

4. Simulationsstudie: Parameterbestimmung bivariater Copulas – Vergleich von Maximum-Likelihood und Minimum-Distanz Schätzern

Nachdem die theoretischen Grundlagen für dieses abschließende Kapitel gelegt sind, soll es nun nachfolgend um die Konstruktion der Simulationsstudie gehen. Die Vorgehensweise orientiert sich an einer ähnlich durchgeführten Studie von Weiß [2009, S.32 ff.]. Die Ergebnisse der dort vorgestellten Simulationsstudie sollen als Vergleichswerte für die empirischen Befunde dieser Arbeit dienen. Doch zunächst wird der Fokus auf die zugrundeliegenden Schätzverfahren gelegt. Dabei werden diese zunächst näher erläutert und charakterisiert, ehe im nächsten Schritt die Zielstellung und die Durchführung beschrieben werden. Abgerundet wird dieses Kapitel mit der Präsentation der Ergebnisse und den daraus resultierenden Schlussfolgerungen.

Die Simulationsstudie wurde mithilfe des Statistikprogramms R praktisch umgesetzt. Es sei bereits an dieser Stelle darauf hingewiesen, dass sich der vollständige R-Code im Anhang dieser Arbeit befindet.

4.1 Schätzverfahren zur Parameterbestimmung

Das Thema Schätzverfahren für Copula-Parameter besitzt in einschlägiger Literatur einen hohen Stellenwert und wurde in der Vergangenheit bereits des öfteren thematisiert. So werden beispielsweise durch Jondeau et al. [2007, S. 254 ff.] die Parameterbestimmung mittels *Maximum-Likelihood-Verfahren* und mittels *empirischer Verteilungsfunktion* als die zwei Hauptverfahren genannt. Auf dem Gebiet der *Maximum-Likelihood-Schätzer* (ML) könne zudem noch zwischen dem *Standardansatz* (*fully parametric*) und dem *zweistufigen Ansatz* (*stepwise parametric – inference functions for margins – IFM Ansatz*) unterschieden werden. Der *IFM Ansatz* folgt dem Gedanken, die Parameter der Randverteilungen getrennt von denen der Copula zu modellieren. Wie bereits in Kapitel 2.1 kennengelernt, lässt sich die gemeinsame Verteilungsfunktion, in die Abhängigkeitsstruktur, repräsentiert durch die Copula und in ihre Randverteilungen, aufspalten. Demnach ergibt sich die log-Likelihood Funktion des gesamten Modells als Summe der Randverteilungen und der Abhängigkeitsstruktur. (vgl. Jondeau et al. [2007, S. 254 f.])

Das Konzept der *empirischen Verteilungsfunktion*, bzw. vielmehr der *empirischen Copula*, basiert auf Forschungen von Deheuvel [1978], der gezeigt hat, dass die *empirische Copula* konsistent gegen die wahre, zugrunde liegende hypothetische Copula konvergiert.

Basierend auf dieser Idee haben sich zahlreiche Goodness-of-Fit-Teststatistiken (GoF) herausgebildet (vgl. Fermanian [2005]). Diese können unter anderem dazu genutzt werden, *Minimum-Distanz* (MD) Schätzer zu entwickeln. Dabei dient die zugrundeliegende Teststatistik als Maß für die Distanz zwischen *empirischer* und *hypothetisch unterstellter* Copula. Dabei soll der *Copula-Parameter* so gewählt werden, dass die Distanz minimiert wird.

In dieser Arbeit wird der Standardansatz der *ML-Schätzverfahren* angewandt, der im nachfolgenden Kapitel näher vorgestellt wird. Daraufhin werden drei *MD-Schätzer* vorgestellt, die auf dem Konzept der empirischen Copula beruhen und drei die auf Kendall's Transformation beruhen.

4.1.1 Maximum Likelihood Schätzer

Für die Parameterbestimmung mittels *ML-Verfahren*, werden die Dichtefunktionen der jeweils zu untersuchenden Copula benötigt. Wie in den zuvor behandelten Kapiteln explizit für die einzelnen *Copula-Klassen* definiert, ergibt sich die Copuladichte einer bivariaten Copula C im Allgemeinen aus $U_i \equiv (U_{i,1}, U_{i,2})^t \in [0; 1]^2$ mit der i -ten Stichprobe der Copula und c sei die Dichte parametrisiert durch θ als

$$c(u_1, u_2 | \theta) = \frac{\partial C(u_1, u_2 | \theta)}{\partial u_1 \cdot \partial u_2}, \quad u_1, u_2 \in [0; 1]. \quad (4.1)$$

Die (*logarithmierte*) *Likelihood-Funktion* L_U für die Parameterbestimmung $\hat{\theta}_n^{ML}$ der Stichprobe $\mathbf{U} \equiv (U_{i,j})$, mit Stichprobengröße $i = 1, \dots, n$ und Dimension $j \in \{1; 2\}$, lautet

$$L_U(\theta) = \sum_{i=1}^n \log c(U_{i,1}, U_{i,2} | \theta). \quad (4.2)$$

Dabei ist $\theta \in \Theta \subset \mathbb{R}^p$ ($p \geq 1$), der Parameterraum der zugrunde liegenden *bivariaten Copula*.

Der *ML-Schätzer* ergibt sich dann durch numerische Maximierung von L_U , sodass für $\hat{\theta}_n^{ML}$

$$\hat{\theta}_n^{ML}(\mathbf{U}) \equiv \arg \max_{\theta \in \Theta} L_{\mathbf{U}}(\theta) \quad (4.3)$$

gilt. (vgl. Weiß [2005, S.35 f.])

Demnach wird der Parameter $\hat{\theta}_n^{ML}$ gesucht, der die *Likelihood-Funktion* $L_{\mathbf{U}}$ maximiert. Wie einführend zum Teilkapitel 4.1 erläutert, gibt es noch weitere *ML-basierte* Schätzverfahren die durch Jondeau et al. [2007, S. 256 ff.] grundlegend betrachtet werden.

4.1.2 Minimum Distanz Schätzer

Das Konzept der *Minimum-Distanz Schätzer* basiert auf diversen GoF-Teststatistiken. In dieser Arbeit werden die *Cramér-von-Mises*, die *Kolmogorov-Smirnov* und die L_1 -Variante der *Cramér-von-Mises Teststatistik* betrachtet. Dabei werden zunächst Schätzer, basierend auf dem *empirischen Copula Prozess* betrachtet, ehe anschließend die auf *Kendall's Transformation* basierenden *MD-Schätzer* eingeführt werden.

Minimum-Distanz Schätzer basierend auf dem empirischen Copula Prozess

Eine bivariate *empirische Copula* einer i.i.d. Stichprobe der Größe n ist auf einem Gitter, gemäß

$$L = \left\{ \left(\frac{i_1}{n}, \frac{i_2}{n} \right)^t \in [0; 1]^2 \mid i_1, i_2 = 0, \dots, n; \right\} \quad (4.4)$$

definiert. Für die empirische Copula gilt

$$C_n(\mathbf{u}) \equiv \frac{1}{n} \sum_{i=1}^n \mathbf{1}(U_{i,1} \leq u_1, U_{i,2} \leq u_2), \mathbf{u} \equiv (u_1, u_2)^t \in [0; 1]^2, \quad (4.5)$$

mit $\mathbf{1}(\cdot)$ als *logische Indikator Funktion*. (vgl. Weiß [2005, S.36]).

Eine ähnliche Definition liefert Nelsen [2006, S. 219 ff.], indem er die Funktion C_n definiert als

$$C_n \left(\frac{i_1}{n}, \frac{i_2}{n} \right) = \frac{\text{Anzahl der Paare } (U_{i,1}, U_{i,2}) \text{ in der Stichprobe für } U_{i,1} \leq u_1, U_{i,2} \leq u_2}{n} \quad (4.6)$$

mit $1 \leq i \leq n$. Im Gegensatz zu „Weiß“ *logischer Indikator Funktion*, spricht Nelsen von der *empirischen Copula Häufigkeit*, die den Wert $\frac{1}{n}$ annimmt, insofern (u_1, u_2) Elemente der Stichprobe sind und den Wert 0 ansonsten.

Alternativ kann die *empirische* Copula C_n , gemäß Weiß [2005, S.36 f.] „[...] aus der 2-dimensionalen cdf \mathbb{H}_n des Zufallsvektors X und dessen Randverteilungen $F_{1,n}$ und $F_{2,n}$ unter Benutzung von

$$C_n(\mathbf{u}) = \mathbb{H}_n(F_{1,n}^{-1}(u_1), F_{2,n}^{-1}(u_2)) \quad (4.7)$$

mit \mathbb{H}_n und $F_{i,n}$ ($i \in [1; 2]$), berechnet aus der i.i.d. Stichprobe der Größe n “ bestimmt werden.

