Quantitatives Portfoliomanagement -
Neue Ansätze in Portfoliooptimierung und Risikosteuerung

INAUGURALDISSERTATION

zur Erlangung des akademischen Grades
Doktor rerum politicarum
der
Technischen Universität Dortmund

vorgelegt von

Diplom-Kaufmann Tobias Berens

Mai 2014

Inhaltsverzeichnis

Tabellenverzeichnis

Abbildungsverzeichnis

Kapitel 1

Einleitung

1.1 Einführung in die Thematik und Motivation

Die Idee, Anlageportfolios durch quantitative Methoden zu steuern, basiert maßgeblich auf der grundlegenden Arbeit zur Portfoliotheorie von Harry Markowitz aus dem Jahr 1952. Sein Beitrag *Portfolio Selection*[1] beschreibt formal die Vorteilhaftigkeit einer diversifizierten Investition, welche durch die Aufteilung der Anlagesumme eines Portfolios auf verschiedene Anlageobjekte erreicht wird. Aus diesem theoretischen Rahmen leitet Markowitz ein Modell ab, mit dessen Hilfe ein Investor diejenigen Wertpapiere selektieren kann, die unter Berücksichtigung von Diversifikationseffekten auf Portfolio-Ebene zu einem optimalen Verhältnis zwischen Rendite und Risiko führen. Dieses Entscheidungsmodell zur Bildung eines sogenannten effizienten Portfolios basiert jedoch auf teils sehr restriktiven Annahmen. Unter anderem wird vorausgesetzt, dass der Investor Kenntnis über zu erwartende Renditen und deren (Ko-)Varianzen besitzt. In der Realität unterliegen jedoch Erwartungswerte und (Ko-)Varianzen von Wertpapierrenditen im Zeitablauf starken Schwankungen. Daher ist es für einen Anleger nur schwer möglich, auf Basis der beobachtbaren Daten verlässliche Prognosen für diese Parameter zu erstellen. Dies ist problematisch, weil in dem Entscheidungsmodell die optimale Gewichtung der einzelnen Anlagen in einem effizienten Portfolio von den geschätzten Parametern abhängt. Aus den Ergebnissen mehrerer Studien lässt sich schließen, dass die optimalen Portfoliogewichte besonders sensitiv auf Veränderungen der erwarteten Rendite reagieren und sich somit Schätzfehler bezüglich der Renditeprognose erheblich auf die Effizienz des Portfolios auswirken (vgl. u. a. Best and Grauer, 1991, Chopra and Ziemba, 1993 sowie Gohout and Specht, 2007). Das globale Minimum-Varianz Optimierungsmodell ist ein

[1] Vgl. Markowitz [1952].

sehr intuitiver Ansatz, welcher weniger anfällig für falsch geschätzte Parameter ist als die klassiche Erwartungswert-Varianz Optimierung.[2] Mit Hilfe dieser Methode können Portfoliogewichte ermittelt werden, die zu einer minimalen Portfolio-Varianz, also einem geringstmöglichen Risiko führen. Die Optimierung basiert ausschließlich auf Varianzen und Kovarianzen der Renditen. Erwartete Renditen stellen keinen relevanten Parameter dar und müssen daher auch nicht geschätzt werden. Die Anwendung dieses Optimierungsmodells auf Aktienportfolios hat in mehreren Studien zu im Vergleich mit unterschiedlichen Benchmarks zu verbesserten Ergebnissen geführt (vgl. Haugen and Baker, 1991; Jagannathan and Ma, 2003; Clarke et al., 2006 sowie Clarke et al., 2013). Gleichwohl hängt die Leistungsfähigkeit des Modells auch hier von einer möglichst verlässlichen Schätzung der Varianzen von Renditen einzelner Wertpapiere und deren Kovarianzen untereinander maßgeblich ab.

Neben der Optimierung im Allgemeinen spielt die Risikosteuerung im Speziellen eine wichtige Rolle im Portfoliomanagement. In den vergangenen Jahren hat sich der Value-at-Risk (VaR) in diesem Kontext als ein zentrales Risikomaß in Literatur und Praxis sowie insbesondere in der Bankenregulierung etabliert. Diese Kennzahl gibt den Verlust einer Risikoposition an, der über eine bestimmte Haltedauer mit einer festgelegten Wahrscheinlichkeit nicht überschritten wird und lässt sich einfach aus dem Quantil der Verlustverteilung ermittelt. Dem Vorteil einer einfachen Anwendung und intuitiven Verständlichkeit dieses Risikomaßes stehen jedoch wesentliche Nachteile gegenüber. Zum einen handelt es sich bei dem VaR nicht um ein kohärentes Risikomaß, da die Eigenschaft der Subadditivtät grundsätzlich nicht gegeben ist.[3] Zum anderen liefert er definitionsgemäß keine Informationen darüber, in welcher Höhe ein Verlust zu erwarten ist, wenn dieser den VaR überschreitet. Diese Schwächen des VaR haben dazu geführt, dass der Basler Ausschuss für Bankenaufsicht in einem kürzlich veröffentlichen Konsultationspapier zu den Eigenkapitalanforderungen für Positionen des Handelsbuches dem Expected Shortfall (ES) als neues Risikomaß eine große Bedeutung zukommen lässt (vgl.

[2] In der Literatur wurden weitere Ansätze für robuste Modelle entwickelt, beispielsweise das Black-Litterman-Verfahren (Black and Litterman, 1992).

[3] Der Begriff Subadditivität beschreibt die Eigenschaft, dass das Risiko eines Portfolios kleiner oder gleich der Summe der Einzelrisiken zu sein hat (vgl. Artzner et al., 1999).

Basel Committee on Banking Supervision, 2013). Der ES ist definiert durch den erwarteten Verlust für die Fälle, in denen er den VaR erreicht oder überschreitet. Neben der Erfüllung der Anforderungen an ein kohärentes Risikomaß, schließt die Berechnung des ES auch den extremen linken Rand der Verlustverteilung ein. Es spiegeln sich im ES somit auch sehr hohe Verluste wider, die mit einer äußerst geringen Wahscheinlichkeit eintreten können. Sowohl für die Ermittlung des VaR, als auch die des ES können unterschiedliche Ansätze verwendet werden. Die klassische parametrische Varianz-Kovarianz Methode basiert auf der Idee, dass Renditen, und somit auch Verluste, über Varianzen und Kovarianzen einer Normalverteilung dargestellt werden. So können VaR und ES sehr einfach über Quantils- und Dichtefunktionen der Normalverteilung ermittelt werden. In der Literatur wurden zahlreiche weitere parametrische Modelle entwickelt, häufig mit dem Ziel, Varianzen der Renditen und deren Abhängigkeiten in einem Portfolio im Zeitablauf dynamisch zu modellieren. Als populäres Beispiel sind hier sogenannte GARCH-Modelle zu nennen, die auf der Arbeit von Bollerslev [1986] basieren.[4] Demgegenüber steht das in der Praxis sehr gebräuchliche Modell der Historischen Simulation.[5] Bei diesem nicht-parametrischen Modell werden historische Renditen einer zu bestimmenden Periode der Größe nach sortiert. Der VaR lässt sich dann einfach aus dem gewünschten Quantil dieser Verteilung ablesen. Als dritte Gattung versuchen semi-parametrische Modelle die wesentlichen Vorzüge der parametrischen Modelle (dynamische Modellierung) und der nicht-parametrischen Modelle (Verzicht auf Verteilungsannahmen) zu verbinden. Die von Barone-Adesi et al. [1999] entwickelte Gefilterte Historische Simulation ist hier als gebräuchliches VaR-Modell zu nennen.

Sowohl die Modelle zur Portfoliooptimierung als auch VaR-Modelle werden unabhängig von ihrem konkreten Aufbau in der Regel auf Basis von historischen Renditen parametrisiert. Es stellt sich die Frage, wie weit die Datenhistorie in die Vergangenheit zurückreichen sollte, damit das jeweilige Modell verlässlich spezifiziert werden kann. Dabei ist zu beachten, dass statistische Lage-, Streuungs- und Zusammenhangsmaße von

[4] Hansen and Lunde [2005] liefern einen umfassenden Überblick über den Einsatz von GARCH-Modellen zur VaR Prognose.

[5] Pérignon and Smith [2010] zeigen in einer Studie, dass von 60 analysierten internationalen Banken 73% Historische Simulation als VaR-Modell einsetzen.

Wertpapier-Renditen sich im Zeitablauf verändern. In der Literatur finden sich vielfältige Studien, die Finanzmärkte insbesondere auf Strukturbrüche in Volatilitäten und Korrelationen untersuchen.[6] Werden solche Brüche ignoriert, kann dies zu fehlerhaft spezifizierten Modellen führen (vgl. unter anderem Hillebrand, 2005). Die Berücksichtigung von Strukturbrüchen bei der Auswahl der Datenhistorie ist jedoch ebenso mit Herausforderungen verbunden. So muss sichergestellt sein, dass ein vorliegender Strukturbruch mit hinreichender Sicherheit erkannt wird. Hierzu wurden in der Literatur verschiedene Verfahren und statistische Tests entwickelt, die in der Lage sind, signifikante Änderungen insbesondere von Streuungs- und Abhängigkeitsmaßen zu identifizieren.[7] Zudem kann zwar eine Verkürzung der Datenhistorie durch die Verschiebung ihres Startpunktes bis zu dem Zeitpunkt eines Strukturbruchs zu einer weniger verzerrten Datengrundlage führen. Dieser Vorteil wird allerdings dadurch konterkariert, dass aufgrund der verringerten Datenmenge die Gefahr von signifikanten Schätzfehlern zunimmt.

Die Beurteilung, ob ein Modell zu präzisen VaR-Prognosen führt, erfolgt grundsätzlich über die Analyse der Eigenschaften von VaR-Überschreitungen.[8] Christoffersen [1998] entwickelte hierzu grundlegende Hypothesen, die in der Literatur standardmäßig bei der Entwicklung von Backtests verwendet werden. Intuitiv ist die Hypothese, dass über eine Periode die tatsächlich gemessene Anzahl an VaR-Überschreitungen nicht signifikant von der statistisch erwarteten Anzahl abweichen darf. Diese Eigenschaft wird als Unconditional Coverage bezeichnet. Eine weitere Hypothese bezieht sich auf die Verteilung der VaR-Überschreitungen auf der Zeitachse. Gemäß der sogenannten Independence Eigenschaft führen korrekte VaR-Prognosen zu im Zeitablauf statistisch unabhängigen VaR-Überschreitungen, da Abhängigkeiten zu deren gehäuften Auftreten führen können. Der dritten und abschließenden Hypothese folgend entsprechen die VaR-Überschreitungen der Conditional Coverage Eigenschaft, wenn sie simultan sowohl die Unconditional Coverage als auch die Independence Eigenschaften erfüllen. Um evaluieren zu können, ob

[6] Einen Überblick liefert Andreou and Ghysels [2009].

[7] Als Beispiele können hier der Test auf konstante Varianz von Wied et al. [2012a], auf konstante Kovarianz von Aue et al. [2009] oder auf konstante Korrelation von Wied et al. [2012b] genannt werden.

[8] Alternativ können auch Verfahren eingesetzt werden, die ganz allgemein die Prognosefähigkeit eines Modells über den Abstand eines realisierten Wertes zum prognostizierten Wert untersuchen. Hierzu sei beispielsweise auf die Arbeit von Giacomini und White [2006] verwiesen.

die aus Prognosemodellen resultierenden Sequenzen von VaR-Überschreitungen den genannten Hypothesen entsprechen, wurden in der Literatur verschiedene statistische Tests entwickelt.[9] Verschiedene Studien haben jedoch gezeigt, dass einige dieser Tests lediglich eine geringe Güte besitzen oder ein fehlerhaftes asymptotisches Verhalten aufweisen.[10] Andere Tests wiederum sind sehr komplex und wenig intuitiv.

Daneben vernachlässigen die genannten drei Hypothesen und die in diesem Kontext entwickelten statistischen Tests die unerwünschte Eigenschaft von VaR-Überschreitungen, dass deren Wahrscheinlichkeit sich im Zeitablauf ändern kann. So ist es möglich, dass ein Modell VaR-Prognosen generiert, welche sich zu langsam oder in einem zu geringen Ausmaß an längere, sehr volatile Marktphasen anpassen und vice versa. Diese Schwankungen können dazu führen, dass VaR-Überschreitungen zeitlich gehäuft auftreten, obwohl sie stochastisch unabhängig voneinander sind. Die Zeitpunkte, an denen ein Verlust größer ist als der prognostizierte VaR, sollten mithin nicht nur unabhängig, sondern auch identisch verteilt sein.

