

Ein Fluktuationstest auf konstante Korrelation

Dissertation
zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund

Der Fakultät Statistik der Technischen Universität Dortmund
vorgelegt von
Dominik Wied

Tag der mündlichen Prüfung: 18.12.2009
Erstgutachter: Prof. Dr. Walter Krämer
Zweitgutachter: JProf. Dr. Uwe Ligges
Drittgutachter: Prof. Dr. Herold Dehling

Inhaltsverzeichnis

1	Einleitung und Motivation	1
2	Mathematische Grundlagen	6
2.1	Schwache Konvergenz in metrischen Räumen	6
2.2	Die verallgemeinerte Delta-Methode	23
3	Die Teststatistik und ihre asymptotische Nullverteilung	28
3.1	Die Teststatistik	28
3.2	Die asymptotische Nullverteilung	33
4	Lokale Güte	54
4.1	Allgemeine Überlegungen	54
4.2	Konstruktionsbedingte Ablehneigenschaften	58
5	Methodenvergleich mit anderen Tests	61
6	Simulationen für endliche Stichproben	65
6.1	Die Nullverteilung	65
6.2	Die Güte	66
6.3	Weitere Simulationsergebnisse	71
7	Anwendung auf Aktienrenditen	72
8	Ausblick	76
A	Einige wichtige Sätze	81
	Symbolverzeichnis	84
	Abbildungsverzeichnis	86
	Literaturverzeichnis	87

1 Einleitung und Motivation

Im Sommer 2007 begann mit der US-Immobilienkrise eine große Weltwirtschaftskrise. Sie äußerte sich zunächst in Verlusten bei Unternehmen der Finanzbranche, aber seit September 2008 mit der Insolvenz von Lehman Brothers auch dramatisch in der Realwirtschaft. So kam es zum Beispiel im 4. Quartal 2008 in Deutschland zu einem Einbruch der Kraftfahrzeugexporte im Vergleich zum 4. Quartal 2007 um 20 Prozent, obwohl es in den ersten beiden Quartalen noch eine Steigerung gegeben hatte. Das Ausmaß ist bisher noch nicht abzusehen. Es gab und gibt eine Phase des weltweiten wirtschaftlichen Abschwungs; die Finanzmärkte ziehen sich anscheinend gegenseitig nach unten.

Dies ist aus ökonometrischer Sicht nicht leicht zu erklären oder zu quantifizieren. Eine spannende Frage in diesem Zusammenhang lautet: Kann der so genannte „Diversification Meltdown“, das heißt die Zunahme der Korrelationen von Renditen in wirtschaftlichen Abschwungphasen im Vergleich zu Aufschwungphasen, empirisch bestätigt werden? Wie können wir am besten testen, ob die Korrelationen konstant oder nicht konstant sind?

Diese Fragen wurde in der ökonometrischen Literatur schon oft behandelt und sind unter anderem relevant für die Portfoliotheorie, die auf Markowitz (1952) zurückgeht. Nach Markowitz sind die Diversifikationseffekte im Portfolio, also die Reduktion von Marktrisiken durch Kombination mehrerer Anlagen, umso niedriger, je stärker die einzelnen Renditen miteinander korrelieren. Die Existenz des „Diversification Meltdown“ bedeutet also, dass die Diversifikationseffekte gerade dann nicht mehr da sind, wenn man sie wirklich bräuchte.

Im Folgenden soll ein kurzer Überblick über Ergebnisse aus der ökonometrischen Literatur hinsichtlich der Frage konstanter Korrelationen bei Renditen gegeben werden. Die Mehrzahl der Autoren findet bei den Studien empirische Anzeichen für den „Diversification Meltdown“, so etwa Erb et al. (1994) bei den monatlichen Renditen der Aktienindizes der damaligen G7-Länder im Zeitraum von 1970 bis 1993. Longin und Solnik (1995) finden Anzeichen instabiler Korrelationen bei Aktienindizes von sieben (teilweise anderen) Ländern im Zeitraum von 1960 bis 1990. Ansteigender Korrelatio-

nen in der Asien-Krise 1997 im Vergleich zur Zeit davor finden Chakrabarti und Roll (2002). Goetzmann et al. (2005) untersuchen in dem vergleichsweise langen Zeitraum von 1872 bis 2000 Monatsrenditen und kommen ebenfalls zu dem Schluss, dass die Korrelationsstruktur instabil war. Am höchsten schienen die Korrelationen in Phasen wirtschaftlicher Integration, das heißt bei wirtschaftlichen Zusammenschlüssen mehrerer Länder, zu sein.

Kaplanis (1988) untersucht Aktienindizes aus zehn Ländern von 1967 bis 1982 und kommt zu dem Schluss, dass die Korrelationsstruktur stabil blieb, wohingegen die Kovarianzstruktur schwankte. Meric und Meric (1989) betrachten Monatsrenditen von 1973 bis 1987 und untersuchen unter anderem Saisonmerkmale in den Korrelationen. Hier zeigt sich, dass diese über die Jahre im Zeitraum von September bis Mai konstant, aber im Zeitraum von Mai bis September eher nicht konstant blieben. Fischer (2007) entdeckt instabile Korrelationen bei Aktienrenditen der Automobilindustrie, wohingegen die Korrelationen bei Aktienrenditen der Telekommunikationsindustrie eher konstant blieben.

Ragea (2003) kommt zu einem Ergebnis, das nicht mit der Theorie des „Diversification Meltdown“ übereinstimmt. Er betrachtet tägliche Aktienrenditen im Zeitraum von 1999 bis 2002. Trotz großer Hektik auf den Finanzmärkten - etwa in Folge der Anschläge vom 11. September - schien die Korrelation konstant zu bleiben.

Viele der Untersuchungen basieren auf statistischen Tests, in deren Nullhypothese die Korrelation konstant ist und die auf einen als bekannt unterstellten festen Bruchpunkt in der Korrelation testen. Die Beobachtungsreihe wird vor dem Test in mehrere Abschnitte aufgeteilt und es wird vorausgesetzt, dass die Korrelation in den einzelnen Teilreihen konstant ist. Solch einen χ^2 -Test konstruiert etwa Jennrich (1970).

Fischer (2007) geht diesen Weg nicht, sondern testet darauf, ob die Korrelationen mit gewissen trigonometrischen funktionalen Zusammenhängen schwanken. Dieser Ansatz ist flexibler, aber es gelingt ihm nicht, die (exakte oder asymptotische) Verteilung der Teststatistiken in geschlossener Form anzugeben.

Wir schlagen in dieser Dissertation einen Fluktuationstest auf konstante Korrelation vor, bei dem die möglichen Bruchpunkte im Vorfeld nicht bekannt sein müssen und

bei dem wir die asymptotische Verteilung der Teststatistik angeben können. Der Test verwirft die Nullhypothese konstanter Korrelation, wenn die empirischen Korrelationen im Zeitverlauf zu sehr schwanken.

Immer wieder wird in Artikeln die Frage diskutiert, inwiefern ein univariates Abhängigkeitsmaß überhaupt in der Lage ist, ökonomische Sachverhalte sinnvoll zu beschreiben, vgl. etwa Forbes und Rigobon (2002). Longin und Solnik (2002) stellen heraus, dass es bei Aktienrenditen häufig eine hohe Randabhängigkeit gibt, die in gewöhnlichen Verteilungsfamilien nur schwer berücksichtigt werden kann. Boyer et al. (1999), Andreou und Ghysels (2003) und Campbell et al. (2008) thematisieren das Problem der korrekten Unterscheidung von bedingter und unbedingter Korrelation. King et al. (1994) stellen heraus, dass Korrelationsschwankungen zwischen verschiedenen Märkten hauptsächlich durch Änderungen in unbeobachtbaren Variablen begründet sind. Wir diskutieren dies jedoch nicht weiter und konzentrieren uns auf die unbedingte Korrelation zwischen zwei Variablen X und Y , die geschätzt wird durch den Bravais-Pearson-Korrelationskoeffizienten, da selbst hierfür die Herleitung eines Fluktuationstests nicht trivial ist.

Die Prüfgröße unseres Tests ist

$$Q_T(X, Y) = \hat{D} \max_{2 \leq j \leq T} \frac{j}{\sqrt{T}} |\hat{\rho}_j - \hat{\rho}_T|,$$

wobei $\hat{\rho}_j$ der Bravais-Pearson-Korrelationskoeffizient basierend auf den ersten j und $\hat{\rho}_T$ der Bravais-Pearson-Korrelationskoeffizient basierend auf allen T Beobachtungen ist. Die Größe \hat{D} ist ein kompliziert aufzuschreibender Schätzer. Es ist zu erkennen, dass die Teststatistik groß wird, wenn die empirischen Korrelationen im Zeitverlauf stark schwanken; dann soll der Test die Nullhypothese konstanter Korrelation ablehnen.

Wir leiten die asymptotische Verteilung dieser Teststatistik her. Dies ist erheblich einfacher als die Herleitung einer exakten Verteilung für endliches T . Zur Berechnung verwenden wir die Theorie der Prozesskonvergenz, die etwa in Billingsley (1968) dargestellt ist.

Ein wichtiges Hilfsmittel ist ein multivariates Invarianzprinzip und die verallgemeinerte Delta-Methode, die in der von uns benötigten Form unseres Wissens nach in der

Literatur noch nicht betrachtet wurde. Sie basiert auf der von uns entwickelten verallgemeinerten Hadamard-Differenzierbarkeit zwischen Funktionenräumen, die sich von der gewöhnlichen Hadamard-Differenzierbarkeit unterscheidet. Es gelingt in der Dissertation, für eine spezielle Funktionenklasse eine allgemeine Formel für die Hadamard-Ableitung herzuleiten; diese wird im Beweis der Nullverteilung angewandt. Wichtig ist auch die Verwendung des Raums $D^+(J)$, der Raum der durch eine Konstante nach unten beschränkten Càdlàg-Funktionen auf einem abgeschlossenen Intervall J . Der Raum fand bisher keine Berücksichtigung in der Literatur. Der Raum der Càdlàg-Funktionen wird mit der Supremumsnorm versehen, um die Hadamard-Differenzierbarkeit verwenden zu können. Damit ist er nicht separabel, was einige Messbarkeitsprobleme mit sich bringt. Diese werden durch Verwendung einer kleineren σ -Algebra als der Borelschen σ -Algebra behoben. Es zeigt sich, dass die asymptotische Verteilung die Verteilung des Maximums des Absolutbetrags einer Brownschen Brücke auf $[0, 1]$ ist. Für deren Verteilungsfunktion gibt es eine explizite Darstellung.

Zusätzlich zu den bereits genannten hat unser Test noch weitere Vorteile. So müssen die Zufallsvektoren nicht unabhängig identisch verteilt sein, gewisse Abhängigkeiten und Schwankungen der ersten und zweiten Momente von X und Y sind erlaubt. Die Zufallsvektoren sind near-epoch dependent (NED) über einem α - oder ϕ -mischendem Dreiecksschema; ein Szenario, das bei GARCH-Prozessen realistisch ist. Die Momente müssen jeweils um eine Konstante schwanken. Wenn das Verhältnis der Varianzen konstant ist - ein bei Aktienrenditen durchaus realistisches Szenario - ist eine stärkere Schwankung erlaubt. Außerdem ist unser Test nichtparametrisch und benutzt daher insbesondere keine Normalverteilungannahme. Somit ist er auch, zum Beispiel, bei der t -Verteilung anwendbar. Dies ist erfreulich, weil die t -Verteilung bei der Modellierung von Aktienrenditen wegen den im Vergleich zur Normalverteilung schwereren Rändern sehr beliebt ist. In endlichen Stichproben hat der Test gute Eigenschaften, sowohl hinsichtlich des empirischen Niveaus als auch hinsichtlich der Güte. Für die Güte leiten wir zudem einige analytische Ergebnisse her.

Die Dissertation ist wie folgt aufgebaut:

Kapitel 2 präsentiert die mathematischen Grundlagen, die für die Herleitung der Null-

verteilung erforderlich sind. Wir schaffen auf Basis verschiedener Literatur ein theoretisches Fundament und leiten einige neue Aspekte her, die in der Literatur bisher unbekannt sind. Im 3. Kapitel beschreiben wir die Teststatistik mit den notwendigen Annahmen und leiten ausführlich die asymptotische Nullverteilung her. Im 4. Kapitel geht es um lokale Güte, das heißt, wir berechnen die theoretische Güte für verschiedene Alternativen. Wir gehen auf den Fall ein, dass das Verhältnis der Varianzen von X und Y konstant ist, aber auch darauf, dass die Varianzen nicht zu stark schwanken dürfen. Das 5. Kapitel vergleicht unseren Test mit einigen anderen Tests auf konstante Korrelation aus der Literatur. Wir schauen, nach welchem Prinzip diese Tests funktionieren und vergleichen die Annahmen. Das Verhalten des Tests in endlichen Stichproben überprüfen wir durch Simulationsstudien in Kapitel 6. Es geht um das empirische Niveau, um Güteeigenschaften und um einige andere Eigenschaften der Teststatistik. Die Simulationsergebnisse werden wenn möglich mit theoretischen Ergebnissen untermauert. Das 7. Kapitel wendet den Test auf die täglichen Aktienrenditen von BASF und BMW von 2003 bis 2008 an. Es zeigt sich, dass der Effekt der Weltwirtschaftskrise gut zu erkennen ist. Nach der Insolvenz von Lehman Brothers im September 2008 war die Korrelation der beiden Renditen erkennbar größer als zuvor, dies wird durch unseren Test auch bestätigt und spricht demnach für den „Diversification Meltdown“. Das 8. Kapitel beinhaltet einen Ausblick. Hier diskutieren wir insbesondere offene Fragen und Probleme, die sich bei unserer Herleitung ergeben haben.

2 Mathematische Grundlagen

In diesem Kapitel stellen wir mathematische Grundlagen bereit, die nötig sind, um die asymptotische Nullverteilung der Teststatistik herzuleiten. Teilweise definieren und erklären wir in der Literatur bereits bekannte Konzepte, die bei der Herleitung wichtig sind. Insbesondere vereinheitlichen wir die Notation und klären Zusammenhänge. Es soll nicht darum gehen, die Thematik im Sinne eines Lehrbuches ausführlich darzustellen, sondern der Schwerpunkt soll auf den Konzepten liegen, die wir später benötigen. Manche Aspekte, wie etwa die verallgemeinerte Delta-Methode, sind komplett neu, andere müssen auf die uns vorliegende Situation angepasst werden, dies betrifft etwa die Wahl von geeigneten σ -Algebren.

In der gesamten Arbeit sind Funktionen als Punkte aufzufassen, bei jeder Konvergenzbetrachtung geht T gegen unendlich, falls nichts anderes gesagt wird. Viele der Sätze sind für Folgen von Zufallsvariablen X_1, X_2, \dots definiert, gelten aber genauso auch für Schemata $(X_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$. Wir erwähnen jeweils explizit, wenn bei bestimmten Sätzen tatsächlich ein Schema gebraucht wird.

2.1 Schwache Konvergenz in metrischen Räumen

Definition 2.1 (Funktionenräume) *Sei J ein abgeschlossenes Intervall in \mathbb{R} . Dann ist*

$$C(J) := \{x : J \rightarrow \mathbb{R}, x \text{ ist stetig.}\}$$

$$D(J) := \{x : J \rightarrow \mathbb{R},$$

$$x(t+) := \lim_{s > t, s \rightarrow t} x(s) \text{ existiert in } \mathbb{R} \text{ und } x(t+) = x(t) \forall t \in J,$$

$$x(t-) := \lim_{s < t, s \rightarrow t} x(s) \text{ existiert in } \mathbb{R} \forall t \in J\}$$

$$D^+(J) := \{x \in D(J), \exists k > 0 \text{ sodass } x(t) \geq k \forall t \in J\},$$

das heißt, $C(J)$ ist der Raum der stetigen Funktionen auf J , $D(J)$ ist der Raum der rechtsstetigen Funktionen auf J , bei denen der linksseitige Grenzwert existiert (auch

Càdlàg-Funktionen genannt), $D^+(J)$ ist der Raum der Funktionen aus $D(J)$, die durch eine Konstante $k > 0$ nach unten beschränkt sind.

Bemerkungen In unseren Anwendungen ist meist $J = [0, 1]$ oder $J = [\epsilon, 1]$ für ein $\epsilon > 0$, die Schreibweise für $J = [0, 1]$ ist dann etwa $D[0, 1]$. Offensichtlich ist $C(J)$ eine Teilmenge von $D(J)$, da jede stetige Funktion auch rechtsstetig ist und dann $x(t+) = x(t-) = x(t) \forall t \in J$ gilt. Die Mengeninklusion ist echt, da etwa die Funktion $x(t) = I_{[\frac{1}{2}, 1]}(t)$ in $D[0, 1]$, aber nicht in $C[0, 1]$ liegt. Genauso ist $D^+(J)$ eine echte Teilmenge von $D(J)$, da etwa die Funktion $x(t) = 0, t \in [0, 1]$ stetig ist und damit in $D(J)$ liegt, allerdings nicht in $D^+(J)$ liegt. $D^+(J)$ ist außerdem eine echte Teilmenge der Menge aller positiven Funktionen aus $D(J)$; so ist etwa die Funktion $x(t) = 1 - t, t \in [0, 1], x(1) = 1$ aus $D(J)$ strikt positiv, liegt aber nicht in $D^+(J)$.

Zusammenfassend ist $D(J)$ Obermenge aller drei Mengen.

Wir definieren, dass eine beliebige Funktion auf J eine Unstetigkeitsstelle erster Art in einem beliebigen $t \in J$ hat, falls $x(t+)$ und $x(t-)$ existieren und falls $x(t)$ dazwischen liegt. Jede Unstetigkeitsstelle einer Funktion aus $D(J)$ ist erster Art; die Forderung $x(t) = x(t+)$ ist eine Konvention. Es wäre genauso auch möglich, $x(t) = x(t-)$ zu fordern.

Alle drei Räume werden in der Dissertation benötigt. In $C(J)$ werden gewisse Grenzwerte liegen, in $D(J)$ findet die Prozesskonvergenz statt und einige der betrachteten Funktionen aus $D(J)$ müssen in $D^+(J)$ liegen, um Konvergenz sicherzustellen (etwa im Beweis von Lemma 3.4). In der Literatur wird unseres Wissens nach der Raum $D^+(J)$ nicht betrachtet und meistens ist auch $J = [0, 1]$. Falls J ein abgeschlossenes Teilintervall von $[0, 1]$ ist, lassen sich topologische und maßtheoretische Aussagen jedoch analog übertragen, das wird später noch deutlich.

Wir halten zunächst fest, dass jede Funktion aus $D(J)$ und damit aus $C(J)$ und $D^+(J)$ nach oben beschränkt ist:

Lemma 2.2 *Sei $x \in D(J)$. Dann gibt es ein $K < \infty$, sodass $|x(t)| \leq K \forall t \in J$.*

Beweis. Wäre x nicht nach oben beschränkt, gäbe es eine Folge t_T , sodass

$\lim_{T \rightarrow \infty} x(t_T) = \infty$. Wegen der Kompaktheit von J besitzt t_T eine konvergente Teilfolge $t_{T'}$ mit Grenzwert t_0 und $\lim_{T \rightarrow \infty} x(t_{T'}) = \infty$. Wegen der Rechtsstetigkeit von x gibt es eine weitere Teilfolge $t_{T''} < t_0$ mit $\lim_{T \rightarrow \infty} x(t_{T''}) = \infty$. Dann kann aber der linksseitige Grenzwert in t_0 nicht existieren. \square

Wir definieren auf $D(J)$ bzw. $D^+(J)$ die Supremumsnorm:

$$\|x\|_{D(J)} := \sup_{t \in J} |x(t)| \text{ für } x \in D(J), \|x\|_{D^+(J)} := \sup_{t \in J} |x(t)| \text{ für } x \in D^+(J).$$

Ist

$$\mathbb{G} := \mathbb{H}_1 \times \dots \times \mathbb{H}_k \text{ (} k\text{-mal, } k \geq 1, \mathbb{H}_i \in \{C[\epsilon, 1], D[\epsilon, 1], D^+[\epsilon, 1], \epsilon \geq 0\}) \quad (1)$$

der Produktraum, wobei die einzelnen Komponenten entweder $C[\epsilon, 1]$, $D[\epsilon, 1]$ oder $D^+[\epsilon, 1]$ sind, so ist für $x = (x_1, \dots, x_k) \in \mathbb{G}$

$$\|x\|_{\mathbb{G}} := \sup_{1 \leq i \leq k} \|x_i\|_{\mathbb{H}_i}$$

die mehrdimensionale Supremumsnorm.

Wir verwenden die Notation $D[\epsilon, 1]^l := D[\epsilon, 1] \times \dots \times D[\epsilon, 1]$ (l mal), dies geht analog für $C[\epsilon, 1]$.

Eine Projektion π_i ist eine Abbildung, die einer Funktion aus $D(J)$ den Wert an der Stelle $i \in J$ zuordnet. Dies funktioniert auch vektorwertig und wir gelangen zur

Definition 2.3 Seien $t_1 < \dots < t_n$ Punkte in J . Dann heißt die Abbildung

$$\begin{aligned} \pi_{t_1, \dots, t_n} : D(J) &\rightarrow \mathbb{R}^n, \\ x &\rightarrow (x(t_1), \dots, x(t_n)) \end{aligned}$$

n -dimensionaler Projektionsvektor.

Bemerkung Bei Verwendung der Supremumsnorm sind Projektionen grundsätzlich stetig. Konvergiert nämlich eine beliebige Folge x_T in der Supremumsnorm gegen x , so konvergiert auch $\pi_{t_1, \dots, t_n}(x_T) = (x_T(t_1), \dots, x_T(t_n))$ in der Supremumsnorm gegen $\pi_{t_1, \dots, t_n}(x) = (x(t_1), \dots, x(t_n))$.

Im Folgenden seien die genannten Räume immer mit diesen Normen versehen. Bei einem allgemeineren normierten Raum \mathbb{D} sei $\|\cdot\|_{\mathbb{D}}$ die zugehörige Norm. Ein normierter Raum wird automatisch auch ein metrischer Raum mit der durch die Norm erzeugten Metrik.

In der Arbeit benötigen wir drei stochastische Konvergenzbegriffe, einmal schwache Konvergenz (das heißt Konvergenz in Verteilung), einmal Konvergenz in Wahrscheinlichkeit und einmal fast sichere Konvergenz. Wir beziehen uns auf einen allgemeinen normierten Raum \mathbb{D} (mit einer σ -Algebra \mathfrak{D}) und wenden dies später an auf \mathbb{G} , versehen mit der Supremumsnorm, und den \mathbb{R}^k , versehen mit der euklidischen Norm. Zur konkreten Wahl der σ -Algebra folgen später einige Ausführungen. Manchmal genügt die Annahme, dass \mathbb{D} nur ein metrischer Raum ist, dies erwähnen wir dann explizit. Zunächst geben wir die Definition der Verteilung.

Definition 2.4 (*Billingsley (1968, S. 22)*) Sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit Werten in einem allgemeinen metrischen Raum \mathbb{D} , versehen mit einer σ -Algebra \mathfrak{D} . Die Verteilung von X ist das Wahrscheinlichkeitsmaß \mathbb{P}_X auf $(\mathbb{D}, \mathfrak{D})$, definiert durch

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\omega : X(\omega) \in A) = \mathbb{P}(X \in A), \quad A \in \mathfrak{D}.$$

Hiermit definieren wir jetzt die schwache Konvergenz.

Definition 2.5 (*Billingsley (1968, S. 23)*) Sei X_1, X_2, \dots eine Folge von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit Werten in einem allgemeinen metrischen Raum \mathbb{D} , versehen mit einer σ -Algebra \mathfrak{D} . Die Folge X_T konvergiert in Verteilung gegen X ,

$$X_T \rightarrow_d X,$$

falls die Verteilungen \mathbb{P}_{X_T} der X_T schwach gegen die Verteilung \mathbb{P}_X von X konvergieren. Dies ist erfüllt, falls

$$\int_{\mathbb{D}} f d\mathbb{P}_{X_T} \rightarrow \int_{\mathbb{D}} f d\mathbb{P}_X$$

für jede stetige und beschränkte Funktion $f : \mathbb{D} \rightarrow \mathbb{R}$ gilt.

Äquivalent zu dieser Definition ist unter anderem, dass

$$\limsup_{T \rightarrow \infty} \mathbb{P}(X_T \in A) \leq \mathbb{P}(X \in A)$$

für jede abgeschlossene Menge A gilt.

Die Konvergenz in Wahrscheinlichkeit benötigt den Begriff der Verteilung nicht, aber dafür die Metrik in dem metrischen Raum.

Definition 2.6 (*Billingsley (1968, S. 24)*) *Unter den Voraussetzungen von Definition 2.5 sei d die Metrik in \mathbb{D} und sei $d(X_T, a)$ eine Zufallsvariable für ein nichtstochastisches $a \in \mathbb{D}$. Dann konvergiert die Folge X_T in Wahrscheinlichkeit gegen a ,*

$$X_T \rightarrow_p a,$$

falls

$$\mathbb{P}(d(X_T, a) \geq \epsilon) \rightarrow 0$$

für jedes $\epsilon > 0$ gilt.

Bemerkung Für $\mathbb{D} = \mathbb{G}$ versehen mit der Supremumsnorm bedeutet Konvergenz in Wahrscheinlichkeit, dass X_T in der Supremumsnorm in Wahrscheinlichkeit konvergiert.

Definition 2.7 (*Davidson (1994, S. 178)*) *Unter den Voraussetzungen von Definition 2.5 konvergiert die Folge X_T \mathbb{P} -fast sicher gegen a ,*

$$X_T \rightarrow_{f.s.} a,$$

falls

$$\lim_{T \rightarrow \infty} X_T(\omega) = a$$

für \mathbb{P} -fast alle $\omega \in \Omega$ gilt.

Die fast sichere Konvergenz impliziert Konvergenz in Wahrscheinlichkeit und diese wiederum Konvergenz in Verteilung. Die Umkehrungen gelten im Allgemeinen nicht,

aber Konvergenz in Verteilung gegen eine Konstante impliziert die Konvergenz in Wahrscheinlichkeit.

Wir betrachten nun allgemeine Zufallsfunktionen $X : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (\mathbb{G}, \mathfrak{G})$. Es handelt sich hierbei um k -dimensionale Zufallsfunktionen, die für festes $t \in \mathbb{R}^k$ einen k -dimensionalen Zufallsvektor darstellen.

Wichtig ist die Frage der Messbarkeit. Wir verwenden für \mathfrak{G} das Produkt der σ -Algebren $\mathfrak{H}_1 \times \dots \times \mathfrak{H}_k$, wobei \mathfrak{H}_i jeweils die σ -Algebra der offenen Bälle bezüglich der durch die Supremumsnorm erzeugten Metrik auf $C[\epsilon, 1]$, $D[\epsilon, 1]$ oder $D^+[\epsilon, 1]$ ist. Diese entspricht der Projektions σ -Algebra, die (o.B.d.A. für $D[\epsilon, 1]$) definiert ist als $\sigma(H)$ mit

$$H = \{\pi_{t_1, \dots, t_k}^{-1}(B) \subseteq D[\epsilon, 1] : B \in \mathfrak{B}^k, t_1, \dots, t_k \in [\epsilon, 1], k \in \mathbb{N}\},$$

wobei \mathfrak{B} die Borelsche σ -Algebra auf \mathbb{R} ist. Zur Vereinfachung der Notation schreiben wir grundsätzlich \mathfrak{G} , aus der konkreten Wahl von \mathbb{G} wird dann klar, welche σ -Algebra zu wählen ist. Weitere Informationen zur Projektions- σ -Algebra sind in Davidson (1994, S. 435) oder in Pollard (1984) in Kapitel 4 zu finden. Dort wird nicht explizit das Intervall $[\epsilon, 1]$ betrachtet, die Aussagen lassen sich aber übertragen, indem die σ -Algebren auf $[\epsilon, 1]$ eingeschränkt werden.