Im Folgenden sei

$$\mathbb{C}_n = \sqrt{n}(C_n - C_{\hat{\theta}}), \quad (4.8)$$

der zu untersuchende *empirische* Prozess. Dabei ist C_n die in (4.5) definierte *empirische* Copula, häufig auch als *Deheuvel's empirische Copula* (vgl. Weiß [2005, S.37]) bezeichnet. Diese wird mit der hypothetisch angenommenen, durch den Parameter $\hat{\theta}$ spezifizierte Copula $C_{\hat{\theta}}$, verglichen. Verallgemeinert gilt für die *Cramér-von-Mises* (*CvM*) und die *Kolmogorov-Smirnov* (*K-S*) Statistik (vgl. Weiß [2005, S.37]), basierend auf dem *empirischen* Prozess \mathbb{C}_n

$$\rho_{emp}^{CvM} \equiv \int_{[0,1]^2} \mathbb{C}_n(\mathbf{u})^2 dC_n(\mathbf{u}), \quad (4.9)$$

$$\rho_{emp}^{KS} \equiv \sup_{\mathbf{u} \in [0,1]^2} |\mathbb{C}_n(\mathbf{u})|. \quad (4.10)$$

Für die empirischen Varianten gilt,

$$\hat{\rho}_{emp}^{CvM}(\mathbf{U}; \theta) \equiv \sum_{i=1}^n \{C_n(\mathbf{U}_i) - C_{\hat{\theta}}(\mathbf{U}_i)\}^2 \quad (4.11)$$

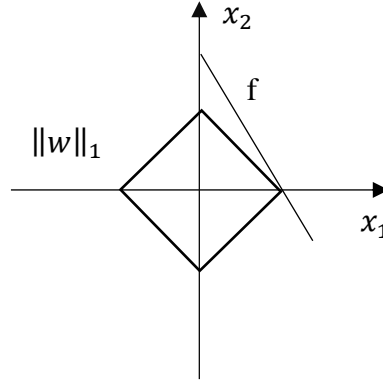
und

$$\hat{\rho}_{emp}^{KS}(\mathbf{U}; \theta) \equiv \max_{i=1, \dots, n} |C_n(\mathbf{U}_i) - C_{\hat{\theta}}(\mathbf{U}_i)|, \quad (4.12)$$

welche später in der Simulationsstudie angewandt werden. Weiterhin soll eine L_1 - Variante der *CvM-Statistik* betrachtet werden.

L_1 -Norm

Die L_1 -Norm, wird häufig als *kleinste absolute Werte-Methode* bezeichnet. (vgl. Dodge [1987, S.239 f.]). Für die L_1 -Norm gilt grafisch gesehen der folgende Zusammenhang.



Für $\rho_{emp}^{CvM L1}$ gilt

$$\rho_{emp}^{CvM L1} \equiv \sqrt{n} \int_{[0,1]^2} |\mathbb{C}_n(\mathbf{u})| dC_n(\mathbf{u}), \quad (4.13)$$

bzw. für die *empirische* Variante $\hat{\rho}_{emp}^{L1}$

$$\hat{\rho}_{emp}^{CvM L1}(\mathbf{U}; \theta) \equiv \sum_{i=1}^n |C_n(\mathbf{U}_i) - C_{\hat{\theta}}(\mathbf{U}_i)|. \quad (4.14)$$

Die zugehörigen *MD-Schätzer* für die Formeln (4.11, 4.12, 4.14) lauten dann

$$\hat{\theta}_n^{emp, CvM}(\mathbf{U}) \equiv \arg \min_{\theta \in \Theta} \hat{\rho}_{emp}^{CvM}(\mathbf{U}; \theta), \quad (4.15)$$

$$\hat{\theta}_n^{emp, KS}(\mathbf{U}) \equiv \arg \min_{\theta \in \Theta} \hat{\rho}_{emp}^{KS}(\mathbf{U}; \theta) \text{ und} \quad (4.16)$$

$$\hat{\theta}_n^{emp, CvM L1}(\mathbf{U}) \equiv \arg \min_{\theta \in \Theta} \hat{\rho}_{emp}^{CvM L1}(\mathbf{U}; \theta). \quad (4.17)$$

Zudem bemerkt Weiß [2005, S. 37] abschließend, dass die Konvergenz des *empirischen* Prozesses \mathbb{C}_n unter regulären Bedingungen gezeigt werden könne, wie z.B. durch Genest et al. [2009].

Minimum-Distanz Schätzer basierend auf Kendall's Transformation

Das zweite Schätzverfahren welches im Rahmen der Simulationsstudie untersucht wird, basiert auf *Kendall's Wahrscheinlichkeits-Integral Transformation*, dass in der Literatur auch unter *Kendall's Abhängigkeitsfunktion* zu finden ist (vgl. Weiß [2005, S.38]).

Die Transformation eines Zufallsvektors X , mit der gemeinsamen kumulativen Verteilungsfunktion (*cdf*) G und den Randverteilungen F_i für $i \in \mathbb{N}_d$ lautet, in Anlehnung an Genest et al. [2009, S. 201],

$$X \mapsto V = G(X) = C(U_1, \dots, U_d). \quad (4.18)$$

Dabei ist C die gemeinsame Verteilungsfunktion des Vektors $U = (U_1, \dots, U_d)^t$. Zudem gilt, dass $U_i = F_i(X_i)$ für $i \in \mathbb{N}_d$. Wie Genest et al. [2009, S. 201] an dieser Stelle bemerken, sei die Bezeichnung *Kendall Transformation* (bzw. *Kendall's Abhängigkeitsfunktion*) aus der Überlegung heraus entstanden, dass der Erwartungswert von V eine affine Transformation der multivariaten Variante des Kendall'schen Abhängigkeitskoeffizienten sei.

In einem nächsten Schritt soll es nun um die Verteilungsfunktion der *Kendall'schen Wahrscheinlichkeits-Integral Transformation* gehen. Dafür wird nachfolgend angenommen, dass K die gemeinsame kumulative Verteilungsfunktion von V ist. Eine nichtparametrische Berechnung von K , basiert auf der *transformierten* Stichprobe der Größe n , sodass $V_i = C_n(U_i)$ und damit

$$K_n(\varpi) \equiv \frac{1}{n} \sum_{i=1}^n \mathbf{1}(V_i \leq \varpi), \varpi \in [0; 1] \quad (4.19)$$

gilt. Weiß [2005, S.38] leitet aus (4.19) eine wichtige Implikation ab, denn „*[w]enn U gemäß $C_{\hat{\theta}}$ verteilt ist, dann ist die parametrische Berechnung von K gegeben durch die Verteilung $K_{\hat{\theta}}$ der Kendall'schen Transformation $C_{\hat{\theta}}(U)$.*“ Durch die Wahl eines entsprechenden Maßes, welches die Distanz zwischen K und $K_{\hat{\theta}}$ misst, ist es möglich dieses Konzept für die Parameterbestimmung zu benutzen.

Im Folgenden wird wieder der *empirische Prozess* betrachtet, der hier als \mathbb{K}_n bezeichnet wird und für den analog zu (4.8) gilt,

$$\mathbb{K}_n \equiv \sqrt{n}(K_n - K_{\hat{\theta}}). \quad (4.20)$$

Mithilfe des unter (4.20) vorgestellten *empirischen Prozesses*, sollen in der Simulationsstudie wiederum drei *GoF-Teststatistiken* genutzt werden. Basierend auf \mathbb{K}_n lauten die Teststatistiken

$$\rho_K^{CvM} \equiv \int_0^1 \mathbb{K}_n(\varpi)^2 dK_{\hat{\theta}}(\varpi), \quad (4.21)$$

$$\rho_K^{KS} \equiv \sup_{\varpi \in [0;1]} |\mathbb{K}_n(\varpi)| \text{ und} \quad (4.22)$$

$$\rho_K^{CvM L1} \equiv \sqrt{n} \int_0^1 |\mathbb{K}_n(\varpi)| dK_{\hat{\theta}}(\varpi). \quad (4.23)$$