Die vorliegende kumulative Dissertation umfasst fünf in sich abgeschlossene Beiträge zu neuen Ansätzen in der Portfoliooptimierung und der Risikosteuerung. Der erste Beitrag der vorliegenden Arbeit (Kapitel 2) befasst sich grundlegend mit den Einsatzmöglichkeiten von statistischen Tests auf eine konstante Kovarianzmatrix sowie auf konstante Varianzen und paarweise Korrelationen im Rahmen der Minimum-Varianz-Portfoliooptimierung. Es werden Problemfelder bei der Anwendung der Tests herausgearbeitet und diskutiert. Zudem wird evaluiert, ob der Einsatz der Tests zu verbesserten Risiko-Rendite-Verhältnissen führt. Der darauf folgende Artikel (Kapitel 3) baut auf diesen Erkenntnissen auf und erarbeitet Lösungsansätze für die zuvor dargelegten Herausforderungen beim Einsatz von Strukturbruchtests in der Portfoliooptimierung. In diesem Rahmen wird eine automatisierte Anlagestrategie erarbeitet und deren Leistungsfähigkeit analysiert. Neben der klassischen Risiko-Rendite-Bertrachtung werden hierbei zusätzlich Auswirkungen auf Transaktionkosten untersucht, die für einen Einsatz in der Pra-

[9] Die Arbeit von Berkowitz et al. [2011] liefert einen Überblick über verschiedene Ansätze für statistische Tests.

[10] Vgl. Berkowitz et al. [2011] und Candelon et al. [2011].

xis von hoher Bedeutung sind. Im Kapitel 4 befasst sich die vorliegende Arbeit mit der Diskussion der gewünschten Eigenschaften von VaR-Überschreitungen. Dies führt zu der Entwicklung neuer statistischer Tests, die neben der Unabhängigkeit von VaR-Überschreitungen explizit auch deren identische Verteilung mit einbeziehen. Diese auf Monte-Carlo-Simulationen beruhenden Verfahren sind sehr intuitiv, einfach zu implementieren und besitzen sehr häufig eine überlegene Testgüte im Vergleich zu den bislang existierenden Ansätzen. In Kapitel 5 werden parametrische, nicht-paarametrische und semi-parametrische Modelle zur Schätzung von VaR- und ES-Prognosen untersucht. Der Fokus dieser Analyse liegt dabei auf der Anwendung unterschiedlicher Strategien zur Bestimmung der Länge einer für die Parametrisierung erforderlichen Datenhistorie. Dabei kommen einfache Ansätze, wie beispielsweise rollierende Datenhistorien mit unterschiedlichen Längen, aber auch komplexere Methoden, wie Strukturbrüche und Strategie-Kombinationen zum Einsatz. Der Beitrag des abschließenden Kapitels 6 setzt sich ebenfalls mit unterschiedlichen Ansätzen zur Schätzung von VaR- und ES-Prognosen auseinander. Der thematische Schwerpunkt liegt hier auf der multivariaten Portfoliosicht, bei der ein univariates GARCH-Modell mit statischen und dynamischen Korrelationsmodellen kombiniert wird. Es wird der Frage nachgegangen, ob sich die Präzision der Modelle erhöht, wenn die Schätzung der Korrelationsmodelle auf einer durch verschiedene Strukturbruchtests definierten Datenhistorie beruht. Als Vergleich dient hierbei ein in Theorie und Praxis sehr häufig verwendetes rollierendes Fenster mit einer festen Länge. Diese ersetzen regelmäßig die ältesten Daten eines Schätzfensters durch aktuelle und berücksichtigen somit implizit zeitliche Parameteränderungen. Aus diesem Grund stellen die rollierenden Fenster eine herausfordernde Vergleichsmethode dar.

Der folgende Abschnitt gibt einen Überblick über Inhalte und Publikationsdetails der einzelnen Artikel der vorliegenden kumulativen Dissertation.

1.2 Publikationsdetails

Neben dieser Einleitung besteht die vorliegende kumulative Dissertation aus fünf in sich abgeschlossenen Beiträgen zu den Themen Portfoliooptimierung und Risikosteuerung. Im Folgenden werden die einzelnen Beiträge kurz inhaltlich zusammengefasst sowie Details zu der Veröffentlichung erläutert.

Beitrag I (Kapitel 2):

On the Application of New Tests for Structural Changes on Global Minimum-Variance Portfolios.

Autoren:

Dominik Wied, Daniel Ziggel und Tobias Berens

Zusammenfassung:

Die Effizienz eines Minimum-Varianz-Portfolios ist abhängig von einer präzisen Schätzung der Kovarianz-Matrix, welche der Optimierung zugrunde liegt. Jedoch sind Abhängigkeitsmaße zwischen den Renditen verschiedener Wertpapiere über längere Zeiträume typischerweise nicht konstant. Dieser Beitrag untersucht daher die Frage, ob sich das Verhältnis zwischen Risiko und Rendite eines optimierten Minimum-Varianz-Portfolios verbessert, wenn bei dessen Konstruktion potenzielle Brüche in der Kovarianz-Matrix berücksichtigt werden. Die Ergebnisse zeigen, dass ein Test auf Konstanz der gesamten Kovarianz-Matrix in Teilen zu verbesserten Ergebnissen des Portfolios führen kann. Dagegen sind paarweise Tests auf konstante Varianzen und Korrelationen nicht ohne weitere Modifikationen auf die Optimierung eines Portfolios anwendbar.

Publikationsdetails:

Veröffentlicht in: Statistical Papers, Vol. 54, Issue 4, 2013, pp. 955-975.

Beitrag II (Kapitel 3):

Automated Portfolio Optimization Based on a New Test for Structural Breaks.

Autoren: Tobias Berens, Dominik Wied und Daniel Ziggel

Zusammenfassung:

Dieser Beitrag präsentiert eine vollständig automatisierte Optimierungsstrategie, welche die klassische Portfoliotheorie nach Markowitz mit Tests auf eine konstante Kovarianz kombiniert. Mehrere Studien zeigen, dass die ausschließlich auf der Kovarianz-Matrix basierende Minimum-Varianz-Portfoliooptimierung bei Aktienportfolios zu sehr guten Ergebnissen im Vergleich zu verschiedenen anderen Ansätzen führt. Da die Struktur einer Kovarianz-Matrix von Aktien-Renditen im Zeitablauf zu Brüchen neigt, wird in diesem Beitrag die Kovarianz-Matrix unter Berücksichtigung der Ergebnisse von Strukturbruchtests geschätzt. Dabei bestimmen die Bruchpunkte die Länge des der Schätzung der Kovarianz-Matrix zugrundeliegenden Datenfensters. Darüber hinaus wird untersucht, ob sich die identifizierten Bruchpunkte dazu eignen, die Zeitpunkte für eine Re-Optimierung festzulegen. Im Rahmen einer Out-Of-Sample Studie wird die Methodik auf zwei unterschiedliche Datensätze angewendet und die Ergebnisse hinsichtlich Risiko-Rendite-Verhältnis sowie Auswirkung auf Transaktionskosten mit unterschiedlichen Alternativmethoden verglichen. Die Studie zeigt, dass der hier präsentierte Ansatz im Durchschnitt zu besseren Resultaten führt als gleichgewichtete Portfolios und einfache Minimum-Varianz-Optimierungen ohne die Berücksichtigung von Strukturbrüchen.

Publikationsdetails:

Veröffentlicht in: Acta Universitatis Danubius. Œconomica, Vol. 10, Issue 2, 2014, pp. 241-262.

Beitrag III (Kapitel 4):

A New Set of Improved Value-at-Risk Backtests.

Autoren: Daniel Ziggel, Tobias Berens, Gregor N.F. Weiß und Dominik Wied

Zusammenfassung: Dieser Beitrag präsentiert eine Gruppe neuer formaler Backtests für Eigenschaften von VaR-Überschreitungen, welche signifikante Vorteile gegenüber bislang veröffentlichten Ansätzen aufweisen. Ein neuer Test auf Unconditional Coverage kann sowohl für einseitiges als auch für zweiseitiges Testen eingesetzt werden, wodurch die Testgüte deutlich erhöht wird. Daneben wird die gewünschte Eigenschaft von unabhängigen und identischen VaR-Überschreitungen diskutiert und ein Test vorgestellt, der explizit auf das Auftreten von zeitlich gehäuften VaR-Überschreitungen testet. Die Anwendung dieser auf Monte-Carlo-Simulationen basierenden Tests in einer Simulationsstudie liefert in vielen Fällen überlegene Ergebnisse gemessen an vergleichbaren Tests. Eine abschließende empirische Studie verdeutlicht die Vorteile der Tests in der Anwendung auf reale Daten.

Publikationsdetails:

Zur Veröffentlichung eingereicht in: Journal of Banking and Finance; nach erster und zweiter Begutachtung Aufforderung zur Überarbeitung des Manuskripts und Wiedereinreichung (revise and resubmit). Ein Extrakt des Artikels wurde in deutscher Sprache in der Fachzeitschrift Risiko Manager veröffentlicht.[11] Der Artikel wurde im Rahmen der SFB Finanzakademie sowie der 7th InternationalConference on Computational and Financial Econometrics (CFE 2013) in London präsentiert.

[11] Ziggel, D., Berens, T., Wied, D., Weiß, G. (2013): Value-at-Risk im Risikomanagement: Der unevaluierte Standard, Risiko Manager, 24/2013, 1 & 7-9.

Beitrag IV (Kapitel 5):

Estimation Window Strategies for Value at Risk Forecasting.

Autor: Tobias Berens

Zusammenfassung: Im Vergleich zur großen Anzahl unterschiedlicher Modelle zur Schätzung von VaR- und ES-Prognosen existieren in der finanzwissenschaftlichen Literatur verhältnismäßig wenige Beiträge zu der Frage, welche Strategie zur Bestimmung der für die Parametrisierung solcher Modelle erforderlichen Datenfenster zu guten Ergebnissen führt. Im Rahmen dieses Beitrags werden unterschiedliche Datenfenster-Strategien auf parametrische, semi-parametrische und nicht-parametrische VaR-Modelle angewendet. Dabei werden sowohl einfache Modelle wie beispielsweise ein rollierendes Datenfenster wie auch komplexere Modelle, die auf Tests zur Identifizierung von Strukturbrüchen in der Varianz von Wertpapier-Renditen angewendet. Zudem wird untersucht, wie sich die Kombination einzelner Strategien auf die Prognosefähigkeit der Modelle auswirkt. Die Evaluierung der VaR-Prognosen erfolgt auf Basis statistischer Tests der Eigenschaften von VaR-Überschreitungen. Konkret wird getestet, ob diese der Unconditional Coverage Eigenschaft entsprechen und sowohl unabhängig als auch identisch verteilt sind. Zusätzlich werden Tests auf korrekte ES-Prognosen und auf Conditional Predictive Ability durchgeführt. Der Beitrag zeigt, dass die Auswahl der Strategie zur Bestimmung des Datenfensters zu signifikanten Unterschieden in den VaR- und ES-Prognosen der VaR-Modelle führt. Dabei ist grundsätzlich zu erkennen, dass die Kombination einzelner Datenfenster-Strategien im Vergleich zu den übrigen Strategien vorteilhaft ist.

Beitrag V (Kapitel 6):

Testing for Structural Breaks in Correlations: Does it Improve Value-at-Risk Forecasting?

Autoren: Tobias Berens, Gregor N.F. Weiß und Dominik Wied

Zusammenfassung:

Im Rahmen der Prognose von VaR und ES werden in diesem Beitrag das Constant Conditional Correlation (CCC) sowie das Dynamic Conditional Correlation (DCC) Modell mit einem paarweisen Test auf konstante Korrelationen, einem Test auf eine konstante Korrelationsmatrix sowie einem Test auf eine konstante Kovarianzmatrix kombiniert. Eine empirische Studie auf Basis multivariater Portfolios analysiert die Prognosefähigkeit sowohl der modifizierten als auch der einfachen Modelle ohne Berücksichtigung von Strukturbrüchen. Dabei erfolgt die Bewertung anhand statistischer Tests der Unconditional Coverage und der Independence Eigenschaft der VaR-Überschreitungen sowie der Vorhersagegenauigkeit der ES-Prognosen. Daneben beinhaltet die Studie ein Vergleich der Ansätze auf Basis aufsichtsrechtlicher Methoden und der Conditional Predictive Ability. Die Ergebnisse der Untersuchung zeigen, dass die mit Strukutbruchtests modifizierten Modelle grundsätzlich in der Lage sind, bessere Prognosen zu generieren.