Die Skorohod-Metrik (vgl. Billingsley (1968, S. 111 ff.)) wird vermieden, da zum Beispiel die Addition unter dieser Metrik keine stetige Abbildung und $D[\epsilon, 1]$ kein topologischer Vektorraum ist (vgl. Gill (1989, S. 99)). Daher könnten wir die Hadamard-Differenzierbarkeit nicht sinnvoll definieren. Es gilt jedoch: Wenn ein Prozess in der Skorohod-Metrik gegen einen fast sicher stetigen Prozess konvergiert, so gilt die Konvergenz auch in der Supremumsnorm, vgl. Billingsley (1968, S. 150 ff.) oder Gill (1989, S. 106).

Der Raum \mathbb{G} ist für $\mathbb{H}_i \in \{D[\epsilon, 1], D^+[\epsilon, 1], \epsilon \geq 0\}$ mit der Supremumsnorm nicht mehr separabel, besitzt also keine dichte, abzählbare Teilmenge, vgl. Billingsley (1968, S. 150). Mit der Skorohod-Metrik wäre die Separabilität gegeben. Dies ist eine unschöne Eigenschaft und der Grund dafür, dass wir die Projektions- σ -Algebra verwenden. Bei Benutzung der größeren Borel- σ -Algebra der offenen Mengen bezüglich der durch die Supremumsnorm erzeugte Metrik würden wegen der Nichtseparabilität Messbarkeits-

probleme entstehen, vgl. Davidson (1994, S. 415 und 456 ff). Die Verwendung der kleineren Projektions- σ -Algebra vermeidet diese Probleme. Auf $C[\epsilon, 1]$ stimmen die beiden σ -Algebren überein, da dieser Raum - versehen mit der Supremumsnorm - separabel ist.

Für \mathbb{G} ausgestattet mit der Supremumsnorm ist $d(X_T, a)$ immer eine Zufallsvariable, weil wir das Supremum über abzählbar viele Zufallsvariablen bilden.

Theorem 2.8 *Sei eine Folge von Zufallsfunktionen $X_T : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (\mathbb{G}, \mathfrak{G})$ (k -dimensional) gegeben. Konvergiert jede Komponente $X_{Tj} : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (\mathbb{H}_j, \mathfrak{H}_j)$ ($j \in \{1, \dots, k\}$) in Wahrscheinlichkeit gegen 0, so konvergiert auch der ganze Vektor in Wahrscheinlichkeit gegen 0.*

Beweis. Es gilt

$$\{\max_{j=1}^k \sup_{z \in [\epsilon, 1]} |X_{Tj}(z)| \geq \epsilon\} \subset \bigcup_{j=1}^k \{\sup_{z \in [\epsilon, 1]} |X_{Tj}(z)| \geq \epsilon\}$$

und damit

$$\mathbb{P}(\sup_{j=1}^k \sup_{z \in [\epsilon, 1]} |X_{Tj}(z)| \geq \epsilon) \leq \sum_{j=1}^k \mathbb{P}(\sup_{z \in [\epsilon, 1]} |X_{Tj}(z)| \geq \epsilon).$$

Die rechte Seite geht gegen 0 für jedes $\epsilon > 0$ und damit folgt die Behauptung. \square

Wir benötigen den Begriff der gleichmäßigen Beschränktheit von Momenten.

Definition 2.9 *Ein Schema von Zufallsvariablen $(X_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ hat gleichmäßig beschränkte α -te absolute Momente, falls eine Konstante $k > 0$ existiert, sodass*

$$\mathbb{E}(|X_{T,i}|^\alpha) \leq k \quad \forall T \in \mathbb{N}, i \in \{1, \dots, T\}.$$

In der Dissertation wird nicht von unabhängigen Zufallsvariablen ausgegangen, sondern von abhängigen. Dafür definieren wir im Folgenden einige Abhängigkeitskonzepte, zunächst analog zu Billingsley (1968, S. 167) die m -Abhängigkeit:

Definition 2.10 (*m*-Abhängigkeit) *Ein Schema von Zufallsvektoren $(X_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ ist *m*-abhängig, falls für jedes $i \in \mathbb{N}$ und alle $n > m$ gilt, dass die Zufallsvektoren $(\dots, X_{T,i-1}, X_{T,i})$ und $(X_{T,i+n}, X_{T,i+n+1} \dots)$ unabhängig sind. Eine Folge unabhängiger Zufallsvektoren heißt 0-abhängig.*

Definition 2.11 (ϕ -mischend, Davidson (1994), S. 209) *Ein Schema von Zufallsvektoren $(X_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ ist ϕ -mischend bzw. gleichmäßig mischend mit Größe ϕ^* , falls für die Größe*

$$\phi_m := \sup_{i,T} \sup_{G \in \mathfrak{G}_{T,-\infty}^i, H \in \mathfrak{G}_{T,i+m}^\infty} |\mathbb{P}(H|G) - \mathbb{P}(H)|$$

gilt, dass

$$\lim_{m \rightarrow \infty} \phi_m = 0 \tag{2}$$

ist. Hierbei ist $\mathfrak{G}_{T,a}^b = \sigma(X_{T,a}, \dots, X_{T,b})$ und $\alpha_m = O(m^{-\phi})$ für mindestens ein $\phi > \phi^$.*

Definition 2.12 (α -mischend, Davidson (1994), S. 209) *Ein Schema von Zufallsvektoren $(X_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ ist α -mischend bzw. stark mischend mit Größe ϕ^* , falls für die Größe*

$$\alpha_m := \sup_{i,T} \sup_{G \in \mathfrak{G}_{T,-\infty}^i, H \in \mathfrak{G}_{T,i+m}^\infty} |\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H)|$$

gilt, dass

$$\lim_{m \rightarrow \infty} \alpha_m = 0 \tag{3}$$

ist. Hierbei ist $\mathfrak{G}_{T,a}^b = \sigma(X_{T,a}, \dots, X_{T,b})$ und $\alpha_m = O(m^{-\phi})$ für mindestens ein $\phi > \phi^$.*

Definition 2.13 (Near-Epoch Dependence, Davidson (1994), S. 261) *Ein Schema von *n*-dimensionalen Zufallsvektoren $(X_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ mit $n \geq 1$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ ist L_p -NED (near-epoch dependent) mit Größe $-\phi^*$ und den Konstanten $(c_{T,i})$ über einem Schema $(V_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$, wenn für den Vektor*

$$Z_{T,i} := \begin{pmatrix} Z_{T,i}^1 & \dots & Z_{T,i}^n \end{pmatrix} := X_{T,i} - \mathbb{E}(X_{T,i} | \mathfrak{F}_{T,i-m}^{i+m})$$

gilt, dass

$$\left(\sum_{j=1}^n \mathbb{E} (|Z_{T,i}^j|)^p \right)^{\frac{1}{p}} \leq c_{T,i} \nu_m \quad (4)$$

gilt, wobei $\mathfrak{F}_{T,i-m}^{i+m} = \sigma(V_{T,i-m}, \dots, V_{T,i+m})$ und ν_m eine Nullfolge ist. Außerdem ist $\nu_m = O(m^{-\phi})$ für mindestens ein $\phi > \phi^*$.

Bemerkung Oft wird gefordert, dass $c_{T,i} \leq 2 \left(\sum_{j=1}^n \mathbb{E} (|Z_{T,i}^j - \mathbb{E}(Z_{T,i}^j)|)^p \right)^{\frac{1}{p}}$ gilt. Die ersten drei Konzepte stellen Bedingungen an den Grad der Abhängigkeit der Zufallsvariablen. Sie besagen, dass Beobachtungen aus der Vergangenheit und Gegenwart Beobachtungen in der fernen Zukunft weniger stark beeinflussen als Beobachtungen in der nahen Zukunft. Unter den drei ersten Konzepten lässt das des α -Mischens am meisten Abhängigkeit zu. Es gilt nämlich, dass ϕ -Mischen α -Mischen wegen

$$|\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H)| = |\mathbb{P}(G)| |\mathbb{P}(H|G) - \mathbb{P}(H)| \leq |\mathbb{P}(H|G) - \mathbb{P}(H)|$$

impliziert. Außerdem ist bei m -abhängigen Zufallsvektoren $\phi_q = 0$ für $q > m$, da für jedes T die σ -Algebren $\mathfrak{F}_{T,-\infty}^i$ sowie $\mathfrak{G}_{T,i+q}^j$ mit $j \geq T + q$ unabhängig sind. Für beliebige $G \in \mathfrak{F}_{T,1}^i$ und $H \in \mathfrak{G}_{T,i+q}^j$ ist dann $\mathbb{P}(G \cap H) = \mathbb{P}(G)\mathbb{P}(H)$. Somit ist dann die Bedingung des ϕ -Mischens erfüllt.

Das Konzept der NED wird meist angewandt, wenn für festes T die Folge $X_{T,i}$ eine Funktion der gesamten Folge $V_{T,i}$ ist. Die NED ist streng genommen keine Eigenschaft der Zufallsvariablen, sondern der Abbildung von $V_{T,i}$ nach $X_{T,i}$. Das Konzept wurde entwickelt, weil Funktionen von mischenden Zufallsvariablen im Allgemeinen nicht mehr mischend sind, vgl. Davidson (1994), S. 261. Wenn $V_{T,i}$ mischend ist, ist $X_{T,i}$ mit der NED-Eigenschaft „approximativ“ mischend. Weitere Informationen hierzu stehen auch in Bradley (2007).

Eine spezielle Zufallsfunktion auf $C[\epsilon, 1]$ ist die Brownsche Bewegung $W_1(z)$, die in Billingsley (1968) in Kapitel 9 behandelt wird und auf Wiener (1923) zurückgeht.

Definition 2.14 (Brownsche Bewegung) *Eine Zufallsfunktion*

$W_1 : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (C[\epsilon, 1], \mathfrak{G})$ heißt Brownsche Bewegung, falls folgende zwei Bedingungen gelten:

- Für festes $z \in [\epsilon, 1]$ ist

$$W_1(z) \stackrel{L}{=} N(0, z). \quad (5)$$

- Für beliebige $t_1 < \dots < t_n$ sind die Zuwächse $W_1(t_1), W_1(t_2 - t_1), \dots, W_1(t_n - t_{n-1})$ stochastisch unabhängig.

Bemerkungen Die Schreibweise

$$X \stackrel{L}{=} Y$$

bedeutet, dass die Verteilung zweier Zufallsvariablen X und Y auf \mathbb{R}^k gleich ist.

Eine Brownsche Bewegung ist ein Gauß-Prozess mit Erwartungswertfunktional 0 und Kovarianzkern $k(s, t) := \text{Cov}(W_1(s), W_1(t)) = \min\{s, t\}$, vgl. Billingsley (1968, S. 64). Falls W_1 auf $[0, 1]$ definiert wird, bedeutet Bedingung (5) für $z = 0$, dass $W_1(0) = 0$ \mathbb{P} -fast sicher gilt.

Da wir Konvergenzen später im Raum $D[\epsilon, 1]$ betrachten, müsste die Brownsche Bewegung auch auf $D[\epsilon, 1]$ definiert werden. Dies funktioniert dadurch, dass wir das hinter der Brownschen Bewegung stehende Wiener-Maß auf $C[\epsilon, 1]$ geeignet auf $D[\epsilon, 1]$ erweitern. Dann gilt $\mathbb{P}(X \in C[\epsilon, 1]) = 1$, vgl. Billingsley (1968, S. 137). Für unsere Anwendung ist dieser Unterschied jedoch nicht weiter relevant.

Auch muss für den Nachweis, dass ein stochastischer Prozess eine Brownsche Bewegung ist, die Stetigkeit nicht gezeigt werden. Es gilt nämlich mit dem Stetigkeitssatz von Kolmogorov-Chentsov, dass es zu einer Zufallsfunktion mit den obigen beiden Eigenschaften stets eine Version, das heißt eine fast sicher gleiche Zufallsfunktion, mit fast sicher stetigen Pfaden gibt, vgl. Karatzas und Shreve (1988, S. 53). Da fast sicher gleiche Zufallsfunktionen dieselbe Verteilung haben, gelten Konvergenzaussagen für die Brownsche Bewegung stets auch für Versionen mit fast sicher stetigen Pfaden.

Dieses Konzept lässt sich auch auf höhere Dimensionen ausdehnen, vgl. Davidson (1994, S. 454):

Definition 2.15 (*n*-dimensionale Brownsche Bewegung) *Eine n-dimensionale Zufallsfunktion $W_n : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (C[\epsilon, 1]^n, \mathfrak{G})$ heißt n-dimensionale Brownsche Bewegung, falls jede Komponente $W_{n,i}(z), i \in \{1, \dots, n\}$ eine eindimensionale Brownsche Bewegung ist und die Komponenten unabhängig sind.*

Aus der Brownschen Bewegung W_1 entsteht die Brownsche Brücke B , die definiert ist als

$$B(z) = W_1(z) - zW_1(1)$$

mit $z \in [\epsilon, 1]$. Sie ist ein Gauß-Prozess mit Erwartungswertfunktional 0 und Kovarianzkern $k(s, t) := Cov(B(s), B(t)) = \min\{s, t\} - st$.

Im Verlauf der Arbeit wird öfter das verallgemeinerte Continuous Mapping Theorem, CMT, benötigt, vgl. van der Vaart (1998, S. 259):

Theorem 2.16 (Verallgemeinertes Continuous Mapping Theorem) *Seien \mathbb{D} und \mathbb{D}' normierte Räume, seien $\mathbb{D}_T \subset \mathbb{D}$ beliebige Teilräume und $g_T : \mathbb{D}_T \rightarrow \mathbb{D}'$ eine beliebige Funktionenfolge mit Grenzwert g_0 , sodass für jede Folge $x_T \in \mathbb{D}_T$ gilt: Falls $\|x_{T'} - x\|_{\mathbb{D}} \rightarrow 0$ entlang einer Teilfolge T' und $x \in \mathbb{D}_0$, dann auch $\|g_{T'}(x_{T'}) - g_0(x)\|_{\mathbb{D}'} \rightarrow 0$. Dann gilt für beliebige Zufallsfunktionen $X_T : \Omega \rightarrow \mathbb{D}_T$ und jede Zufallsfunktion X mit Werten in \mathbb{D}_0 , sodass $g_0(X)$ eine Zufallsfunktion in \mathbb{D}' ist:*

1. Falls $X_T \rightarrow_d X$, dann auch $g_T(X_T) \rightarrow_d g_0(X)$,
2. Falls $X_T \rightarrow_p X$, dann auch $g_T(X_T) \rightarrow_p g_0(X)$,
3. Falls $X_T \rightarrow_{f.s.} X$, dann auch $g_T(X_T) \rightarrow_{f.s.} g_0(X)$.

Ein wichtiger Spezialfall ist, dass $g_T = g_0$ für alle T , vgl. hierzu auch Billingsley (1968, Korollar 1 zu Theorem 5.1).

Im Verlauf der Arbeit sind ferner folgende Sätze von Interesse:

Theorem 2.17 *Es seien Folgen von Zufallsvariablen X_1, X_2, \dots und Y_1, Y_2, \dots auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit Werten in einem normierten Raum \mathbb{D} ,*

versehen mit einer σ -Algebra \mathfrak{D} , gegeben, für die gilt: $X_T \rightarrow_d X$ für eine Zufallsvariable X und $Y_T \rightarrow_p y$ für eine Konstante y . Dann gilt:

1. $X_T + Y_T \rightarrow_d X + y$,
2. $X_T \cdot Y_T \rightarrow_d X \cdot y$.

Wir nennen dies im Folgenden immer Satz von Cramér. Zum Beweis verwenden wir das folgende Resultat aus Billingsley (1968, S. 27) über Produkträume.

Lemma 2.18 *Unter den Voraussetzungen von Theorem 2.17 gilt*

$$(X_T, Y_T) \rightarrow_d (X, y).$$

Bemerkungen (X_T, Y_T) ist durch Verwendung des Produkts der σ -Algebren eine Zufallsvariable auf $\mathbb{D} \times \mathbb{D}$, obwohl \mathbb{D} nicht notwendigerweise separabel sein muss, vgl. van der Vaart (1998, S. 258).

Die Aussage des Lemmas gilt im Allgemeinen nicht, falls Y_T nur in Verteilung gegen eine Zufallsvariable Y konvergiert. Betrachte etwa den Fall, dass X_T und Y_T gemeinsam normalverteilt sind mit $X_T \stackrel{L}{=} Y_T \stackrel{L}{=} N(0, 2)$ für alle T sowie $Cov(X_T, Y_T) = (-1)^T$.

Beweis. (zu Theorem 2.17) Wende das CMT auf den zweidimensionalen Vektor an. Nutze dabei aus, dass die Addition und die Multiplikation unter der Supremumsnorm stetige Abbildungen sind. \square

Sehr wichtig für die Herleitung unseres Tests sind Invarianzprinzipien, das heißt Sätze über die Konvergenz gegen die Brownsche Bewegung. Dies sind Verallgemeinerungen des Zentralen Grenzwertsatzes auf Zufallsfunktionen.

Wenn Konvergenz gegen die Brownsche Bewegung oder gegen eine andere Zufallsfunktion gezeigt werden soll, ist prinzipiell zu zeigen, dass die endlichdimensionalen Randverteilungen das Gewünschte tun und dass der Prozess straff (engl. tight) ist. Ein solches Konvergenzkriterium ist als Spezialfall von Billingsley (1968, Theorem 15.6) das folgende.

Theorem 2.19 *Es sei eine Folge von Zufallsfunktionen X_1, X_2, \dots und eine Zufallsfunktion X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit $\mathbb{P}(X \in C[\epsilon, 1]) = 1$ gegeben. Es gelte*

$$\left(X_T(t_1) \ \dots \ X_T(t_k) \right)' \rightarrow_d \left(X(t_1) \ \dots \ X(t_k) \right)' \quad (6)$$

als Zufallsvektor im \mathbb{R}^k für alle $t_1 < \dots < t_k$. Außerdem gebe es $\gamma \geq 0, \alpha > \frac{1}{2}, C > 0$ mit

$$\mathbb{E}(|X_T(t) - X_T(t_1)|^\gamma |X_T(t_2) - X_T(t)|^\gamma) \leq C|t_2 - t_1|^{2\alpha} \quad (7)$$

für beliebige $t_1 \leq t \leq t_2$. Dann gilt

$$X_T \rightarrow_d X.$$

Bemerkung Die Gleichung (6) beschreibt die Konvergenz der endlich-dimensionalen Randverteilungen, (7) beschreibt die Bedingung der Straffheit. Durch letztere wird verhindert, dass Wahrscheinlichkeitsmasse ins Unendliche „verschwindet“. Ist in einem Vektor jede einzelne Komponente straff, so gilt dies auch für den gesamten Vektor, vgl. Billingsley (1968, S. 41).

Der Satz ist in Billingsley (1968) für die Skorohod-Metrik gezeigt, gilt aber wegen der fast sicheren Stetigkeit der Pfade auch bei Verwendung der Supremumsnorm. Wir zitieren ein Invarianzprinzip von Wooldridge und White (1988), das wir in der Dissertation benötigen. Es ist unter anderem für ein Schema von eindimensionalen Zufallsvariablen anwendbar.

Theorem 2.20 (Invarianzprinzip) *Sei $(U_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ ein Schema von eindimensionalen Zufallsvariablen, $(c_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ ein Schema von positiven Konstanten und $(K_T(z))_{T \in \mathbb{N}}$ eine Folge von rechtsstetigen, in z monoton steigenden Funktionen mit ganzzahligen Werten mit $K_T(0) = 0$ für alle T und $K_T(z_2) - K_T(z_1) \rightarrow \infty$ für $T \rightarrow \infty$ und $z_2 > z_1$. Sei außerdem $X_T(z) = \sum_{i=1}^{K_T(z)} U_{T,i}$. Folgende Bedingungen seien erfüllt:*

1. $\mathbb{E}(U_{T,i}) = 0 \ \forall T \in \mathbb{N}, i \in \mathbb{Z}$.

2. Es gibt ein $r > 2$, sodass das r -te absolute Moment von $\left(\frac{U_{T,i}}{c_{T,i}}\right)_{T \in \mathbb{N}, 1 \leq i \leq T}$ gleichmäßig beschränkt ist.
3. $(U_{T,i})$ ist L_2 -NED mit Größe $-\gamma$ für $\frac{1}{2} \leq \gamma \leq 1$ und den Konstanten $(c_{T,i})$ über einem Schema $(V_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$, welches α -mischend der Größe $\phi^* := -\frac{r}{r-2}$ ist.
4. $\sup_{z \in [0,1], \delta \in (0,1-z]} \left(\limsup_{T \rightarrow \infty} \frac{v_T^2(z, \delta)}{\delta} \right) < \infty$, wobei $v_T^2(z, \delta) = \sum_{i=K_T(z)+1}^{K_T(z+\delta)} c_{T,i}^2$.
5. $\max_{1 \leq i \leq K_T(1)} c_{T,i} = O(K_T(1)^{\gamma-1})$.
6. Für festes $z \in [0, 1]$ gilt $\mathbb{E}(X_T(z)^2) \rightarrow z$ für $T \rightarrow \infty$.

Dann gilt auf $D[0, 1]$

$$X_T \rightarrow_d W_1.$$

Bemerkung Das Invarianzprinzip gilt auch für $r = 2$, wenn das Schema $V_{T,i}$ sogar ϕ -mischend der Größe $\phi^* := -\frac{r}{2r-2}$ und zusätzlich $\left(\frac{U_{T,i}^2}{c_{T,i}^2}\right)$ gleichförmig integrierbar ist. Falls $(U_{T,i})$ selber m -abhängig, α - oder ϕ -mischend ist, so ist die 3. Bedingung in jedem Fall erfüllt. Man kann dann $(U_{T,i}) = (V_{T,i})$ wählen, sodass Bedingung (4) gemäß den Rechenregeln für bedingte Erwartungswerte erfüllt ist.

Ein ähnliches Invarianzprinzip gibt es auch für den multivariaten Fall, das auf dem Cramér-Wold-Theorem basiert, vgl. Davidson (1994, S. 492).

Theorem 2.21 (Multivariates Invarianzprinzip) Sei $(U_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ ein Schema von zentrierten n -dimensionalen Zufallsvariablen. Es sei $(K_T(z))_{T \in \mathbb{N}}$ eine Folge von rechtsstetigen, in z monoton steigenden Funktionen mit ganzzahligen Werten und $X_T(z) = \sum_{i=1}^{K_T(z)} U_{T,i}$. Folgende Bedingungen seien erfüllt:

1. Für jeden festen n -dimensionalen Vektor λ mit $\lambda' \lambda = 1$ gibt es ein skalares Schema $(c_{T,i}^\lambda)$, sodass die Bedingungen von Theorem 2.20 für die Schemata $(\lambda' U_{T,i})$ und $(c_{T,i}^\lambda)$ gelten.
2. Für festes $z \in [0, 1]$ gilt $\mathbb{E}(X_T(z) X_T(z)') \rightarrow z \cdot \Sigma$ für eine positiv definite Matrix Σ .

Dann gilt auf $D[0, 1]^n$

$$\Sigma^{-\frac{1}{2}} X_T \rightarrow_d W_n.$$

Bemerkung Das Theorem ist in Davidson (1994) etwas allgemeiner definiert für weitere Transformationen der Brownschen Bewegung, doch die hier vorgestellte Version genügt völlig. Sie entsteht dadurch, dass in dem ursprünglichen Satz $p = m$ gesetzt und als Homöomorphismus die Identität gewählt wird. Außerdem wird der Satz von Cramér angewandt.

Um Tests basierend auf Invarianzprinzipien konstruieren zu können, ist es oft notwendig, Ausdrücke der Form

$$\Sigma = \lim_{T \rightarrow \infty} \sum_{i=1}^T \sum_{j=1}^T \mathbb{E}(U_{T,i} U'_{T,j})$$

konsistent schätzen zu können. Eine Möglichkeit hierfür ist die Wahl von Kernschätzern, wie es zum Beispiel in Davidson und de Jong (2000) beschrieben wird. Der Schätzer funktioniert so, dass er Autokovarianzen von näher aneinander liegenden Zeitpunkten stärker gewichtet als Autokovarianzen von weiter auseinander liegenden Zeitpunkten. Wir gehen davon aus, dass $U_{T,i} = \begin{pmatrix} U_{T,i}^1 & \dots & U_{T,i}^n \end{pmatrix}$ die Bedingungen von 2.21 erfüllt und von einem Parameter $\theta_0 \in \Theta \subset \mathbb{R}^n$ mit einer konvexen, kompakten Teilmenge Θ aus \mathbb{R}^n abhängt. Der Kernschätzer hat dann die Form

$$\Sigma_T := \lim_{T \rightarrow \infty} \sum_{i=1}^T \sum_{j=1}^T U_{T,i}(\hat{\theta}_T) U'_{T,j}(\hat{\theta}_T) k\left(\frac{i-j}{\gamma_T}\right),$$

wobei $\hat{\theta}_T$ eine konsistente Schätzfolge für θ_0 , die Funktion $k(\cdot)$ die Kernfunktion und γ_T die Bandbreite ist. An diese drei Größen werden gewisse Bedingungen gestellt, die in dem folgenden Satz zusammengefasst sind.

Theorem 2.22 *Unter den Annahmen*

1. $k(\cdot) \in \mathfrak{K}$, wobei

$$\mathfrak{K} = \{k(\cdot) : \mathbb{R} \rightarrow [-1, 1] \mid k(0) = 1, k(x) = k(-x) \forall x \in \mathbb{R}, \\ \int_{-\infty}^{\infty} |k(x)| dx < \infty, \int_{-\infty}^{\infty} \psi(t) dt < \infty,$$

$k(\cdot)$ ist stetig in 0 und an höchstens endlich vielen Punkten unstetig}

mit

$$\psi(t) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(x) \exp(itx) dx.$$

2. Es gibt ein $r > 2$, für das gilt: $(U_{T,i})$ ist L_2 -NED mit Größe $-\frac{1}{2}$ und den Konstanten $(c_{T,i})$ über einem Schema $(V_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$, welches α -mischend der Größe $\phi^* := -\frac{r}{r-2}$ ist. Außerdem gibt es ein Schema $(d_{T,i})$, sodass

$$\sup_{T,i} \frac{\left(\sum_{j=1}^n \mathbb{E}(|U_{T,i}^j|)^r \right)^{\frac{1}{r}} + c_{T,i}}{d_{T,i}} < \infty$$

mit

$$\sup_{T \in \mathbb{N}} \sum_{i=1}^T d_{T,i}^2 < \infty.$$

3. Es gilt

$$\lim_{T \rightarrow \infty} \left(\gamma_T^{-1} + \gamma_T \max_{1 \leq i \leq T} d_{T,i}^2 \right) = 0$$

4. Sei $M = \frac{\partial X_i(\theta)}{\partial \theta}$, dann gilt für jede Komponente M_{jk} von M :

- (a) $\sqrt{T}(\hat{\theta}_T - \theta_0)$ ist stochastisch beschränkt.
- (b) $T^{-\frac{1}{2}} \sum_{i=1}^T \mathbb{E} M_{jk}$ ist stetig in θ_0 für alle T .