Für eine einfachere Handhabung in der Simulationsstudie, werden wiederum die *empirischen Versionen*

$$\hat{\rho}_K^{CvM}(\mathbf{V}; \theta) \equiv \sum_{i=1}^n \{K_n(\mathbf{V}_i) - K_{\hat{\theta}}(\mathbf{V}_i)\}^2 \quad (4.24)$$

$$\hat{\rho}_K^{KS}(\mathbf{V}; \theta) \equiv \max_{i=1, \dots, n} |K_n(\mathbf{V}_i) - K_{\hat{\theta}}(\mathbf{V}_i)| \text{ und} \quad (4.25)$$

$$\hat{\rho}_K^{CvM L1}(\mathbf{V}; \theta) \equiv \sum_{i=1}^n |K_n(\mathbf{V}_i) - K_{\hat{\theta}}(\mathbf{V}_i)| \quad (4.26)$$

von (4.21 - 4.23) betrachtet. Demnach lassen sich die *Minimum-Distanz Schätzer* aufstellen, für die

$$\hat{\theta}_n^{K,CvM}(\mathbf{U}) \equiv \arg \min_{\theta \in \Theta} \hat{\rho}_K^{CvM}(\mathbf{V}; \theta), \quad (4.27)$$

$$\hat{\theta}_n^{K,KS}(\mathbf{U}) \equiv \arg \min_{\theta \in \Theta} \hat{\rho}_K^{KS}(\mathbf{V}; \theta) \text{ und} \quad (4.28)$$

$$\hat{\theta}_n^{K,CvM L1}(\mathbf{U}) \equiv \arg \min_{\theta \in \Theta} \hat{\rho}_K^{CvM L1}(\mathbf{V}; \theta). \quad (4.29)$$

gilt. In der Simulationsstudie wird dann der *Maximum-Likelihood Schätzer* (4.3) mit den *Minimum-Distanz Schätzern* basierend auf dem *empirischen Copula-Prozess* (4.15 – 4.17) und denen basierend auf *Kendall's Transformation* (4.27 – 4.29) verglichen und in Bezug auf deren Güte untersucht. Wie durch Weiß [2005, S. 39 f.] und Genest et al. [2009, S. 202 ff.] dargelegt, können auch für *Rosenblatt Transformationen* Minimum Distanz Schätzer definiert werden, die hier nicht näher untersucht werden. Im nächsten Teilkapitel geht es um die Vorstellung der Simulationsstudie.

4.2 Simulationsstudie

Nachdem insgesamt 7 Schätzverfahren vorgestellt wurden, sollen diese im Rahmen einer Simulationsstudie auf ihre Genauigkeit bzw. Performance beim Schätzen *bivariater* Copula-Modelle untersucht werden. Dabei soll der *Mean-Squared-Error (MSE)* als Vergleichskriterium für die Performance der Schätzer dienen, da dieser den geschätzten Parameterwert mit dem wahren bzw. unterstellten Parameter vergleicht und diesen Abstand quadratisch bewertet. In einführender statistischer Literatur, wie Elpelt und Hartung [2007, S. 40], wird der *MSE* gemäß

$$MSE(\hat{\theta}) = \mathbb{E}(\theta - \hat{\theta})^2 \quad (4.30)$$

definiert. Dabei entspricht $\mathbb{E}(\hat{\theta}) - \theta$ dem Bias bzw. der Verzerrung des Schätzers $\hat{\theta}$.

Weiterhin sei der *MSE* ein Kriterium für die Güte eines Schätzers, weshalb dieser bei einem guten Schätzer möglichst klein sein soll.

Das Ziel der Simulationsstudie ist es, eine Aussage bezüglich der Güte der vorgestellten Schätzverfahren abzuleiten. Darüber hinaus sind dann auch Erkenntnisse darüber möglich, welchen Einfluss Parameterwert und Stichprobengröße auf die Schätzverfahren haben.

4.2.1 Design der Simulationsstudie

Im Folgenden werden vier bivariate Copulas mit zugehörigen Parametrisierungen betrachtet:

- *Gauss-Copula* für $\theta \in \{-0.9; -0.7; \dots; 0.7; 0.9\}$
- *t-Copula* für $\theta \in \{-0.9; -0.7; \dots; 0.7; 0.9\}$ und $v \in \{3; 5; 10\}$
- *Gumbel-Copula* für $\theta \in \{1.5; 2.5; \dots; 8.5; 9.5\}$
- *Clayton-Copula* für $\theta \in \{0.5; 1.5; \dots; 8.5; 9.5\}$.

Für jede der *bivariaten* Copula-Familien C_{θ}^i mit wahren Parameter θ für Familie $i \in \{1; 2; 3; 4\}$, werden die folgenden Schritte K -mal wiederholt. Wobei K auf 500 gesetzt wird. Für die Stichprobengröße n , gilt $n \in \{100; 250; 500; 1000\}$.

- (1) Simulation einer Stichprobe $(u_1, \dots, u_n)^t$ der Größe n der zugrundeliegenden Copula C_θ^i .
- (2) Berechnung der Parameterschätzer $\hat{\theta}_S$ mit $S \in \{1; 2; \dots; 6; 7\}$ basierend auf *ML-Verfahren* (4.3), *MD-Schätzer empirischer Copula Prozess* (4.15 – 4.17), *MD-Schätzer Kendall's Transformation* (4.27 – 4.29).
- (3) Vergleich des jeweiligen Parameterschätzers $\hat{\theta}_S$ mit dem dazugehörigen wahren Parameter θ . Als Vergleichskriterium dient hierbei der *MSE* (4.30).

Insgesamt werden in der Simulationsstudie 826.000 Parameterschätzer berechnet, dies entspricht 118.000 Werten pro Schätzverfahren. Aufgrund der Vielzahl an zur Verfügung stehenden Simulationsergebnissen, sollen insbesondere die Einflüsse der Parameterwerte und die der Stichprobengröße auf die Schätzverfahren geklärt werden.

Die Berechnungen wurden mithilfe von R Version 3.1.3, basierend auf der Optimierungsfunktion *optim* und der Methode *L-BFGS-B (Limited-memory- Broyden–Fletcher–Goldfarb–Shanno-box constraints)* unter Festlegung unterer und oberer Schranken, durchgeführt. Zudem wurde der R-Code aufgrund der Vielzahl an zu berechnenden Werte parallelisiert, damit ein verbessertes Laufzeitverhalten vorliegt.

4.2.2 Resultat und Schlussfolgerung

In den folgenden Abbildungen 24, 25, 26 und 27 werden grafisch die Ergebnisse der Simulationsstudie präsentiert. Dabei werden in jeder Abbildung eine Copula-Familie C_θ^i mit deren Simulationsergebnissen der Parameterschätzer $\hat{\theta}_S$ für alle zu betrachtenden Stichprobengrößen dargestellt. Als Vergleichskriterium dient der unter (4.30) vorgestellte *mittlere quadratische Fehler (MSE)*, der für alle K -Werte berechnet wurde. Im Anschluss daran wurde der Mittelwert der MSE's eines Parameterschätzers $\hat{\theta}_S$ für $\Phi \in \theta$ berechnet und grafisch festgehalten. Es sei darauf hingewiesen, dass im Falle der Schätzer für die Klasse der t-Copulas, lediglich die Ergebnisse für den Freiheitsgrad $df = 10$ präsentiert werden. Die Grafiken für die Freiheitsgrade $df = 3$ und $df = 5$ befinden sich im Anhang. In Abbildung 24 werden die MSE's der Parameterschätzer auf der Ordinate unter Abhängigkeit des wahren Parameters θ auf der Abszisse unter der Hypothese, dass