Publikationsdetails:

Zur Veröffentlichung eingereicht in: Journal of Empirical Finance.

Kapitel 2

On the application of new tests for structural changes on global minimum-variance portfolios

Veröffentlicht in:
Statistical Papers, Vol. 54, Issue 4, 2013, pp. 955-975 (zusammen mit Dominik Wied und Daniel Ziggel).

2.1 Introduction

The mean-variance approach proposed by Markowitz [1952] has become the foundation of large parts of modern finance theory. Its simplicity and intuitive arrangement cause its common use in both industry and science. In the beginning it was usually supposed that the parameters of interest, i.e. expected returns, volatilities and correlations can be estimated accurately. Nowadays, this assumption is, at least, questionable. As shown in various works, it is not an appropriate simplification for expected returns in most practical situations (Chopra and Ziemba, 1993, Kempf and Memmel, 2002, Merton, 1980). Moreover, it is well known, in particular in empirical finance, that variances and correlations among many time series cannot be assumed to remain constant over longer periods of time (Krishan et al., 2009). A good example is the recent financial crisis, in which capital market volatilities and correlations raised quite dramatically. As a consequence, risk figures increased significantly as diversification effects were overestimated (Bissantz et al., 2011a, Bissantz et al., 2011b).

It is well known that the expected return is the most important parameter within the Markowitz model, cf. e.g. Gohout and Specht [2007]. Chopra and Ziemba [1993] demonstrate that, for mean tolerated risk levels, wrong return estimators have an eleven times larger impact than wrong risk estimators. Best and Grauer [1991] investigate the sensitivity of mean-variance-efficient portfolios to changes in the means of individual assets. The results state that portfolio weights are extremely sensitive to changes in asset means and a surprisingly small increase in the mean of a single asset drives half the securities from the portfolio.

One possible solution to overcome the problem of portfolio weights, which seem overly sensitive to estimation errors of returns, is the usage of the global minimum-variance portfolio at the left-most tip of the mean-variance efficient frontier. This portfolio has the unique property that portfolio weights are independent of the forecasted or expected returns on the individual assets as risk is minimized without an expected return input. Besides the advantage that no returns have to be estimated, the global minimum-variance portfolio allows the investor a risk optimal strategy. This is of special interest as passive (equity) investing has gained popularity (Clarke et al., 2006). Moreover, the covariance matrix can usually be estimated and forecasted much more reliable, resulting in an advantage for both practical and theoretical problems (Golosnoy et al., 2011).

In this context, several studies find that mean-variance optimization does not appear to result in a meaningful diversification improvement over non-optimized portfolios, despite the added complexity. This finding is consistent with the extensive literature documenting the puzzling under-performance of global minimum-variance approaches (Chow et al., 2011). Nevertheless, using historical backtests, Haugen and Baker [1991] and Clarke et al. [2006] demonstrate that minimum-variance strategies improve upon their cap-weighted counterparts by supplying better returns with reduced volatility, suggesting a latent potential of this approach. In order to open up this potential, the remaining market parameters (i.e. correlations and volatilities) have to be modelled time-dependent and flexible.

One of such methods is the regime switching model. This model allows the market parameters to depend on the market mode which switches among a finite number of states. In the simplest form, the market could be divided as “bullish” and “bearish” with very different market parameters. Of course, it is useful to define more intermediate states between these extremes, cf. e.g. Garcia and Perron [1996]. Zhou and Yin [2003] develop a continuous-time version mean-variance portfolio selection model with regime switching and attempt to derive closed-form solutions for both, efficient portfolios and efficient frontier. Although this approach is promising, the question arises how to define the states properly. Moreover, market parameters change frequently and are complexly interwoven. This kind of problem is addressed with control charts in e.g. Golosnoy et al. [2011]. The present paper makes use of several recently proposed fluctuation tests for structural changes in the market parameters.

For example, Aue et al. [2009] and Wied et al. [2012b] propose formal completely nonparametric tests for unconditional dependence measures. They do not build upon prior knowledge as to the timing of potential shifts. More precisely, Aue et al. [2009] propose a test to detect changes in the (multivariate) covariance structure, while Wied et al. [2012b] present a method to test for changes in the correlation structure between two assets. They are based on cumulated sums of second order empirical cross moments (in the style of Ploberger et al., 1989) and reject the null of constant covariance or correlation structure if these cumulated sums fluctuate too much. A similar fluctuation test for univariate variances is introduced and applied to financial time series by Wied et al. [2012a].

In this paper, we investigate if a combination of these new fluctuation tests and the classical minimum-variance approach improves global minimum-variance portfolios. To this end, we perform an empirical out-of-sample study and compare the performance of the resulting portfolios. Moreover, we investigate the resulting portfolio shiftings as a further quality measure.

The remainder of the paper is organized as follows. In Section 2 we present a summary of the required theory and introduce the investigated tests for structural breaks. A description of the empirical analysis is given in Section 3, while the results can be found in Section 4. Finally, we end with a discussion of the results in Section 5 and a conclusion in Section 6.

2.2 Methods

In this section, we briefly present the mean-variance approach proposed by Markowitz [1952] for constructing a portfolio with minimal variance. To this end, we assume that there are d risky assets with random returns, $R^1, \ldots, R^d$, such that $R := (R^1, \ldots, R^d)$. Let μ be the vector of expectations and Σ the covariance matrix of R which is assumed to be positively definite (such that there is no risk-free asset). The vector μ and the matrix Σ are both assumed to be constant over time.

A portfolio is a mixture of the n assets with portfolio weights $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbb{R}^d$ such that $\mathbf{a}'1_d = 1$. In the mean-variance theory we want to solve the optimization problem

$$\min_{\mathbf{a} \in \mathbb{R}^d} \mathbf{a}'\Sigma\mathbf{a} \text{ s.t. } \mathbf{a}'\mu = \mu_{\mathsf{P}}, \mathbf{a}'1_d = 1, \tag{2.1}$$

where μ_{P} is a constant chosen by the analyst.

In general, the solution of this problem depends on the value of μ_{P}. However, it is also possible to solve the problem globally with the weighting vector

$$\mathbf{a}_{min} = \frac{1}{1_d'\Sigma^{-1}1_d}\Sigma^{-1}1_d$$

which yields the lowest possible variance $(1_d'\Sigma^{-1}1_d)^{-1}$. In practice, some more assumptions on $\mathbf{a}$ are often imposed, e.g. the entries of $\mathbf{a}$ have to be nonnegative (such that no short sales are allowed) or have to be bounded (such that we have maximal limits). In these cases, the optimization problem (2.1) still has a well-defined solution which can be calculated or approximated with numerical optimization.

The global minimum-variance portfolio has the unique property that portfolio weights are independent of the expected returns, which are very difficult to estimate. Hence, this portfolio relies solely on the covariance matrix which can usually be estimated more accurately.

To get a feasible solution of (2.1) in practice, it is necessary to estimate Σ based on realizations of R. If we assume Σ to be constant over time, it is useful to use the largest quantity of data available for estimation. If we, however, make the more realistic assumption that Σ is subject to structural changes, we have to take these changes into account. In this paper, we compare two nonparametric approaches for testing for the presence of structural breaks: The first one by Aue et al. [2009] tests for changes in the complete matrix. Since they assume throughout their paper that the vector of expectations is equal to 0, the whole test and the test statistic base on the second-order cross moments. Complementarily to this, the tests proposed by Wied et al. [2012b] and Wied et al. [2012a] separately test for changes in correlations and variances. Since the covariance matrix Σ can be written as

$$\Sigma(i,j) = \sqrt{\mathsf{Var}(R^i)} \cdot \sqrt{\mathsf{Var}(R^j)} \cdot \mathsf{Cor}(R^i, R^j),\ i,j \in \{1,\ldots,d\}, \tag{2.2}$$

we can thus steer each entry of Σ separately.

Basically, all three nonparametric tests work in a similar way: Given the null hypothesis of constant covariance matrix, correlation or variance and T realizations of R these fluctuation tests compare the successively estimated covariance matrix (transformed into a vector with the vec-operator), (pairwise) correlation coefficient or (element-wise) variance with the respective value calculated from all T observations. The null hypothesis is rejected whenever these differences become too large over time. To be more precisely, the test statistic is a functional, for example the maximum - functional, of the series

$$P(j) := \frac{j^2}{T}(\hat{q}_j - \hat{q}_T)'\hat{\Sigma}_q^{-1}(\hat{q}_j - \hat{q}_T),$$

where $\hat{q}_j$ is the quantity of interest calculated from the first j observations, $\hat{q}_T$ is the quantity of interest calculated from the first T observations and $\hat{\Sigma}_q$ is an estimator (from all T observations) for the asymptotic covariance matrix of $\hat{q}_j$ under the null. Both expressions $\frac{j^2}{T}$ and $\hat{\Sigma}_q^{-1}$ serve for standardization. In particular, with $\frac{j^2}{T}$ less weight is laid on the differences at the beginning, where the parameters cannot be well estimated. The expression $\hat{\Sigma}_q^{-1}$ captures serial dependence and fluctuations of the time series. The process $P(j)$ converges against a Gaussian process and thus, in practice we compare the functionals of

$P(j)$ with the respective quantiles of this functional. In the correlation case, we get

$$\max_{2\leq j\leq T}|\sqrt{P(j)}| \to_d \sup_{0\leq z\leq 1}|B(z)|$$

and in the variance case, we get

$$\max_{1\leq j\leq T}|\sqrt{P(j)}| \to_d \sup_{0\leq z\leq 1}|B(z)|.$$

In these cases, $B(z)$ is a one-dimensional Brownian bridge with quantiles 1.358 (95%) and 1.628 (99%).

In the case of a covariance matrix, we have

$$\max_{1\leq j\leq T}|P(j)| \to_d \sup_{0\leq z\leq 1}\sum_{i=1}^{d}(B_i(z))^2,$$

where d is the number of upper-diagonal elements in the covariance matrix and $B_i(\cdot)$ are independent Brownian bridges. We approximate the quantiles of this limit distribution by simulating Brownian bridges on a fine grid. For this, the representation of a Brownian bridge as a limit of a random walk is used. Note that we do not use the “second” approximation for growing d, which is discussed in Remark 2.1 in Aue et al. [2009], as this does not seem to be appropriate here (cf. Aue et al., 2009, p. 4064). Based on these simulations, we obtain 53.583 (95%) and 56.961 (99%) as quantiles for 18 assets.

The tests are basically applicable to financial time series with its specific characteristics such as serial dependence and missing normality. For example, all tests can be applied if the returns can be modeled by a GARCH process. An important property is the fact that the location of the possible change points need not be specified a priori. In general, these fluctuation tests are sufficiently powerful and Aue et al. [2009] prove consistency of the covariance matrix test against fixed alternatives while Wied et al. [2012b] and Wied et al. [2012a] obtain local power results against smooth alternatives characterized by a continuous function g.

Once the presence of a parameter change is detected, a suitable estimate of its location can be obtained by the statistic proposed in Galeano and Wied [2014] (the original idea goes back at least to Vostrikova, 1981), i.e. by the point at which $P(j)$ (or a transformation of $P(j)$) takes its maximum. For example, in the correlation case we get $\widehat{k} = \arg\max_{2 \leq j \leq T} \sqrt{P(j)}$. Since we use these break point estimators in our study, we have decided to focus on the maximum-functional instead of considering for example the Cramér-von Mises functional as e.g. do Aue et al. [2009], equation (2.8).

2.3 Empirical investigation

2.3.1 Data

In order to investigate if a combination of the above mentioned fluctuation tests and the classical minimum-variance optimization yields reasonable results, we perform an out-of-sample study and compare the results with several alternative methods. We use two different data sets. More precisely, we use daily log-returns based on final quotes of 18 sector subindices based on the STOXX EUROPE 600 (total return indices) and log-returns based on final quotes of 18 stocks (treated as total return indices), which were listed on the DAX 30 for the period between 01.01.1973 and 30.06.2011 (10044 data points). For the subindices, data are available for the time span 01.01.1992 - 30.06.2011, which equates to 5087 data points. All data sets are obtained from *Thomson Reuters Datastream*.