(c) Es gibt eine Menge $\mathfrak{N} \subset \Theta$, eine offene Umgebung von θ_0 , sodass

$$\limsup_{T \rightarrow \infty} \sum_{i=1}^T \mathbb{E} \sup_{\theta \in \mathfrak{N}} M_{jk}^2 < \infty$$

gilt, sowie

$$\sup_{\theta \in \mathfrak{N}} \left| \sum_{i=1}^T MM' \right|$$

stochastisch beschränkt ist und außerdem $T^{-\frac{1}{2}} \gamma_T \rightarrow 0$ gilt.

gilt

$$\Sigma - \Sigma_T \rightarrow_p 0.$$

Bemerkung Bedingung 1 an den Kern wird zum Beispiel vom Tukey-Hanning Kern,

$$K(x) = \begin{cases} \frac{1+\cos(\pi x)}{2}, & |x| \leq 1 \\ 0, & \text{sonst} \end{cases}$$

erfüllt. Bei Bedingung 2 ist auch $r = 2$ erlaubt, wenn $(V_{T,i})_{T \in \mathbb{N}, i \in \mathbb{Z}}$ sogar ϕ -mischend der Größe $\phi^* := -\frac{r}{2r-2}$ ist und außerdem für jedes $1 \leq j \leq n$ die Zufallsvariable

$$\frac{(U_{T,i}^j)^2}{d_{T,i}^2}$$

gleichmäßig integrierbar ist. Die Bedingungen 1 bis 3 genügen für die Konsistenz, wenn θ_0 bekannt ist; Bedingung 4 ist nötig, falls θ_0 konsistent geschätzt wird. Die Bedingungen der stochastischen Beschränktheit sind zum Beispiel erfüllt, wenn die Folgen in Verteilung gegen reelle Zufallsvektoren konvergieren. Bedingung 4.3 könnte noch modifiziert werden, doch damit beschäftigen wir uns nicht weiter.

2.2 Die verallgemeinerte Delta-Methode

In Anlehnung an van der Vaart (1998, S. 296 ff.) definieren wir die verallgemeinerte Hadamard-Differenzierbarkeit. Zunächst sei als Spezialfall von (1)

$$\mathbb{G}_1 := \mathbb{H}_1 \times \dots \times \mathbb{H}_k \text{ (} k\text{-mal, } k \geq 1, \mathbb{H}_i \in \{D[\epsilon, 1], D^+[\epsilon, 1], \epsilon \geq 0\}), \quad (8)$$

$$\mathbb{G}_2 := \mathbb{H}_1 \times \dots \times \mathbb{H}_l \text{ (} l\text{-mal, } l \geq 1, \mathbb{H}_i \in \{D[\epsilon, 1], D^+[\epsilon, 1], \epsilon \geq 0\}). \quad (9)$$

Der Raum \mathbb{G}_1^* ist ein weiterer Funktionenraum, der dieselbe Gestalt wie \mathbb{G}_1 hat. Der einzige Unterschied ist, dass einige Komponenten eventuell gleich $D^+[\epsilon, 1]$ und nicht gleich $D[\epsilon, 1]$ sind. Damit ist \mathbb{G}_1^* eine Teilmenge von \mathbb{G}_1 .

Im weiteren Verlauf der Arbeit wird der Spezialfall betrachtet, dass $\mathbb{G}_1 = D[\epsilon, 1]^k$ ist.

Definition 2.23 (Verallgemeinerte Hadamard-Differenzierbarkeit) *Seien $\mathbb{G}_1^* \subseteq \mathbb{G}_1$, \mathbb{G}_2 aus (8) und (9) sowie $\theta \in \mathbb{G}_1^*$ und $\epsilon > 0$ gegeben. Eine Funktion $f : \mathbb{G}_1^* \rightarrow \mathbb{G}_2$ heißt verallgemeinert Hadamard-differenzierbar in θ , falls eine stetige, lineare Abbildung $f'_\theta : \mathbb{G}_1 \rightarrow \mathbb{G}_2$ (die verallgemeinerte Hadamard-Ableitung) existiert, sodass*

$$\lim_{T \rightarrow \infty} \left\| \frac{f(\theta + r_T h_T) - f(\theta)}{r_T} - f'_\theta(h) \right\|_{\mathbb{G}_2} = 0$$

für alle $r_T \in D[\epsilon, 1]$ mit $r_T(z) \neq 0 \forall z \in [\epsilon, 1] \forall T$, $h_T, h \in \mathbb{G}_1$ mit $\|r_T\|_{D[\epsilon, 1]} \rightarrow 0$ und $\|h_T - h\|_{\mathbb{G}_1} \rightarrow 0$, sodass $\theta + r_T h_T \in \mathbb{G}_1^*$ für alle T .

Der Unterschied zu der gewöhnlichen Hadamard-Differenzierbarkeit, wie sie etwa in van der Vaart (1998) definiert ist, liegt im Wesentlichen darin, dass hier r_T ein Element aus $D[\epsilon, 1]$ und nicht bloß eine Zahlenfolge aus \mathbb{R} ist. Daher benötigen wir die stärkere Voraussetzung, dass r_T in der Supremumsnorm auf $D[\epsilon, 1]$ gegen 0 geht. Diese Verallgemeinerung ermöglicht die verallgemeinerte Delta-Methode, die wir später bei der Herleitung der asymptotischen Verteilung benötigen. Die gewöhnliche Hadamard-Differenzierbarkeit ist außerdem für Abbildungen zwischen allgemeinen normierten Räumen definiert, ein Beispiel ist die Funktion

$$f : D[-\infty, \infty] \times BV_1[-\infty, \infty] \rightarrow \mathbb{R},$$

$$f(x_1, x_2) = \int_{\mathbb{R}} \phi(x_1) dx_2$$

für eine feste zweimal stetig differenzierbare Funktion $\phi : [0, 1] \rightarrow \mathbb{R}$ und den Raum $BV_1[-\infty, \infty]$ den Raum der betragsmäßig durch 1 beschränkten Funktionen aus $D[-\infty, \infty]$ mit beschränkter Variation, vgl. van der Vaart (1998, S. 298). Die Hadamard-Ableitung an der Stelle $\theta = (\theta_1, \theta_2) \in D[-\infty, \infty] \times BV_1[-\infty, \infty]$ ist gegeben durch

$$f' : D[-\infty, \infty] \times BV_1[-\infty, \infty] \rightarrow \mathbb{R},$$

$$f'_\theta(h_1, h_2) = h_2\phi \circ \theta_1|_{-\infty}^\infty - \int h_{2-}d(\phi \circ \theta_1) + \int \phi'(\theta_1)h_1dx_2$$

und spielt eine große Rolle bei der Untersuchung des asymptotischen Verhaltens der Mann-Whitney-U-Teststatistik. Die Funktion h_{2-} ist hierbei die linksstetige Version der rechtsstetigen Funktion h_2 .

Es gibt auch noch andere Differenzierbarkeitsbegriffe auf Funktionenräumen, etwa Fréchet- oder Gateaux-Differenzierbarkeit, doch diese behandeln wir in dieser Arbeit nicht weiter.

In der Dissertation interessiert uns bei der verallgemeinerten Hadamard-Differenzierbarkeit vor allem folgender Typ von Funktionen:

$$f(\theta)(\cdot) = \psi(\theta(\cdot)) \tag{10}$$

für zweimal stetig differenzierbare Funktionen ψ von einer offenen Teilmenge von \mathbb{R}^k nach \mathbb{R}^l . Die Funktion f bildet also $\theta(\cdot)$ ab. Die Zuordnungsvorschrift ist somit relativ simpel - die Funktion θ wird nicht etwa integriert oder ähnliches - und hierfür gibt es eine allgemeine Formel für die verallgemeinerte Hadamard-Ableitung.

Theorem 2.24 *Sei f wie in (10), sodass $\psi = (\psi^1 \dots \psi^l)'$, und θ wie in 2.23 gegeben. Für jedes $i \in \{1, \dots, l\}$ sei die Hessematrix von ψ^i in einer offenen Umgebung der Menge $\{\theta(z), z \in [\epsilon, 1]\}$ stetig und beschränkt. Dann ist f in θ verallgemeinert Hadamard-differenzierbar und die Hadamard-Ableitung hat dieselbe Gestalt wie die gewöhnliche Ableitung von ψ , das heißt*

$$f'_\theta(h)(\cdot) = D\psi_{\theta(\cdot)}(h(\cdot)).$$

Beweis. Wir betrachten zunächst nur eine Komponente von ψ , o.B.d.A. ψ^1 und halten $z \in [\epsilon, 1]$ fest. Mit dem Satz von Taylor, vgl. Kaballo (1997, S. 128), und den Größen aus 2.23 gilt

$$\begin{aligned} \psi^1(\theta(z) + r_T(z)h_T(z)) &= \psi^1(\theta(z)) + D\psi_{\theta(z)}^1(r_T(z)h_T(z)) + \frac{1}{2} \langle r_T(z)h_T(z), \\ &\quad H\psi_{\theta(z)+\tau(r_T(z)h_T(z))r_T(z)h_T(z)}^1(r_T(z)h_T(z)) \rangle > \\ &= \psi^1(\theta(z)) + r_T(z)D\psi_{\theta(z)}^1(h_T(z)) + \frac{1}{2}(r_T(z))^2 \\ &\quad \langle h_T(z), H\psi_{\theta(z)+\tau(r_T(z)h_T(z))r_T(z)h_T(z)}^1(h_T(z)) \rangle > \end{aligned}$$

mit geeignetem $\tau(r_T(z)h_T(z)) \in [0, 1]$, wobei $D\psi_{\theta(z)}^1$ und $H\psi_{\theta(z)+\tau(r_T(z)h_T(z))r_T(z)h_T(z)}^1(h_T(z))$ stetige lineare Abbildungen sind. Für hinreichend große T ist

$$\left\| \langle h_T(z), H\psi_{\theta(z)+\tau(r_T(z)h_T(z))r_T(z)h_T(z)}^1(h_T(z)) \rangle \right\|_{D[\epsilon,1]} < \infty,$$

da $\|r_T(z)\|_{D[\epsilon,1]} \rightarrow 0$ und $\|h_T - h\|_{\mathbb{G}_1} \rightarrow 0$. Nach Subtraktion von $\psi^1(\theta(z))$, Division durch $r_T(z)$ und Bilden des Supremums ($z \in [\epsilon, 1]$) folgt dann für ψ^1 direkt die Behauptung mit der Definition der verallgemeinerten Hadamard-Differenzierbarkeit und der Stetigkeit von $D\psi_{\theta(z)}^1$.

Für die anderen Komponenten gilt Entsprechendes und die Behauptung für die gesamte Funktion folgt dann mit der Wahl von

$$D\psi_{\theta}(h) = \left(D\psi_{\theta}^1(h) \quad \dots \quad D\psi_{\theta}^l(h) \right)'$$

und der Definition der mehrdimensionalen Supremumsnorm. □

Bemerkung Aus dem Theorem folgt direkt durch Wahl von r_T als konstante Funktion, dass eine Funktion f der dort angegebenen Form auch gewöhnlich Hadamard-differenzierbar ist.

Nun folgt die Delta-Methode, die in dieser Form in der Literatur unseres Wissens nach noch nicht behandelt wurde:

Theorem 2.25 (Delta-Methode) *Es gelten die Voraussetzungen aus Definition 2.23, sodass $f : \mathbb{G}_1^* \rightarrow \mathbb{G}_2$ verallgemeinert Hadamard-differenzierbar in θ ist. Seien $M_T : \Omega \rightarrow \mathbb{G}_1^*$ Zufallsfunktionen, sodass*

$$r_T \cdot (M_T - \theta) \rightarrow_d M$$

für eine Folge $r_T \in D[\epsilon, 1]$ mit $\|\frac{1}{r_T}\|_{D[\epsilon, 1]} \rightarrow 0, r_T(z) \neq 0 \forall z, \forall T$, und eine Zufallsfunktion M , die Werte in \mathbb{G}_1 annimmt. Dann gilt auch

$$r_T \cdot (f(M_T) - f(\theta)) \rightarrow_d f'_\theta(M),$$

wobei f'_θ die verallgemeinerte Hadamard-Ableitung von f im Punkt θ bezeichnet.

Beweis. Wir definieren für jedes T eine Funktion

$$g_T(h) = r_T \cdot \left(f\left(\theta + \frac{1}{r_T}h\right) - f(\theta) \right),$$

die definiert ist auf $\mathbb{G}_T := \{h : \theta + \frac{1}{r_T}h \in \mathbb{G}_1^*\}$. Da f verallgemeinert Hadamard-differenzierbar ist, gilt

$$\lim_{T \rightarrow \infty} \|g_T(h_T) - f'_\theta(h)\|_{\mathbb{G}_2} = 0$$

für jede Folge h_T mit $\|h_T - h\|_{\mathbb{G}_1}$ und $h \in \mathbb{G}_1$. Mit dem verallgemeinerten CMT folgt dann

$$r_T \cdot (f(M_T) - f(\theta)) = g_T(r_T \cdot (M_T - \theta)) \rightarrow_d f'_\theta(M).$$

Das beweist die Behauptung. □

Der Unterschied zu der Delta-Methode aus van der Vaart (1998, S. 297) liegt im Wesentlichen darin, dass hier r_T ein Element aus $D[\epsilon, 1]$ und nicht bloß eine Zahlenfolge aus \mathbb{R} ist. Daher benötigen wir die stärkere Voraussetzung, dass $\frac{1}{r_T}$ in der Supremumsnorm auf $D[\epsilon, 1]$ gegen 0 geht.

Folgendes Korollar wird später wichtig sein:

Korollar 2.26 *Gilt in Theorem 2.25, dass M_T sogar \mathbb{P} -fast sicher gegen θ konvergiert, dann genügt die Forderung $M_T : \Omega \rightarrow \mathbb{G}_1$.*

Beweis. Mit dem Theorem von Egoroff, vgl. Theorem A.5, gibt es für jedes $\eta > 0$ ein $T(\eta)$ und eine Menge $\Omega_\eta \subset \Omega$ mit $\mathbb{P}(\Omega_\eta) \geq 1 - \eta$, sodass $M_T(\omega) \in \mathbb{G}_1^*$ für alle $\omega \in \Omega_\eta$ gilt. Mit der alternativen Charakterisierung der Konvergenz in Verteilung gilt dann mit dem vorherigen Theorem für eine abgeschlossene Menge $A \in \mathbb{G}_2$ für $T \geq T(\eta)$

$$\begin{aligned} & \mathbb{P}(\omega \in \Omega | (r_T \cdot (f(M_T) - f(\theta)))(\omega) \in A) \\ &= \mathbb{P}(\omega \in \Omega_\eta | (r_T \cdot (f(M_T) - f(\theta)))(\omega) \in A) \\ &+ \mathbb{P}(\omega \in \Omega/\Omega_\eta | (r_T \cdot (f(M_T) - f(\theta)))(\omega) \in A) \\ &\leq \mathbb{P}(\omega \in \Omega_\eta | (r_T \cdot (f(M_T) - f(\theta)))(\omega) \in A) + \eta \end{aligned}$$

und daher

$$\begin{aligned} & \limsup_{T \rightarrow \infty} \mathbb{P}(\omega \in \Omega | (r_T \cdot (f(M_T) - f(\theta)))(\omega) \in A) \\ &\leq \mathbb{P}(\omega \in \Omega_\eta | (f'_\theta(M))(\omega) \in A) + \eta \\ &\leq \mathbb{P}(\omega \in \Omega | (f'_\theta(M))(\omega) \in A) + \eta. \end{aligned}$$

Da η beliebig ist, folgt hiermit die Behauptung. □

3 Die Teststatistik und ihre asymptotische Nullverteilung

3.1 Die Teststatistik

Für den Test auf konstante Korrelation sei ein Schema von bivariaten Zufallsvektoren $(X_{T,i}, Y_{T,i}), T \in \mathbb{N}, i \in \mathbb{Z}$, auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ gegeben. Zur besseren Lesbarkeit lassen wir den Index T im Folgenden soweit es geht weg. Die Zufallsvektoren müssen nicht unabhängig identisch verteilt sein; wir gehen später auf die genauen Annahmen ein. Das Hypothesenpaar ist

$$H_0 : \rho_i = \rho_0 \forall i \in \{1, \dots, T\} \text{ vs. } H_1 : \exists i \in \{1, \dots, T-1\} \text{ mit } \rho_i \neq \rho_{i+1}.$$

Hierbei ist

$$\rho_i = \frac{\text{Cov}(X_i, Y_i)}{\sqrt{\text{Var}(X_i)}\sqrt{\text{Var}(Y_i)}}$$

und ρ_0 eine konstante Zahl $\in [-1, 1]$. Definiere zunächst folgende Ausdrücke:

$$\begin{aligned} \tau(z) &= [2 + z(T-2)], z \in [0, 1] \\ \bar{X}_k &= \frac{1}{k} \sum_{i=1}^k X_i \\ \bar{Y}_k &= \frac{1}{k} \sum_{i=1}^k Y_i \\ \overline{(X^2)}_k &= \frac{1}{k} \sum_{i=1}^k X_i^2 \\ \overline{(Y^2)}_k &= \frac{1}{k} \sum_{i=1}^k Y_i^2 \\ \overline{(XY)}_k &= \frac{1}{k} \sum_{i=1}^k X_i Y_i \\ \hat{\rho}_k &= \frac{\sum_{i=1}^k (X_i - \bar{X}_k)(Y_i - \bar{Y}_k)}{\sqrt{\sum_{i=1}^k (X_i - \bar{X}_k)^2} \sqrt{\sum_{i=1}^k (Y_i - \bar{Y}_k)^2}}. \end{aligned} \tag{11}$$

Formel (11) beschreibt den empirischen Korrelationskoeffizienten, berechnet aus den ersten k Beobachtungen. Weiterhin gelten folgende Annahmen:

(A1) Sei

$$U_i := \left(X_i^2 - \mathbb{E}(X_i^2), Y_i^2 - \mathbb{E}(Y_i^2), X_i - \mathbb{E}(X_i), Y_i - \mathbb{E}(Y_i), X_i Y_i - \mathbb{E}(X_i Y_i) \right)'$$

und

$$S_j := \sum_{i=1}^j U_i,$$

dann gilt

$$\lim_{T \rightarrow \infty} \mathbb{E} \left(\frac{1}{T} S_T S_T' \right) =: D_1 >_L 0,$$

wobei $>_L$ positiv definit (größer im Sinne der Löwner-Ordnung) meint.

(A2) Es gibt ein $r > 2$, sodass die r -ten absoluten Momente der Komponenten von U_i gleichmäßig beschränkt sind.

(A3) Für das r aus Bedingung (A2) erfüllt (X_i, Y_i) die L_2 -NED-Bedingung mit Größe $-\frac{r-1}{r-2}$ und den Konstanten (c_i) über einem Schema $(V_i)_{i \in \mathbb{Z}}$, welches α -mischend der Größe $\phi^* := -\frac{r}{r-2}$ ist. Es gilt

$$c_i \leq 2 \left(\sum_{j=1}^5 \mathbb{E} (|U_i^{*,j}|)^2 \right)^{\frac{1}{2}}.$$

(A4) Unter der Nullhypothese und für beliebiges $\epsilon > 0$ gilt: Für

$$f(i) \in \{\mathbb{E}(X_i), \mathbb{E}(Y_i), \mathbb{E}(X_i^2), \mathbb{E}(Y_i^2), \mathbb{E}(X_i Y_i)\}$$

ist $f(i) = c_f + d_{fi}$ mit

$$\lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \sum_{i=1}^T |d_{fi}| = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T d_{fi}^2 = 0$$

für jedes $z \in [0, 1]$ und

$$c_{\mathbb{E}(X_i)} =: \mu_x, c_{\mathbb{E}(Y_i)} =: \mu_y, c_{\mathbb{E}(X_i^2)} =: m_x^2, c_{\mathbb{E}(Y_i^2)} =: m_y^2, c_{\mathbb{E}(X_i Y_i)} =: m_{xy}.$$

Insbesondere konvergiert damit das arithmetische Mittel der theoretischen Momente $\mathbb{E}(X_i), \mathbb{E}(Y_i), \mathbb{E}(X_i^2), \mathbb{E}(Y_i^2), \mathbb{E}(X_i Y_i)$ für $T \rightarrow \infty$ schneller als \sqrt{T} jeweils gegen eine feste Zahl. Wir fordern auch, dass alle auftretenden Momente gleichmäßig beschränkt sind. Definiere noch

$$\sigma_x := \sqrt{m_x^2 - \mu_x^2}, \sigma_y := \sqrt{m_y^2 - \mu_y^2}, \sigma_{xy} := m_{xy} - \mu_x \mu_y$$

und

$$\rho_0^* := \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}.$$

Bemerkung Wegen Annahme (A2) ist $\mathbb{E}(|X_i|^{4+\alpha^*}) < \infty, \mathbb{E}(|Y_i|^{4+\alpha^*}) < \infty \forall i \in \{1, \dots, T\}$ für ein $\alpha^* > 0$, an die 5. Momente wird jedoch keine Anforderung gestellt. Annahme (A1) wird in Lemma 3.1 noch weiter spezifiziert. Die Abhängigkeitsstruktur (A3) könnte auch anders aussehen, sodass nicht α -mischende, sondern ϕ -mischende Zufallsvariablen verwendet werden und dann $r = 2$ genügt, aber die zweiten absoluten Momente der Komponenten von U_i gleichmäßig integrierbar sein müssen. Die Bedingung an die Existenz der Momente ist dann also schwächer, andererseits wird mit ϕ -Mischen weniger Abhängigkeit zugelassen.

Annahme (A3) garantiert, dass

$$U_i^* := \left(X_i^2, Y_i^2, X_i, Y_i, X_i Y_i \right)'$$

L_2 -NED ist mit Größe $-\frac{1}{2}$, vgl. hierzu Davidson (1994, S.273). Annahme (A4) besagt nicht, dass die angegebenen Momente asymptotisch gleich sind, sondern besagt, dass die Momente nicht zu stark schwanken dürfen. Aus Annahme (A4) folgt insbesondere, dass

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T d_{fi} = 0.$$

Die Voraussetzungen sind zum Beispiel erfüllt, wenn sich die Varianzen an etwas weniger als \sqrt{T} Zeitpunkten von einem Mittel unterscheiden und ansonsten konstant sind. Selbst wenn die Momente in Annahme (A4) asymptotisch gleich sind, muss nicht zwangsläufig

$$\rho_0 = \rho_0^*$$

gelten. Die Forderung

$$\frac{\sigma_{xy} + d_{1i}}{\sqrt{(\sigma_x^2 + d_{2i})(\sigma_y^2 + d_{3i})}} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}$$

lässt sich nämlich äquivalent umschreiben zu

$$2\sigma_x^2 \sigma_y^2 d_{1i} + \sigma_x^2 \sigma_y^2 d_{1i}^2 = \sigma_{xy}^2 \sigma_x^2 d_{3i} + \sigma_{xy}^2 \sigma_y^2 d_{2i} + \sigma_{xy}^2 d_{2i} d_{3i}$$

und dies ist im Allgemeinen nicht erfüllt. Sei etwa bei zentrierten Zufallsvariablen

$$\mathbb{E}(X_i^2) = \sigma_x^2 + \frac{d_2}{T},$$

$$\mathbb{E}(Y_i^2) = \sigma_y^2 + \frac{d_3}{T},$$

$$\mathbb{E}(X_i Y_i) = \sigma_{xy} + \frac{d_1}{T},$$

dann gilt dies, wenn

$$d_1 = \sigma_{xy}, \quad d_2 = \sigma_x^2, \quad d_3 = \sigma_y^2,$$

aber nicht, wenn

$$\frac{d_1}{\sqrt{d_2 d_3}} \neq \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}.$$

Wir werden später sehen, dass die Annahmen noch etwas abgeschwächt werden können, vgl. hierzu die Ausführungen in Kapitel 4.2 bei der Situation (17). Für die Herleitung in diesem Kapitel brauchen wir aber noch alle Annahmen.

Für die Anwendung des Tests auf Aktienrenditen sind die Annahmen sinnvoll, weil wir

keine Varianzstationarität fordern. Außerdem benötigen wir keine Normalverteilung, sodass der Test asymptotisch auch mit der bei Aktienrenditen beliebten t -Verteilung valide ist. Es wäre durchaus vertretbar, Erwartungswertstationarität zu fordern, weil diese plausibel bei Renditen ist, wenn man die (bei Tagesrenditen ohnehin kaum vorhandene) Risikoprämie vernachlässigt. Unsere Annahme bezüglich der Erwartungswerte ist jedoch schwächer und damit allgemeiner. Auch die Annahme der Near-Epoch Dependence ist sinnvoll. Aktienrenditen sind über die Zeit im Allgemeinen nicht unabhängig, für große Zeitabstände ist die Abhängigkeit jedoch vernachlässigbar. Unter gewissen Bedingungen weist ein GARCH(p,q)-Prozess, der häufig zur Modellierung von Aktienrenditen verwendet wird, eine NED-Struktur auf. Die Bedingung (A3) ist dann erfüllt, insbesondere, weil das ν_m aus der NED-Definition exponentiell abfällt. Somit gilt die Bedingung, dass die NED-Größe gleich $\frac{1}{2}$ ist. Nähere Informationen hierzu finden sich in Hansen (1991) oder Davidson (2002).

Die Teststatistik Q_T ist definiert als

$$Q_T(X, Y) = \hat{D} \max_{2 \leq j \leq T} \frac{j}{\sqrt{T}} |\hat{\rho}_j - \hat{\rho}_T| = \max_{2 \leq j \leq T} |\hat{D} \frac{j}{\sqrt{T}} (\hat{\rho}_j - \hat{\rho}_T)| =: \max_{2 \leq j \leq T} |P_T(j)|,$$

wobei \hat{D} eine komplizierte Darstellung hat, vgl. Korollar 3.7.

Die Idee des Tests ist es, die Nullhypothese konstanter Korrelation zu verwerfen, wenn die empirischen Korrelationen im Zeitverlauf zu sehr schwanken. Zur Messung dieser Schwankung dient der Term $\max_{2 \leq j \leq T} |\hat{\rho}_j - \hat{\rho}_T|$. Bei konstanter Korrelation sollten die aus den ersten j Beobachtungen geschätzten Korrelationen jeweils nicht zu sehr von der Gesamtkorrelation abweichen. Wir betrachten den ersten Wert nicht, da die empirische Korrelation dort nicht sinnvoll definiert ist.

Der Gewichtungsfaktor $\frac{j}{\sqrt{T}}$ bewirkt, dass Abweichungen am Anfang im Vergleich zu Abweichungen am Ende herunter gewichtet werden. Damit wird ausgeglichen, dass die Korrelationen am Anfang konstruktionsbedingt stärker schwanken als am Ende. Der Gewichtungsfaktor ist linear in j .

Der Schätzer \hat{D} hängt nicht von j ab und dient als Normierungsfaktor, um eine asymptotische Verteilung (für $T \rightarrow \infty$) unter der Nullhypothese in geschlossener Form berechnen zu können. Die Berechnung einer asymptotischen Verteilung ist erheblich einfacher als die Herleitung einer exakten Verteilung für endliches T .