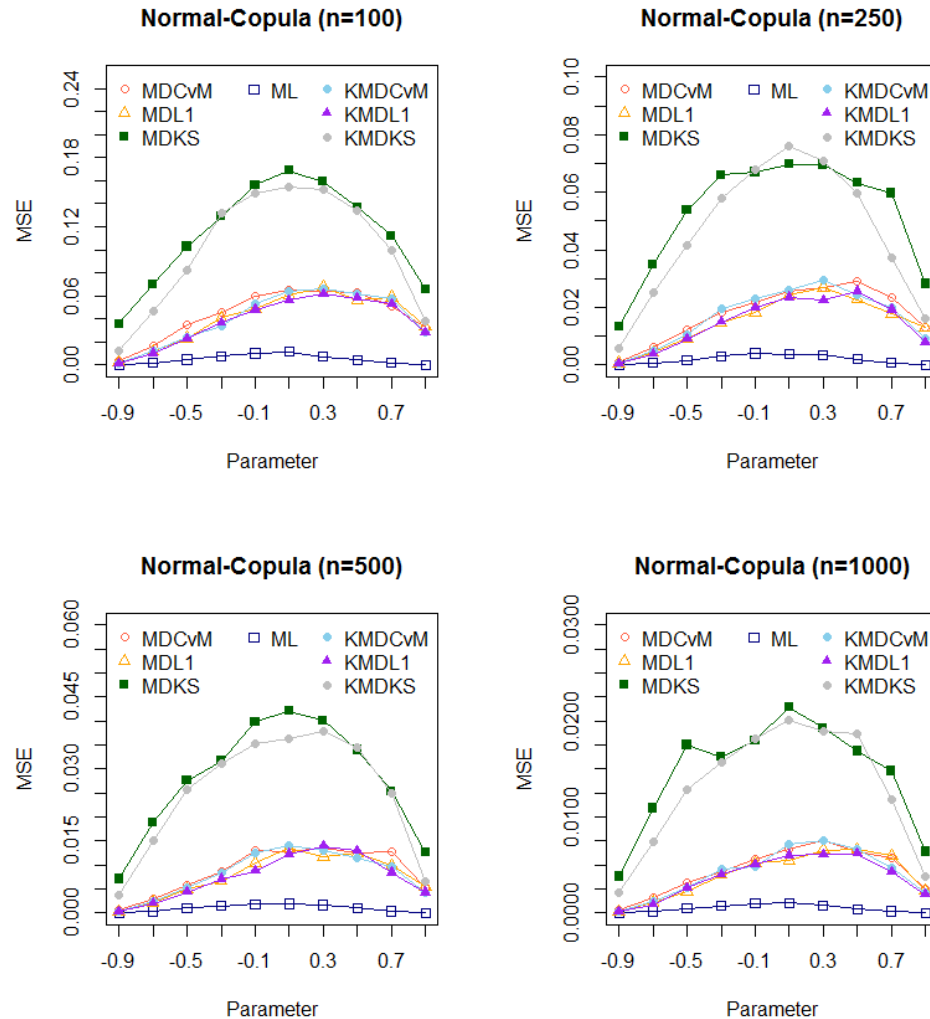


Abbildung 24: Darstellung des MSE für ML-Schätzer, MD-Schätzer basierend auf dem empirischen Prozess und MD-Schätzer basierend auf Kendall's Transformation mit ansteigender Stichprobengröße für die Normal-Copula

die Copula-Familie korrekt bestimmt wurde, dargestellt. Zunächst ist grundlegend festzustellen, dass der ML-Schätzer in jedem Falle einen geringeren mittleren quadratischen Fehler aufweist als alle anderen Parameterschätzer. Die MD-Schätzer performen demnach alle weniger gut. Doch es muss hervorgehoben werden, dass insbesondere die Kolmogorov-Smirnoff Teststatistik sowohl für die auf dem empirischen Prozess basierenden Schätzer als auch für die auf Kendall's Transformation basierten wesentlich höhere Abweichungen besitzen. Alle anderen MD-Schätzer befinden sich von ihrer Güte, basierend auf dem MSE als Kriterium und in dieser Simulation, auf nahezu einem gleichen Niveau. Zudem wird durch ein erhöhten Stichprobenumfang eine

verbesserte Güte der Schätzer erreicht, welche sich durch eine Abnahme des MSE mit bemerkbar macht.

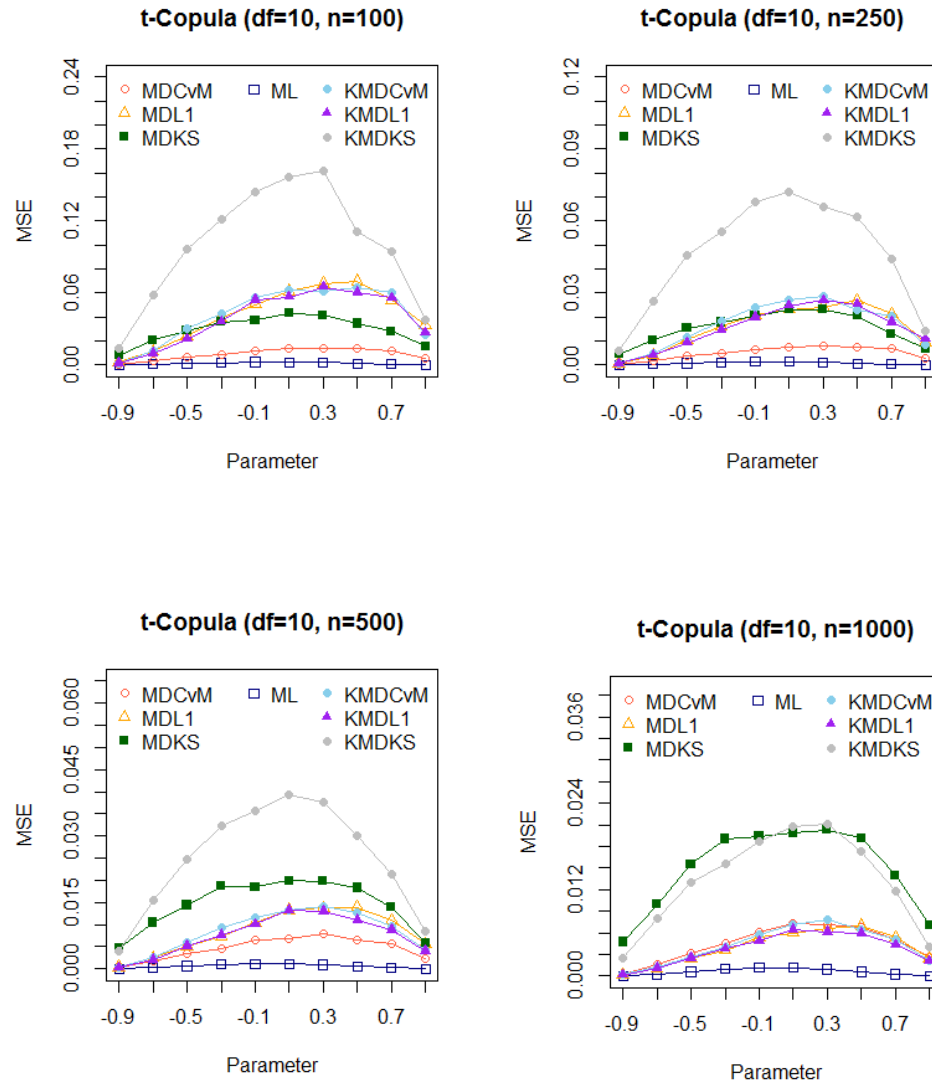


Abbildung 25: Darstellung des MSE für ML-Schätzer, MD-Schätzer basierend auf dem empirischen Prozess und MD-Schätzer basierend auf Kendall's Transformation mit ansteigender Stichprobengröße für die t-Copula mit Freiheitsgrad 10

In Abbildung 25 können ähnliche Effekte festgestellt werden, was Anpassungsgüte in Abhängigkeit von der Stichprobengröße anbelangt. Eine Auffälligkeit ist die Veränderung des MDKS in Abhängigkeit der restlichen MD-Schätzer. Bei einer Stichprobengröße von 100 liegt der MSE leicht unter den der restlichen MD-Schätzer. Doch durch sukzessive Erhöhung des Stichprobenumfangs, nähert sich das Niveau des MSE des MDKS-Schätzers immer mehr an das des KMDKS-Schätzers. Sowohl die auf CvM- als auch die auf L1-