2.3.2 Parameter Estimation

As already mentioned, for a fixed point in time, calculation of the global minimum-variance portfolio depends only on the estimated covariance matrix. Hence, we compare the results of several estimation procedures. First, we use the empirical covariance matrix given by the last 250/500/1000 data points. For sake of simplicity, we denote combinations of these empirical estimators and the minimum-variance optimization as plain Markowitz optimizations. In addition to that, we use the new fluctuation tests. Here, the estimation procedure is as follows:

1. Initialize $i = 1$, $k = 1000$ and $m =$ number of observed returns.

2. Perform the fluctuation test for the data $\{x_i, ..., x_k\}$.

3. If the test rejects the null, define $i = l$, where l maximizes the corresponding functional of $P(j)$ and go back to step 2. Otherwise use the data $\{x_p, ..., x_k\}$ in order to calculate the empirical estimator of the respective parameter, where $p = min\{i, k - 19\}$.

4. Set $k = k+n$, where n is the number of days between two optimizations. If $k > m$, stop. Otherwise go back to step 2.

Note that the modification $p = min\{i, k - 19\}$ ensures that at least 20 data points are used for parameter estimation. This proceeding is in line with Wied et al. [2012a]. As mentioned above, we use the fluctuation tests in two different ways. On the one hand, we use the test of Aue et al. [2009]. Hence, the procedure provides the covariance matrix directly. On the other hand, we separately apply the tests proposed by Wied et al. [2012b] and Wied et al. [2012a]. The resulting covariance matrix is then given by (2.2).

We choose $\alpha = 1\%$ and $\alpha = 5\%$ as significance levels for the fluctuation tests. The choice of $\alpha = 1\%$ is due to the fact that in this case the number of possible false signals should be relatively small. Nevertheless, several applications show that $\alpha = 5\%$ yields convincing results in practice (Ziggel and Wied, 2012).

2.3.3 Optimization

In addition to parameter estimation, there are several adjusting screws concerning the optimization which have an impact on the results. First, we have to define an interval for re-optimizations. To this end, we define $n = 21, 63$ and 252, respectively. These choices correspond to monthly, quarterly and yearly re-optimization. The same intervals will also be used in order to perform a re-balancing of the equally weighted portfolio which serves as a benchmark. These frequencies allow us to neglect the problem of sequential testing. Nevertheless, it would be worthwhile to implement a theoretical analysis for smaller frequencies about this issue using ideas of Chu et al. [1996] and Wied and Galeano [2013], but this lies beyond the scope of the present paper.

Aside from the interval for re-optimizations, we analyze the impact of some additional constraints to the portfolio weights. In a first run, we define $|a_i| \leq 1, \forall i$ which particularly allows for short selling. In the next step, we exclude short selling by requiring $0 \leq a_i \leq 1, \forall i$.

2.3.4 Miscellaneous

We use several quality criteria in order to judge the performance. Of course, we investigate the resulting variance. Nevertheless, we also compare the resulting returns and Sharpe-ratios. To this end, we assume 1.1% as risk free return for the latter. This corresponds to the average return of German government bonds with less than 3 years to maturity in the year 2011. Besides, we measure the portfolio turnover in order to draw conclusions for a usage in practice. In line with DeMiguel et al. [2009], we define the average absolute change in the weights as

$$Turnover(R) = \frac{1}{TD-1} \sum_{i=1}^{TD-1} \sum_{j=1}^{d} |a_{i+1,j} - a_{i+,j}|,$$

where TD is the number of the trading days and d the number of assets. Besides, $a_{i+,j}$ is the portfolio weight of asset j before a rebalancing or re-optimization at time $i+1$. In addition, we call $Turnover(A)$ the absolute amount of changes, that means $Turnover(A) = Turnover(R) \cdot (TD-1)$.

To evaluate the impact of diverging turnovers, we compute adjusted returns and Sharpe-ratios by including transaction costs. Therefore, we assume a constant relative bid-ask spread s_c (bid-ask spread divided by bid-ask midpoint) which diminishes the return R. To quantify the spread, we have analyzed daily bid and ask quotes of the 18 stocks listed on the DAX 30 and for all stocks listed on the STOXX EUROPE 600 for the time span 01.07.2010-30.06.2011. The average relative spread of the analyzed stocks amounts to 0.15% (DAX) and 0.22% (STOXX). As a simple approximation, we determine s_c

to be 0.2% in both cases. The loss of return due to transaction costs is calculated by $Turnover(A) \cdot \frac{s_c}{2}$.

MATLAB R2009b is used for all computations. While the global optimization problem can be solved analytically, numerical optimization methods are necessarily under additional conditions on the weighting vector **a** (see Section 2). We perform these methods with the "fmincon" function included in the "Optimization Toolbox". Since the usage of just one starting point in the optimization can lead to a local minimum, we use multiple starting points. More precisely, we use starting points which lie on the boundary of the feasible region, the equally weighted portfolio and some random starting points. However, the optimizations have proceeded stable and the starting points have had only minor impact on the results.

2.4 Results

In this section we present the results of our out-of-sample study which can be found in Tables 2.1 to 2.8. We start with the dataset including 18 sector subindices based on the STOXX EUROPE 600. As described in Section 3, the equally weighted portfolios serve as a benchmark. It is noticeable that the interval for re-balancings has only a negligible effect on these results. In all cases the p.a. volatility is around 19.2%, while the average p.a. return is slightly above 8%. Besides, the portfolio turnover is very low and has no relevant impact on adjusted returns and Sharpe-ratios.

Table 2.1: Results for the Equally Weighted Portfolios - Indices
Results for the equally weighted portfolios including 18 sector subindices based on the STOXX EUROPE 600. Values in brackets include transaction costs.

Interval Re-Balancing	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
21	0.3620	(0.3604)	8.08%	(8.05%)	19.29%	0.03	5.66
63	0.3633	(0.3622)	8.07%	(8.05%)	19.20%	0.05	3.24
252	0.3638	(0.3633)	8.04%	(8.03%)	19.08%	0.10	1.62

The Markowitz optimizations based on the empirical covariance matrix improve upon the equally weighted portfolios. The average volatility decreases by 3.99% to 5.78%, while the return simultaneously increases by 0.1% to 3.04%. Nevertheless, the portfolio turnover increases by about ten times on average leading to return losses of 0.03% to 0.59%. With respect to the setup options of the plain Markowitz optimizations, the choice of the days of data history as well as the re-optimization interval has only little impact to the volatility results. Nevertheless, in terms of returns, turnover and the Sharpe-ratio, the choice of 1,000 days as the data point history seems to be preferable. Besides, the choice of the weight limits has a marked effect on the results. The allowance for short selling reduces the volatility by more than 1% on average.

Table 2.2: Results for the Plain Markowitz Optimizations - Indices

Results for optimizations using the empirical covariance matrix including 18 sector subindices based on the STOXX EUROPE 600. Values in brackets include transaction costs.

# Data	Interval Re-Opt.	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
Panel A: $\|a_i\| < 1$								
250	21	0.5819	(0.5388)	9.02%	(8.43%)	13.61%	0.61	118.53
250	63	0.6283	(0.6013)	9.46%	(9.10%)	13.30%	1.14	73.18
250	252	0.5338	(0.5196)	8.51%	(8.32%)	13.89%	2.44	38.95
500	21	0.6800	(0.6549)	10.24%	(9.90%)	13.44%	0.35	68.29
500	63	0.7014	(0.6852)	10.52%	(10.30%)	13.43%	0.69	43.85
500	252	0.6074	(0.5985)	9.66%	(9.53%)	14.09%	1.61	25.77
1000	21	0.7295	(0.7144)	11.01%	(10.80%)	13.58%	0.22	42.12
1000	63	0.7292	(0.7197)	11.12%	(10.99%)	13.74%	0.42	26.63
1000	252	0.6149	(0.6093)	9.98%	(9.89%)	14.43%	1.10	17.58
Panel B: $0 < a_i < 1$								
250	21	0.5128	(0.5014)	8.56%	(8.40%)	14.55%	0.17	33.27
250	63	0.4935	(0.4860)	8.29%	(8.18%)	14.57%	0.34	21.99
250	252	0.4754	(0.4715)	8.18%	(8.12%)	14.88%	0.81	13.03
500	21	0.5211	(0.5141)	8.68%	(8.58%)	14.55%	0.10	20.26
500	63	0.5158	(0.5113)	8.66%	(8.59%)	14.65%	0.22	14.07
500	252	0.5142	(0.5110)	8.75%	(8.70%)	14.88%	0.59	9.36
1000	21	0.5584	(0.5542)	9.35%	(9.29%)	14.77%	0.07	13.08
1000	63	0.5582	(0.5553)	9.41%	(9.36%)	14.88%	0.15	9.47
1000	252	0.5618	(0.5597)	9.58%	(9.55%)	15.09%	0.44	7.04

By using the test of Aue et al. [2009] in order to estimate the covariance matrix, the resulting level of volatilities is slightly higher than those of the empirical covariance matrix. Nevertheless, the returns increase by about 2% on average which leads to superior

Sharpe-ratios. Besides, the turnover is much lower and is reduced by about two thirds compared to the plain Markowitz optimization approach. Transaction costs reduce the returns by only 0.07% on average. The application of different significance levels (5% and 1%) makes no notable difference in the results. Considering the volatility, the choice of 21 days for the re-optimization interval lowers the volatility by about 1% compared to 252 days. Astonishingly, no clear statement can be made with regards to the limits of asset weights because the results are quite inconclusive. On average, the differences between both options are negligible. In terms of the returns, the results of the different test and optimization options are comparable to each other and can be located between 10.88% and 11.11%.

Table 2.3: Results for the Markowitz Optimizations in Combination with the Test of Aue et al. [2009] - Indices

Results for optimizations using the test of Aue et al. [2009] including 18 sector subindices based on the STOXX EUROPE 600. Values in brackets include transaction costs.

α	Interval Re-Opt.	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
Panel A: $\lvert a_i \rvert < 1$								
95%	21	0.6901	(0.6790)	10.77%	(10.62%)	14.02%	0.16	30.25
95%	63	0.6750	(0.6681)	10.89%	(10.79%)	14.51%	0.30	19.26
95%	252	0.6063	(0.6023)	10.60%	(10.54%)	15.67%	0.78	12.52
99%	21	0.7021	(0.6920)	10.97%	(10.82%)	14.05%	0.15	29.75
99%	63	0.6936	(0.6865)	11.06%	(10.96%)	14.37%	0.30	19.30
99%	252	0.6355	(0.6313)	11.01%	(10.95%)	15.60%	0.77	12.33
Panel B: $0 < a_i < 1$								
95%	21	0.6658	(0.6628)	11.15%	(11.10%)	15.09%	0.05	9.79
95%	63	0.6578	(0.6555)	11.08%	(11.04%)	15.17%	0.11	7.32
95%	252	0.6433	(0.6416)	10.99%	(10.96%)	15.37%	0.36	5.69
99%	21	0.6747	(0.6714)	11.30%	(11.25%)	15.12%	0.05	9.71
99%	63	0.6665	(0.6643)	11.20%	(11.16%)	15.15%	0.11	7.24
99%	252	0.6393	(0.6372)	10.96%	(10.93%)	15.43%	0.35	5.61

An application of the tests of Wied et al. [2012b] and Wied et al. [2012a] yields favorable results compared to the benchmark of equally weighted portfolios. The results are however considerably worse with respect to each measured performance indicator compared to the plain Markowitz optimization as well as to the optimization including the test of Aue et al. [2009]. High turnovers lead to a substantial loss of returns by 0.35% on average. Nevertheless, it should be noted that the choice of the weight limits has a considerable impact on the resulting volatility. Surprisingly, the more restrictive option of $0 \leq a_i \leq 1, \forall i$ shows lower volatilities.

Table 2.4: Results for the Markowitz Optimizations in Combination with the Test of Wied et al. [2012b] and Wied et al. [2012a] - Indices
Results for optimizations using the tests of Wied et al. [2012b] and Wied et al. [2012a] including 18 sector subindices based on the STOXX EUROPE 600. Values in brackets include transaction costs.

