: Chapman and Hall/CRC. ISBN: 978-0-429-49225-9. DOI: `10.1201/9780429492259`.

Vesco, Paola et al. (July 4, 2022). "United They Stand: Findings from an Escalation Prediction Competition". In: *International Interactions* 48.4, pp. 860–896. ISSN: 0305-0629, 1547-7444. DOI: `10.1080/03050629.2022.2029856`.

Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: `10.1038/s41592-019-0686-2`.

Wang, Hongrui et al. (June 2017). "Bayesian Forecasting and Uncertainty Quantifying of Stream Flows Using Metropolis–Hastings Markov Chain Monte Carlo Algorithm". In: *Journal of Hydrology* 549, pp. 476–483. ISSN: 00221694. DOI: `10.1016/j.jhydrol.2017.03.073`.

Wang, J.P. et al. (Dec. 2018). "Copula-Based Earthquake Early Warning Decision-Making Strategy". In: *Soil Dynamics and Earthquake Engineering* 115, pp. 324–330. ISSN: 02677261. DOI: `10.1016/j.soildyn.2018.08.031`.

Wang, Xiaoqian et al. (Oct. 2023). "Forecast Combinations: An over 50-Year Review". In: *International Journal of Forecasting* 39.4, pp. 1518–1547. ISSN: 01692070. DOI: `10.1016/j.ijforecast.2022.11.005`.

Wilks, Daniel S. (2011). *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series v. 100. Amsterdam ; Boston: Elsevier/Academic Press. 676 pp. ISBN: 978-0-12-385022-5.

Wolffram, Daniel et al. (Aug. 11, 2023). "Collaborative Nowcasting of COVID-19 Hospitalization Incidences in Germany". In: *PLOS Computational Biology* 19.8. Ed. by James M McCaw, e1011394. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1011394`.

Woo, Gerald et al. (2024). "Unified Training of Universal Time Series Forecasting Transformers". Version 1. In: DOI: `10.48550/ARXIV.2402.02592`.

World Meteorological Organization (WMO) (2017). *Guidelines for Nowcasting Techniques*. Geneva: WMO. 82 pp. ISBN: 978-92-63-11198-2. URL: `https://library.wmo.int/idurl/4/55666`.

Wu, Joseph T. et al. (Mar. 2021). “Nowcasting Epidemics of Novel Pathogens: Lessons from COVID-19”. In: *Nature Medicine* 27.3, pp. 388–395. ISSN: 1078-8956, 1546-170X. DOI: `10.1038/s41591-021-01278-w`.

Yang, Wentong, Liyuan Zhang, and Yu Gao (Feb. 2023). “Drought and Flood Risk Assessment for Rainfed Agriculture Based on Copula-Bayesian Conditional Probabilities”. In: *Ecological Indicators* 146, p. 109812. ISSN: 1470160X. DOI: `10.1016/j.ecolind.2022.109812`.

Yazdandoost, Farhad, Mina Zakipour, and Ardalan Izadi (Sept. 2021). “Copula Based Post-Processing for Improving the NMME Precipitation Forecasts”. In: *Heliyon* 7.9, e07877. ISSN: 24058440. DOI: `10.1016/j.heliyon.2021.e07877`.

Zass, Ron and Amnon Shashua (Sept. 7, 2007). “Doubly Stochastic Normalization for Spectral Clustering”. In: *Advances in Neural Information Processing Systems 19*. Ed. by Bernhard Schölkopf, John Platt, and Thomas Hofmann. The MIT Press, pp. 1569–1576. ISBN: 978-0-262-25691-9. DOI: `10.7551/mitpress/7503.003.0201`.