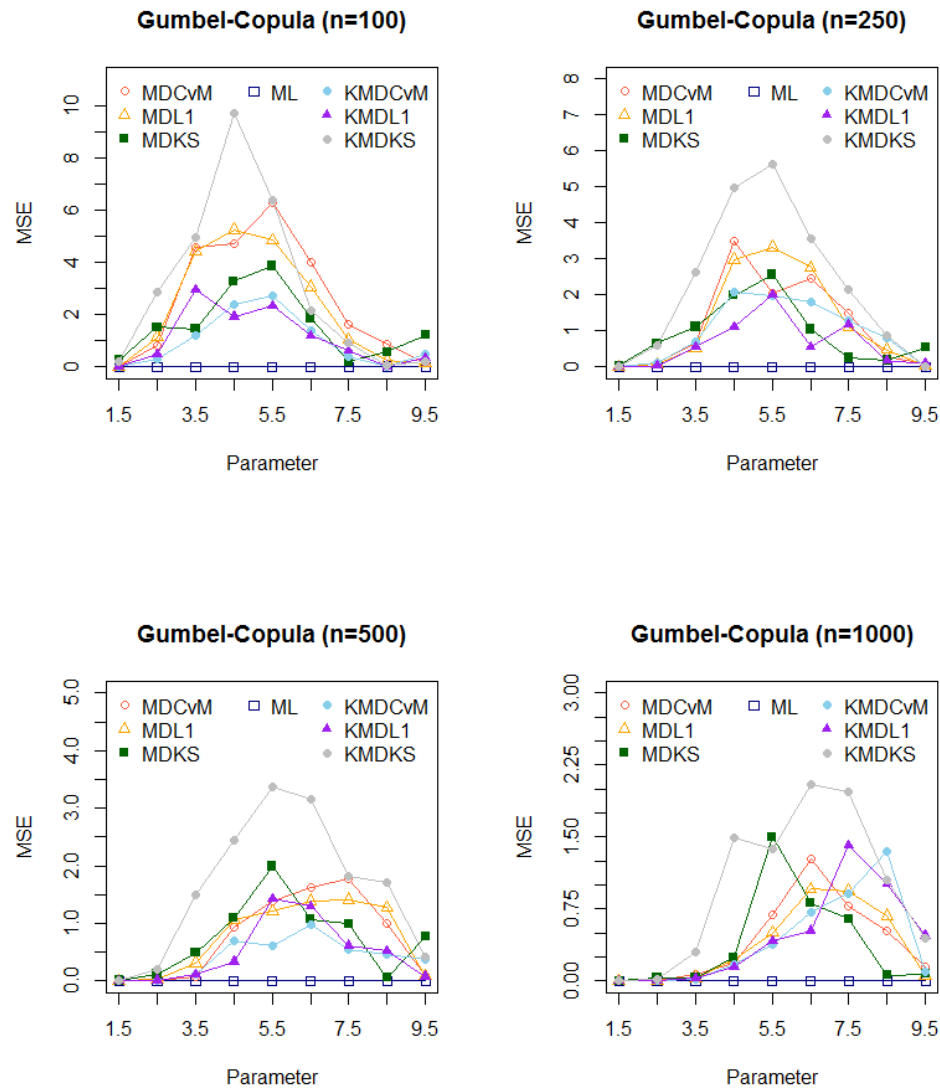


Abbildung 26: Darstellung des MSE für ML-Schätzer, MD-Schätzer basierend auf dem empirischen Prozess und MD-Schätzer basierend auf Kendall's Transformation mit ansteigender Stichprobengröße für die Gumbel-Copula

basierten Schätzer liefern eine ähnliches MSE-Niveau für die unterschiedlichen Stichproben. Die in Abbildung 26 dargestellten Ergebnisse für die Familie der Gumbel-Copulas, fällt insbesondere durch einen erhöhten MSE im Vergleich zu den parametrischen Copulas auf. Dabei fällt besonders auf, dass die MSE's der MD-Schätzer sprunghaft ansteigen, was insbesondere bei einer Stichprobengröße von 1000 ab 3.5 der Fall ist.

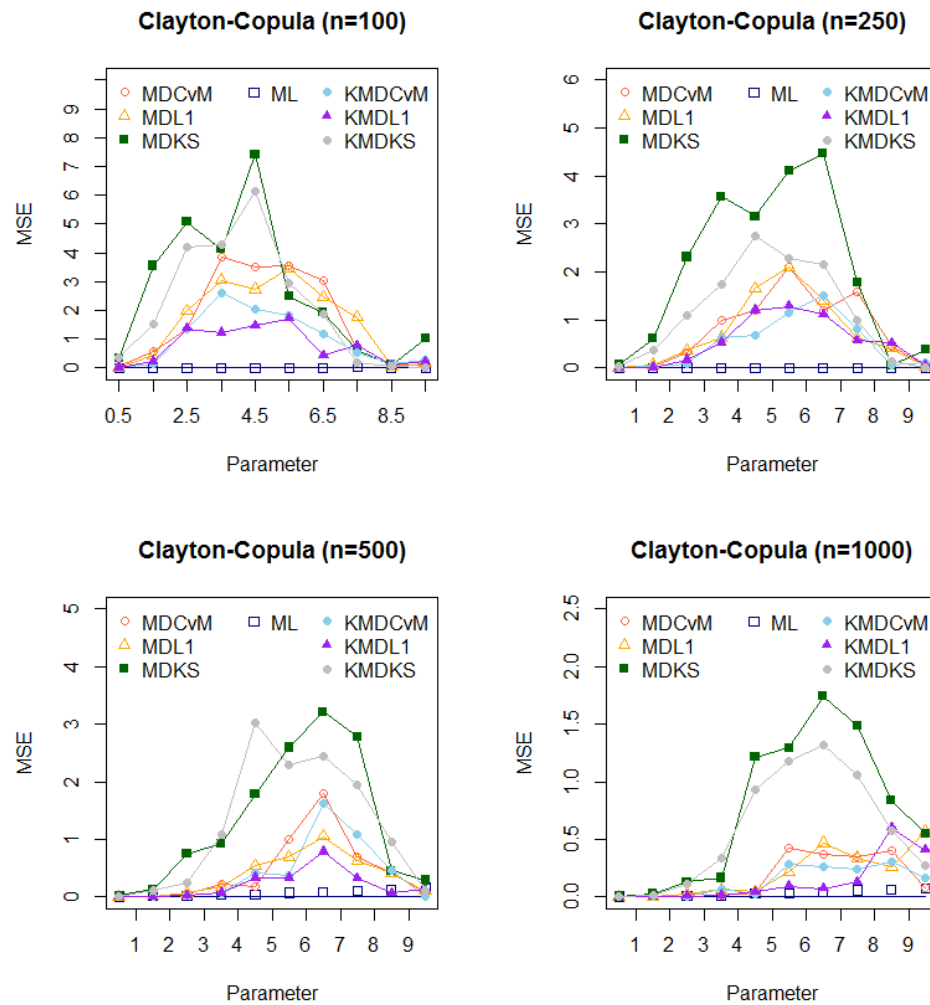


Abbildung 27: Darstellung des MSE für ML-Schätzer, MD-Schätzer basierend auf dem empirischen Prozess und MD-Schätzer basierend auf Kendall's Transformation mit ansteigender Stichprobengröße für die Clayton-Copula

Die in Abbildung 27 dargestellten Verläufe des MSE's in Abhängigkeit von Stichprobengröße und unterstelltem Parameter ähneln sehr stark denen der Gumbel-Copulas. In dieser Simulationsstudie kann festgestellt werden, wie auch schon von Weiß [2005, S. 42 ff.] propagiert, dass die mittleren quadratischen Fehler der Archimedischen-Copulas deutlich höher sind als die für parametrische Copulas.

Ergebnisse unter dem Gesichtspunkt der Stichprobengröße

In allen dargelegten Ergebnissen, lässt sich wie zu erwarten, eine Verbesserung der Güte mit steigendem Stichprobenumfang feststellen. Dies ist durch Eigenschaft asymptotischen Konvergenz der betrachteten Parameterschätzer bedingt. Darüber hinaus liefert der ML-Schätzer in allen Simulationen die geringsten MSE's und auch bezüglich Berechnungszeit ist der ML-Schätzer den restlichen sechs MD-Schätzern deutlich überlegen.

Die Verbesserung der Güte basierend auf dem MSE als Kriterium, kann gut mithilfe eines Box-Plots veranschaulicht werden.

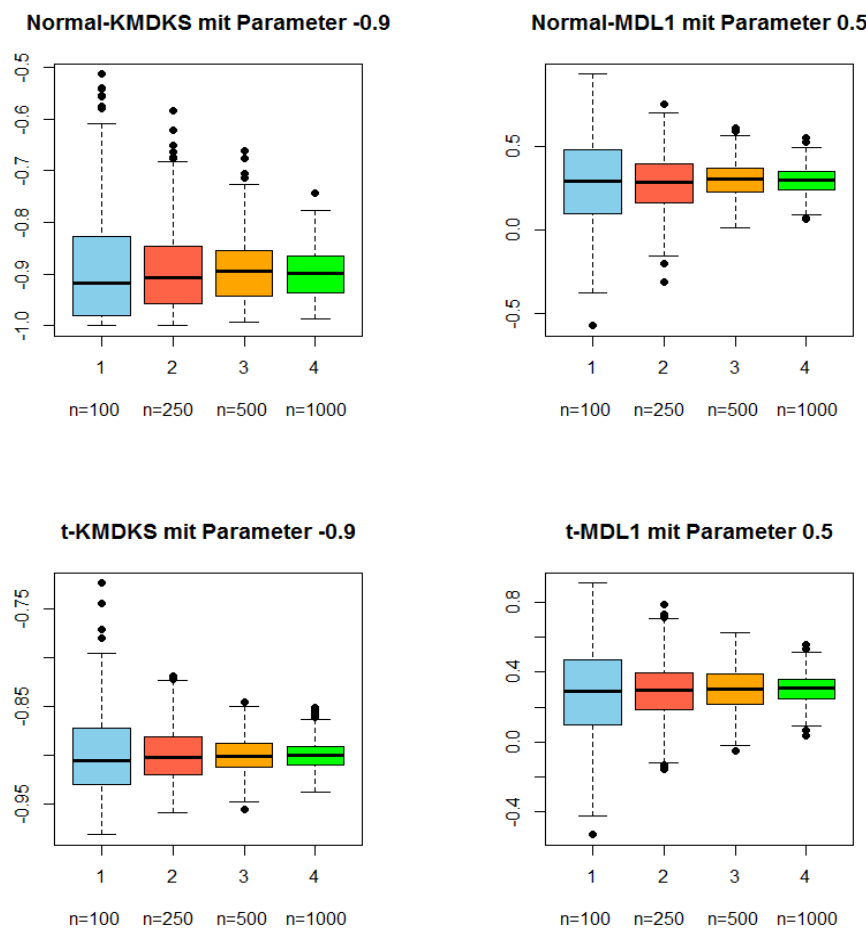


Abbildung 28: Boxplots für Normal- und t-Copulas und verschiedene Stichprobenumfänge