α	Interval Re-Opt.	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
Panel A: $\lvert a_i \rvert < 1$								
95%	21	0.5813	(0.5142)	11.34%	(10.16%)	17.61%	1.23	239.12
95%	63	0.5958	(0.5682)	12.33%	(11.81%)	18.85%	1.64	104.82
95%	252	0.3728	(0.3629)	8.73%	(8.53%)	20.48%	2.50	40.06
99%	21	0.3539	(0.3149)	7.18%	(6.51%)	17.17%	0.70	135.93
99%	63	0.4615	(0.4452)	8.96%	(8.68%)	17.03%	0.88	56.22
99%	252	0.4075	(0.4029)	8.33%	(8.24%)	17.73%	1.10	17.63
Panel B: $0 < a_i < 1$								
95%	21	0.3882	(0.3545)	7.28%	(6.74%)	15.92%	0.56	108.29
95%	63	0.6144	(0.5953)	10.70%	(10.40%)	15.62%	0.95	60.86
95%	252	0.3418	(0.3359)	6.72%	(6.62%)	16.44%	1.23	19.61
99%	21	0.4675	(0.4569)	8.82%	(8.64%)	16.51%	0.18	35.67
99%	63	0.4900	(0.4831)	9.23%	(9.11%)	16.59%	0.36	23.22
99%	252	0.5138	(0.5098)	9.12%	(9.06%)	15.62%	0.71	11.35

We continue with the second dataset including the returns of 18 stocks, which were listed on the DAX 30 for the time span 01.01.1973 - 30.06.2011. The benchmark of equally weighted portfolios shows that the re-balancing interval has only very little effect on the volatility as well as on the return. The volatility amounts to about 18.8%, while the returns are around 11.4%. Transaction costs are negligible.

Compared to the equally weighted portfolios, the results of the plain Markowitz optimizations show an improvement again. The average volatility decreases by 2.21%, while the average return increases by 0.48%. Consequently, the average Sharpe-ratio increases by about 0.10 points. The portfolio turnover is about six times higher, while transaction costs decrease the returns by averaged 0.14%. Concerning the setup options, a lower re-

Table 2.5: Results for the Equally Weighted Portfolios - Stocks

Results for the equally weighted portfolios including 18 stocks listed on the DAX 30. Values in brackets include transaction costs.

Interval Re-Balancing	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
21	0.5520	(0.5496)	11.50%	(11.45%)	18.84%	0.04	18.49
63	0.5475	(0.5464)	11.42%	(11.37%)	18.84%	0.07	10.59
252	0.5366	(0.5362)	11.14%	(11.13%)	18.70%	0.14	4.90

optimization interval is accompanied by lower volatilities and returns, while the influence to the Sharpe-ratio is inconclusive. Furthermore, the results show only a little impact of the choice of the data history and surprisingly the weight limits.

Table 2.6: Results for the Plain Markowitz Optimizations - Stocks

Results for optimizations using the empirical covariance matrix including 18 stocks listed on the DAX 30. Values in brackets include transaction costs.

# Data	Interval Re-Opt.	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
Panel A: $\lvert a_i \rvert < 1$								
250	21	0.6489	(0.6205)	11.39%	(10.94%)	15.85%	0.42	181.30
250	63	0.5888	(0.5722)	11.03%	(10.75%)	16.87%	0.77	110.70
250	252	0.6413	(0.6334)	12.29%	(12.15%)	17.45%	1.56	54.48
500	21	0.6178	(0.6016)	10.89%	(10.63%)	15.84%	0.24	103.90
500	63	0.6229	(0.6124)	11.29%	(11.13%)	16.37%	0.46	65.59
500	252	0.6578	(0.6527)	12.37%	(12.28%)	17.13%	1.04	35.86
1000	21	0.6553	(0.6457)	11.58%	(11.42%)	15.99%	0.14	62.10
1000	63	0.6519	(0.6460)	11.65%	(11.55%)	16.18%	0.27	39.08
1000	252	0.6861	(0.6824)	12.50%	(12.44%)	16.62%	0.67	23.52
Panel B: $0 < a_i < 1$								
250	21	0.6567	(0.6435)	11.83%	(11.62%)	16.34%	0.20	85.58
250	63	0.6235	(0.6154)	11.56%	(11.43%)	16.78%	0.37	53.32
250	252	0.6489	(0.6446)	12.36%	(12.28%)	17.35%	0.86	30.11
500	21	0.6345	(0.6264)	11.47%	(11.34%)	16.34%	0.13	53.53
500	63	0.6453	(0.6399)	11.75%	(11.66%)	16.51%	0.24	34.26
500	252	0.6669	(0.6641)	12.51%	(12.46%)	17.10%	0.61	21.33
1000	21	0.6611	(0.6559)	11.93%	(11.84%)	16.38%	0.08	34.46
1000	63	0.6568	(0.6538)	11.91%	(11.86%)	16.45%	0.15	21.81
1000	252	0.6783	(0.6765)	12.54%	(12.51%)	16.86%	0.40	13.93

The extension by the test of Aue et al. [2009] outperforms the plain Markowitz optimization and results to small improvements of the average returns and Sharpe-ratios. But in contrast to the application to the subindices dataset, the volatility remains unchanged. Additionally, the portfolio turnover and hence transaction costs are much lower. There are just minor changes of the performance measures due to the choice of the setup options except the re-optimization interval, where the return increases with larger gaps.

Table 2.7: Results for the Markowitz Optimizations in Combination with the Test of Aue et al. [2009] - Stocks
Results for optimizations using the test of Aue et al. [2009] including 18 stocks listed on the DAX 30. Values in brackets include transaction costs.

α	Interval Re-Opt.	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
Panel A: $\lvert a_i \rvert < 1$								
95%	21	0.6493	(0.6439)	11.74%	(11.65%)	16.38%	0.09	36.71
95%	63	0.6602	(0.6565)	11.93%	(11.87%)	16.41%	0.16	22.81
95%	252	0.6868	(0.6851)	12.72%	(12.69%)	16.91%	0.39	13.66
99%	21	0.6552	(0.6496)	11.85%	(11.76%)	16.41%	0.08	35.88
99%	63	0.6632	(0.6600)	11.98%	(11.92%)	16.40%	0.16	22.32
99%	252	0.6952	(0.6937)	12.80%	(12.77%)	16.82%	0.37	12.90
Panel B: $0 < a_i < 1$								
95%	21	0.6588	(0.6553)	11.99%	(11.93%)	16.53%	0.06	23.60
95%	63	0.6637	(0.6515)	12.10%	(12.06%)	16.57%	0.10	14.45
95%	252	0.6820	(0.6807)	12.65%	(12.63%)	16.93%	0.26	9.10
99%	21	0.6623	(0.6585)	12.05%	(11.99%)	16.54%	0.05	23.14
99%	63	0.6636	(0.6614)	12.08%	(12.04%)	16.54%	0.10	14.21
99%	252	0.6846	(0.6836)	12.64%	(12.62%)	16.85%	0.25	8.67

Employing the tests of Wied et al. [2012b] and Wied et al. [2012a], the results show a slight decrease of the Sharpe-ratio compared to the equally weighted benchmark portfolio which is caused by a small improvement of the average return and an increase of the volatility. However, this approach does not achieve the convincing results of the remaining optimization methods. This goes along with the highest portfolio turnover and transaction costs of all alternatives. In line with the corresponding optimization on the basis of the subindices data, the allowance for short sales leads to a substantial higher volatility on average.

Table 2.8: Results for the Markowitz Optimizations in Combination with the Test of Wied et al. [2012b] and Wied et al. [2012a] - Stocks
Results for optimizations using the tests of Wied et al. [2012b] and Wied et al. [2012a] including 18 stocks listed on the DAX 30. Values in brackets include transaction costs.

α	Interval Re-Opt.	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
Panel A: $\lvert a_i \rvert < 1$								
95%	21	0.4929	(0.4420)	10.04%	(9.12%)	18.14%	0.86	367.74
95%	63	0.4418	(0.4186)	11.74%	(11.18%)	24.09%	1.55	221.80
95%	252	0.5128	(0.5033)	14.21%	(13.97%)	25.56%	2.76	96.66
99%	21	0.4593	(0.4358)	11.21%	(10.69%)	22.01%	0.48	206.76
99%	63	0.4456	(0.4281)	10.56%	(10.19%)	21.23%	1.03	147.80
99%	252	0.5524	(0.5464)	12.51%	(12.38%)	20.65%	1.45	50.83
Panel B: $0 < a_i < 1$								
95%	21	0.5020	(0.4847)	10.52%	(10.19%)	18.76%	0.30	130.15
95%	63	0.5463	(0.5324)	10.47%	(10.24%)	17.16%	0.65	93.48
95%	252	0.6264	(0.6212)	12.28%	(12.19%)	17.86%	1.07	37.59
99%	21	0.4115	(0.4046)	9.52%	(9.38%)	20.46%	0.13	56.86
99%	63	0.6377	(0.6314)	11.94%	(11.83%)	17.00%	0.29	42.02
99%	252	0.6529	(0.6496)	12.50%	(12.44%)	17.46%	0.67	23.38

2.5 Discussion

In line with previous works of Haugen and Baker [1991] and Clarke et al. [2006], our empirical study supports the finding that plain Markowitz optimized portfolios deliver superior results in terms of portfolio variance as well as portfolio returns compared to equally weighted portfolios. On the basis of two different datasets, we show that equally weighted portfolios are clearly outperformed by this optimization strategy. Moreover, the benefit of lower volatilities and higher returns is only marginally offset by increasing transaction costs due to considerable higher portfolio turnovers.

The extension of the plain Markowitz optimization by the test of Aue et al. [2009] leads to inconclusive results. With respect to the two used datasets, the results show increased returns and volatilities on average. However, it is remarkable that the portfolio turnover is much lower compared to the classical optimization. Basically, this is reasoned by the fact that the test yields only a few rejections of the null hypothesis of a constant covariance matrix. For example, a portfolio optimization including the DAX 30 dataset under the option setup of a 5% significance level and a re-optimization interval of 21 days leads to only four rejections within 10,043 data points.

The small number of rejections might be the result of a lack of accuracy of the critical values in connection to the setup of our study. The critical values are approximated by simulating Brownian bridges on a fine grid as described in Section 2. However, an additional simulation study, whose results are available from the authors upon request, indicates that this approximation does not perform well if the sample size is small. We simulated the actual critical values for $d = 18$ by generating standard normal distributed values and calculating the respective test statistic. For example, for a sample of $1,000$ data points the 0.95-quantile is 24.20 while the asymptotic critical value is 53.58. Probably, a suitable derivation of finite sample critical values is a non-trivial task because in practice the underlying distribution of the asset returns is unclear; especially the assumption of the standard normal distribution is doubtful. Nevertheless, we used this procedure to show the effect of using critical values that are to some degree more suitable for the finite samples of our dataset. As a simple and rough approximation we concern a sample size of $\lceil \frac{5,087}{2} \rceil = 2,544$ for the STOXX EUROPE 600 subindices dataset and $\lceil \frac{10,043}{2} \rceil = 5,022$ with respect to the DAX 30 dataset. The actual critical values for the 0.95-quantile (0.99-quantile) are estimated to 34.77 (36.43) and 41.32 (43.40). Applying the test with the modified critical values leads to a higher number of rejections, e.g. seven instead of four considering the example above (DAX 30, 5% significance level, 21 days interval). Compared to Table 2.3, the improved results of Table 2.9 show exemplary that the adjustment of our very simple approach is a step in the right direction. A more sophisticated procedure for calculating critical values may perform even better.

Certainly, the dates at which the null is rejected are of interest. Returning to the example mentioned above (DAX 30, 5% significance level, 21 days interval) these dates are 26.01.1983, 25.07.1989, 05.11.1996, and 19.02.2001 for the critical values based on the asymptotic analysis. In contrast to that, the dates at which the null is rejected are

Table 2.9: Results for the Markowitz Optimizations in Combination with the Test of Aue et al. [2009] - Modified Critical Values

Results for optimizations using the test of Aue et al. [2009] in combination with the modified critical values including 18 sector subindices based on the STOXX EUROPE 600. Values in brackets include transaction costs.