3.2 Die asymptotische Nullverteilung

Zur Herleitung der asymptotischen Verteilung schreiben wir die Teststatistik als

$$\sup_{0 \leq z \leq 1} |K_T(z)|$$

mit

$$K_T(z) = \hat{D} \frac{\tau(z)}{\sqrt{T}} (\hat{\rho}_{\tau(z)} - \hat{\rho}_T). \quad (12)$$

Hier läuft j nicht mehr von 2 bis T , sondern $\tau(z)$ läuft von 0 bis 1, was äquivalent ist. Das ist sinnvoll, weil z jetzt in einem abgeschlossenen Intervall liegt und das Intervall für wachsende T nicht immer größer wird, im Gegensatz zu $[2, T]$. Wir können so auf dem Funktionenraum $D[0, 1]$ arbeiten.

Es zeigt sich, dass die Grenzverteilung der Teststatistik gleich der Verteilung des Maximums des Absolutbetrags einer Brownschen Brücke auf $[0, 1]$ ist. Wir werden also zeigen, dass

$$\sup_{0 \leq z \leq 1} |K_T(z)|$$

in Verteilung gegen

$$\sup_{0 \leq z \leq 1} |B(z)|$$

konvergiert. Diese Verteilung heißt auch Kolmogoroff-Smirnoff-(KS-)Verteilung und hat die folgende Verteilungsfunktion F :

$$F(t) = \begin{cases} 0, & t < 0, \\ 1 + 2 \sum_{i=1}^{\infty} (-1)^i \exp(-2i^2 t^2), & t \geq 0. \end{cases}$$

α	0.8	0.9	0.95	0.98	0.99
q_α	1.073	1.224	1.358	1.517	1.628

Tabelle 1: Quantile q_α für verschiedene Niveaus α der KS-Verteilung

Tabelle 1 zeigt ausgewählte Quantile:

Im Folgenden beweisen wir einige Lemmata; für jedes Lemma sei vorausgesetzt, dass die Nullhypothese wahr ist. Zunächst:

Lemma 3.1 Für beliebiges $\epsilon > 0$ gilt auf $D[0, 1]^5$:

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(\cdot)} U_i^{**} := \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(\cdot)} \begin{pmatrix} X_i^2 - m_x^2 \\ Y_i^2 - m_y^2 \\ X_i - \mu_x \\ Y_i - \mu_y \\ X_i Y_i - m_{xy} \end{pmatrix} \rightarrow_d D_1^{\frac{1}{2}} W_5(\cdot),$$

wobei

$$D_1 = D_1' = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{j=1}^T \sum_{i=1}^T \begin{pmatrix} \text{Cov}(X_i^2, X_j^2) & \text{Cov}(X_i^2, Y_j^2) & \text{Cov}(X_i^2, X_j) & \text{Cov}(X_i^2, Y_j) & \text{Cov}(X_i^2, X_j Y_j) \\ \cdot & \text{Cov}(Y_i^2, Y_j^2) & \text{Cov}(Y_i^2, X_j) & \text{Cov}(Y_i^2, Y_j) & \text{Cov}(Y_i^2, X_j Y_j) \\ \cdot & \cdot & \text{Cov}(X_i, X_j) & \text{Cov}(X_i, Y_j) & \text{Cov}(X_i, X_j Y_j) \\ \cdot & \cdot & \cdot & \text{Cov}(Y_i, Y_j) & \text{Cov}(Y_i, X_j Y_j) \\ \cdot & \cdot & \cdot & \cdot & \text{Cov}(X_i Y_i, X_j Y_j) \end{pmatrix} \in \mathbb{R}^{5 \times 5}.$$

Beweis. Es gilt

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} \begin{pmatrix} X_i^2 - m_x^2 \\ Y_i^2 - m_y^2 \\ X_i - \mu_x \\ Y_i - \mu_y \\ X_i Y_i - m_{xy} \end{pmatrix} &= \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} \begin{pmatrix} X_i^2 - \mathbb{E}(X_i^2) \\ Y_i^2 - \mathbb{E}(Y_i^2) \\ X_i - \mathbb{E}(X_i) \\ Y_i - \mathbb{E}(Y_i) \\ X_i Y_i - \mathbb{E}(X_i Y_i) \end{pmatrix} + \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} \begin{pmatrix} \mathbb{E}(X_i^2) - m_x^2 \\ \mathbb{E}(Y_i^2) - m_y^2 \\ \mathbb{E}(X_i) - \mu_x \\ \mathbb{E}(Y_i) - \mu_y \\ \mathbb{E}(X_i Y_i) - m_{xy} \end{pmatrix} \\ &=: A_1 + A_2. \end{aligned}$$

Betrachte nun die erste Komponente von A_2 :

$$A_{2,1}(z) = \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} (\mathbb{E}(X_i^2) - m_x^2).$$

Es gilt

$$\|A_{2,1}(z)\|_{D[\epsilon,1]} \leq \frac{1}{\sqrt{T}} \sum_{i=1}^T |\mathbb{E}(X_i^2) - m_x^2|.$$

Mit Annahme (A4) konvergiert die rechte Seite gegen 0. Mit denselben Argumenten konvergieren die anderen vier Komponenten von A_2 ebenfalls gegen 0, damit konvergiert A_2 insgesamt in der Supremumsnorm gegen 0.

Weiterhin ist

$$\begin{aligned} A_1 &= \frac{1}{\sqrt{T}} \sum_{i=1}^{\lceil Tz \rceil} \begin{pmatrix} X_i^2 - \mathbb{E}(X_i^2) \\ Y_i^2 - \mathbb{E}(Y_i^2) \\ X_i - \mathbb{E}(X_i) \\ Y_i - \mathbb{E}(Y_i) \\ X_i Y_i - \mathbb{E}(X_i Y_i) \end{pmatrix} + \frac{1}{\sqrt{T}} \sum_{i=\lceil Tz+1 \rceil}^{\lceil Tz+(1-z)2 \rceil} \begin{pmatrix} X_i^2 - \mathbb{E}(X_i^2) \\ Y_i^2 - \mathbb{E}(Y_i^2) \\ X_i - \mathbb{E}(X_i) \\ Y_i - \mathbb{E}(Y_i) \\ X_i Y_i - \mathbb{E}(X_i Y_i) \end{pmatrix} \quad (13) \\ &=: A_3 + A_4. \end{aligned}$$

Wir zeigen, dass A_4 für jedes feste z in Wahrscheinlichkeit gegen 0 konvergiert. Falls

$$\lceil Tz + (1-z)2 \rceil < \lceil Tz + 1 \rceil,$$

ist A_4 ohnehin gleich 0. Für

$$[Tz + (1 - z)2] \geq [Tz + 1]$$

argumentieren wir mit der Markov-Ungleichung. A_4 besteht aus maximal zwei Summanden und der Erwartungswert der ersten Komponente A_{41} lässt sich wie folgt abschätzen:

$$\mathbb{E} \left(\left| \frac{1}{\sqrt{T}} \sum_{i=[Tz+1]}^{[Tz+(1-z)2]} (X_i^2 - \mathbb{E}(X_i^2)) \right| \right) \leq \frac{1}{\sqrt{T}} 2 \sup_{i \in \mathbb{N}} \mathbb{E}(|X_i^2 - \mathbb{E}(X_i^2)|).$$

Wegen der gleichmäßigen Beschränktheit des zweiten Moments der X_i (Annahme (A2)) ist $\sup_{i \in \mathbb{N}} \mathbb{E}(|X_i^2 - \mathbb{E}(X_i^2)|)$ endlich. Für $T \rightarrow \infty$ konvergiert die rechte Seite also gegen 0. Mit der Markov-Ungleichung folgt dann für beliebiges $\epsilon > 0$

$$\begin{aligned} & \mathbb{P} \left(\left| \frac{1}{\sqrt{T}} \sum_{i=[Tz+1]}^{[Tz+(1-z)2]} (X_i^2 - \mathbb{E}(X_i^2)) \right| > \epsilon \right) \\ & \leq \frac{1}{\epsilon} \frac{1}{\sqrt{T}} 2 \sup_{i \in \mathbb{N}} \mathbb{E}(|X_i^2 - \mathbb{E}(X_i^2)|) \rightarrow 0 \quad (T \rightarrow \infty). \end{aligned}$$

Für die anderen Komponenten gilt das Entsprechende.

Es folgt mit Theorem 2.8, dass auch alle endlich-dimensionalen Verteilungen in Wahrscheinlichkeit und damit in Verteilung gegen 0 konvergieren. Die Straffheit des Prozesses zeigen wir angelehnt an die Methode auf Seite 138 in Billingsley (1968), die Theorem 2.19 verwendet. Wir zeigen die Straffheit jeder einzelnen Komponente (exemplarisch für die erste), woraus dann die Straffheit des Vektors folgt.

$$B := \mathbb{E} (|A_{41}(t) - A_{41}(t_1)|^{1+\frac{\alpha}{2}} \cdot |A_{41}(t_2) - A_{41}(t)|^{1+\frac{\alpha}{2}}) \leq \frac{1}{T^{1+\frac{\alpha}{2}}} C$$

für $\epsilon \leq t_1 \leq t \leq t_2 \leq 1$ und für eine Konstante C wegen der gleichmäßigen Beschränktheit. Falls $[Tt_2] - [Tt_1] = 0$, so ist $B = 0$, falls $[Tt_2] - [Tt_1] \geq 1$, so gilt

$$\frac{1}{T^{1+\frac{\alpha}{2}}} C \leq ([Tt_2] - [Tt_1])^{1+\frac{\alpha}{2}} \frac{1}{T^{1+\frac{\alpha}{2}}} C = C \left(\frac{[Tt_2] - [Tt_1]}{T} \right)^{1+\frac{\alpha}{2}}$$

und die Bedingung von Theorem 2.19 ist erfüllt.

Damit konvergiert A_4 als Prozess in Verteilung gegen die Nullfunktion und damit auch in Wahrscheinlichkeit gegen die Nullfunktion.

Wende auf A_3 das Invarianzprinzip aus Theorem 2.21 an. Die Bedingungen hierfür sind erfüllt: Seien

$$U_i = \left(X_i^2 - \mathbb{E}(X_i^2), Y_i^2 - \mathbb{E}(Y_i^2), X_i - \mathbb{E}(X_i), Y_i - \mathbb{E}(Y_i), X_i Y_i - \mathbb{E}(X_i Y_i) \right)'$$

und $V_i = D_1^{-\frac{1}{2}} \frac{1}{\sqrt{T}} U_i$ sowie $K_T(z) = [Tz]$ und $X_T(z) = \sum_{i=1}^{[Tz]} V_i$. Die Matrix $D_1^{-\frac{1}{2}}$ existiert, da D_1 wegen der positiven Definitheit insbesondere invertierbar ist. Wegen der Linearität des Erwartungswertes sind die V_i zentriert. Betrachte zunächst Bedingung 2 aus Theorem 2.21. Für $z = 0$ ist die Bedingung sicher erfüllt mit beliebigem Σ , wenn die Summe von $i = 1$ bis 0 als 0 definiert wird, für $z > 0$ gilt

$$\mathbb{E}(X_T(z)X_T(z)') = \frac{[Tz]}{T} D_1^{-\frac{1}{2}} \mathbb{E} \left(\frac{1}{[Tz]} S_{[Tz]} S'_{[Tz]} \right) D_1^{-\frac{1}{2}} \rightarrow z I_5 \quad (14)$$

mit Annahme (A1). Die Matrix Σ wird als die fünfdimensionale Einheitsmatrix I_5 gewählt. Die Grenzmatrix D_1 hat die angegebene Gestalt, da in D_1 laut Annahme (A1) Erwartungswerte von Produkten von Zufallsvariablen, die um ihren Erwartungswert vermindert werden, und damit Kovarianzen, stehen.

Für Bedingung 1 aus Theorem 2.21 sei ein fünfdimensionaler Vektor λ mit $\lambda' \lambda = 1$ fest vorgegeben. Wähle dann $c_{T,i}^\lambda = (\lambda' D_1^{-1} T^{-1} \lambda)^{\frac{1}{2}}$.

Für das Schema $\lambda' V_i$ gilt Bedingung 1 aus Theorem 2.20 wegen der Zentriertheit der U_i . Da $\frac{\lambda' V_i}{c_{T,i}^\lambda}$ eine Linearkombination der U_i ist und sich die quadrierten Gewichte zu 1 addieren, gilt Bedingung 2 wegen Annahme (A2) der gleichmäßigen Beschränktheit der Momente der U_i . Bedingung 3 ist erfüllt, weil $\lambda' V_i$ eine Linearkombination der U_i ist und sich an der NED-Bedingung dadurch nichts ändert. Bei Bedingung 4 ist

$$v_T^2(z, \delta) = ([T(z + \delta)] - (1 + [Tz])) \lambda' D_1^{-1} T^{-1} \lambda = \frac{[T(z + \delta)] - (1 + [Tz])}{T} \lambda' D_1^{-1} \lambda.$$

Der Vorfaktor konvergiert gegen δ und damit ist die Bedingung

$$\sup_{z \in [0,1], \delta \in (0,1-z]} \left(\limsup_{T \rightarrow \infty} \frac{v_T^2(z, \delta)}{\delta} \right) < \infty$$

erfüllt, weil D_1^{-1} eine feste Matrix und damit $\lambda'D_1^{-1}\lambda$ eine feste Zahl ist. Bedingung 5 ist erfüllt für jedes γ zwischen $\frac{1}{2}$ und 1, da $c_{T,i}^\lambda = O(T^{-\frac{1}{2}})$ gilt. Bedingung 6 gilt, weil für festes $z \in [0, 1]$ mit Gleichung (14)

$$\mathbb{E}(\lambda'X_T(z))^2 = \lambda'\mathbb{E}(X_T(z)X_T(z)')\lambda \rightarrow \lambda'\lambda z = z$$

gilt. Es folgt also

$$D_1^{-\frac{1}{2}}A_3 \rightarrow_d W_5(\cdot).$$

Mit dem Satz von Cramér konvergiert A_3 als Prozess gegen $D_1^{\frac{1}{2}}W_5(\cdot)$. Da nun A_2 und A_4 in Wahrscheinlichkeit gegen 0 konvergieren, folgt mit dem Satz von Cramér die Behauptung. □

Wir benötigen im Folgenden die Aussage von Lemma 3.1 nur für das Intervall $[\epsilon, 1]$ für beliebiges $\epsilon > 0$.

Lemma 3.2 *Für ein beliebiges $\epsilon > 0$ gilt für $z \in [\epsilon, 1]$:*

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} U_i^{**} = \frac{\tau(z)}{\sqrt{T}} \begin{pmatrix} \overline{(X^2)}_{\tau(z)} & - & m_x^2 \\ \overline{(Y^2)}_{\tau(z)} & - & m_y^2 \\ \bar{X}_{\tau(z)} & - & \mu_x \\ \bar{Y}_{\tau(z)} & - & \mu_y \\ \overline{(XY)}_{\tau(z)} & - & m_{xy} \end{pmatrix}.$$

Beweis. Dividiere durch $\tau(z) \neq 0$. □

Lemma 3.3

$$\frac{\tau(\cdot)}{\sqrt{T}} \begin{pmatrix} \overline{(X^2)}_{\tau(\cdot)} - (\bar{X}_{\tau(\cdot)})^2 & - & \sigma_x^2 \\ \overline{(Y^2)}_{\tau(\cdot)} - (\bar{Y}_{\tau(\cdot)})^2 & - & \sigma_y^2 \\ \overline{(XY)}_{\tau(\cdot)} - \bar{X}_{\tau(\cdot)}\bar{Y}_{\tau(\cdot)} & - & \sigma_{xy} \end{pmatrix} \rightarrow_d D_2 D_1^{\frac{1}{2}} W_5(\cdot)$$

auf $D[\epsilon, 1]^3$ für ein beliebiges $\epsilon > 0$, wobei

$$D_2 = \begin{pmatrix} 1 & 0 & -2\mu_x & 0 & 0 \\ 0 & 1 & 0 & -2\mu_y & 0 \\ 0 & 0 & -\mu_y & -\mu_x & 1 \end{pmatrix} \in \mathbb{R}^{3 \times 5}.$$

Beweis. Wir wenden die verallgemeinerte Delta-Methode an mit der Funktion

$$f : D[\epsilon, 1]^5 \rightarrow D[\epsilon, 1]^3,$$

$$f(x_1, x_2, x_3, x_4, x_5) = \begin{pmatrix} x_1 - x_3^2 \\ x_2 - x_4^2 \\ x_5 - x_3x_4 \end{pmatrix}$$

für ein beliebiges $\epsilon > 0$.

Die verallgemeinerte Hadamard-Ableitung ist mit Theorem 2.24 für $\theta = (\theta_1 \ \theta_2 \ \theta_3 \ \theta_4 \ \theta_5)' \in D[\epsilon, 1]^5$

$$f'_\theta : D[\epsilon, 1]^5 \rightarrow D[\epsilon, 1]^3,$$

$$f'_\theta(h) = f'_\theta(h_1, h_2, h_3, h_4, h_5) = \begin{pmatrix} 1 & 0 & -2\theta_3 & 0 & 0 \\ 0 & 1 & 0 & -2\theta_4 & 0 \\ 0 & 0 & -\theta_4 & -\theta_3 & 1 \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \end{pmatrix}.$$

Die Bedingungen des Theorems sind sicher erfüllt, da die Funktion eine Zusammensetzung von Polynomen ist. Die Anwendung der Delta-Methode (Theorem 2.25) liefert dann die Behauptung. Hier ist $\mathbb{G}_1 = D[\epsilon, 1]^5$, $\mathbb{G}_2 = D[\epsilon, 1]^3$ sowie

$$M_T(z) = \left(\overline{(X^2)}_{\tau(z)}, \overline{(Y^2)}_{\tau(z)}, \bar{X}_{\tau(z)}, \bar{Y}_{\tau(z)}, \overline{(XY)}_{\tau(z)} \right)'.$$

Außerdem werden die Definitionen

$$\sigma_x^2 := m_x^2 - \mu_x^2, \sigma_y^2 := m_y^2 - \mu_y^2, \sigma_{xy} := m_{xy} - \mu_x\mu_y$$

benutzt. □

Lemma 3.4

$$\frac{\tau(\cdot)}{\sqrt{T}}(\hat{\rho}_{\tau(\cdot)} - \rho_0^*) \rightarrow_d D_3 D_2 D_1^{\frac{1}{2}} W_5(\cdot)$$

auf $D[\epsilon, 1]$ für ein beliebiges $\epsilon > 0$, wobei

$$D_3 = \begin{pmatrix} -\frac{1}{2} \frac{\sigma_{xy}}{\sigma_y} \sigma_x^{-3} & -\frac{1}{2} \frac{\sigma_{xy}}{\sigma_x} \sigma_y^{-3} & \frac{1}{\sigma_x \sigma_y} \end{pmatrix} \in \mathbb{R}^{1 \times 3}.$$

Beweis. Wir wenden die verallgemeinerte Delta-Methode an mit der Funktion

$$f : D^+[\epsilon, 1]^2 \times D[\epsilon, 1] \rightarrow D[\epsilon, 1],$$

$$f(x_1, x_2, x_3) = \frac{x_3}{\sqrt{x_1 x_2}}$$

für ein beliebiges $\epsilon > 0$. Die verallgemeinerte Hadamard-Ableitung ist mit Theorem 2.24 für $\theta = \left(\theta_1 \quad \theta_2 \quad \theta_3 \right)' \in D^+[\epsilon, 1]^2 \times D[\epsilon, 1]$ mit

$$\theta_1(z) \geq c_1 > 0, \theta_2(z) \geq c_2 > 0 \quad \forall z \in [\epsilon, 1]$$

gegeben durch

$$f'_\theta : D[\epsilon, 1]^3 \rightarrow D[\epsilon, 1],$$

$$f'_\theta(h) = f'_\theta(h_1, h_2, h_3) = \begin{pmatrix} -\frac{1}{2} \frac{\theta_3}{\sqrt{\theta_2}} \theta_1^{-\frac{3}{2}} & -\frac{1}{2} \frac{\theta_3}{\sqrt{\theta_1}} \theta_2^{-\frac{3}{2}} & \frac{1}{\sqrt{\theta_1 \theta_2}} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}.$$

Die Bedingungen des Theorems sind erfüllt, da in den Nennern der zweiten partiellen Ableitungen von

$$\psi : \mathbb{R}^3 \rightarrow \mathbb{R},$$

$$\psi(x_1, x_2, x_3) = \frac{x_3}{\sqrt{x_1 x_2}}$$

nur x_1 und x_2 auftauchen; durch Verwendung des Raumes $D^+[\epsilon, 1]^2$ ist dann garantiert, dass die Ableitungen beschränkt bleiben.

Die Anwendung der Delta-Methode (Theorem 2.25 und Korollar 2.26) liefert jetzt die Behauptung: Hier ist $\mathbb{G}_1^* = D^+[\epsilon, 1]^2 \times D[\epsilon, 1]$, $\mathbb{G}_1 = D[\epsilon, 1]^3$, $\mathbb{G}_2 = D[\epsilon, 1]$ sowie

$$M_T(z) = \left(\overline{(X^2)}_{\tau(z)} - (\bar{X}_{\tau(z)})^2, \overline{(Y^2)}_{\tau(z)} - (\bar{Y}_{\tau(z)})^2, \overline{(XY)}_{\tau(z)} - \bar{X}_{\tau(z)}\bar{Y}_{\tau(z)} \right)'.$$

Außerdem ist $\theta(z) = \begin{pmatrix} \sigma_x^2 & \sigma_y^2 & \sigma_{xy} \end{pmatrix} \forall z \in [\epsilon, 1]$. Die ersten beiden Komponenten sind nach Voraussetzung größer als 0. Wegen des starken Gesetzes der großen Zahlen, vgl. Theorem A.1, ist dann die Bedingung von Korollar 2.26 erfüllt. \square

Für das spätere Theorem 3.10 ist es wichtig, durch Anwendung der Delta-Methode $\hat{\rho}_{\tau(z)}$ und ρ_0^* zu trennen. Daher ist es nicht möglich, $\tau(z)$ wie in Lemma 3.1 in den Vektor als Faktor zu schreiben. Wäre dies möglich, wäre die Theorie der verallgemeinerten Delta-Methode nicht notwendig, da der Vorfaktor dann eine reelle Zahl und keine Funktion wäre.

Korollar 3.5 *Es gilt*

$$(D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}} \frac{\tau(\cdot)}{\sqrt{T}} (\hat{\rho}_{\tau(\cdot)} - \rho_0^*) \rightarrow_d W_1(\cdot)$$

auf $D[\epsilon, 1]$ für ein beliebiges $\epsilon > 0$.

Beweis. Aus Lemma 3.4 ist bekannt, dass

$$\frac{\tau(\cdot)}{\sqrt{T}} (\hat{\rho}_{\tau(\cdot)} - \rho_0^*) \rightarrow_d D_3 D_2 D_1^{\frac{1}{2}} W_5(\cdot).$$

Zeige zunächst, dass

$$W^*(z) := D_3 D_2 D_1^{\frac{1}{2}} W_5(z) \stackrel{L}{=} (D_3 D_2 D_1 D_2' D_3')^{\frac{1}{2}} W_1(z) =: W^{**}(z).$$

Hierzu müssen die beiden Axiome der Brownschen Bewegung nachgewiesen werden.

- Für festes z ist $W^*(z) \stackrel{L}{=} W^{**}(z)$, da Linearkombinationen von Normalverteilungen wieder normalverteilt sind (mit entsprechender Änderung der Varianz) und die Komponenten von $W_5(z)$ unabhängig und identisch verteilt sind.

- Für $t_1 < \dots < t_n$ sind die Zuwächse $W^*(t_1), W^*(t_2 - t_1), \dots, W^*(t_n - t_{n-1})$ unabhängig, da für $i \in \{1, \dots, 5\}$ die Komponenten $W_{5,i}(t_1), W_{5,i}(t_2 - t_1), \dots, W_{5,i}(t_n - t_{n-1})$ jeweils unabhängig sind. Außerdem sind $W^{**}(t_1), W^{**}(t_2 - t_1), \dots, W^{**}(t_n - t_{n-1})$ unabhängig, da $W_1(t_1), W_1(t_2 - t_1), \dots, W_1(t_n - t_{n-1})$ unabhängig sind. Dabei wird jeweils ausgenutzt, dass die Unabhängigkeit bei Verknüpfungen erhalten bleibt.

Durch Multiplikation mit $(D_3 D_2 D_1 D_2' D_3')^{\frac{1}{2}}$ und Anwendung des Satzes von Cramér folgt die Behauptung. \square

Lemma 3.6 *Es gibt einen konsistenten Schätzer \hat{D} für $(D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}}$, der nur von T , nicht aber von z abhängt.*

Beweis. Es ist zunächst

$$E = D_2 D_1 D_2' = \begin{pmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{pmatrix}$$

mit

$$\begin{aligned} E_{11} &= D_{1,11} - 4\mu_x D_{1,13} + 4\mu_x^2 D_{1,34}, \\ E_{12} &= E_{21} = D_{1,12} - 2\mu_x D_{1,23} - 2\mu_y D_{1,14} + 4\mu_x \mu_y D_{1,34}, \\ E_{22} &= D_{1,22} - 4\mu_y D_{1,24} + 4\mu_y^2 D_{1,44}, \\ E_{13} &= E_{31} = -\mu_y D_{1,13} + 2\mu_x \mu_y D_{1,33} - \mu_x D_{1,14} + 2\mu_x^2 D_{1,34} + D_{1,15} - 2\mu_x D_{1,35}, \\ E_{23} &= E_{32} = -\mu_y D_{1,23} + 2\mu_x \mu_y D_{1,44} - \mu_x D_{1,24} + 2\mu_y^2 D_{1,34} + D_{1,25} - 2\mu_y D_{1,45}, \\ E_{33} &= -\mu_y^2 D_{1,33} + 2\mu_x \mu_y D_{1,34} - 2\mu_y D_{1,35} + \mu_x^2 D_{1,44} + D_{1,55} - 2\mu_x D_{1,45}. \end{aligned}$$

Weiterhin:

$$\begin{aligned} D_3 D_2 D_1 D_2' &= \begin{pmatrix} D_{3,1} E_{11} + D_{3,2} E_{21} + D_{3,3} E_{31} \\ D_{3,1} E_{12} + D_{3,2} E_{22} + D_{3,3} E_{32} \\ D_{3,1} E_{13} + D_{3,2} E_{23} + D_{3,3} E_{33} \end{pmatrix}' \\ &=: \begin{pmatrix} F_1 & F_2 & F_3 \end{pmatrix}. \end{aligned}$$

und

$$(D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}} = (F_1 D_{3,1} + F_2 D_{3,2} + F_3 D_{3,3})^{-\frac{1}{2}}.$$

$(D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}}$ ist also eine stetige Verknüpfung von Momenten der X_i und Y_i , die in den Matrizen D_3 und E stehen. Wegen des Satzes von Cramér sind also konsistente Schätzer für sämtliche Elemente aus D_3 und E zu finden. In D_3 stehen stetige Verknüpfungen der Grenzkovarianzen und Grenzvarianzen, in E stehen Summen von Elementen aus D_1 , eventuell multipliziert mit anderen Faktoren. Diese Faktoren sind entweder konstant oder durch Mittelwerte zu schätzen, sodass das Problem gelöst ist, wenn wir die Matrix D_1 konsistent schätzen können. Dazu werden wir das in Theorem 2.22 beschriebene Kernschätzverfahren anwenden. Wir wählen als Bandbreite $\gamma_T = [\log T]$ und als Kern $k(\cdot)$ den Tukey-Hanning-Kern. Es wären auch andere Wahlen möglich, doch diese ist vergleichsweise einfach.