Anhand von Abbildung 28 lässt sich erkennen, dass sich der Interquartilsabstand, also dem Bereich in dem die mittleren 50% der Daten liegen, mit Zunahme des Stichprobenumfangs mehr als halbiert. Das bedeutet, dass die Streuung der Werte abnimmt und diese sich näher

um den Median anordnen. Zudem verringert sich die Größe der Antennen (*Whiskers*) mit zunehmender Stichprobengröße, was bedeutet, dass die Spannweite der Daten abnimmt.

Ergebnisse unter dem Gesichtspunkt der Parameterlage

In Abbildung 24 und 25 verläuft der MSE der sechs MD-Schätzer in Abhängigkeit vom Parameter θ glockenförmig. Dies ist möglicherweise mithilfe des Wertebereichs von θ erklärbar. Da die perfekt negative Abhängigkeit durch $\theta = -1$ und die perfekt positive durch $\theta = 1$ determiniert ist, können die Parameterwerte je näher sie $\theta = 0$ kommen absolut gesehen eine höhere Spannweite in den simulierten Daten aufweisen. Dies lässt sich empirisch für die Normal- und die t-Copula feststellen. Der Verlauf der ML-Schätzer ist im Vergleich zu den MD-Schätzern relativ robust gegenüber der Parameterlage.

Unter Betrachtung von Abbildung 26 und 27 im Vergleich zu Abbildung 24 und 25 lassen sich empirisch enorme Unterschiede in der Güte der MD-Schätzer feststellen. Wie zu sehen, ist der MSE für kleine Parameterwerte θ bei einer Stichprobengröße von 1000 bis etwa $\theta = 3$ relativ gering. Im Folgenden steigt der MSE recht stark an, was den Schluss zulässt, dass die MD-Schätzer für Archimedische Copulas eher als instabil und weniger robust einzuschätzen sind. In dieser Familie hat der Effekt der Parameterlage einen großen Einfluss auf die Güte der MD-Schätzer.

Ergebnisse unter dem Gesichtspunkt der zugrundeliegenden Teststatistik

Wie in den Abbildungen 24-27 zu sehen ist, scheinen sich die Schätzer in drei Güte Kategorien zu unterscheiden. Zum einen ist der ML-Schätzer zu nennen, der in allen Simulationsergebnissen die niedrigsten MSE-Werte hat. Die zweite Gruppe besteht aus den CvM- und L1-Teststatistiken, die über die Simulationsstudie hinweg ähnliche MSE-Werte aufweisen. Hingegen besteht die dritte Güte Kategorie nur aus den Kolmogorov-Smirnoff Teststatistiken. Eine Vermutung für die eher schlechte Güte kann sein, dass die KS-Statistik, wie in (4.12) und (4.25) definiert, immer nur den maximalen Betrag zwischen empirischer Copula bzw. Kendall transformierter Copula und der hypothetisch unterstellten Copula minimiert. Im Gegensatz dazu minimieren die CvM-Statistiken die quadratische Abweichung und die L1-Statistiken die betragsmäßige Abweichung **aller** Stichproben. Demnach werden alle Werte in der Stichprobe gleich gewichtet und unter der Maßgabe die Abstände minimiert. Hingegen gewichten die KS-Schätzer nur das minimale Maximum.

Dadurch kann offensichtlich eine bessere Approximation der CvM- und L1-simulierten Werte an die hypothetisch unterstellte Copula erfolgen als eine vergleichbare KS-Statistik.

Ergebnisse unter dem Gesichtspunkt der statistischen Signifikanz der betrachteten Simulationsergebnisse

Das Ziel der verwendeten Parameterschätzer $\hat{\theta}_S$ ist es, den wahren Parameter θ durch möglichst präzise Schätzer $\hat{\theta}$ gut zu approximieren. Man kann daher auch den Mittelwert der Schätzergebnisse als Schätzer für den wahren Parameter θ definieren. Es lässt sich der folgende zweiseitige Hypothesentest aufstellen

$$H_0: \hat{\theta}_S^n = \theta \text{ versus } H_1: \hat{\theta}_S^n \neq \theta$$

mit $n \in \{100; 250; 500; 1000\}$ und Schätzverfahren $S \in \{1,2,3,4,5,6,7\}$.

In dieser Arbeit wird die Normal-Copula stellvertretend für die Klasse der parametrischen Copulas betrachtet und die Gumbel-Copula als Mitglied der Archimedischen. Als Entscheidungskriterium wird der p-Wert eines zweiseitigen Hypothesentests zum Signifikanzniveau $\alpha = 5\%$ herangezogen. Demnach wird die Nullhypothese, dass der Mittelwert der Stichprobe gleich dem hypothetisch unterstellten ist, genau dann abgelehnt, wenn $p < \alpha$. Es zeigt sich in Abbildung 29 und 30, dass mit zunehmender

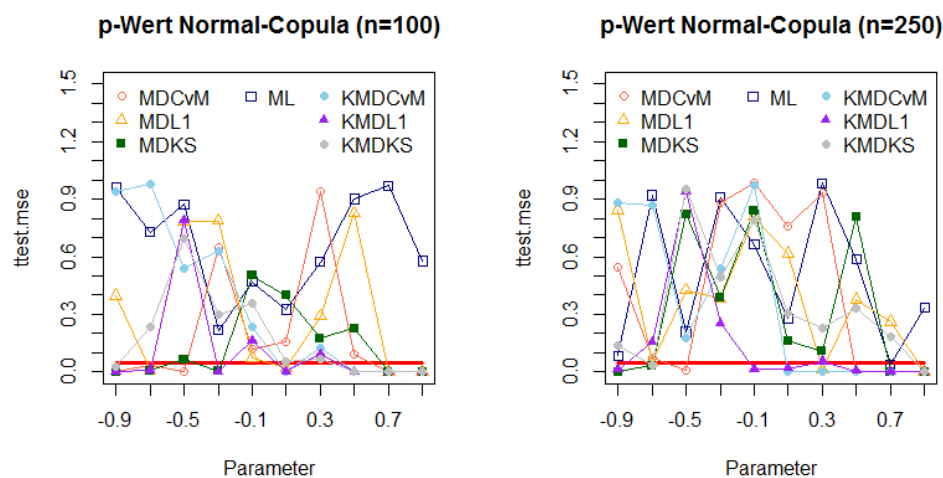


Abbildung 29: p-Werte eines zweiseitigen t-Test zum Signifikanzniveau 5% für die Normal-Copula mit $n=100$ und $n=250$

Stichprobengröße, die Nullhypothese immer schwieriger für ein Signifikanzniveau von 5%

abgelehnt werden kann. Zudem fällt auf, dass der ML-Schätzer für keine Stichprobengröße unter Maßgabe von $\alpha = 5\%$ abgelehnt werden kann.