α	Interval Re-Opt.	Sharpe Ratio		Return		Volatility	Turnover (R)	Turnover (A)
Panel A: $\|a_i\| < 1$								
95%	21	0.7886	(0.7744)	11.94%	(11.75%)	13.75%	0.20	39.41
95%	63	0.8080	(0.7986)	12.34%	(12.21%)	13.91%	0.42	26.63
95%	252	0.6501	(0.6440)	10.61%	(10.52%)	14.63%	1.11	17.81
99%	21	0.7724	(0.7591)	11.63%	(11.45%)	13.63%	0.19	36.62
99%	63	0.7655	(0.7570)	11.76%	(11.64%)	13.93%	0.38	24.06
99%	252	0.6611	(0.6557)	11.05%	(10.97%)	15.05%	1.03	16.51
Panel B: $0 < a_i < 1$								
95%	21	0.6103	(0.6065)	10.16%	(10.10%)	14.84%	0.06	11.37
95%	63	0.6083	(0.6054)	10.19%	(10.15%)	14.95%	0.14	8.65
95%	252	0.5799	(0.5776)	9.87%	(9.84%)	15.13%	0.44	6.96
99%	21	0.6035	(0.5996)	10.07%	(10.01%)	14.86%	0.06	11.55
99%	63	0.6084	(0.6055)	10.21%	(10.16%)	14.97%	0.14	8.75
99%	252	0.6146	(0.6126)	10.53%	(10.50%)	15.34%	0.40	6.46

28.08.1975, 03.02.1981, 10.10.1986, 13.11.1990, 25.08.1995, 15.02.1999 and 03.12.2001 for the modified critical values based on a sample size of 5,022. Most of these dates seem to be reasonable. The Latin American debt crisis of the early 1980s in combination with the savings and loan crisis of the 1980s in the United States explain some rejection dates. Besides, in each case one rejection date corresponds to the German reunification. Finally, in both cases the last rejection date can be explained by the burst of the dot-com bubble. Nevertheless, in both cases no change point is detected during the market turmoils of the financial crisis at the end of the last decade or the current European sovereign-debt crisis. Hence, it is very likely that the accurate number of changes in the covariance matrix is somewhat higher.

As described in Section 2.4, the results of the optimization in combination with the tests proposed by Wied et al. [2012b] and Wied et al. [2012a] are relatively poor compared to the remaining optimization approaches. This could be the result of the special character of these statistical tests. In contrast to the test for changes in the entire covariance matrix, the test for changes of variances is applied to each of the $d = 18$ different time series, whereas the test for changes in correlations is applied to each of the $18(18-1)/2 = 153$ upper-diagonal elements of the correlation matrix. Due to the high number of statistical tests, it is very likely that after every re-optimization interval one or more tests (wrongly) reject the null hypothesis. For example, Table 2.10 and 2.11 show the number of rejections of the tests including the DAX 30 dataset under the option setup of a 1% significance level and a re-optimization interval of 21 days.

Table 2.10: Number of rejections - Test of Wied et al. [2012a]

Number of rejections of the null hypothesis of constant volatility for each asset for an optimization using the test of Wied et al. [2012a] including 18 stocks listed on the DAX 30 under the option setup of 1% significance level and a test interval of 21 days.

$d =$	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
Number of Rejections	10	8	7	7	9	13	12	9	8	7	12	3	7	9	9	11	13	7

Table 2.11: Number of rejections - Test of Wied et al. [2012b]

Number of rejections of the null hypothesis of constant correlation for each bivariate combination of the assets for an optimization using the test of Wied et al. [2012b] including 18 stocks listed on the DAX 30 under the option setup of 1% significance level and a test interval of 21 days.

$d =$	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
1	0	5	6	6	8	6	6	6	2	5	6	6	8	5	7	10	11	7
2		0	4	5	8	8	6	10	5	3	5	6	8	5	6	9	7	7
3			0	10	5	9	7	10	4	2	4	5	3	6	8	11	9	7
4				0	6	11	7	8	9	5	7	6	7	4	5	7	7	5
5					0	11	10	6	2	2	4	5	7	5	7	15	9	10
6						0	7	7	6	6	5	9	10	5	6	10	7	11
7							0	8	5	4	5	8	6	6	10	11	8	8
8								0	4	3	4	6	6	3	8	13	7	10
9									0	2	4	6	6	3	4	6	6	4
10										0	3	5	7	1	3	5	4	4
11											0	4	3	4	5	8	7	8
12												0	5	2	9	9	6	6
13													0	3	8	7	8	7
14														0	3	7	7	7
15															0	15	10	9
16																0	13	9
17																	0	8
18																		0

In summary, this setup leads to 161 rejections of the volatility test and $1,001$ rejections of the correlation test within 431 test intervals in total or 2.7 rejections at each test interval on average. As a consequence, the data history changes after each interval which might lead to substantial fluctuations within the covariance matrix and hence an increased portfolio turnover. Apparently, these large shifts have negative effects on the performance of the model.

In order to remedy this drawback, it may be advantageous to leave out the test for changes in the bivariate correlations. Bissantz et al. [2011b] show that the impact of fluctuations and estimation errors is ten times larger for volatilities than for correlations. Consequently, the detection of change points of volatilities is obviously much more important than the correlation based test. By omitting that test, the number of tests for each interval is reduced to $d = 18$. First studies show an improvement into the desired direction. On average, volatility is reduced by 1.29% (STOXX EUROPE 600) and by 2.68% (DAX 30). But the benchmark volatility and Sharpe-ratio levels of the plain Markowitz optimizations are still not attained. More details are available from the authors upon request. Further studies of the suggested solution may provide a deeper analysis.

In addition to the investigated strategies, it might be possible to pursue a further strategy, i.e. to let the fluctuation tests themselves determine reasonable dates for a re-optimization. To be more precisely, a re-optimization of the portfolio would only be performed if a fluctuation test rejects the null hypothesis. However, in this paper we refused this further strategy for two different reasons. With respect to the test of Aue et al. [2009], this strategy suffers from the seldom rejections of the null hypothesis. We would then re-optimize very infrequently which is not useful in practice. Regarding the tests proposed by Wied et al. [2012b] and Wied et al. [2012a], the opposite problem arises, namely the problem of multiple testing and undesired frequent re-optimizations. Consequently, this kind of application would require different theoretical adjustments of the procedures.

Surprisingly, the allowance for short selling does not lead to lower volatilities in all cases (e.g., see Table 2.6). Although it is not intuitive that imposing the constraint of non-negative portfolio weights leads to an improved efficiency, this finding is in line with the empirical study of Jagannathan and Ma [2003]. These authors argue that constraints for portfolio weights increase specification error, but can also reduce sampling error. The gain or loss in efficiency depends on the trade-off between both error types.

Since we employ portfolios consisting of very liquid German and European blue chip stocks, transaction costs are marginal due to the small bid-ask spread of 0.2%. For example, assuming $Turnover(A)$ to be 100 for the portfolio consisting of the dataset of 18 stocks listed on the DAX 30, the loss of annual log-returns amounts only to 0.25%. However, the impact of high turnovers may be significantly higher when datasets of less liquid assets are used. It would be worthwhile for further research to address a more detailed analysis of the trade-off between improved volatility and return of an optimized portfolio on the one side and costs relating to increased portfolio turnover on the other side.

2.6 Conclusion

The aim of this paper is to investigate whether a classical Markowitz mean-variance portfolio can be improved by the use of change point tests for dependence measures. To the best of our knowledge, we are the first to apply, on the one hand, the recently proposed test of Aue et al. [2009] for a constant covariance matrix and, on the other hand, the tests of Wied et al. [2012b] and Wied et al. [2012a] for constant variances and correlations to a minimum-variance optimization. We find out that portfolio optimizations considering change points of the covariance matrix yield considerable results and outperform plain Markowitz optimizations in several cases. In conducting the empirical study, we gain interesting insights in the behavior of these tests in combination with a portfolio optimization. This allows us to carve out the benefits as well as some challenging drawbacks of these new approaches. Moreover, we make some notes which might be helpful to future works.

Kapitel 3

Automated Portfolio Optimization Based on a New Test for Structural Breaks

Veröffentlicht in:

Acta Universitatis Danubius. Œconomica, Vol. 10, Issue 2, 2014, pp. 241-262 (zusammen mit Dominik Wied und Daniel Ziggel).

3.1 Introduction

The model by Markowitz [1952] represents a milestone in development of modern techniques concerning portfolio optimization. Nevertheless, it is well known that there are some serious challenges for the application of optimization techniques to portfolio management practice. In particular, the error-prone estimation of the expected returns is crucial for reasonable results of the optimization (Best and Grauer, 1991, Chopra and Ziemba, 1993). The global minimum-variance portfolio approach circumvents this problem. It determines the portfolio weights independently from expected returns. The optimization depends solely on the covariance matrix which can be estimated much more reliable than expected returns (Golosnoy et al., 2011). It leads to a minimum-variance portfolio that lies on the left-most tip of the efficient frontier. Considering equity portfolios, numerous historical backtests show that minimum-variance optimization provides higher returns and lower risk compared to capitalization-weighted portfolios (e.g. Haugen and Baker, 1991, Jagannathan and Ma, 2003, Clarke et al., 2006, Clarke et al., 2013).

However, some crucial challenges remain by this approach. In order to compose an efficient minimum-variance portfolio a precise estimation of the covariance matrix is essential. Surprisingly, in finance literature and practice the covariance matrix is often estimated on the basis of a constant historical (rolling) time-window of more or less arbitrary length (e.g. Haugen and Baker [1991]: 24 month; Jagannathan and Ma [2003]: 60 month and 1260 days; Pojarliev and Polasek [2003]: 800 days; Clarke et al. [2006]: 60 month and 250 days; DeMiguel et al. [2012]: 250 and 750 days; Behr et al. [2013]: 120 month), although several studies show that variances and correlations of asset returns are not constant over time (e.g. Longin and Solnik, 1995). To this end, this common approach may suffer from serious sampling errors.

Besides parameter estimation, the question arises when a rebalancing or a reoptimization should be performed. In finance literature and in practice it is common to choose a fixed reoptimization frequency (e.g. Baltutis and Dockner [2007]: weekly; Lenoir and Tuchschmid [2001] and Clarke et al. [2006]: monthly; Haugen and Baker [1991]: quarterly; Chan et al. [1999] and Jagannathan and Ma [2003]: annually; MSCI Minimum Volatility World Index: semi-annually). Usually, previous studies fail to motivate the determination of the frequency in detail despite the fact that portfolio rebalancing is crucial for portfolio performance. Behr and Miebs [2008] showed that minimum-variance portfolios are highly sensitive to revision frequencies. Baltutis and Dockner [2007] found out that under high frequency revision the turnover of the portfolio increased undesirably not necessarily reducing its realized volatility significantly.

By improving on the naive approach of periodic rebalancing, the financial literature provides numerous paper dealing with the issue of (optimal) portfolio revisions. These works proposed rebalancing strategies based on different approaches like e.g. tolerance bands around the desired target allocation (e.g. Masters, 2003 and Donohue and Yip, 2003), dynamic programming (Sun et al., 2006), and quadratic heuristics (Markowitz and van Dijk, 2003 and Kritzman et al., 2009)[1].

To the best of our knowledge, there are just a few paper using explicitly chan-

[1]See Sun et al. [2006] and Kritzman et al. [2009] for a discussion of these rebalancing strategies.

ges in the covariance matrix as a trigger to perform a reoptimization. Baltutis [2009], Golosnoy and Schmid [2007] and Golosnoy et al. [2011] use control charts for monitoring changes in the the covariance matrix and global minimum variance portfolio weights. In addition, Baltutis [2009] proposed a concept where an update of the portfolio weights is based on testing for statistically significant shifts in the covariance matrix which have already occurred in a realized sample.

In these contexts, we follow Baltutis [2009] by using a statistical test for structural breaks in the covariance matrix, but apply the recently proposed fluctuation test by Aue et al. [2009] for a constant covariance matrix to daily asset returns. Additionally, the break points detected by this test are used not only for automatically inducing dates for reoptimizations, but also for determining proper samples for parameter estimation. Wied et al. [2013b] introduce basic concepts of combining the minimum-variance approach with various fluctuation tests for volatility and dependence measures. Within the optimization context, they investigated a combination of the fluctuation tests for constant volatility and for constant correlations (Wied et al., 2012a; Wied et al., 2012b) as well as a fluctuation test for constancy of the entire covariance matrix (Aue et al., 2009). They find out that the usage of the test for constancy of the entire covariance matrix is the most promising approach.