Zunächst geben wir einen Schätzer für D_1 in Abhängigkeit von

$\theta_0 = \begin{pmatrix} m_x^2 & m_y^2 & \mu_x & \mu_y & m_{xy} \end{pmatrix} \in (\mathbb{R} \cap (0, \infty))^2 \times \mathbb{R}^3$ an. Die Matrix D_1 offensichtlich dasselbe asymptotische Verhalten wie

$$\Sigma_T^* = \sum_{i=1}^T \sum_{j=1}^T \mathbb{E} V_i V_j'$$

mit $V_i = \frac{1}{\sqrt{T}} U_i$ und U_i aus Annahme (A1). Ein konsistenter Schätzer hierfür ist gemäß Theorem 2.22

$$\hat{\Sigma}_T^* = \sum_{i=1}^T \sum_{j=1}^T V_i V_j' k\left(\frac{i-j}{\gamma_T}\right).$$

Die erste Bedingung von Theorem 2.22 ist erfüllt durch Wahl des Tukey-Hanning-Kerns, die zweite durch die NED-Bedingung, die Wahl von $d_{T,i} = T^{-\frac{1}{2}}$ und die Beschränktheit von $c_{T,i}$, die dritte durch die Wahl der Bandbreite. Nachrechnen unter Ausnutzen von Bedingung (A4) zeigt, dass $\hat{\Sigma}_T^*$ dasselbe asymptotisch Verhalten wie

$$\hat{\Sigma}_T = \sum_{i=1}^T \sum_{j=1}^T V_i^{**} V_j^{**'} k\left(\frac{i-j}{\gamma_T}\right)$$

mit $V_i^{**} = \frac{1}{\sqrt{T}}U_i^{**}$ und U_i^{**} aus Lemma 3.1 hat. Exemplarisch betrachten wir das Element der ersten Zeile und zweiten Spalte von $\hat{\Sigma}_T^*$, genannt $\hat{\Sigma}_{T,12}^*$, die anderen Elemente werden dann analog betrachtet.

$$\hat{\Sigma}_{T,12}^* = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^{\lfloor \log T \rfloor} (X_i^2 - \mathbb{E}(X_i^2))(Y_{i+j}^2 - \mathbb{E}(Y_{i+j}^2))k\left(\frac{i-j}{\gamma_T}\right)$$

und außerdem

$$\hat{\Sigma}_{T,12} = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^{\lfloor \log T \rfloor} (X_i^2 - m_x^2)(Y_{i+j}^2 - m_y^2)k\left(\frac{i-j}{\gamma_T}\right).$$

Die Differenz von $\hat{\Sigma}_{T,12}^*$ und $\hat{\Sigma}_{T,12}$ besteht aus endlich vielen Summanden der Gestalt

$$D = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^{\lfloor \log T \rfloor} d_{X^2_i Y_{i+j}^2} k\left(\frac{i-j}{\gamma_T}\right).$$

Unter Ausnutzen von Bedingung (A4) und der gleichmäßigen Beschränktheit der Momente gilt, dass $\mathbb{E}(|D|) \rightarrow 0$, daraus folgt dann die Zwischenbehauptung.

Der Vektor U_i^{**} hängt von θ_0 ab, als konsistente Schätzfolge wählen wir

$$\theta_T = \left(\overline{(X^2)}_T \quad \overline{(Y^2)}_T \quad \bar{X}_T \quad \bar{Y}_T \quad \overline{(XY)}_T \right).$$

Durch Nachrechnen zeigt sich, dass auch Bedingung 4 aus Theorem 2.22 erfüllt ist. Bedingung 4.i folgt direkt aus Lemma 3.1, weil die schwache Konvergenz gegen die Brownsche Bewegung insbesondere die Verteilungskonvergenz in $z = 1$ impliziert. Sei jetzt M die Ableitungsmatrix von V_i nach θ mit den Komponenten M_{jk} , dann ist $M = -\frac{1}{\sqrt{T}}I_5$. Der Ausdruck aus Bedingung 4.ii ist dann für jede Komponente gleich 1 oder 0, also sicher stetig, die beiden Ausdrücke aus Bedingung 4.iii sind ebenfalls gleich 1 oder 0, also sicher beschränkt. Damit folgt dann die Behauptung. \square

Eine explizite Darstellung des Schätzers folgt in

Korollar 3.7 *Der Schätzer \hat{D} lässt sich schreiben als*

$$\hat{D} = (\hat{F}_1 \hat{D}_{3,1} + \hat{F}_2 \hat{D}_{3,2} + \hat{F}_3 \hat{D}_{3,3})^{-\frac{1}{2}}$$

mit

$$\begin{pmatrix} \hat{F}_1 & \hat{F}_2 & \hat{F}_3 \end{pmatrix} = \begin{pmatrix} \hat{D}_{3,1}\hat{E}_{11} + \hat{D}_{3,2}\hat{E}_{21} + \hat{D}_{3,3}\hat{E}_{31} \\ \hat{D}_{3,1}\hat{E}_{12} + \hat{D}_{3,2}\hat{E}_{22} + \hat{D}_{3,3}\hat{E}_{32} \\ \hat{D}_{3,1}\hat{E}_{13} + \hat{D}_{3,2}\hat{E}_{23} + \hat{D}_{3,3}\hat{E}_{33} \end{pmatrix}',$$

$$\hat{E}_{11} = \hat{D}_{1,11} - 4\hat{\mu}_x\hat{D}_{1,13} + 4\hat{\mu}_x^2\hat{D}_{1,33},$$

$$\hat{E}_{12} = \hat{E}_{21} = \hat{D}_{1,12} - 2\hat{\mu}_x\hat{D}_{1,23} - 2\hat{\mu}_y\hat{D}_{1,14} + 4\hat{\mu}_x\hat{\mu}_y\hat{D}_{1,34},$$

$$\hat{E}_{22} = \hat{D}_{1,22} - 4\hat{\mu}_y\hat{D}_{1,24} + 4\hat{\mu}_y^2\hat{D}_{1,44},$$

$$\hat{E}_{13} = \hat{E}_{31} = -\hat{\mu}_y\hat{D}_{1,13} + 2\hat{\mu}_x\hat{\mu}_y\hat{D}_{1,33} - \hat{\mu}_x\hat{D}_{1,14} + 2\hat{\mu}_x^2\hat{D}_{1,34} + \hat{D}_{1,15} - 2\hat{\mu}_x\hat{D}_{1,35},$$

$$\hat{E}_{23} = \hat{E}_{32} = -\hat{\mu}_y\hat{D}_{1,23} + 2\hat{\mu}_x\hat{\mu}_y\hat{D}_{1,44} - \hat{\mu}_x\hat{D}_{1,24} + 2\hat{\mu}_y^2\hat{D}_{1,34} + \hat{D}_{1,25} - 2\hat{\mu}_y\hat{D}_{1,45},$$

$$\hat{E}_{33} = \hat{\mu}_y^2\hat{D}_{1,33} + 2\hat{\mu}_x\hat{\mu}_y\hat{D}_{1,34} - 2\hat{\mu}_y\hat{D}_{1,35} + \hat{\mu}_x^2\hat{D}_{1,44} + \hat{D}_{1,55} - 2\hat{\mu}_x\hat{D}_{1,45},$$

$$\hat{D}_1 = \begin{pmatrix} \hat{D}_{1,11} & \hat{D}_{1,12} & \hat{D}_{1,13} & \hat{D}_{1,14} & \hat{D}_{1,15} \\ \hat{D}_{1,21} & \hat{D}_{1,22} & \hat{D}_{1,23} & \hat{D}_{1,24} & \hat{D}_{1,25} \\ \hat{D}_{1,31} & \hat{D}_{1,32} & \hat{D}_{1,33} & \hat{D}_{1,34} & \hat{D}_{1,35} \\ \hat{D}_{1,41} & \hat{D}_{1,42} & \hat{D}_{1,43} & \hat{D}_{1,44} & \hat{D}_{1,45} \\ \hat{D}_{1,51} & \hat{D}_{1,52} & \hat{D}_{1,53} & \hat{D}_{1,54} & \hat{D}_{1,55} \end{pmatrix} = \sum_{i=1}^T \sum_{j=1}^T k\left(\frac{t-s}{\gamma_T}\right) V_i V_j',$$

$$V_i = \frac{1}{\sqrt{T}} U_i^{***}, \gamma_T = [\log T],$$

$$U_i^{***} = \left(X_i^2 - \overline{(X^2)}_T \quad Y_i^2 - \overline{(Y^2)}_T \quad X_i - \bar{X}_T \quad Y_i - \bar{Y}_T \quad X_i Y_i - \overline{(XY)}_T \right)',$$

$$k(x) = \begin{cases} \frac{1+\cos(\pi x)}{2}, & |x| \leq 1 \\ 0, & \text{sonst} \end{cases},$$

$$\hat{D}_{1,11} = \frac{1}{T} \sum_{i=1}^T X_i^4 - \left(\frac{1}{T} \sum_{i=1}^T X_i^2 \right)^2 + 2 \cdot \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} X_i^2 X_{i+j}^2 - \left(\frac{1}{T} \sum_{i=1}^T X_i^2 \right)^2 \right),$$

$$\begin{aligned}
\hat{D}_{1,12} &= \frac{1}{T} \sum_{i=1}^T X_i^2 Y_i^2 - \frac{1}{T} \sum_{i=1}^T X_i^2 \frac{1}{T} \sum_{i=1}^T Y_i^2 + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} X_i^2 Y_{i+j}^2 + \frac{1}{T} \sum_{i=1}^{T-j} Y_i^2 X_{i+j}^2 - 2 \cdot \frac{1}{T} \sum_{i=1}^T X_i^2 \frac{1}{T} \sum_{i=1}^T Y_i^2 \right), \\
\hat{D}_{1,13} &= \frac{1}{T} \sum_{i=1}^T X_i^3 - \frac{1}{T} \sum_{i=1}^T X_i^2 \frac{1}{T} \sum_{i=1}^T X_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} X_i^2 X_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} X_i X_{i+j}^2 - 2 \cdot \frac{1}{T} \sum_{i=1}^T X_i^2 \frac{1}{T} \sum_{i=1}^T X_i \right), \\
\hat{D}_{1,14} &= \frac{1}{T} \sum_{i=1}^T X_i^2 Y_i - \frac{1}{T} \sum_{i=1}^T X_i^2 \frac{1}{T} \sum_{i=1}^T Y_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} X_i^2 Y_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} Y_i X_{i+j}^2 - 2 \cdot \frac{1}{T} \sum_{i=1}^T X_i^2 \frac{1}{T} \sum_{i=1}^T Y_i \right), \\
\hat{D}_{1,15} &= \frac{1}{T} \sum_{i=1}^T X_i^2 X_i Y_i - \frac{1}{T} \sum_{i=1}^T X_i^2 \frac{1}{T} \sum_{i=1}^T X_i Y_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} X_i^2 X_{i+j} Y_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} X_i Y_i X_{i+j}^2 - 2 \cdot \frac{1}{T} \sum_{i=1}^T X_i^2 \frac{1}{T} \sum_{i=1}^T X_i Y_i \right), \\
\hat{D}_{1,22} &= \frac{1}{T} \sum_{i=1}^T Y_i^4 - \left(\frac{1}{T} \sum_{i=1}^T Y_i^2 \right)^2 + 2 \cdot \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} Y_i^2 Y_{i+j}^2 - \left(\frac{1}{T} \sum_{i=1}^T Y_i^2 \right)^2 \right), \\
\hat{D}_{1,23} &= \frac{1}{T} \sum_{i=1}^T Y_i^2 X_i - \frac{1}{T} \sum_{i=1}^T Y_i^2 \frac{1}{T} \sum_{i=1}^T X_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} Y_i^2 X_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} X_i Y_{i+j}^2 - 2 \cdot \frac{1}{T} \sum_{i=1}^T Y_i^2 \frac{1}{T} \sum_{i=1}^T X_i \right), \\
\hat{D}_{1,24} &= \frac{1}{T} \sum_{i=1}^T Y_i^3 - \frac{1}{T} \sum_{i=1}^T Y_i^2 \frac{1}{T} \sum_{i=1}^T Y_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} Y_i^2 Y_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} Y_i Y_{i+j}^2 - 2 \cdot \frac{1}{T} \sum_{i=1}^T Y_i^2 \frac{1}{T} \sum_{i=1}^T Y_i \right),
\end{aligned}$$

$$\begin{aligned}
\hat{D}_{1,25} &= \frac{1}{T} \sum_{i=1}^T X_i Y_i^3 - \frac{1}{T} \sum_{i=1}^T Y_i^2 \frac{1}{T} \sum_{i=1}^T X_i Y_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} Y_i^2 X_{i+j} Y_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} X_i Y_i Y_{i+j}^2 - 2 \cdot \frac{1}{T} \sum_{i=1}^T Y_i^2 \frac{1}{T} \sum_{i=1}^T X_i Y_i \right), \\
\hat{D}_{1,33} &= \frac{1}{T} \sum_{i=1}^T X_i^2 - \left(\frac{1}{T} \sum_{i=1}^T X_i \right)^2 + 2 \cdot \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} X_i X_{i+j} - \left(\frac{1}{T} \sum_{i=1}^T X_i \right)^2 \right), \\
\hat{D}_{1,34} &= \frac{1}{T} \sum_{i=1}^T X_i Y_i - \frac{1}{T} \sum_{i=1}^T X_i \frac{1}{T} \sum_{i=1}^T Y_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} Y_i X_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} X_i Y_{i+j} - 2 \cdot \frac{1}{T} \sum_{i=1}^T X_i \frac{1}{T} \sum_{i=1}^T Y_i \right), \\
\hat{D}_{1,35} &= \frac{1}{T} \sum_{i=1}^T X_i X_i Y_i - \frac{1}{T} \sum_{i=1}^T X_i \frac{1}{T} \sum_{i=1}^T X_i Y_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} X_i X_{i+j} Y_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} X_i Y_i X_{i+j} - 2 \cdot \frac{1}{T} \sum_{i=1}^T X_i \frac{1}{T} \sum_{i=1}^T X_i Y_i \right), \\
\hat{D}_{1,44} &= \frac{1}{T} \sum_{i=1}^T Y_i^2 - \left(\frac{1}{T} \sum_{i=1}^T Y_i \right)^2 + 2 \cdot \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} Y_i Y_{i+j} - \left(\frac{1}{T} \sum_{i=1}^T Y_i \right)^2 \right), \\
\hat{D}_{1,45} &= \frac{1}{T} \sum_{i=1}^T Y_i X_i Y_i - \frac{1}{T} \sum_{i=1}^T Y_i \frac{1}{T} \sum_{i=1}^T X_i Y_i + \\
&\quad \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} Y_i X_{i+j} Y_{i+j} + \frac{1}{T} \sum_{i=1}^{T-j} X_i Y_i Y_{i+j} - 2 \cdot \frac{1}{T} \sum_{i=1}^T Y_i \frac{1}{T} \sum_{i=1}^T X_i Y_i \right), \\
\hat{D}_{1,55} &= \frac{1}{T} \sum_{i=1}^T X_i^2 Y_i^2 - \left(\frac{1}{T} \sum_{i=1}^T X_i Y_i \right)^2 + \\
&\quad 2 \cdot \sum_{j=1}^{[\log T]} k_j \left(\frac{1}{T} \sum_{i=1}^{T-j} X_i Y_i X_{i+j} Y_{i+j} - \left(\frac{1}{T} \sum_{i=1}^T X_i Y_i \right)^2 \right),
\end{aligned}$$

$$k_j = \frac{1 + \cos(\pi j)}{2},$$

$$\hat{\mu}_x = \bar{X}_T,$$

$$\hat{\mu}_y = \bar{Y}_T,$$

$$\hat{D}_{3,1} = -\frac{1}{2} \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_y} \hat{\sigma}_x^{-3},$$

$$\hat{D}_{3,2} = -\frac{1}{2} \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x} \hat{\sigma}_y^{-3},$$

$$\hat{D}_{3,3} = \frac{1}{\hat{\sigma}_x \hat{\sigma}_y}$$

und

$$\hat{\sigma}_x^2 = \overline{(X^2)}_T - (\bar{X}_T)^2,$$

$$\hat{\sigma}_y^2 = \overline{(Y^2)}_T - (\bar{Y}_T)^2,$$

$$\hat{\sigma}_{xy} = \overline{(XY)}_T - \bar{X}_T \bar{Y}_T.$$

Durch Anwendung des Satzes von Cramér und Kombination von Korollar 3.5 und Lemma 3.6 erhalten wir jetzt sofort:

Korollar 3.8

$$W_T(\cdot) := \hat{D} \frac{\tau(\cdot)}{\sqrt{T}} (\hat{\rho}_{\tau(\cdot)} - \rho_0^*) \rightarrow_d W_1(\cdot)$$

auf $D[\epsilon, 1]$ für ein beliebiges $\epsilon > 0$.

Der nächste Schritt ist der Nachweis der Konvergenz auf dem Intervall $[0, 1]$. Die Delta-Methode hilft hier nicht, da diese ohne das ϵ als feste Schranke zur 0 nicht anwendbar wäre: Es gilt

$$\left\| \frac{1}{r_T} \right\|_{D[0,1]} = \left\| \frac{\sqrt{T}}{\tau(z)} \right\|_{D[0,1]} = \frac{\sqrt{T}}{2} \rightarrow \infty \neq 0 \text{ für } T \rightarrow \infty,$$

sodass eine Voraussetzung für die Anwendung nicht erfüllt ist. Auf $[\epsilon, 1]$ ist die Voraussetzung erfüllt, da

$$\left\| \frac{1}{r_T} \right\|_{D[\epsilon, 1]} = \left\| \frac{\sqrt{T}}{\tau(z)} \right\|_{D[\epsilon, 1]} \leq \frac{\sqrt{T}}{\epsilon T} \rightarrow 0 \text{ für } T \rightarrow \infty.$$

Wir werden aber sehen, dass es genügt zu zeigen, dass das Supremum von W_T auf dem Intervall $[0, \epsilon]$ von der Größenordnung ϵ ist, also für beliebig kleines ϵ beliebig klein wird.

Lemma 3.9

$$W_T(\cdot) := \hat{D} \frac{\tau(\cdot)}{\sqrt{T}} (\hat{\rho}_{\tau(\cdot)} - \rho_0^*) \rightarrow_d W_1(\cdot)$$

auf $D[0, 1]$.

Beweis. Wir definieren die folgenden Funktionen:

$$W_T^\epsilon(z) = \begin{cases} W_T(z), & z \geq \epsilon \\ 0 & z < \epsilon \end{cases},$$

$$W^\epsilon(z) = \begin{cases} W_1(z), & z \geq \epsilon \\ 0 & z < \epsilon \end{cases}.$$

Es gilt mit Korollar 3.8

$$W_T^\epsilon(\cdot) \rightarrow_d W^\epsilon(\cdot)$$

für $T \rightarrow \infty$ auf $D[0, 1]$ und außerdem

$$W^\epsilon(\cdot) \rightarrow_d W_1(\cdot)$$

für rationale $\epsilon \rightarrow 0$ auf $D[0, 1]$.

Die Konvergenz von $W_T(\cdot)$ auf $D[0, 1]$ folgt mit Theorem A.4, wenn wir zeigen können, dass

$$\lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P}(\sup_{z \in [0, 1]} |W_T^\epsilon(z) - W_T(z)| \geq \eta) = \lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P}(\sup_{z \in [0, \epsilon]} |W_T(z)| \geq \eta) = 0$$

für alle $\eta > 0$. Nun gilt mit

$$D_2(z) = \sqrt{\frac{1}{\tau(z)} \sum_{i=1}^{\tau(z)} (X_i - \bar{X}_{\tau(z)})^2 \frac{1}{\tau(z)} \sum_{i=1}^{\tau(z)} (Y_i - \bar{Y}_{\tau(z)})^2}$$

die Gleichung

$$\begin{aligned} & \sup_{z \in [0, \epsilon]} |W_T(z)| \\ &= \hat{D} \sup_{z \in [0, \epsilon]} \left| \frac{\frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} (X_i - \bar{X}_{\tau(z)}) (Y_i - \bar{Y}_{\tau(z)}) - \frac{\rho_0^*}{\sqrt{T}} \tau(z) D_2(z)}{D_2(z)} \right| \\ &=: \hat{D} \sup_{z \in [0, \epsilon]} \left| \frac{D_1(z)}{D_2(z)} \right|. \end{aligned}$$

Aufgrund eines starken Gesetzes der großen Zahlen, vgl. Theorem A.1, und des CMT strebt D_2 gegen $\sigma_x \sigma_y$ \mathbb{P} -fast sicher für festes $z > 0$ für $T \rightarrow \infty$. Analoges gilt für $\frac{1}{\tau(z)} \sum_{i=1}^{\tau(z)} (Y_i - \bar{Y}_{\tau(z)})^2$ mit dem Grenzwert σ_y^2 sowie \bar{X}_T und \bar{Y}_T mit dem Grenzwert μ_x bzw. μ_y . Sei nun $\delta > 0$ beliebig. Mit dem Theorem von Egoroff, vgl. Theorem A.5, gibt es eine Menge $\Omega_\delta \subset \Omega$ mit $\mathbb{P}(\Omega_\delta) \geq 1 - \delta$ und eine Zahl $M(\delta) > 0$, sodass

$$\left| \frac{1}{\tau(z)} \sum_{i=1}^{\tau(z)} (Y_i - \bar{Y}_{\tau(z)})^2 - \sigma_y^2 \right| < \delta,$$

$|D_2(z) - \sigma_x \sigma_y| < \delta$, $|\bar{X}_{\tau(z)} - \mu_x| < \delta$ sowie $|\bar{Y}_{\tau(z)} - \mu_y| < \delta$ auf Ω_δ für $\tau(z) \geq M(\delta)$. Aufgrund der Aussage für D_2 gilt für $z \geq \frac{M(\delta)}{T}$ für hinreichend große T

$$\sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} \left| \frac{1}{D_2(z)} \right| \leq \frac{1}{\sigma_x \sigma_y - \delta} < \infty.$$

Betrachte nun $D_1(z)$. Nach Einfügen einer künstlichen $0 \left(= -\frac{\tau(z)}{\sqrt{T}} \sigma_{xy} + \frac{\tau(z)}{\sqrt{T}} \sigma_{xy} \right)$ in $D_1(z)$ lässt sich D_1 abschätzen durch

$$\sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} |D_1(z)| \leq A + B$$

mit

$$A = \sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} \left| \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} (X_i - \bar{X}_{\tau(z)})(Y_i - \bar{Y}_{\tau(z)}) - \frac{\tau(z)}{\sqrt{T}} \sigma_{xy} \right| \text{ und}$$

$$B = \sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} \left| \frac{\tau(z)}{\sqrt{T}} \sigma_{xy} - \frac{\tau(z)}{\sqrt{T}} \sigma_{xy} \left(\frac{1}{\sigma_x \sigma_y} D_2(z) \right) \right|.$$

Nach Anwendung der Dreiecksungleichung, des Invarianzprinzips und des CMT gilt

$$A \leq \sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} D_4(z),$$

wobei $D_4(z)$ aus vier Summanden $D_{4i}(z), i \in \{1, \dots, 4\}$, besteht, für die jeweils gilt:

$$\sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} |D_{4i}(z)| \rightarrow_d A_i(\delta) \sup_{z \in [0, \epsilon]} |W_1(z)|$$

mit Konstanten $A_i(\delta)$. Weiterhin ist mit der 3. Binomischen Formel

$$B = \sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} \left| \frac{\tau(z)}{\sqrt{T}} \sigma_{xy} \left(\frac{\sigma_x \sigma_y - D_2(z)}{\sigma_x \sigma_y} \right) \right|$$

$$\leq \sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} \left| \frac{\tau(z)}{\sqrt{T}} \frac{\sigma_{xy}}{2\sigma_x \sigma_y} \left(\frac{1}{\tau(z)} \sum_{i=1}^{\tau(z)} (X_i - \bar{X}_{\tau(z)})^2 \frac{1}{\tau(z)} \sum_{i=1}^{\tau(z)} (Y_i - \bar{Y}_{\tau(z)})^2 - \sigma_x^2 \sigma_y^2 \right) \right|$$

$$\leq \sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} \left| \frac{\tau(z)}{\sqrt{T}} \frac{\sigma_{xy}}{2\sigma_x \sigma_y} \left(\frac{1}{\tau(z)} \sum_{i=1}^{\tau(z)} (X_i - \bar{X}_{\tau(z)})^2 (\sigma_y + \delta) - \sigma_x^2 \sigma_y^2 \right) \right|.$$

Nach Anwendung der Dreiecksungleichung, geeigneten Divisionen, des Invarianzprinzips und des CMT gilt

$$B \leq \sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} D_5(z),$$

wobei $D_5(z)$ aus vier Summanden $D_{5i}(z), i \in \{1, \dots, 4\}$, besteht, für die jeweils gilt:

$$\sup_{z \in [\frac{M(\delta)}{T}, \epsilon]} |D_{5i}(z)| \rightarrow_d B_i(\delta) \sup_{z \in [0, \epsilon]} |W_1(z)|$$

mit Konstanten $B_i(\delta)$. Es gilt aufgrund der Beschränktheit des Korrelationskoeffizienten und $\frac{\tau(z)}{\sqrt{T}}$ auf dem Intervall $\left[0, \frac{M(\delta)}{T}\right]$

$$\sup_{z \in \left[0, \frac{M(\delta)}{T}\right]} |W_T(z)| \leq \frac{C_2(\delta)}{\sqrt{T}}$$

für eine Konstante $C_2(\delta)$; dieser Ausdruck geht gegen 0 für $T \rightarrow \infty$.

Da $W(0) = 0$ \mathbb{P} -fast überall, gilt

$$\lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P}\left(\sup_{z \in [0, \epsilon]} |W_T(z)| \geq \eta\right) = 0$$

auf Ω_δ . Da $\delta > 0$ beliebig war, folgt das Lemma. \square

$W_T(z)$ hängt von ρ_0^* ab und ist deshalb noch keine sinnvolle Teststatistik. Durch Verwendung der Brownschen Brücke erhalten wir eine von ρ_0^* unabhängige und somit sinnvolle Teststatistik.

Theorem 3.10

$$B_T(\cdot) := \hat{D} \frac{\tau(\cdot)}{\sqrt{T}} (\hat{\rho}_{\tau(\cdot)} - \hat{\rho}_T) \rightarrow_d B(\cdot)$$

auf $D[0, 1]$.