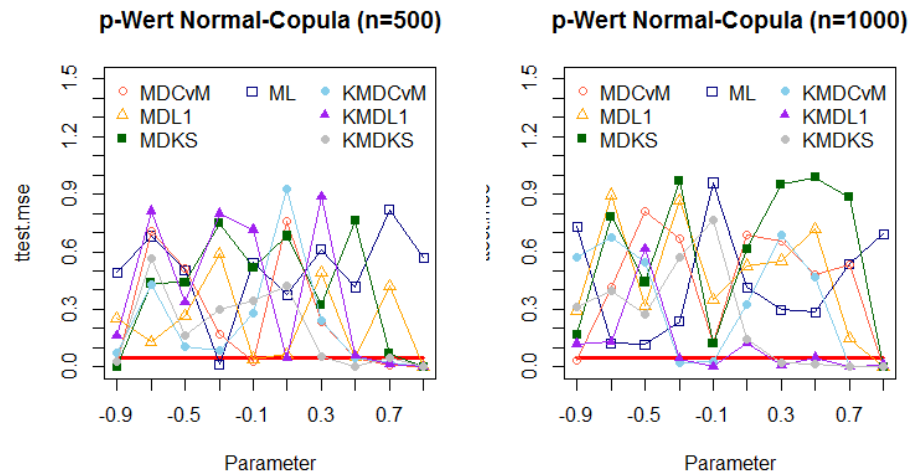


Abbildung 30: p-Werte eines zweiseitigen t-Test zum Signifikanzniveau 5% für die Normal-Copula mit $n=500$ und $n=1000$

Ein etwas differenziertes Bild ergibt sich unter Betrachtung des Verlaufs des p-Wertes für die Gumbel-Copula.

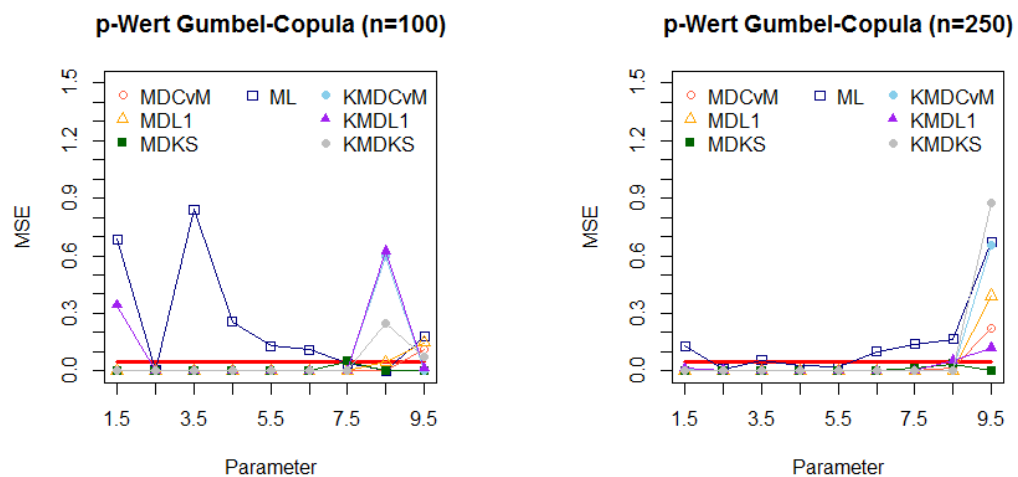


Abbildung 31: p-Werte eines zweiseitigen t-Test zum Signifikanzniveau 5% für die Gumbel-Copula mit $n=100$ und $n=250$

In Abbildung 31 und 32 lassen sich nur sehr begrenzte Aussagen bezüglich der Signifikanz treffen. Es ist zu erkennen, dass es mit zunehmender Stichprobengröße durchaus schwieriger ist, für den ML-Schätzer die Nullhypothese abzulehnen. Für die MD-Schätzer zeigt sich, dass praktisch keine Signifikanz der Mittelwerte vorliegt.

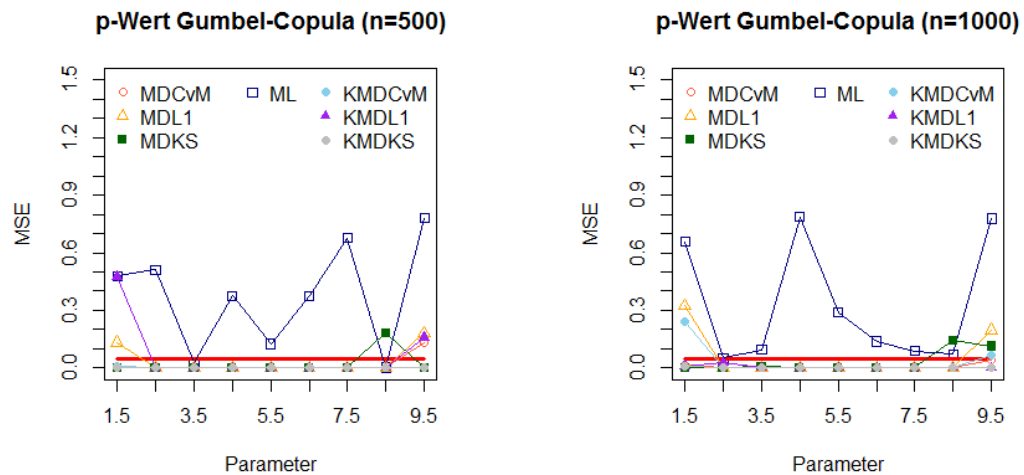


Abbildung 32: p-Werte eines zweiseitigen t-Test zum Signifikanzniveau 5% für die Gumbel-Copula mit $n=500$ und $n=1000$

Es ist anhand der Abbildungen 29-32 klar nachzuvollziehen, dass die Signifikanz der ML-Schätzer deutlich höher ist als die der MD-Schätzer. Des Weiteren wird die Signifikanz umso robuster je größer die Stichprobengröße ist.

Zusammenfassung der Ergebnisse

Die Ergebnisse der hier durchgeführten Simulationsstudie liefern insbesondere für die Normal- und t-Copula ähnliche Ergebnisse wie in Weiß [2005, S. 42 ff.].

Die zentrale Schlussfolgerung lautet, dass der ML-Schätzer was Berechnungszeit und Güte anbelangt, der effizienteste der hier vorgestellten Schätzer ist. Es scheint daher nicht sinnvoll zu sein die verfahrens- und rechenaufwändigeren MD-Schätzer zu benutzen.

Jedoch muss abstrahiert werden, denn speziell für die Normal- und t-Copulas konnte eine relativ passable Güte, insbesondere durch Erhöhung der Stichprobengröße erreicht werden. Jedoch bräuchte man einen deutlich größeren Stichprobenumfang um annähernd in die Gütebereiche des ML-Schätzers zu gelangen. Da dies eine umso großvolumigere Berechnung erfordert, die sehr zeitintensiv ist, gibt es weder ökonomische noch rationale

Gründe diese Schätzverfahren zu verwenden. Darüber hinaus schwankt die Güte der MD-Schätzer insbesondere bei den Archimedischen-Copulas recht stark in Abhängigkeit von der Parameterlage. Zur Unterstützung dieser Aussagen können die Simulationsergebnisse wie geschehen durch einen Hypothesentest gestärkt werden.

VI Literaturverzeichnis

- Alegre, A., Roch, O. (2006) Testing the bivariate distribution of daily equity returns using copulas. An application to the Spanish stock market. *Comput. Stat. Data Anal.* 51(2): 1312–1329
- Beaudoin D., Genest C., Rémillard B. (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insur Math Econ* 44:199–213
- Beck, A., Lesko, M. (2006) Copula-Funktionen zur Ermittlung des Gesamtbankrisikoprofils. *Betriebswirtschaftliche Blätter* 05/2006: 289-293
- Beck, A., Lesko, M., Schlottmann, F., Wimmer, K. (2006) Copulas im Risikomanagement. *Zeitschrift für das gesamte Kreditwesen* 14/2006: 29-33
- Deheuvels, P. (1978) Caractérisation complète des Lois Extrêmes Multivariées et de la Convergence des Types Extrêmes. *Pub l’Institut Stat l’Université Paris* 23:1–36
- Demarta, S., McNeil, A.J. (2004) The t Copula and Related Copulas.: 2-5
- De Kort, J. (2007) Modeling tail dependence using copulas — literature review.: 22-27
- De Luca, G., Riveccio, G. (2012) Multivariate tail dependence coefficients.
- Dodge, Y. (1987) An introduction to &-norm based statistical data analysis.: 239-240
- Dos Santos Silva, R., Freitas Lopes, H. (2008) Copula, marginal distributions and model selection: a Bayesian note.
- Embrechts, P., McNeil, A., Straumann, D. (1998) Correlation and Dependency in Risk Management: Properties and Pitfalls.: 176-223
- Embrechts, P., Frey, R., McNeil, A.J. (2005) *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press
- Fermanian, J. (2005) Goodness-of-fit tests for copulas. *J. Multivar. Anal.* 95:119–152
- Frey, R., McNeil, A.J., Nyfeler, M.A. (2001) Copulas and Credit Models.: 2-3