However, despite the demonstrated potential of this approach they point out several serious drawbacks and challenges which have to be solved in further investigations in order to make this approach applicable for practitioners. In this paper, we take up these points and present useful methodological adjustments in order to develop algorithms and techniques for applications. Furthermore, we discuss the implementation of this new approach as an automated investment system for strategic asset allocations. Our empirical study shows that tests for structural breaks in the covariance matrix improve the results of a global minimum-variance optimization on average.

3.2 Portfolio Optimization

As the model by Markowitz [1952] is well known, we give only a very brief summary. It assumes the existence of d assets with normally distributed returns. Optimal selection of the portfolio weights $\omega = (\omega_1, \cdots, \omega_d)$ is intended, where ω_i is the fraction which is invested into asset i. For most applications it is required that $\omega_i \geq 0$, which avoids short selling, and $\sum_{i=1}^{d} \omega_i = 1$, which ensures an investor to be fully invested. The crucial parameter for a global minimum-variance optimization is the risk of the portfolio, which is defined by the variance σ_P^2. Hence, the portfolio weights are determined independently from expected returns and the optimization depends solely on the covariance matrix. The resulting portfolio lies on the left-most tip of the efficient frontier. These considerations result in the following optimization problem:

$$\begin{aligned} & min \quad \sigma_P^2 \\ \text{s.t.} \quad & \sum_{i=1}^{d} \omega_i = 1, \end{aligned} \tag{3.1}$$

where $\sigma_P^2 = \omega \Sigma \omega'$ and Σ is the covariance matrix. Moreover, sometimes the additional constraint $\omega_i \geq 0, \forall i$, is imposed.

As mentioned above, the global minimum-variance optimization depends solely on the covariance matrix. In this context, however, the question arises which time window should be used in order to estimate the covariance matrix. In the following section, we present a new approch to tackle this issue.

3.3 Tests for Breaks in the Covariance Structure

Aue et al. [2009] present a nonparametric fluctuation test for a constant d-dimensional covariance matrix of the random vectors $X_1, \ldots, X_T$ with $X_j = (X_{j,1}, \ldots, X_{j,d})$. The basic idea of the procedure is to compare the empirical covariance matrix calculated from the first observations with the one from all observations and to reject the null hypothesis if this difference becomes too large over time. Denote $\mathsf{vech}(\cdot)$ the operator which stacks the columns on and below the diagonal of a $d \times d$ matrix into a vector and A' the transpose of a matrix A. Then, we consider the term

$$S_k = \frac{k}{\sqrt{T}} \left(\frac{1}{k} \sum_{j=1}^{k} \mathsf{vech}(X_j X_j') - \frac{1}{T} \sum_{j=1}^{T} \mathsf{vech}(X_j X_j') \right) \tag{3.2}$$

which measures the fluctuations of the estimated covariance matrices calculated by means of the first k observations and use the maximum of the results for $k = 1, \cdots, T$. Here, the factor $\frac{k}{\sqrt{T}}$ serves for standardization; intuitively it corrects for the fact that the covariance matrices cannot be well estimated with a small sample size. If the maximum is standardized correctly, the resulting test statistic converges against a well know distribution and the null of a constant covariance matrix is rejected, if the test statistic is larger than the respective critical value.

For sake of readability we will not describe the entire test statistic at this point and refer to appendix C or Aue et al. [2009]. Nevertheless, the limit distribution under the null hypothesis is the distribution of

$$\sup_{0 \le t \le 1} \sum_{l=1}^{d(d+1)/2} B_l^2(t), \tag{3.3}$$

where $(B_l(t), t \in [0, 1]), l = 1, \ldots, d(d+1)/2$ are independent Brownian bridges.

The test basically works under mild conditions on the time series under consideration. One does not need to assume a particular distribution such as the normal distribution and the test allows for some serial dependence which makes it possible to consider e.g. GARCH models. Moreover, the test is consistent against fixed alternatives and has considerable power in finite samples. Regarding moments of the random variables, note that the correct application of the test needs constant expectations. The asymptotic result is derived under the assumption of zero expectation; if we had constant non-zero expectation, it would be necessary to subtract the arithmetic mean. While this assumption is sufficiently fulfilled for daily return series, the derivation of the asymptotic null distribution also needs the assumption of finite fourth moments. Theoretically, this assumption could be violated (Mandelbrot, 1962). However, in the following, we do not further consider this potential problem as this lies beyond our scope.

3.4 Empirical Study

The aim of this empirical study is to compare the out-of-sample performance of a global minimum-variance optimization combined with the test for a constant covariance matrix (hereinafter referred to as covariance-test optimization) to various relevant asset allocation strategies. First, we decide for a equally weighted asset allocation strategy as a natural benchmark.[2] For this, we obtain market values for each of the (sub)indices from *Thomson Reuters Datastream* and the portfolio weights are rebalanced each 21/63/252 traiding days, which corresponds approximately to monthly, quarterly and yearly rebalancings. The benchmark of most interest is the classical global minimum-variance portfolio where the optimization is based on constant rolling time-windows for calculation of the empirical covariance matrix (hereinafter referred to as plain optimization).

As this study is focused on strategic asset allocation, we use time series from indices or subindices rather than from single stocks. The pros and cons of active portfolio management are extensively discussed in numerous studies (e.g. Wermers, 2000, Jacobsen, 2011). However, we agree with Sharpe [1991] who pointed out that the return on the average actively managed dollar will equal the return on the average passively managed dollar. Including costs for the active management it will be even less. This statement is underpinned by Standard & Poor's [2012] who showed that 65% of all U.S. large cap equity funds do not outperform the S&P 500 index over the last five years. Moreover, indices are much more robust against unsystematic market risks and movements and can easily be replicated by means of ETFs. Note, as we deal with indices in a strategic asset allocation environment we can avoid questions arising from large investable sets (compare for example Michaud, 1989, Bai et al., 2009, Arnold et al., 2013).[3] Hence, we apply each of

[2] We also investigated cap-weighted portfolios. Nevertheless, the results of the equally wighted portfolios were slightly better. The results for cap-weighted portfolios are available from the authors upon request.

[3] Furthermore, high-dimensional portfolios can be reduced to managable sizes for example by factor analysis (Krzanowski, 2000, Hui, 2005).

these approaches to two samples consisting of five and ten indices, respectively. In detail, the empirical study is designed as follows:

3.4.1 Data

To carry out the out-of-sample study we compute log-returns from two different datasets. To avoid undesirable effects, both datasets have to fulfill the requirements of single currency and uniform time zone. For the first portfolio, we use daily total return quotes from five stock indices of main European countries that are founding members of the eurozone (AEX, CAC 40, DAX 30, FTSE MIB, IBEX 35). The quotes cover a period from the introduction of the Euro at January 1, 1999 to July 31, 2012 leading to 3481 trading days. For the second portfolio, we used daily total return quotes from the ten S&P 500 sector subindices (Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunication Services, Utilities). This quotes cover the total period provided by S&P starting at the initial publication on January 1, 1995 to July 31, 2012 leading to 4429 trading days. All quotes are obtained from *Thomson Reuters Datastream*.

3.4.2 Parameter Estimation

The optimization of a global minimum-variance portfolio based solely on the covariance matrix. Consequently, the performance differences between plain optimizations and covariance-test optimizations are due to the varying length of time-windows for parameter estimation. For the plain optimizations we define constant rolling time-windows of 250, 500 and 1000 trading days. The time-window of the covariance-test optimization is determined by following procedure:

1. Initialize $i = 1$ and $k = 1000$.
2. Apply the test of a constant covariance matrix to the data $\{x_i, \ldots, x_k\}$.
3. If the test rejects the null, set $p = k$, otherwise set $p = i$.

4. Adjust the time-window by $i = min\{p, k-126+1\}$ in case of the five-dimensional portfolio or $i = min\{p, k-252+1\}$ in case of the ten-dimensional portfolio.

5. Use the data $\{x_i, \ldots, x_k\}$ for estimating the empirical covariance matrix.

6. Set $k = k + n$, where n is the number of trading days between two tests and optimizations and go back to step 2.

Note, a reliable estimation of the covariance matrix requires a sufficient sample size. To this end, the modifications $i = min\{p, k-126+1\}$ and $i = min\{p, k-252+1\}$ ensure that the estimation is based on data of the last (half) year, depending on the dimensionality of the portfolio. As before, we choose $n = 21, 63$ and 252.

The determination of critical values is a crucial issue for the application of the test for a constant covariance matrix. Aue et al. [2009] approximated critical values by simulating Brownian bridges on a fine grid. Wied et al. [2013b] showed that this approximation does not perform well if the sample size is small. In this case, the critical values are overestimated and hence lead to low numbers of rejections. We take up this point and propose an alternative approach which is suitable for a practical application of the test. To this end, we generate d-dimensional standard normal distributed random variables. Then, we apply the test for a constant covariance matrix to the sample. This procedure is carried out 10000 times. After that, we determine the $(1-\alpha)$-quantile of the resulting test statistics as the critical value. In line with Wied et al. [2013b], we compute the critical values for $\alpha = 1\%$ and $\alpha = 5\%$. Depending on the chosen length of the sample, the critical value varies within a relatively wide range. Therefore, regarding the five-dimensional (ten-dimensional) portfolio, we estimate critical values for 18 (12) different sample sizes which are congruent to time-windows of 126 (250) to 1400 trading days (Table 3.1).

Using these critical values as grid points, we compute critical values for time-windows of any required length by linear interpolation. Although it seems only to be a small modification, it leads to a much more realistic determination of the dates where structural breaks in the covariance matrix occur. Moreover, it allows us to establish an automated investment strategy, which automatically determines dates for reoptimizations.

Table 3.1: Critical Values
Critical values for the five and the ten dimensional portfolio estimated by use of a Monte-Carlo-Simulation.

Sample Size	five-dimensional Portfolio		ten-dimensional Portfolio	
	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
126	4.25	4.63	-	-
138	4.39	4.80	-	-
150	4.54	4.96	-	-
175	4.74	5.19	-	-
200	4.92	5.45	-	-
225	5.11	5.65	-	-
250	5.24	5.84	8.60	8.94
275	5.37	6.01	8.97	9.35
300	5.48	6.10	9.36	9.77
350	5.69	6.41	10.01	10.48
400	5.89	6.68	10.60	11.18
500	6.11	6.99	11.49	12.12
600	6.31	7.25	12.28	13.05
700	6.47	7.41	12.88	13.83
800	6.57	7.52	13.41	14.35
1000	6.76	7.76	14.26	15.27
1200	6.86	7.90	14.95	16.07
1400	6.99	8.12	15.47	16.61

As we have just mentioned, the more precise estimation technique for critical values allows us to investigate an automated investment system, where the test is performed on a daily basis and the optimization is conducted only if the test rejects the null. Hence, an investor does not need to decide for a particular time-window in order to estimate the covariance matrix and reoptimization interval. Only the significance level has to be determined in advance. In more detail, we set $n = 1$ and modify the last step of the previous procedure as follows:

6. If the test rejects the null, set $k = k + 63$, otherwise set $k = k + 1$. Then go back to step 2.

By conducting the fluctuation test at each day, clustered rejections are very likely due to the small changes in the sample. The condition $k = k + 63$ in case of a null rejection assures that the sample for the subsequent test includes an adequate amount of new data.

3.4.3 Optimization Setup

The portfolio performance is strongly affected by the frequency of reoptimizations. In line with the test intervals of the previous section, we optimize every 21, 63, and 252 traiding days in the first setting. In this case, the asset weights are reoptimized after each test, regardless whether the null is rejected or not. Because of the identical intervals, this procedure allows for a direct comparison between the plain optimization and the covariance test optimization. In contrast to that, if the constancy of the covariance is tested on a daily basis, optimizations will be conducted only when a structural break is detected. In this context, portfolio weights remain unchanged in the sense that no trading takes place until the test again rejects the null. Hence, the portfolio weights will drift from the initially determined portfolio weights due to the variation in asset returns. Note, however, the simulations for the equally weighted portfolios suggest that the rebalancing frequency is only of minor importance. Besides, we consider two different constraints concerning the portfolio weights. First, we assume $0 \leq w_i \leq 1, \forall i$, which in particular excludes short selling (hereinafter referred to as long portfolios). In addition to that, we assume $|w_i| \leq 1, \forall i$, throughout the second run (hereinafter referred to as short portfolios). The optimizations are performed by using the $fmincon$-function of *MATLAB R2012a*.[4]

3.4.4 Performance Measurement

The portfolio performance is analyzed from various perspectives. First of all, the measurement of the risk in terms of volatility takes a prominent part of the evaluation, as portfolio variances are optimized. Nevertheless, we investigate the impact on the resulting returns and the relationship between risk and return in terms of the Sharpe-ratio, too. For its computation we assume 1.1% as risk free return which corresponds to the average return of German goverment bonds with less than 3 years to maturity in 2011.