Beweis. Es gilt

$$B_T(z) = W_T(z) - \frac{\tau(z)}{T} W_T(1) =: h\left(W_T(z), \frac{\tau(z)}{T}\right)$$

mit

$$h : D[0, 1] \times D[0, 1] \rightarrow D[0, 1],$$

$$h(x, y) = x - y\pi_1(x)$$

mit π_1 der Projektion auf 1. h ist stetig, weil die Projektion auf 1 stetig ist und weil die Addition stetig ist. Da $W_T(\cdot)$ als Prozess in Verteilung für $T \rightarrow \infty$ gegen $W_1(\cdot)$

konvergiert (Lemma 3.8) und $\frac{\tau(z)}{T}$ als reelle Zahl für $T \rightarrow \infty$ gegen z konvergiert, gilt mit dem CMT

$$B_T \rightarrow_d B,$$

mit

$$B(z) = W_1(z) - zW_1(1),$$

der eindimensionalen Brownsche Brücke. □

Korollar 3.11

$$\sup_{z \in [0,1]} \left| \hat{D} \frac{\tau(z)}{\sqrt{T}} (\hat{\rho}_{\tau(z)} - \hat{\rho}_T) \right| \rightarrow_d \sup_{z \in [0,1]} |B(z)|.$$

Beweis. Wende das CMT mit der stetigen Funktion

$$h : D[0, 1] \rightarrow \mathbb{R},$$

$$h(x) = \sup_{t \in [0,1]} |x(t)|$$

an. □

Die Konvergenz findet im Gegensatz zu Theorem 3.10 hier nicht mehr auf $D[0, 1]$, sondern auf \mathbb{R} statt.

4 Lokale Güte

4.1 Allgemeine Überlegungen

Im weiteren soll die Alternative

$$\rho_{i,T} = \rho_0 + \frac{1}{\sqrt{T}}g\left(\frac{i}{T}\right)$$

für $i \in \{1, \dots, T\}$ und eine beschränkte, durch Treppenfunktionen approximierbare Funktion g betrachtet werden. Hierbei gilt $g(z) \neq 0$ für mindestens ein $z \in [\epsilon, 1]$, außerdem hängen die Werte von g nicht von T ab. Zur Vereinfachung nehmen wir an, dass die ersten und zweiten Momente der X_i und Y_i konstant bleiben und gleich μ_x , μ_y , m_x^2 bzw. m_y^2 sind. Die Kovarianz ändert sich über die Zeit:

$$(A5) \quad \mathbb{E}(X_i Y_i) = \sigma_{xy} + \frac{1}{\sqrt{T}}g\left(\frac{i}{T}\right), \\ \mathbb{E}(X_i^2) = m_x^2, \mathbb{E}(Y_i^2) = m_y^2, \mathbb{E}(X_i) = \mu_x, \mathbb{E}(Y_i) = \mu_y.$$

g könnte beispielsweise eine Treppenfunktion sein, die an einer Stelle z_0 einen Sprung von 0 auf g_0 hat, das heißt

$$g(z) = \begin{cases} 0, & z \leq z_0, \\ g_0, & z > z_0. \end{cases} \quad (15)$$

Die Kovarianz ändert sich dann nach dem Zeitpunkt $[Tz_0]$.

Alle Annahmen aus Kapitel 3 bis auf Annahme (A4) sollen weiter gelten. Annahme (A4) ist insofern verletzt, als dass $\mathbb{E}(X_i Y_i) = \sigma_{xy} + d_i$ mit

$$\lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} d_i = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^{\tau(z)} g\left(\frac{i}{T}\right) = \int_0^z g(u) du \neq 0$$

für mindestens ein $z \in [0, 1]$ ist. Die Darstellung als Integral folgt aus Lemma A.3. Daran wird deutlich, dass es sinnvoll ist, gerade $\frac{1}{\sqrt{T}}$ als Änderungsgeschwindigkeit zu wählen. Nur dadurch ist der Grenzwert eine von 0 und ∞ verschiedene Zahl.

Die Idee ist, dass sich Lemma 3.1 prinzipiell - wenn auch mit anderem Grenzwert - übertragen lässt.

Lemma 4.1 *Mit den Bezeichnungen aus Lemma 3.1 gilt auf $D[\epsilon, 1]^5$:*

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(\cdot)} U_i := \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(\cdot)} \begin{pmatrix} X_i^2 - m_x^2 \\ Y_i^2 - m_y^2 \\ X_i - \mu_x \\ Y_i - \mu_y \\ X_i Y_i - m_{xy} \end{pmatrix} \rightarrow_d D_1^{\frac{1}{2}} W_5(\cdot) + A,$$

wobei $A = \left(0 \ 0 \ 0 \ 0 \ \int_0^z g(u) du \right)'$.

Beweis. Führe den Beweis von Lemma 3.1 durch, wobei das dortige A_2 nun die folgende Gestalt hat:

$$A_2 = \frac{1}{\sqrt{T}} \sum_{i=1}^{\tau(z)} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{T}} g\left(\frac{i}{T}\right) \end{pmatrix}.$$

Die fünfte Komponente konvergiert als Prozess gegen $\int_0^\cdot g(u) du$, A_1 konvergiert als Prozess gegen $D_1^{\frac{1}{2}} W_5(\cdot)$. Da $\int_0^\cdot g(u) du$ eine deterministische Funktion ist, folgt mit dem Satz von Cramér die Behauptung. \square

Auch alle anderen Beweise bis Korollar 3.5 lassen sich analog übertragen. Es folgt

$$\begin{aligned} \frac{\tau(\cdot)}{\sqrt{T}} (\hat{\rho}_{\tau(\cdot)} - \rho_0) &\rightarrow_d D_3 D_2 D_1^{\frac{1}{2}} W_5(\cdot) + D_3 D_2 A \\ &\stackrel{L}{=} (D_3 D_2 D_1 D_2' D_3')^{\frac{1}{2}} W_1(\cdot) + D_3 D_2 A \end{aligned}$$

und

$$\begin{aligned} (D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}} \frac{\tau(\cdot)}{\sqrt{T}} (\hat{\rho}_{\tau(\cdot)} - \rho_0) &\rightarrow_d W_1(\cdot) + (D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}} D_3 D_2 A \\ &\stackrel{L}{=} W_1(\cdot) + (D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}} \cdot \frac{\int_0^\cdot g(u) du}{\sigma_x \sigma_y}. \end{aligned}$$

Der Schätzer \hat{D} aus der Teststatistik konvergiert wie in Kapitel 3 in Wahrscheinlichkeit gegen $(D_3 D_2 D_1 D'_2 D'_3)^{-\frac{1}{2}}$, da es für die Konvergenz irrelevant ist, welcher Grenzwert konkret vorliegt. Die neue Annahme (A5) ist auch kein Widerspruch dazu. Es ergibt sich nun wie im Beweis von Theorem 3.10

$$\hat{D} \frac{\tau(\cdot)}{\sqrt{T}} (\hat{\rho}_{\tau(\cdot)} - \hat{\rho}_T) \rightarrow_d B(\cdot) + C_1(\cdot),$$

mit

$$C_1(z) = \frac{(D_3 D_2 D_1 D'_2 D'_3)^{-\frac{1}{2}}}{\sigma_x \sigma_y} \left(\int_0^z g(u) du - z \int_0^1 g(u) du \right),$$

eine deterministische, von z abhängige Funktion.

Damit gilt auch

Theorem 4.2 *Unter den Annahmen (A1), (A2), (A3) und (A5) gilt*

$$\sup_{z \in [0,1]} \left| \hat{D} \frac{\tau(z)}{\sqrt{T}} (\hat{\rho}_{\tau(z)} - \hat{\rho}_T) \right| \rightarrow_d \sup_{z \in [0,1]} |B(z) + C_1(z)| \quad (16)$$

mit

$$C_1(z) = \frac{(D_3 D_2 D_1 D'_2 D'_3)^{-\frac{1}{2}}}{\sigma_x \sigma_y} \left(\int_0^z g(u) du - z \int_0^1 g(u) du \right).$$

Diese Verteilung ist unhandlich, aber wir können relativ einfach eine Aussage über die Güte für beliebig großes g machen. Hierzu setzen wir in Annahme (A5) $g(z) = Mh(z)$ für eine Funktion h und eine Konstante M . Dann gilt

Theorem 4.3 *Sei $P_{H_1}(M)$ die Ablehnwahrscheinlichkeit für gegebenes M , falls die Alternative wahr ist. Sei $\epsilon > 0$ und h beliebig, aber nichtkonstant. Dann gibt es ein M_0 , sodass*

$$\lim_{T \rightarrow \infty} P_{H_1}(M) > 1 - \epsilon$$

für alle $M > M_0$.

Beweis. Analog zu der Rechnung vor dem Theorem gilt

$$\begin{aligned} \sup_{z \in [0,1]} \left| \hat{D} \frac{\tau(z)}{\sqrt{T}} (\hat{\rho}_{\tau(z)} - \rho_T) \right| &\rightarrow_d \sup_{z \in [0,1]} |B(z) + MC_1(z)| \\ &= M \sup_{z \in [0,1]} \left| \frac{B(z)}{M} + C_1(z) \right|, \end{aligned}$$

wobei C_1 für mindestens ein z ungleich 0 ist. Daher gilt also

$$M \sup_{z \in [0,1]} \left| \frac{B(z)}{M} + C_1(z) \right| \geq MC_2$$

für eine Konstante C_2 . Die Teststatistik wird demnach beliebig groß, insbesondere auch größer als jedes Quantil der Verteilung unter H_0 . \square

Für den Beweis ist entscheidend, dass h nichtkonstant ist, da die Teststatistik sonst gleich $\sup_{z \in [0,1]} |B(z)|$ wäre. Da wir Mh von 0 bis 1 integrieren, werden asymptotisch sowohl sehr frühe als auch sehr späte Strukturbrüche erkannt, sofern M hinreichend groß ist.

Wenn kein M verwendet wird und die Störung g zu klein ist, verwirft der Test nicht in jedem Fall. Sei etwa

$$\sup_{z \in [0,1]} |C_1(z)| < q_{0.95} \approx 1.358.$$

Dann ist

$$\begin{aligned} \mathbb{P}(\sup_{z \in [0,1]} |B(z) + C_1(z)| \geq q_{0.95}) &\leq \mathbb{P}(\sup_{z \in [0,1]} |B(z)| + \sup_{z \in [0,1]} |C_1(z)| \geq q_{0.95}) \\ &= \mathbb{P}(\sup_{z \in [0,1]} |B(z)| \geq q_{0.95} - \sup_{z \in [0,1]} |C_1(z)|) < 1, \end{aligned}$$

das heißt, der Test behält mit positiver Wahrscheinlichkeit die Nullhypothese bei.

Generell ist zu erwarten, dass Strukturbrüche in der Mitte besser erkannt werden als am Rand. Die Varianz der Brownschen Brücke an der Stelle z ist gleich $z(1-z)$ und damit ist sie für $z = \frac{1}{2}$ am größten. Das Maximum des Absolutbetrags sollte also eher in der Mitte angenommen werden. Dies gilt tendenziell wohl auch schon für endliche Stichproben, insbesondere aber asymptotisch.

4.2 Konstruktionsbedingte Ablehneigenschaften

Ein Problem ist, dass der Test auch dann verwirft, wenn die Korrelation zwar konstant bleibt und die Erwartungswerte gleich 0 sind, aber die Varianzen sich zu stark ändern, das heißt etwa

$$\begin{aligned} \text{Var}(X_i) &= \sigma_x^2 + T^{-\frac{1}{2}+\epsilon} g \left(\frac{i}{T} \right) \\ \text{Var}(Y_i) &= \sigma_y^2 + T^{-\frac{1}{2}+\epsilon} h \left(\frac{i}{T} \right) \end{aligned}$$

für ein $\epsilon > 0$ und g und h wie zum Beispiel in (15). Eine der beiden Varianzen kann durchaus auch konstant sein. Dann ist Annahme (A4) verletzt und in der Matrix A in Lemma 4.1 steht in einigen Einträgen im Grenzwert unendlich. Somit ist keine sinnvolle Konvergenz mehr möglich.

Auch für $\epsilon = 0$, das heißt für eine Änderungsgeschwindigkeit der Varianz von $\frac{1}{\sqrt{T}}$, gilt die Konvergenz im Allgemeinen nicht, weil dann Annahme (A4) verletzt ist. So kann etwa für $h = 0$ und $\sigma_{xy} = 0$ eine analoge Rechnung wie in Kapitel 4.1 durchgeführt werden und es ist zu erkennen, dass die Grenzverteilung nicht gleich dem gewöhnlichen Maximum des Absolutbetrags einer Brownschen Brücke ist.

Anders, das heißt beruhigender, sieht es aus, wenn die Erwartungswerte gleich 0 sind, das Verhältnis beider Varianzen gleich ist und sich die Varianzen jeweils gemäß dem Prinzip in (A5) ändern. Dies ist in dem folgenden Szenario dargestellt:

$$\begin{aligned} \text{Var}(X_i) &= a_2 + a_2 \frac{1}{\sqrt{T}} g \left(\frac{i}{T} \right) \\ \text{Var}(Y_i) &= a_3 + a_3 \frac{1}{\sqrt{T}} g \left(\frac{i}{T} \right) \\ \text{Cov}(X_i, Y_i) &= a_1 + a_1 \frac{1}{\sqrt{T}} g \left(\frac{i}{T} \right). \end{aligned} \tag{17}$$

mit $a_1 \leq \sqrt{a_2 a_3}$ und einer Gestalt von g wie zum Beispiel in (15).

Die Korrelation ist konstant gleich

$$\frac{a_1}{\sqrt{a_2 a_3}}.$$

Ein solches Szenario ist bei Aktienrenditen durchaus plausibel, wenn wir annehmen, dass die Varianzen in etwa gleich stark von äußeren Einflüssen (etwa Turbulenzen auf den Finanzmärkten) abhängen.

Wir können dieselbe Rechnung wie in Kapitel 4.1 durchführen und erhalten

$$\begin{aligned} & \sup_{z \in [0,1]} \left| \hat{D} \frac{\tau(z)}{\sqrt{T}} (\hat{\rho}_{\tau(z)} - \rho_T) \right| \rightarrow_d \sup_{z \in [0,1]} \left| B(z) + (D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}} \right. \\ & \cdot \left(-\frac{a_1}{2\sqrt{a_3 a_2^{\frac{3}{2}}}} \cdot a_2 - \frac{a_1}{2\sqrt{a_2 a_3^{\frac{3}{2}}}} \cdot a_3 + \frac{1}{\sqrt{a_2 a_3}} \cdot a_1 \right) \left(\int_0^z g(u) du - z \int_0^1 g(u) du \right) \left. \right| \\ & = \sup_{z \in [0,1]} |B(z)|, \end{aligned}$$

also die bekannte Grenzverteilung.

Kleinere Abweichungen von der Nullhypothese erkennt der Test nicht. Wenn etwa die Varianzen konstant sind und die Änderung der Kovarianz asymptotisch schneller als $\frac{1}{\sqrt{T}}$ verschwindet, also zum Beispiel

$$Cov(X_i, Y_i) = \sigma_{xy} + \frac{1}{T} g\left(\frac{i}{T}\right)$$

gilt, dann ist Annahme (A4) erfüllt und die Verteilung der Teststatistik konvergiert gegen das Maximum des Absolutbetrags einer Brownschen Brücke. Das liegt daran, dass für die Herleitung der asymptotischen Nullverteilung nicht explizit benutzt wird, dass die Korrelationen konstant sind. Annahme (A4) garantiert lediglich, dass die Korrelationen „konstant genug“ sind. Für die praktische Anwendung ist das jedoch nicht tragisch, da kleinere Abweichungen von der Nullhypothese oftmals gar nicht von Interesse sind.

Ein interessantes Phänomen ist, dass X und Y nicht perfekt korreliert sein dürfen. Dann konvergiert die Teststatistik nicht wie gewünscht, weil D_1 nicht positiv definit ist bzw.

$$(D_3 D_2 D_1 D_2' D_3')^{-\frac{1}{2}}$$

unendlich ist. In endlichen Stichproben könnte es auch Probleme geben, wenn X und Y zwar nicht perfekt, aber trotzdem sehr stark korreliert sind.

Dann ist die Determinante von E relativ klein, weil die Spalten recht ähnlich sind und E nahe an der Singularität. Schätzfehler werden dann beim Invertieren deutlich verstärkt und

$$(D_3 D_2 D_1 D'_2 D'_3)^{-\frac{1}{2}}$$

wird schlechter geschätzt.

5 Methodenvergleich mit anderen Tests

Generell gibt es in der Literatur wenige Arbeiten zum Testen auf konstante Korrelation. Vergleichsweise bekannter ist ein χ^2 -Test von Jennrich (1970), der bis Ende März 2009 laut ISI Web of Knowledge 65 mal zitiert wurde und breite Anwendungen sowohl in der Finanzwirtschaft als auch in der Biometrie fand. Wir vergleichen unseren Test zunächst exemplarisch mit diesem Test.

Jennrich konstruiert den Test für p -dimensionale Korrelationsmatrizen ($p \geq 2$), der Test ist insbesondere auch für den bivariaten Fall anwendbar. Die gesamte Zeitreihe wird von vornherein in zwei bekannte Teile der Längen n_1 und n_2 unterteilt und es wird angenommen, dass die Korrelationen in den beiden Teilen jeweils konstant sind. Der Test überprüft dann, ob sich die Korrelationen in den Teilreihen unterscheiden. Die Zufallsvektoren in den beiden Teilen sind jeweils u.i.v. normalverteilt und auch die beiden Teile sind untereinander unabhängig. Die Verteilung der beiden Teile muss jedoch nicht gleich sein, insbesondere dürfen die Varianzen verschieden sein. Das liegt daran, dass Jennrich in der Herleitung des Tests ausnutzt, dass der empirische Korrelationskoeffizient in den beiden Teilen für sich genommen geeignet skaliert jeweils asymptotisch normalverteilt ist. Für die Herleitung der Verteilung der Teststatistik nutzt er aus, dass unter der Nullhypothese konstanter Korrelation $\sqrt{b} \cdot D$ mit

$$b = \frac{n_1 n_2}{n_1 + n_2}$$

asymptotisch zentriert normalverteilt mit einer Kovarianzmatrix Γ ist. D ist der Vektor, der die Superdiagonalelemente von $D_1 - D_2$ (D_1 : Korrelationsmatrix des ersten Teils, D_2 : Korrelationsmatrix des zweiten Teils) enthält. Dann ist

$$bD'\hat{\Gamma}^{-1}D$$

für einen konsistenten Schätzer $\hat{\Gamma}$ asymptotisch χ^2 -verteilt mit $\frac{p(p-1)}{2}$ Freiheitsgraden. Jennrichs Hauptarbeit liegt darin, ein leicht berechenbares Verfahren für $\hat{\Gamma}^{-1}$ zu finden. Im bivariaten Fall lautet die Teststatistik

$$\frac{1}{2}tr(Z^2) - diag(Z)'S^{-1}diag(Z),$$

wobei Z und S Matrizen sind, in denen Funktionen der empirischen Korrelationen stehen. Das heißt,

$$Z = \sqrt{b}\bar{R}^{-1}(R_1 - R_2) \text{ und}$$

$$S = I_2 + \bar{R} \cdot \bar{R}^{-1}$$

mit der zweidimensionalen Einheitsmatrix I_2 , den beiden empirischen Korrelationsmatrizen R_1 und R_2 in den beiden Teilen sowie

$$\bar{R} = \frac{1}{n_1 + n_2}(n_1 R_1 + n_2 R_2).$$

Die Teststatistik ist asymptotisch χ_1^2 -verteilt. Im Gegensatz zum Fluktuationstest kommen in der Teststatistik nur die Korrelationen, nicht noch andere Momente vor. Eine wichtige Anwendung seines Tests sieht Jennrich in der multivariaten Faktoranalyse. Andere Tests nach einem ähnlichen Prinzip stammen zum Beispiel von Tang (1995) und von Pearson und Wilks (1933), alle diese Tests gehen von festen Bruchpunkten und Normalverteilung aus.

Die Annahmen unseres Fluktuationstests sind teilweise schwächer und der Test ist damit deutlich realitätsnäher. Ein entscheidender Vorteil ist, dass der Bruchpunkt im Vorfeld nicht bekannt sein muss und dass wir keine Normalverteilung voraussetzen. Außerdem erlauben wir eine gewisse Abhängigkeit sowie eine gewisse Änderung der Varianzen.

In der ökonomischen Literatur wird der Test von Jennrich öfter angewandt zum Testen auf konstante Korrelation bei Aktienrenditen, etwa in Kaplanis (1988). Dieses Vorgehen ist nicht einwandfrei, weil die Annahme der Normalverteilung bei Aktienrenditen oft zweifelhaft ist. Diese besitzen im Allgemeinen schwerere Ränder, sodass etwa die Annahme einer t -Verteilung realistischer ist. Der Test von Jennrich hat bei t -Verteilung ein zu hohes empirisches Niveau (vgl. Kapitel 6.1), sodass er dabei auch tendenziell zu häufig ablehnt. Die Testergebnisse können also verfälscht sein.

Andere Autoren aus der ökonomischen Literatur gehen nicht von festen Bruchpunkten aus. So konstruiert Fischer (2007) einen Lagrange-Multiplikatoren-Test, bei dem sich die Korrelation in der Alternative gemäß einem Sinus- und Cosinus-Prinzip

ändert. Fischer nimmt an, dass

$$\rho_i = \beta_0 + \beta_1 \sin\left(\frac{2f\pi i}{T}\right) + \beta_2 \cos\left(\frac{2f\pi i}{T}\right)$$

für eine Zahl $f \in \mathbb{R}$ und testet darauf, dass $\beta_1 = \beta_2 = 0$. Bei ihm sind die Zufallsvektoren u.i.v.. Er verwendet bei der Herleitung der Verteilung der Teststatistik die t -Verteilung, sodass er keine Normalverteilung voraussetzen braucht. Der Test von Fischer ist durchaus realistisch, da bei ihm die Bruchpunkte nicht im Vorfeld bekannt sein müssen, aber ihm gelingt es nicht, die exakte oder asymptotische Verteilung seiner Teststatistik auszurechnen. Auch die u.i.v. Annahme ist recht restriktiv.

Ein Verfahren zum Test auf konstante bedingte Korrelation schlagen Longin und Solnik (1995) vor. Sie treffen eine bestimmte Modellannahme an die Aktienrenditen, und zwar modellieren sie diese als multivariaten GARCH-Prozess. Das Testprinzip ist dann ähnlich zu dem in Fischer (2007): Sie verwenden Terme, die vom Modell der konstanten Korrelation abweichen, etwa einen Term für einen Zeittrend, und testen dann mit regressionsanalytischen Methoden, ob die zugehörigen Parameter signifikant verschieden von 0 sind.

Ein ähnliches Prinzip wie unser Fluktuationstest hat der Fluktuationstest auf Strukturbruch im K -dimensionalen linearen Modell von Ploberger et al. (1989), eine Modifikation des Ansatzes von Sen (1980). Hier werden anstelle der sukzessiv geschätzten Korrelationen die sukzessiv geschätzten Parametervektoren $\hat{\beta}_j$ betrachtet und die Teststatistik ist ein Funktional von

$$\sup_{K \leq j \leq T} \|\hat{\beta}_j - \hat{\beta}_T\|,$$

wobei $\|\cdot\|$ die euklidische Norm im \mathbb{R}^K meint. Geeignet standardisiert ist die asymptotische Verteilung auch hier das Maximum des Absolutbetrags einer Brownschen Brücke. Dieser Fluktuationstest ist verwandt mit dem CUSUM (und CUSUM of squares) Test, der auf rekursiven Residuen beruht, das heißt auf standardisierten Vorhersagefehlern (vgl. zum Beispiel Deng und Perron (2008)). Dufour (1982) zeigte, dass diese rekursiven Residuen aus den sukzessiven ersten Differenzen $\hat{\beta}_j - \hat{\beta}_{j-1}$ erhalten werden können.

Wegen

$$\hat{\beta}_j - \hat{\beta}_T = \sum_{i=j+1}^T (\hat{\beta}_{i-1} - \hat{\beta}_i)$$

basieren beide Tests auf denselben Bausteinen. Bei diesen Tests im linearen Modell gelingt es, schwache Abhängigkeit, das heißt mischende Zufallsvariablen, zu verwenden, was uns bislang nicht gelang.

Der CUSUM-Test mit rekursiven Residuen erkennt späte Strukturbrüche allerdings kaum (vgl. Ploberger und Krämer (1992)), im Gegensatz zu unserem Fluktuationstest. Dies ist anders bei dem CUSUM-Test, der auf OLS-Residuen basiert, vgl. Ploberger und Krämer (1992).

6 Simulationen für endliche Stichproben

In diesem Kapitel untersuchen wir das Verhalten des Fluktuationstests in endlichen Stichproben. Wir überprüfen mit Monte-Carlo-Simulationsstudien, ob das Niveau bei wahrer Nullhypothese eingehalten wird, untersuchen die Güteeigenschaften, schauen, an welcher Stelle das Maximum des Absolutbetrags von

$$P_T(j) := \hat{D} \frac{j}{\sqrt{T}} (\hat{\rho}_j - \hat{\rho}_T)$$

jeweils realisiert wird und stellen typische Verläufe von $\hat{\rho}_j$ und $P_T(j)$ über die Zeit dar. Wenn möglich, binden wir die Ergebnisse an die theoretischen Ergebnisse aus Kapitel 4 an. In Kapitel 6.2 erfolgen weitere theoretische Überlegungen.

Bei allen Simulationen wird ein theoretisches Niveau von 0.05 angelegt. Wenn nicht anders genannt, werden jeweils 100000 Wiederholungen durchgeführt und haben die Zufallsvektoren (X_i, Y_i) jeweils konstante Varianz $Var(X_i) = Var(Y_i) = 1$.

6.1 Die Nullverteilung

Zunächst überprüfen wir, inwiefern der Test das Niveau bei normalverteilten Zufallsvariablen einhält. Tabelle 2 zeigt Ergebnisse einer Monte-Carlo-Studie bei wahrer Nullhypothese und (über die Zeit) unabhängigen zentrierten bivariat normalverteilten Zufallsvektoren (X_i, Y_i) mit verschiedenen Korrelationen.

Insgesamt greift die Asymptotik gut, so liegt das wahre Niveau spätestens bei $T = 1000$ nahe an 0.05. Bei betragsmäßig hoher Korrelation von $\rho = 0.9$ weicht das wahre Niveau am Anfang noch deutlich von 0.05 ab, was nach den Ausführungen am Ende von Kapitel 4.2 zu erwarten war.

Es ist auch von Interesse, was passiert, wenn die Zufallsvektoren nicht normalverteilt sind. Schon vorher ist zu erwarten, dass die Normalverteilung bei der Asymptotik „im Vorteil ist“, weil die Grenzverteilung stark mit der Normalverteilung zusammenhängt. Tabelle 3 zeigt Ergebnisse einer Monte-Carlo-Studie bei wahrer Nullhypothese und (über die Zeit) unabhängigen zentrierten bivariat t_5 -verteilten Zufallsvektoren (X_i, Y_i) mit verschiedenen Korrelationen. Wir wählen die t_5 -Verteilung, weil die t -

Verteilung bei Aktienrenditen sehr beliebt ist und weil aufgrund der Annahme an die Momente mindestens 5 Freiheitsgrade nötig sind, um den Fluktuationstest anwenden zu können. Bei der t_5 -Verteilung ist außerdem am besten zu sehen, inwiefern der Test robust gegen Verteilungen mit schweren Rändern ist. Zum Vergleich simulieren wir auch das Niveau bei dem Test von Jennrich und teilen dabei die Zeitreihe jeweils in gleich große Stücke der Länge $\frac{T}{2}$ auf.