- Genest, C., Ghoudi, K., Rivest, L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82(3): 543–552
- Hartung, J., Elpelt, B. (2007) *Multivariate Statistik: Lehr- und Handbuch der angewandten Statistik*.: 40f.
- Joe H. (1997) *Multivariate Models and Dependence Concepts*, Chapman & Hall/CRC.: 33f.
- Jondeu, E., Poon, S.-H., Rockinger, M. (2007): *Financial Modeling Under Non Gaussian-Distributions*, Springer Verlag London: 240-250
- Klaassen, A.J., Wellner, J.A. (1997) Efficient Estimation in the Bivariate Normal Copula Model: Normal Margins Are Least Favourable. *Bernoulli*, Vol. 3, No. 1. (Mar., 1997): 55-77.
- Mahfoud, M. (2012) Bivariate Archimedean copulas: an application to two stock market indices.: 14-22
- Mashal, R., Zeevi, A. (2002) Beyond correlation: extreme co-movements between financial assets. Unpublished, Columbia University
- Meyer, C. (2009) *The Bivariate Normal Copula*: 2-7
- Nelsen, R. B. (2006) *An Introduction to Copulas*. Springer New York: 109 ff.
- Nelsen, R. B. (2015) Properties and applications of copulas: A brief survey. Department of Mathematical Sciences Lewis & Clark College: 1-6
- Peng, L., Wang, R. (2014) Interval estimation for bivariate t-copulas via Kendall's tau: 1-10
- Roussas, G. G. (1997) *A Course in Mathematical Statistics*. Academic Press, 2 edition: 242-243
- Schmid, F., Trede, M. (2006) *Finanzmarktstatistik*. Springer Verlag Berlin: 97-110
- Schmidt, R. (2002) Tail dependence for elliptically contoured distributions. *Math. Methods of Operations Research*, 55: 301-327

VII Anhang

Plots

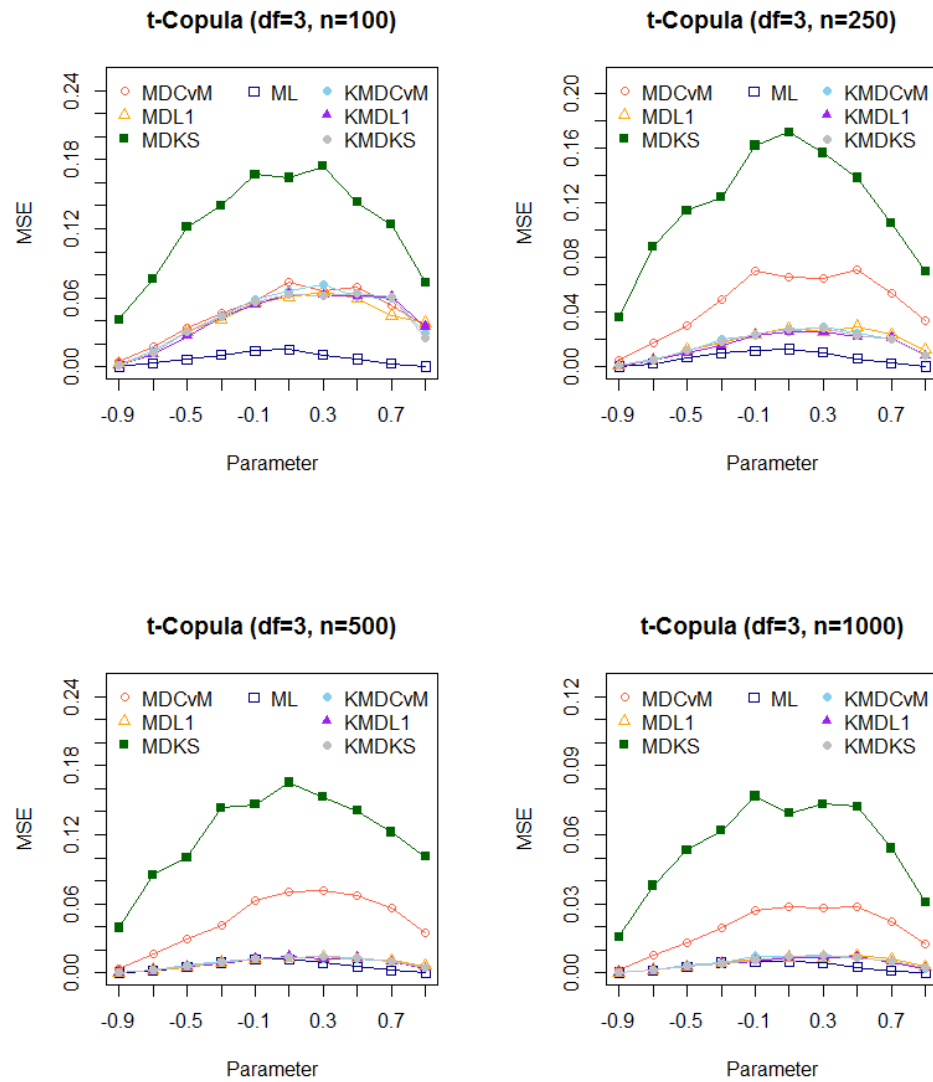


Abbildung 33: Darstellung des MSE für ML-Schätzer, MD-Schätzer basierend auf dem empirischen Prozess und MD-Schätzer basierend auf Kendall's Transformation mit ansteigender Stichprobengröße für die t-Copula mit Freiheitsgrad 3

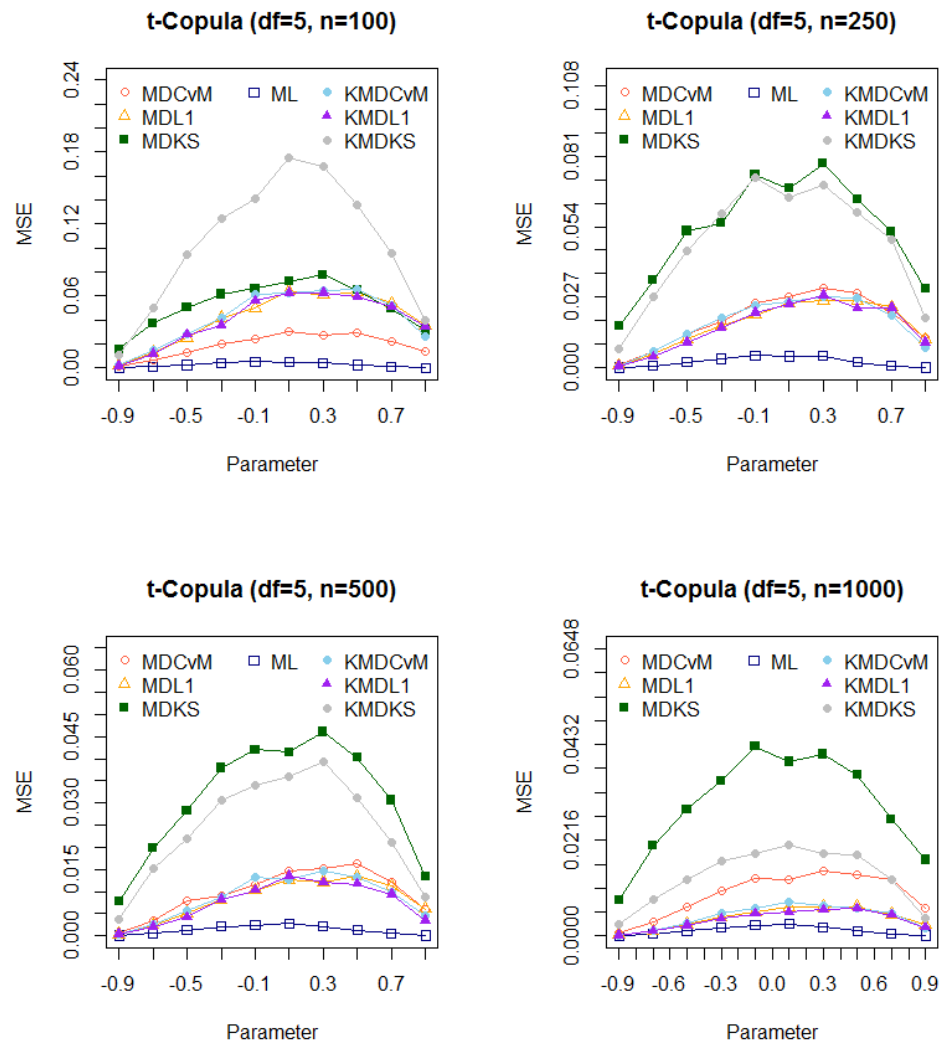


Abbildung 34: Darstellung des MSE für ML-Schätzer, MD-Schätzer basierend auf dem empirischen Prozess und MD-Schätzer basierend auf Kendall's Transformation mit ansteigender Stichprobengröße für die t-Copula mit Freiheitsgrad 5