Unser Test ist bei $\rho = 0$ und $|\rho| = 0.5$ leicht konservativ und bei $|\rho| = 0.9$ ist zumindest ab $T = 1000$ das Niveau kleiner als 0.05. Anscheinend ist der Test also robust gegen schwere Ränder. Auffällig ist, dass das Niveau bei steigendem T zunächst sinkt, sich also von 0.05 entfernt, obwohl es für $T \rightarrow \infty$ gegen 0.05 konvergiert. Die Konvergenz gegen das wahre Niveau ist also anscheinend nicht monoton.

Der Test von Jennrich hält das Niveau allerdings in keinem Fall ein und steigt sogar mit wachsendem T , es ist teilweise um den Faktor 4 zu groß. Der Test ist also anscheinend nicht robust gegen Abweichungen von der Normalverteilung.

Was passiert bei Verletzung der Annahmen, etwa bei schwankenden Varianzen? Eine Simulation der Situation (17), die die gleichgerichtete Änderung der beiden Varianzen modelliert, zeigt: Bei unabhängiger bivariater zentrierter Normalverteilung sowie t_5 -Verteilung, $a_1 = 1, a_2 = 2, a_3 = 2, z_0 = \frac{1}{2}, g_0 = 1$ und $T = 1000$ wird das Niveau eingehalten; bei der Normalverteilung liegt es bei 0.050, bei der t -Verteilung bei 0.042. Dies ist also ein erfreuliches Ergebnis.

Bei stärkerem Schwanken der Varianz ist die Konvergenz allerdings nicht mehr gewährleistet. Bei unabhängiger bivariater zentrierter Normalverteilung, $T = 1000$, konstanter Korrelation 0 und Sprung der Varianz von 1 auf 5 nach der Hälfte ist das Niveau 0.223, also viel zu groß.

6.2 Die Güte

Wir vergleichen die Güte des Fluktuationstests in endlichen Stichproben bei fünf verschiedenen Alternativen mit (über die Zeit) unabhängigen zentrierten bivariat t_5 -verteilten Zufallsvektoren (X_i, Y_i) . Einerseits werden die verschiedenen Alternativen

T	$\rho = -0.9$	$\rho = -0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
200	0.192	0.052	0.039	0.052	0.193
500	0.071	0.046	0.041	0.046	0.071
1000	0.053	0.046	0.043	0.046	0.053
2000	0.050	0.047	0.045	0.047	0.049
10000	0.048	0.047	0.048	0.047	0.048

Tabelle 2: Simulationsergebnisse bei wahrer Nullhypothese (Normalverteilung, konstante Varianz) für das empirische Niveau

T	$\rho = -0.9$	$\rho = -0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
200	0.144 (0.164)	0.054 (0.170)	0.039 (0.172)	0.053 (0.171)	0.142 (0.165)
500	0.064 (0.185)	0.040 (0.188)	0.035 (0.189)	0.041 (0.189)	0.064 (0.186)
1000	0.048 (0.198)	0.038 (0.200)	0.034 (0.200)	0.039 (0.200)	0.049 (0.199)
2000	0.043 (0.213)	0.038 (0.214)	0.036 (0.214)	0.038 (0.214)	0.043 (0.213)
10000	0.041 (0.228)	0.041 (0.228)	0.040 (0.228)	0.040 (0.228)	0.040 (0.228)

Tabelle 3: Simulationsergebnisse bei wahrer Nullhypothese (t_5 -Verteilung, konstante Varianz): Empirisches Niveau beim Fluktuationstest und beim Test von Jennrich

beim Fluktuationstest untereinander verglichen, andererseits erfolgt ein Vergleich mit dem Test von Jennrich. Um die starke Niveauüberschreitung bei letzterem Test zu berücksichtigen, simulieren wir für verschiedene Werte von T bei $\rho = 0$ das tatsächliche 0.95-Quantil und führen die Gütebetrachtung damit durch. Die Korrelation $\rho = 0$ wird repräsentativ gewählt, weil Tabelle 3 erkennen lässt, dass das Niveau des Tests zwar von T , aber kaum von ρ abhängt. Beim Test von Jennrich wird die Zeitreihe stets in zwei gleichgroße Stücke der Länge $\frac{T}{2}$ aufgeteilt.

$$(B1) \quad \rho_i = 0.5, i \in \{1, \dots, \frac{T}{2}\} \text{ und } \rho_i = 0.7, i \in \{\frac{T}{2} + 1, \dots, T\}$$

$$(B2) \quad \rho_i = 0.5, i \in \{1, \dots, \frac{T}{4}\} \text{ und } \rho_i = 0.7, i \in \{\frac{T}{4} + 1, \dots, T\}$$

$$(B3) \quad \rho_i = -0.5, i \in \{1, \dots, \frac{T}{2}\} \text{ und } \rho_i = 0.5, i \in \{\frac{T}{2} + 1, \dots, T\}$$

$$(B4) \quad \rho_i = -0.5, i \in \{1, \dots, \frac{T}{4}\} \text{ und } \rho_i = 0.5, i \in \{\frac{T}{4} + 1, \dots, T\}$$

$$(B5) \quad \rho_i = 0.5, i \in \{1, \dots, \frac{T}{4}\}, \rho_i = 0.7, i \in \{\frac{T}{4} + 1, \dots, \frac{3}{4}T\}, \\ \rho_i = 0.5, i \in \{\frac{3}{4}T + 1, \dots, T\}$$

Alternative (B1) beschreibt einen Sprung der Korrelation von 0.5 auf 0.7 in der Mitte der Zeitreihe, Alternative (B2) denselben Sprung nach dem ersten Viertel der Zeitreihe. Bei Alternative (B3) erfolgt ein Sprung von -0.5 auf 0.5 in der Mitte, bei Alternative (B4) erfolgt derselbe Sprung nach dem ersten Viertel. Der Sprung ist bei den letzten beiden Alternativen größer, allerdings sind die Korrelationen betragsmäßig gleich. Es ist von vornherein nicht auszuschließen, dass die betragsmäßige Gleichheit das Testergebnis (im Sinne einer niedrigeren Güte) beeinflussen könnte. So hat etwa der CUSUM-Test basierend auf OLS-Residuen eine niedrige Güte, wenn der die Änderung beschreibende Vektor orthogonal auf dem so genannten Hauptregressor steht, vgl. Ploberger und Krämer (1992). Alternative (B5) beschreibt einen Sprung der Korrelation von 0.5 auf 0.7 nach dem ersten Viertel und einen Fall von 0.7 auf 0.5 nach dem dritten Viertel. Wir nehmen jeweils an, dass die Änderung gemäß dem Prinzip in (A5) ändern. Die Ergebnisse sind in Tabelle 4 zu sehen, zunächst sollen die Ergebnisse des Fluktuationstests interpretiert werden. Sehr hoch ist seine Güte bei den Alternativen (B3)

Alternative	$T = 200$	$T = 500$	$T = 1000$	$T = 2000$
(B1)	0.316 (0.321)	0.556 (0.648)	0.811 (0.898)	0.963 (0.992)
(B2)	0.270 (0.124)	0.437 (0.237)	0.685 (0.410)	0.907 (0.679)
(B3)	0.954 (0.999)	0.995 (1)	0.999 (1)	1 (1)
(B4)	0.879 (0.750)	0.987 (0.971)	0.997 (0.998)	0.999 (1)
(B5)	0.113 (0.050)	0.174 (0.052)	0.338 (0.054)	0.667 (0.053)
0.95-Quantil Jennrich	8.117	8.877	9.462	9.911

Tabelle 4: Güteeigenschaften des Fluktuationstests bei fünf verschiedenen Alternativen und der t_5 -Verteilung, Vergleich mit dem Test von Jennrich (Werte davon in Klammern)

und (B4), bei (B1) und (B2) mittelprächtig, bei (B5) eher niedrig, aber trotzdem noch deutlich über dem Niveau 0.05. Ein Strukturbruch in der Mitte der Zeitreihe wird in endlichen Stichproben anscheinend besser erkannt. Dem Test macht es offenbar nichts aus, dass die Korrelationen sich nur im Vorzeichen unterschieden, aber betragsmäßig gleich sind. Analytisch untermauern wir dieses Ergebnis durch Berechnung der Funktion C_1 aus (16). Die Funktion g , die die Änderung beschreibt, ist hier nicht beschränkt und somit sind die theoretischen Ergebnisse zur Güte streng genommen nicht anwendbar. Jedoch dienen die Rechnungen hier nur zur Illustration zum Verhalten im kleinen Stichproben. Die Größe $k = (D_3 D_2 D_1 D'_2 D'_3)^{-\frac{1}{2}}$ (vgl. (16)) ist wegen der Wahl der t_5 -Verteilung und der Unabhängigkeit über die Zeit endlich. Somit ergibt sich für die einzelnen Alternativen

$$(B1) \quad C_1(z) = \sqrt{T} k^{\frac{1}{5}} \begin{cases} -\frac{1}{2}z, & z \leq \frac{1}{2}, \\ \frac{1}{2}z - \frac{1}{2}, & z > \frac{1}{2}. \end{cases}$$

Alternative	(B3)	(B4)	(B1)	(B2)	(B5)
Maximum	$\frac{1}{4}\sqrt{T}$	$\frac{3}{16}\sqrt{T}$	$\frac{1}{20}\sqrt{T}$	$\frac{3}{80}\sqrt{T}$	$\frac{1}{40}\sqrt{T}$

Tabelle 5: Maxima der Gütefunktionen

$$(B2) \quad C_1(z) = \sqrt{T}k\frac{1}{5} \begin{cases} -\frac{3}{4}z, & z \leq \frac{1}{4}, \\ \frac{1}{4}z - \frac{1}{4}, & z > \frac{1}{4}. \end{cases}$$

$$(B3) \quad C_1(z) = \sqrt{T}k \begin{cases} -\frac{1}{2}z, & z \leq \frac{1}{2}, \\ \frac{1}{2}z - \frac{1}{2}, & z > \frac{1}{2}. \end{cases}$$

$$(B4) \quad C_1(z) = \sqrt{T}k \begin{cases} -\frac{3}{4}z, & z \leq \frac{1}{4}, \\ \frac{1}{4}z - \frac{1}{4}, & z > \frac{1}{4}. \end{cases}$$

$$(B5) \quad C_1(z) = \sqrt{T}k\frac{1}{5} \begin{cases} -\frac{1}{2}z, & z \leq \frac{1}{4}, \\ \frac{1}{2}z - \frac{1}{4}, & \frac{1}{4} < z \leq \frac{3}{4}, \\ -\frac{1}{2}z + \frac{1}{2}, & z > \frac{3}{4}. \end{cases}$$

Die ersten vier Funktionen sind nach unten gerichtete Dreiecksfunktionen, die fünfte Funktion ist eine Zackenfunktion. Die Reihenfolge der betragsmäßigen Maxima ist in Tabelle 5 dargestellt, es ist zu erkennen, dass die Reihenfolge der Maxima jeweils für festes T der Reihenfolge der empirischen Güte entspricht. Dies ist plausibel, weil ein höheres betragsmäßiges Maximum aufgrund der Gestalt in (16) eher zu einer Ablehnung der Nullhypothese führt. Die Tatsache, dass die Maxima jeweils Vielfache von \sqrt{T} sind, spiegelt sich darin wider, dass die Güte für steigende T zunimmt.

Der Vergleich des Fluktuationstests mit dem Test von Jennrich zeigt, dass letzterer überlegen ist bei den Alternativen, die einen Strukturbruch in der Mitte der Zeitreihe beschreiben. Kommt der Strukturbruch jedoch schon früher (Alternative (B2) bzw. zum Teil Alternative (B4)), so hat unser Test eine höhere Güte. Alternative (B5), das heißt ein Sprung der Korrelation nach oben und dann wieder nach unten, wird von dem Test von Jennrich sogar überhaupt nicht entdeckt. Das zeigt den Vorteil der

Tatsache, dass beim Fluktuationstest die möglichen Zeiten des Strukturbruchs a priori nicht bekannt sein müssen. Muss wie bei Jennrich ein Zeitpunkt vorher festgelegt werden und ist diese Festlegung falsch, so verliert der Test an Güte, im Gegensatz zum Fluktuationstest.

6.3 Weitere Simulationsergebnisse

Wir interessieren uns dafür, an welcher Stelle das Maximum von $|P_T(j)|$ in endlichen Stichproben bei wahrer Nullhypothese angenommen wird. Dazu wählen wir $T = 1000$ und unabhängige zentrierte bivariat normalverteilte Zufallsvektoren (X_i, Y_i) mit $\rho = 0$. Das Histogramm der Verteilung der Maximalstelle bei 5000 Wiederholungen ist in Abbildung 1 zu sehen. Es ist zu erkennen, dass das Maximum wie auch erwartet eher in der Mitte angenommen wird. Dies bestärkt die Simulationsergebnisse aus Kapitel 6.2, dass ein Strukturbruch in der Mitte besser erkannt wird als am Anfang und bestärkt außerdem die Vermutung am Ende von Kapitel 4.1.

Da der Wert der Teststatistik stark vom Verhalten von $\hat{\rho}_j$ abhängt, wollen wir dieses bei wahrer Nullhypothese untersuchen. Wir simulieren dazu 1000 Vektoren der Länge $T = 1000$ aus einer (über die Zeit) unabhängigen bivariaten zentrierten Normalverteilung mit $\rho = 0$ und tragen jeweils das empirische 0.025- und das empirische 0.975-Quantil gegen j ab. Dadurch erhalten wir jeweils ein Konfidenzintervall für $\hat{\rho}_j$ (Abbildung 2). Zur Veranschaulichung des Verhaltens von $P_T(j)$ simulieren wir einen anderen Vektors der Länge $T = 1000$ aus einer (über die Zeit) unabhängigen bivariaten zentrierten Normalverteilung mit $\rho = 0$ und tragen ihn ebenfalls gegen j ab (Abbildung 3).

Es ist zu erkennen, dass sich $\hat{\rho}_j$ im Laufe der Zeit stabilisiert, wobei die Werte für kleine j durchaus stark schwanken. Der Verlauf von $P_T(j)$ ähnelt vom Prinzip dem Verlauf einer Brownschen Brücke, wie er in Abbildung 4 mit einem Abtastabstand von $\frac{1}{1000}$ dargestellt ist.

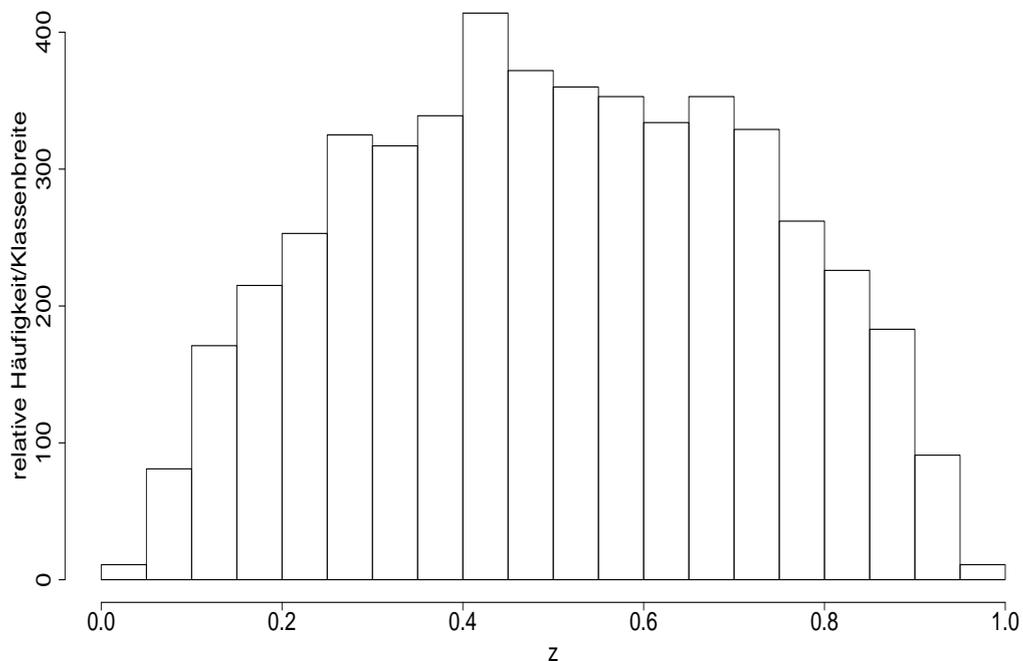


Abbildung 1: Häufigkeitsverteilung der Stelle des Maximums von $|P_T|$

7 Anwendung auf Aktienrenditen

Wir wenden den Fluktuationstest auf tägliche Aktienrenditen von BASF und BMW im Zeitraum Anfang 2003 bis Ende 2008 an, um zu überprüfen, ob die Korrelationen konstant blieben. Nach der Theorie des „Diversification Meltdown“ sollten sie nicht konstant, sondern im Zuge der Finanzkrise höher sein als zuvor. Es gibt $T = 1559$ Datenpunkte und wir erwarten, dass die Asymptotik bereits gut greift. Wir nehmen an, dass die Annahmen erfüllt sind.

Abbildung 5 veranschaulicht das Verhalten der Korrelationen der beiden Renditen im Zeitablauf. In der Grafik sind die „Rolling correlations“ mit einem Fenster von 40 Tagen dargestellt, das heißt, dass in jedem Zeitpunkt jeweils die Korrelationen basierend

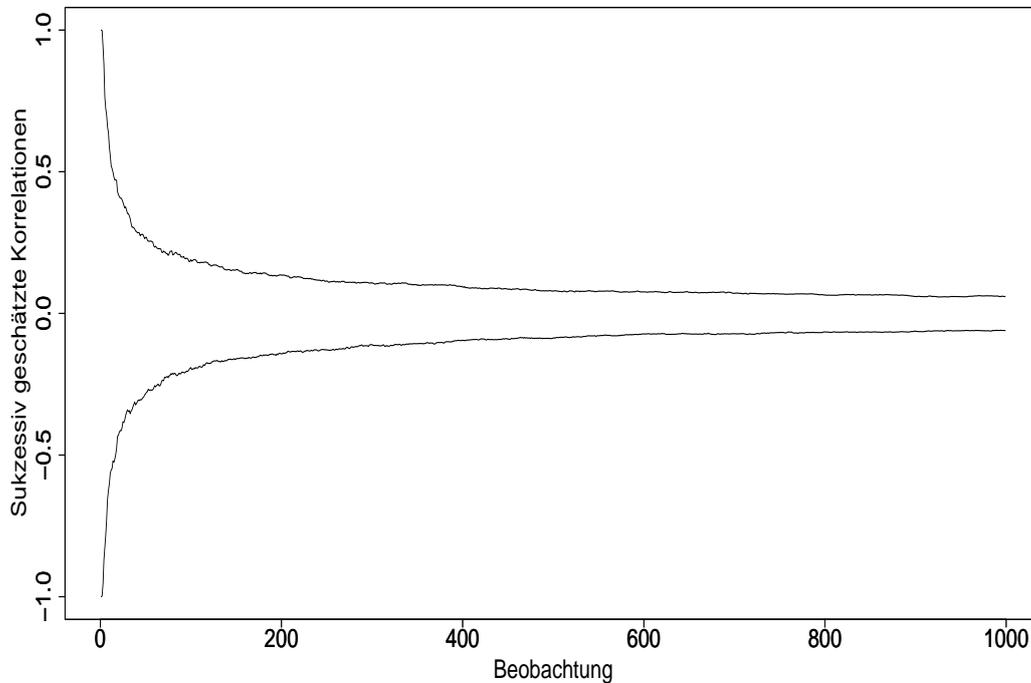


Abbildung 2: Zeitlicher Verlauf von $\hat{\rho}_j$ (Konfidenzband)

auf den letzten 40 Tagen berechnet werden. Anhand der Grafik würde man noch nicht unbedingt vermuten, dass die Korrelationen nicht konstant sind.

Der Test verwirft die Nullhypothese jedoch eindeutig mit einem Wert der Teststatistik von 2.469, was einem p -Wert kleiner als 0.001 entspricht. Wir wollen das Testergebnis durch eine nähere Analyse der Teststatistiken untermauern. Man erkennt aus einer grafischen Darstellung von $\hat{\rho}_j$, dass die Korrelationen offenbar im Zeitablauf schwanken (vgl. Abbildung 6). Bei wahrer Nullhypothese würde sich $\hat{\rho}_j$ im Zeitablauf stabilisieren, wie die Simulationen zeigen (vgl. Abbildung 2). Hier jedoch ist $\hat{\rho}_j$ über einen recht langen Zeitraum relativ konstant und steigt am Ende deutlich an. Der Wert $|P_T(j)|$ nimmt das Maximum weit am Ende der Zeitreihe, das heißt am 10. September 2008, an. Das Maximum ist so groß, dass die Nullhypothese konstanter Korrelation verwor-

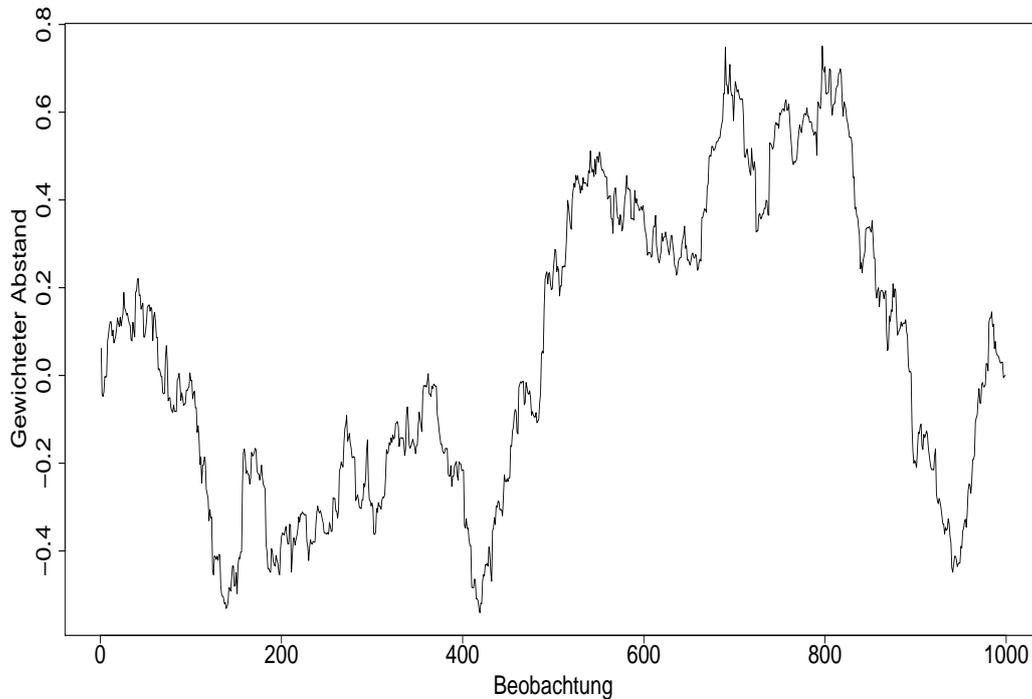


Abbildung 3: Zeitlicher Verlauf von $P_T(j)$ (eine simulierte Realisation)

fen wird.

Das empirische Ergebnis passt zeitlich zur Weltwirtschaftskrise, die seit Ende 2008 drastische Auswirkungen auf die Realwirtschaft hat, und spricht für den „Diversification Meltdown“. Seit diesem Zeitraum gibt es einen globalen Wirtschaftsabschwung, der anscheinend dafür sorgte, dass die Renditen stärker korrelieren als in den Jahren zuvor. Bis zum Zeitraum um den 15. September 2008 (der Tag der Insolvenzbeantragung von Lehman Brothers, der gemeinhin als bisheriger Höhepunkt der Krise angesehen wird, vgl. Sorking (15.09.2008)), schwankten die Korrelationen relativ gleichmäßig um etwa 0.5, es gab keine strukturellen Änderungen. Mit Beginn der Krise änderte sich das jedoch und die Korrelationen blieben über 0.5. Der 10. September ist der dritte Handelstag vor dem 15. September, also wird das Maximum von $|P_T(j)|$ kurz vor dem

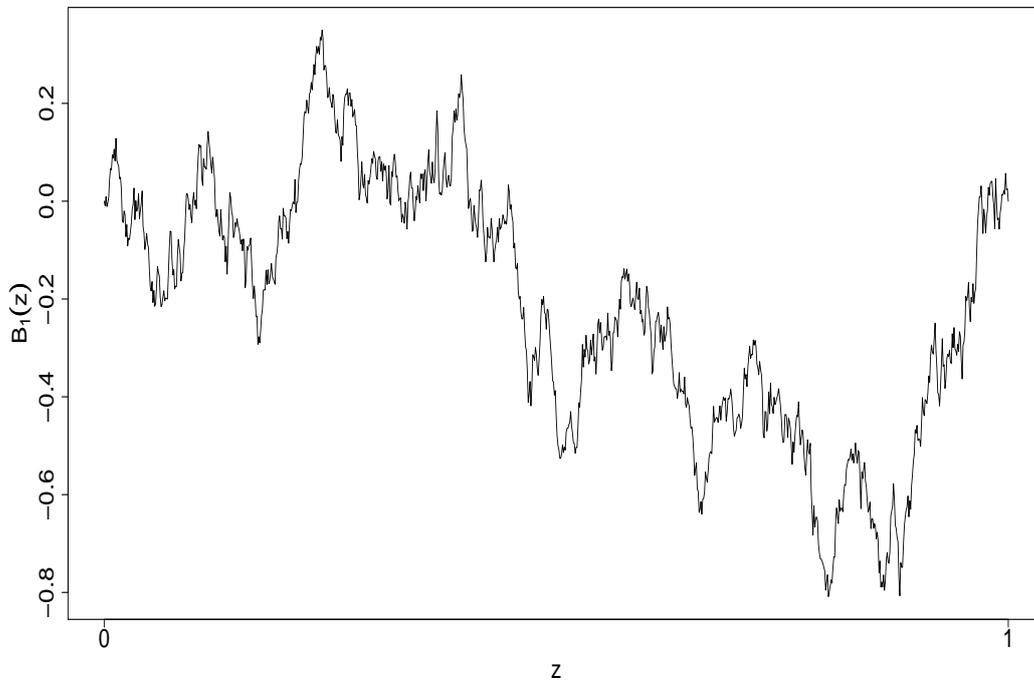


Abbildung 4: Realisation einer Brownschen Brücke mit Abtastabstand $\frac{1}{1000}$

Höhepunkt der Krise angenommen. Durch den in j wachsenden Gewichtungsfaktor $\frac{j}{\sqrt{T}}$ fällt der Unterschied zwischen der bis dahin relativ konstanten Korrelation und der Gesamtkorrelation $\hat{\rho}_T$ (die größer als die Korrelation vor dem Höhepunkt und unabhängig von j ist) dort besonders ins Gewicht. Nach dem Höhepunkt bewegen sich die sukzessive geschätzten Korrelationen wieder mehr in Richtung der Gesamtkorrelation und der Wert $|P_T(j)|$ wird wieder kleiner. Sehr schön ist diese Tatsache in der Grafik (Abbildung 7) des Verlaufs von $|P_T(j)|$ zu sehen.

Diese substanzwissenschaftliche Erklärung passt auch zu dem folgenden Ergebnis: Wenn wir den Datensatz einschränken auf den Zeitraum Anfang 2003 bis zum 10. September 2008 oder auf den Zeitraum 11. September 2008 bis Ende 2008, so verwirft der Test die Nullhypothese konstanter Korrelation jeweils in den beiden Teilen nicht.

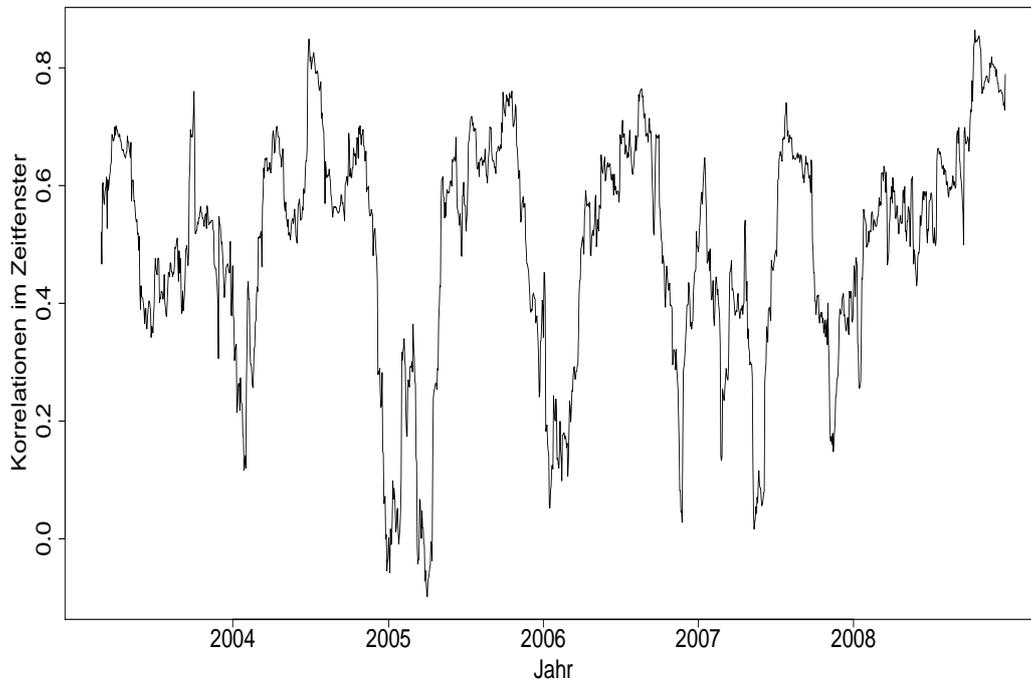


Abbildung 5: Korrelationen der Aktienrenditen von BASF und BMW von Anfang 2003 bis Ende 2008 in einem Zeitfenster von jeweils 40 Tagen

Die Werte der Teststatistik sind 0.573 ($T = 1483$, p -Wert 0.898) bzw. 0.584 ($T = 76$, p -Wert 0.885). Die geschätzte Korrelation springt deutlich von 0.497 auf 0.863. Das Testergebnis ist mit einiger Vorsicht zu genießen, weil die Anzahl der Beobachtungen von 11. September 2008 bis Ende 2008 relativ klein ist. Trotzdem scheint die Erklärung sinnvoll zu sein.

8 Ausblick

Es gibt noch einige offene Fragen und Problemstellungen, die in dieser Dissertation nicht betrachtet wurden, die aber in Zukunft noch bearbeitet werden können.

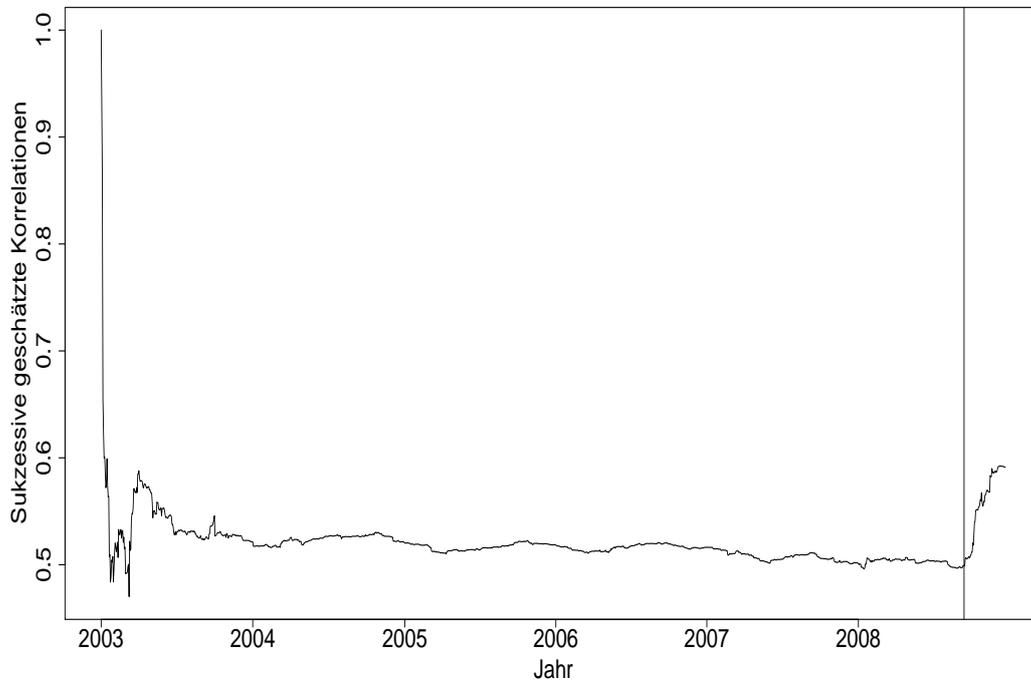


Abbildung 6: Sukzessive geschätzte Korrelationen der Aktienrenditen von BASF und BMW von Anfang 2003 bis Ende 2008, senkrechte Linie: 15. September 2008 (Insolvenzantrag Lehman Brothers)

Noch ungeklärt ist, warum der Test bei der t_5 -Verteilung für kleine $|\rho|$ konservativ ist und sich für wachsende T zunächst vom wahren Niveau entfernt. Es ist die Frage, wie sich stärkere Ränder genau bei der Asymptotik bemerkbar machen.

In irgendeiner Form sollte ausgeglichen werden, dass die Varianz der Brownschen Brücke für $z = \frac{1}{2}$ am größten ist. Damit könnte die Güte des Fluktuationstests bei sehr frühen oder sehr späten Abweichungen verbessert werden. Ein Ansatz wäre es, in der Teststatistik einen anderen Gewichtungsfaktor vor $|\hat{\rho}_j - \hat{\rho}_T|$ zu wählen, also nicht $\frac{j}{\sqrt{T}}$. Er sollte nicht linear in j sein, sondern für kleine und große j schneller wachsen als für mittlere j . Es ist jedoch nicht zu erwarten, dass die Grenzverteilung dann noch

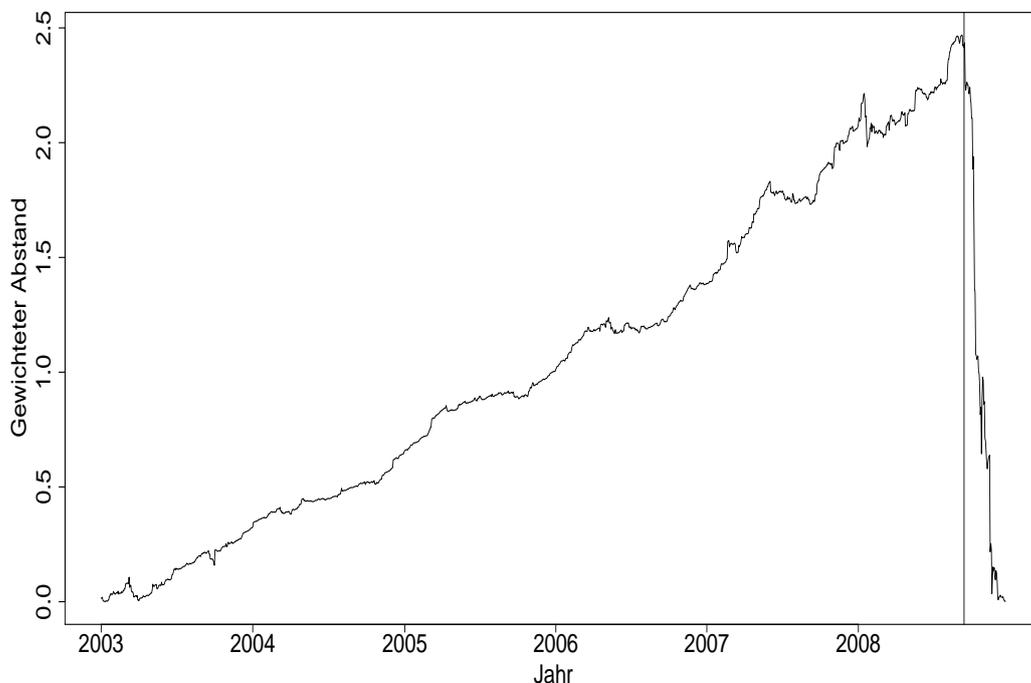


Abbildung 7: Zeitlicher Verlauf von $|P_T(j)|$ bei den Aktienrenditen von BASF und BMW von Anfang 2003 bis Ende 2008, senkrechte Linie: 15. September 2008 (Insolvenzantrag Lehman Brothers)

das Maximum des Absolutbetrags einer Brownschen Brücke ist.

In diesem Zusammenhang wäre die Konstruktion eines sequentiellen Verfahrens für Echtzeitdaten nützlich. Wir haben den Test auf Daten in der Vergangenheit angewandt, aber vielleicht kann man mit ähnlichen Prinzipien in tagesaktuellen Datensätzen von Aktienrenditen frühe Warnzeichen für eine bevorstehende Krise entdecken - zumindest dann, wenn die Korrelationen auch schon vor dem Höhepunkt einer Krise ansteigen. Hierzu ist es vermutlich notwendig, die Güte am Rand zu erhöhen. In etwas anderen Kontexten wird ein sequentielles Verfahren dieser Art in Fried und Imhoff (2004) beschrieben.

Schön wäre es, den Test auf Korrelationsmatrizen mit Dimension > 2 erweitern zu können. Jennrich (1970) hat seinen Test dafür konstruiert, bei uns ist bisher nur der bivariate Fall abgedeckt. Sicher wäre es möglich, mit unserem Test paarweise Tests durchzuführen und dann eine Niveaurektur durchzuführen. Es sollte aber auch möglich sein, direkt eine einzige Teststatistik zu konstruieren, deren Verteilung gegen das Maximum des Absolutbetrags einer höherdimensionalen Brownschen Brücke konvergiert.

Man könnte auch alternative Teststatistiken verwenden, das heißt, nicht $\max_{2 \leq j \leq T} |P_T(j)|$, sondern etwa

$$\max_{2 \leq j \leq T} (P_T(j)) - \min_{2 \leq j \leq T} (P_T(j)),$$

wie es in ähnlicher Form in Krämer und Schotman (1992) gemacht wird. Man nimmt also nicht das Maximum des Absolutbetrags von $P_T(j)$, sondern die Spannweite von $P_T(j)$. Da die Spannweite eine stetige Funktion von $D[\epsilon, 1]$ nach \mathbb{R} ist, ist die asymptotische Verteilung dieser alternativen Teststatistik die Spannweite einer Brownschen Brücke. Die zugehörige Verteilungsfunktion ist bekannt, vgl. Kennedy (1976) und Siddiqui (1976). Es ist zu erwarten, dass der Test eine höhere Güte bei Alternativen wie (B5) besitzt.

Noch ungeklärt ist, wie die Zeiten möglicher Strukturbrüche statistisch sinnvoll geschätzt werden können. In dem Praxisbeispiel dieser Dissertation sind wir implizit von nur einem Strukturbruch ausgegangen, den Zeitpunkt davon haben wir mit dem Zeitpunkt der Maximalstelle von $P_T(j)$ geschätzt. Die Frage ist, ob dieser Schätzer unter der Annahme eines einzigen Strukturbruchs konsistent ist und welche Schätzer man verwenden kann, falls es mehrere Strukturbrüche gibt.

Schließlich ist es sinnvoll, sich über Robustheitsfragen Gedanken zu machen. Wir haben für den Fluktuationstest das am häufigsten verwendete Korrelationsmaß, nämlich den Bravais-Pearson-Korrelationskoeffizienten verwendet. Dieser ist jedoch nicht robust gegenüber Ausreißern. Es wäre schön, in Zukunft auch einen Test für robustere Korrelationskoeffizienten zu konstruieren, die etwa in Croux und Dehon (2008) oder Kendall (1938) betrachtet werden. Die Frage ist, inwiefern sich die Theorie der Delta-Methode darauf übertragen lässt. Praktisch wäre es, wenn die Korrelationsmaße möglichst einfa-

che Funktionen von X und Y mit leicht zu bestimmender Hadamard-Ableitung wären.

A Einige wichtige Sätze

Wir zitieren hier einige wichtige Sätze, die wir in der Dissertation benötigen, die aber im Hauptteil den Lesefluss stören würden.

Nützlich ist ein starkes Gesetz der großen Zahlen für NED-Zufallsvariablen in Anlehnung an Davidson (1994, Theorem 20.20.(i)):

Theorem A.1 *Sei $r > 2$ und $(X_i)_{i \in \mathbb{Z}}$ eine Folge von reellwertigen NED-Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit Größe $-\frac{1}{2}$ und Konstanten $(c_i)_{i \in \mathbb{Z}}$ über einer Folge $(V_i)_{i \in \mathbb{Z}}$, die α -mischend der Größe $-\frac{r}{r-2}$ ist. Es gelte $c_i \leq C(\mathbb{E}(X_i^2))^{\frac{1}{2}}$ für eine Konstante C , die Erwartungswerte seien 0 und die zweiten absoluten Momente seien gleichmäßig beschränkt. Dann konvergiert $\frac{1}{T} \sum_{i=1}^T X_i$ fast sicher gegen 0.*

Bemerkung Aus der Beschränktheit der zweiten absoluten Momente folgt die Gültigkeit der Konvergenzbedingung in Davidson (1994). Im Falle der Gültigkeit von A.1 folgt automatisch auch die stochastische Konvergenz gegen 0, dann sprechen wir auch vom schwachen Gesetz der großen Zahlen.

Falls $(V_i)_{i \in \mathbb{Z}}$ sogar ϕ -mischend der Größe $-\frac{r}{2r-2}$ ist, so gilt der Satz auch für $r = 2$.

Wir benötigen auch die Markov-Ungleichung, vgl. Fisz (1976, S. 98):

Theorem A.2 (Markov-Ungleichung) *Sei X eine reellwertige Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit existierendem Erwartungswert. Dann gilt für jedes beliebige $a > 0$*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

Wir werden einen Summengrenzwert verwenden:

Lemma A.3 *Sei g eine beschränkte, durch Treppenfunktionen approximierbare reell-*

wertige Funktion. Dann gilt für $\epsilon > 0$ gleichmäßig für $z \in [\epsilon, 1]$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^{\lfloor Tz \rfloor} g\left(\frac{i}{T}\right) = \int_0^z g(u) du.$$

Beweis. Wende Lemma 4 aus Krämer et al. (1988) für \mathbb{R} mit $z_t = R = 1$ und $h_T(z) = g(z)$ an. \square

Das nächste Theorem macht eine Aussage über die schwache Konvergenz bei einem Schema.

Theorem A.4 (Billingsley (1968, Theorem 4.2)) Seien $X^1, X^2, \dots, X^k, \dots$ und $Y_1, Y_2, \dots, Y_T, \dots$ sowie für jedes k $X_1^k, X_2^k, \dots, X_T^k, \dots$ Folgen von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit Werten in einem allgemeinen metrischen Raum \mathbb{D} , versehen mit einer σ -Algebra \mathfrak{D} . Sei d die Metrik in \mathbb{D} und sei $d(X_T^k, Y_T)$ eine Zufallsvariable für alle T und k . Außerdem gelte für jedes k $X_T^k \rightarrow_d X^k$ für $T \rightarrow \infty$ und $X^k \rightarrow_d X$ für $k \rightarrow \infty$. Zudem gelte

$$\lim_{k \rightarrow \infty} \limsup_{T \rightarrow \infty} \mathbb{P}(d(X_T^k, Y_T) \geq \epsilon) = 0$$

für jedes $\epsilon > 0$. Dann gilt für $T \rightarrow \infty$

$$Y_T \rightarrow_d X.$$

In Billingsley (1968) ist der Satz für den Fall bewiesen, dass \mathbb{D} separabel ist. Die Separabilität benötigt man jedoch nur dafür, dass $d(X_T^k, Y_T)$ für alle T eine Zufallsvariable ist. Setzt man keine Separabilität voraus, ist es notwendig und hinreichend, letzteres explizit zu fordern. Für \mathbb{G} ausgestattet mit der Supremumsnorm ist $d(X_T, Y_T)$ immer eine Zufallsvariable, weil wir das Supremum über abzählbar viele Zufallsvariablen bilden.

Schließlich kommt der Satz von Egoroff über den Zusammenhang von fast sicherer und gleichmäßiger Konvergenz.

Theorem A.5 (*Davidson (1994, Theorem 18.4)*) *Unter den Voraussetzungen von Definition 2.5 konvergiert die Folge X_T \mathbb{P} -fast sicher gegen a genau dann, wenn es für jedes $\delta > 0$ eine Menge $\Omega_\delta \subset \Omega$ mit $\mathbb{P}(\Omega_\delta) \geq 1 - \delta$ gibt, sodass*

$$X_T(\omega) \rightarrow X(\omega)$$

gleichmäßig auf Ω_δ gilt.

Symbolverzeichnis

Symbol	Erklärung
$C(J)$	Raum der stetigen Funktionen auf dem abgeschlossenen Intervall $J \in \mathbb{R}$
$D(J)$	Raum der càdlàg-Funktionen auf dem abgeschlossenen Intervall $J \in \mathbb{R}$
$D^+(J)$	Raum der Funktionen aus $D(J)$, die durch eine Konstante nach unten beschränkt sind
\mathbb{G}	$\mathbb{H}_1 \times \dots \times \mathbb{H}_k$ (k -mal, $k \geq 1$, $\mathbb{H}_i \in \{C[\epsilon, 1], D[\epsilon, 1], D^+[\epsilon, 1], \epsilon \geq 0\}$)
$\ \cdot\ _{\mathbb{H}_i}$	Supremumsnorm auf \mathbb{H}_i
$\ \cdot\ _{\mathbb{G}}$	mehrdimensionale Supremumsnorm auf \mathbb{G}
\mathbb{G}_1	$\mathbb{H}_1 \times \dots \times \mathbb{H}_k$ (k -mal, $k \geq 1$, $\mathbb{H}_i \in \{D[\epsilon, 1], D^+[\epsilon, 1], \epsilon \geq 0\}$) (Spezialfall von \mathbb{G})
\mathbb{G}_2	$\mathbb{H}_1 \times \dots \times \mathbb{H}_l$ (l -mal, $l \geq 1$, $\mathbb{H}_i \in \{D[\epsilon, 1], D^+[\epsilon, 1], \epsilon \geq 0\}$) (Spezialfall von \mathbb{G})
$C[\epsilon, 1]^l$	$C[\epsilon, 1] \times \dots \times C[\epsilon, 1]$ (l mal)
$D[\epsilon, 1]^l$	$D[\epsilon, 1] \times \dots \times D[\epsilon, 1]$ (l mal)
\mathbb{D}	allgemeiner metrischer bzw. normierter Raum
$\ \cdot\ _{\mathbb{D}}$	zugehörige Norm
\mathfrak{D}	nicht näher definierte σ -Algebra auf \mathbb{D}
\rightarrow_d	Konvergenz in Verteilung
\rightarrow_p	Konvergenz in Wahrscheinlichkeit
\mathfrak{G}	σ -Algebra der offenen Bälle bezüglich der durch die Supremumsnorm erzeugte Metrik auf $C[\epsilon, 1]$, $D[\epsilon, 1]$ oder $D^+[\epsilon, 1]$ auf \mathbb{G}

Symbol	Erklärung
$X \stackrel{L}{=} Y$	Gleichheit der Verteilung zweier Zufallsvariablen X und Y auf \mathbb{R}^k
W_1	eindimensionale Brownsche Bewegung
W_n	n -dimensionale Brownsche Bewegung
B	eindimensionale Brownsche Brücke
$\tau(z)$	$[2 + z(T - 2)], z \in [0, 1]$
ρ_0^*	$\frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}$
$P_T(j)$	$\hat{D} \frac{j}{\sqrt{T}} (\hat{\rho}_j - \hat{\rho}_T)$
Q_T	$\max_{2 \leq j \leq T} P_T(j) $

Abbildungsverzeichnis

1	Häufigkeitsverteilung der Stelle des Maximums von $ P_T $	72
2	Zeitlicher Verlauf von $\hat{\rho}_j$ (Konfidenzband)	73
3	Zeitlicher Verlauf von $P_T(j)$ (eine simulierte Realisation)	74
4	Realisation einer Brownschen Brücke mit Abtastabstand $\frac{1}{1000}$	75
5	Korrelationen der Aktienrenditen von BASF und BMW von Anfang 2003 bis Ende 2008 in einem Zeitfenster von jeweils 40 Tagen	76
6	Sukzessive geschätzte Korrelationen der Aktienrenditen von BASF und BMW von Anfang 2003 bis Ende 2008, senkrechte Linie: 15. September 2008 (Insolvenzantrag Lehman Brothers)	77
7	Zeitlicher Verlauf von $ P_T(j) $ bei den Aktienrenditen von BASF und BMW von Anfang 2003 bis Ende 2008, senkrechte Linie: 15. September 2008 (Insolvenzantrag Lehman Brothers)	78

Literatur

- E. Andreou und E. Ghysels. Tests for breaks in the conditional co-movements of assets returns. *Statistica Sinica*, 13:1045–1073, 2003.
- P. Billingsley. *Convergence of probability measures*. Wiley, New York, 1968.
- B. Boyer, M. Gibson, und M. Loretan. *Pitfalls in tests for changes in correlation*. International Finance Discussion paper 597, Federal Reserve Board, 1999.
- R.C. Bradley. *Introduction to strong mixing conditions, volume 1*. Kendrick Press, 2007.
- R. Campbell, C. Forbes, K. Koedijk, und P. Kofman. Increasing correlations or just fat tails? *Journal of Empirical Finance*, 15:287–309, 2008.
- R. Chakrabarti und R. Roll. East Asia and Europe during the 1997 Asian collapse: A clinical study of a financial crisis. *Journal of Financial Markets*, 5:1–30, 2002.
- C. Croux und C. Dehon. *Robustness versus efficiency for nonparametric correlation measures*. Diskussionspapier 2, Université Libre des Bruxelles, Ecares, 2008.
- J. Davidson. *Stochastic limit theory*. Oxford University Press, 1994.
- J. Davidson. Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics*, 106:243–269, 2002.
- J. Davidson und R.M. de Jong. Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica*, 68(2):407–424, 2000.
- A. Deng und P. Perron. The limit distribution of the CUSUM of squares test under general mixing conditions. *Econometric Theory*, 24:809–822, 2008.
- J.M. Dufour. Recursive stability analysis of linear regression relationships. *Journal of Econometrics*, 19:31–76, 1982.

- C.B. Erb, C.R. Harvey, und T.E. Viskanta. Forecasting international equity correlations. *Financial Analysts Journal*, 50(6):32–45, 1994.
- M. Fischer. *Are correlations constant over time?* Sonderforschungsbereich 649: Ökonomisches Risiko, 2007.
- M. Fisz. *Wahrscheinlichkeitsrechnung und mathematische Statistik*. VEB, Berlin, 1976.
- K. Forbes und R. Rigobon. No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance*, 57:2223–2261, 2002.
- R. Fried und M. Imhoff. On the online detection of monotonic trends in time series. *Biometrical Journal*, 46:90–102, 2004.
- R.D. Gill. Non- and semi-parametric maximum likelihood estimators and the von Mises method - Part 1. *Scandinavian Journal of Statistics*, 16:97–128, 1989.
- W.N. Goetzmann, L. Li, und K.G. Rouwenhorst. Long-term global market correlations. *Journal of Business*, 78(1):1–38, 2005.
- B.E. Hansen. GARCH(1,1) processes are near-epoch dependent. *Economics Letters*, 36:181–186, 1991.
- R.I. Jennrich. An asymptotic chi-square test for the equality of two correlation matrices. *Journal of the American Statistical Association*, 65:904–912, 1970.
- W. Kaballo. *Einführung in die Analysis II*. Spektrum, Heidelberg, 1997.
- E.C. Kaplanis. Stability and forecasting of the comovement measures of international stock market returns. *Journal of International Money and Finance*, 7(1):63–75, 1988.
- I. Karatzas und S.E. Shreve. *Brownian motion and stochastic calculus*. Series in Statistics, Springer, New York, 1988.
- M.G. Kendall. A new measure of rank correlation. *Biometrika*, 30:81–93, 1938.

- D. Kennedy. The distribution of the maximum Brownian excursion. *Journal of Applied Probability*, 13:371–376, 1976.
- M. King, E. Sentana, and S. Wadhvani. Volatility and links between national stock markets. *Econometrica*, 62(4):901–933, 1994.
- W. Krämer, W. Ploberger, and R. Alt. Testing for structural change in dynamic models. *Econometrica*, 56(6):1355–1369, 1988.
- W. Krämer und P. Schotman. Range vs. maximum in the OLS-based version of the CUSUM test. *Economics letters*, 40:379–381, 1992.
- F. Longin und B. Solnik. Is the correlation in international equity returns constant: 1960-1990? *International Money and Finance*, 14(1):3–26, 1995.
- F. Longin und B. Solnik. Extreme correlation of international equity markets. *Journal of Finance*, 56:649–675, 2002.
- H.M. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- I. Meric und G. Meric. Potential gains from international portfolio diversification and the intertemporal stability of international stock market relationships. *Journal of Banking and Finance*, 13(4/5):627–640, 1989.
- E.S. Pearson und S.S. Wilks. Methods of statistical analysis appropriate for k samples of two variables. *Biometrika*, 25:353–378, 1933.
- W. Ploberger und W. Krämer. The CUSUM-test with OLS residuals. *Econometrica*, 60(2):271–285, 1992.
- W. Ploberger, W. Krämer, und K. Kontrus. A new test for structural stability in the linear regression model. *Journal of Econometrics*, 40:307–318, 1989.
- D. Pollard. *Convergence of stochastic processes*. Series in Statistics, Springer, New York, 1984.

- V. Ragea. Testing correlation stability during hectic financial markets. *Financial Markets and Portfolio Management*, 17(3):289–308, 2003.
- P.K. Sen. Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 52:203–218, 1980.
- M. Siddiqui. The asymptotic distribution of the range and other functions of partial sums of stationary processes. *Water Resources Research*, 12:1271–1276, 1976.
- A.R. Sorking. Lehman files for bankruptcy; Merrill is sold. *The New York Times*, 15.09.2008.
- G.Y.N. Tang. Intertemporal stability in international stock markets relationships: A revisit. *The Quarterly Review of Economics and Finance*, 35:579–593, 1995.
- A.W. van der Vaart. *Asymptotic statistics*. Cambridge University Press, Cambridge, 1998.
- N. Wiener. Differential space. *Journal of Mathematical Physics*, 2:131–174, 1923.
- J.M. Wooldridge und H. White. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory*, 4:210–230, 1988.