Reoptimization (and rebalancing) of portfolio asset weights naturally leads to incre-

[4]Note, we checked the performance of the $fmincon$-function by means of several examples and comparison to the $quadprog$-function. All results indicate that there are no conversion problems within this optimization task. Nevertheless, to minimize the risk of detecting local minima, we use an adequate number of different starting points for the optimization. These starting points include the defined weighting boundaries as well as the equal weighted portfolio and random weights.

asing trading volume. Hence, we measure this turnover in absolute and relative Terms. Following DeMiguel et al. [2009], we define the sum of absolute changes in the weights as

$$Turnover(A) = \sum_{i=1}^{RD-1} \sum_{j=1}^{d} |a_{i+1,j} - a_{i^+,j}|, \tag{3.4}$$

where RD is the number of the reoptimization (rebalancing) days and d the number of assets. The portfolio weight of asset j before a rebalancing or reoptimization at time $i+1$ is defined as $a_{i^+,j}$. Besides, we call $Turnover(R)$ the average amount of changes at each RD, that means $Turnover(R) = \frac{1}{RD-1} \cdot Turnover(A)$.

In order to attribute a financial impact to the trading volume, we transform turnover to transaction costs and analyzes the effects. In line with Wied et al. [2013b] we compute adjusted returns and Sharpe-ratios by subtracting transaction costs from the return R. These costs are defined by $Turnover(A) \cdot \frac{s_c}{2}$ where the constant relative bid-ask spread s_c represents the bid-ask spread divided by bid-ask midpoint. We quantify the spread on the basis of the average relative bid-ask spread of the stocks listed on the European indices (5 asset portfolio) and stocks listed on the S&P 500 (10 asset portfolio) for the time-span August 1, 2011 to July 31, 2012. The spread of the analyzed stocks amounts to about 0.15% (European indices) and about 0.05% (S&P 500). Moreover, we refine this methodology used in Wied et al. [2013b] and introduce critical relative bid-ask spreads. To this end, consider two portfolio selection methods where a superior method outperforms an inferior method in terms of Sharpe-ratio (excluding transaction costs) and the absolute turnovers are different. Then, the critical relative bid-ask spread is defined as the spread at which for both portfolios the Sharpe-ratios adjusted by transaction costs are equal. In this context, we use the average Sharpe-ratio of the equally weighted portfolios as benchmark in order to calculate critical spreads for optimized portfolios.

3.5 Results

In the following, we present the results of the out-of-sample study.

3.5.1 European Stock Indices Portfolio

We start with the dataset including the five European stock indices. The results of the equally weighted portfolios are presented in Table 3.2.Volatilities, returns, and Sharpe-ratios remain in a narrow range and show only small variations due to the rebalancing interval. On average, an annualized return of 3.73% and an annualized volatility of 22.67% results to a Sharpe-ratio of 0.1161. The low turnover leads to neglectable transaction costs.

Table 3.2: Results for the Equally Weighted European Stock Indices Portfolio
Results for the equally weighted portfolio consisting of the five European stock indices. Interval refers to the frequency at which a rebalancing is conducted. Values in parentheses refer to Sharpe-ratios and returns adjusted by transaction costs.

Interval	Sharpe Ratio		Return p.a.		Volatility p.a.	Turnover (R)	Turnover (A)
21	0.1164	(0.1158)	3.74%	(3.73%)	22.70%	0.02	1.83
63	0.1162	(0.1159)	3.74%	(3.73%)	22.69%	0.03	1.06
252	0.1155	(0.1154)	3.71%	(3.71%)	22.61%	0.04	0.39
Average	0.1161	(0.1157)	3.73%	(3.72%)	22.67%	0.03	1.09

As expected, the volatility of the plain optimization portfolios (Tables 3.3 and 3.4, Panel A) is reduced significantly by averaged 1.08% for the long portfolios. Furthermore, the portfolio return is improved by 0.61% on average. Nevertheless, the reoptimizations generate a much higher trading volume and the related transaction costs decrease the returns by 0.02% to 0.15%. The allowance for short selling reduces volatilities even more. However, compared to the long portfolios, the returns and Sharpe-ratios tend to be lower and do not even achieve the level of the equally-weighted portfolios on average. Furthermore, the turnover increased by more than two times. Consequently, the average critical spread is negative. On average, the choice of the time-window length has a bigger impact to returns and Sharpe-ratios than the choice of the reoptimization interval. Conversely, the volatility is slightly more affected by the choice of the reoptimization interval.

From a theoretical point of view the allowance for short selling should lead to lower volatilities because it implies less stringent constraints for the optimization. As shown by Table 3.3 and 3.4 for example, applying the optimization to financial market data, a loosening of constraints could lead to a less efficient portfolio in some cases. This finding is

Table 3.3: Results for the Optimized European Stock Indices Portfolio and $0 < \omega_i < 1$
Results for the portfolio consisting of five European stock indices under the constraint $0 < \omega_i < 1$. For Panel A, # Data refers to the sample size used for the optimization. For Panel B and C, α refers to the significance level for the test for a constant covariance matrix. The interval refers to the frequency at which optimizations and tests are conducted. Values in parentheses refer to Sharpe-ratios and returns adjusted by transaction costs.

# Data / α	Interval	Sharpe Ratio		Return p.a.		Volatility p.a.	Turnover (R)	Turnover (A)	Critical Spread
Panel A: Plain Optimizations									
	21	0.1687	(0.1615)	4.66%	(4.51%)	21.11%	0.17	19.86	1.16%
250	63	0.1958	(0.1901)	5.27%	(5.15%)	21.30%	0.41	15.96	2.24%
	252	0.1437	(0.1404)	4.27%	(4.20%)	22.09%	1.05	9.41	1.44%
	21	0.1505	(0.1465)	4.29%	(4.20%)	21.18%	0.09	11.19	1.41%
500	63	0.1664	(0.1633)	4.65%	(4.58%)	21.34%	0.22	8.71	2.75%
	252	0.1663	(0.1643)	4.70%	(4.66%)	21.68%	0.61	5.48	4.82%
	21	0.1192	(0.1170)	3.69%	(3.64%)	21.71%	0.05	6.19	0.26%
1000	63	0.1168	(0.1151)	3.65%	(3.61%)	21.80%	0.12	4.73	0.09%
	252	0.1261	(0.1251)	3.88%	(3.86%)	22.07%	0.33	2.98	2.28%
	Average	0.1504	(0.1470)	4.34%	(4.27%)	21.59%	0.34	9.39	1.83%
Panel B: Optimization + Test for a Constant Covariance Matrix									
	21	0.2127	(0.2028)	5.52%	(5.32%)	20.79%	0.23	26.83	1.53%
5%	63	0.2447	(0.2378)	6.23%	(6.08%)	20.94%	0.49	19.10	2.93%
	252	0.1315	(0.1275)	4.01%	(3.92%)	22.13%	1.27	11.47	0.65%
	21	0.2167	(0.2074)	5.63%	(5.44%)	20.91%	0.21	25.34	1.70%
1%	63	0.2601	(0.2534)	6.59%	(6.45%)	21.12%	0.48	18.63	3.40%
	252	0.1555	(0.1522)	4.46%	(4.39%)	21.63%	1.03	9.31	2.03%
	Average	0.2035	(0.1969)	5.41%	(5.27%)	21.25%	0.62	18.45	2.04%
Panel C: Optimization + Daily Test for a Constant Covariance Matrix									
5%	1	0.1946	(0.1882)	5.21%	(5.07%)	21.10%	0.69	17.82	1.94%
1%	1	0.1301	(0.1261)	3.95%	(3.86%)	21.91%	0.66	11.30	0.59%
	Average	0.1623	(0.1572)	4.58%	(4.47%)	21.51%	0.68	14.56	1.27%

in line with the empirical study of Jagannathan and Ma [2003] who argue that constraints for portfolio weights increase specification error, but can also reduce sampling error. The trade-off between both error types determines the gain or loss in efficiency.

The results of the covariance-test optimizations are presented in Panel B of the Tables 3.3 and 3.4. Considering the long (short) portfolios, the returns increase by 1.07% (0.72%) while the volatility decrease by 0.34% (0.76%) on average compared to the plain optimi-

Table 3.4: Results for the Optimized European Stock Indices Portfolio and $|\omega_i| < 1$

Results for the portfolio consisting of five European stock indices under the constraint $|\omega_i| < 1$. For Panel A, # Data refers to the sample size used for the optimization. For Panel B and C, α refers to the significance level for the test for a constant covariance matrix. The interval refers to the frequency at which optimizations and tests are conducted. Values in parentheses refer to Sharpe-ratios and returns adjusted by transaction costs.

# Data / α	Interval	Sharpe Ratio		Return p.a.		Volatility p.a.	Turnover (R)	Turnover (A)	Critical Spread
Panel A: Plain Optimizations									
	21	0.0603	(0.0443)	2.33%	(2.00%)	20.37%	0.36	42.67	-0.54%
250	63	0.0766	(0.0647)	2.69%	(2.44%)	20.74%	0.83	32.38	-0.51%
	252	0.1468	(0.1399)	4.30%	(4.15%)	21.79%	2.17	19.54	0.71%
	21	0.1315	(0.1217)	3.85%	(3.65%)	20.92%	0.23	26.98	0.25%
500	63	0.1399	(0.1325)	4.07%	(3.91%)	21.24%	0.53	20.75	0.51%
	252	0.1839	(0.1792)	5.11%	(5.01%)	21.80%	1.49	13.40	2.36%
	21	0.0570	(0.0515)	2.33%	(2.21%)	21.51%	0.13	15.38	-1.74%
1000	63	0.0616	(0.0572)	2.44%	(2.35%)	21.81%	0.32	12.41	-2.06%
	252	0.0870	(0.0841)	3.05%	(2.98%)	22.42%	0.96	8.65	-1.69%
	Average	0.1050	(0.0972)	3.35%	(3.19%)	21.40%	0.78	21.35	-0.30%
Panel B: Optimization + Test for a Constant Covariance Matrix									
	21	0.1466	(0.1226)	4.03%	(3.55%)	20.00%	0.53	62.86	0.19%
5%	63	0.1337	(0.1167)	3.83%	(3.49%)	20.45%	1.17	45.64	0.16%
	252	0.1360	(0.1284)	4.07%	(3.91%)	21.87%	2.40	21.57	0.42%
	21	0.1634	(0.1405)	4.37%	(3.91%)	20.02%	0.51	59.90	0.32%
1%	63	0.1363	(0.1210)	3.88%	(3.56%)	20.36%	1.05	40.82	0.20%
	252	0.1497	(0.1436)	4.26%	(4.13%)	21.11%	1.88	16.94	0.88%
	Average	0.1443	(0.1288)	4.07%	(3.76%)	20.64%	1.26	41.29	0.36%
Panel C: Optimization + Daily Test for a Constant Covariance Matrix									
5%	1	0.0928	(0.0793)	3.01%	(2.73%)	20.55%	1.40	36.40	-0.27%
1%	1	-0.0192	-(0.0295)	0.67%	(0.45%)	22.16%	1.76	29.95	-2.04%
	Average	0.0368	(0.0249)	1.84%	(1.59%)	21.35%	1.58	33.17	-1.15%

zation portfolios. This leads to an improvement of the average Sharpe-ratio by 0.0531 (0.0393). For both, long and short portfolios, the application of the tests for structural breaks leads to almost a doubling of the average turnover. Nevertheless, the average critical spreads are higher compared to the plain optimization. The significance level of 1% leads to superior returns, whereas the impact of the significance level on the volatility is inconsistent.

Panel C of the Tables 3.3 and 3.4 present the results for the covariance-test optimizations where the test is performed on a daily basis. It is remarkable that the significance level of 5% leads to much better results compared to a level of 1%. Using 5%, long portfolios are comparable to the corresponding covariance-test optimizations. With respect to the short portfolio, this applies also for the volatility, whereas returns and Sharpe-ratios are worse.

3.5.2 S&P500 Subindices Portfolio

Below, we continue with the results for the portfolio consisting of ten Standard & Poor's 500 subindices. The results of the equally weighted portfolios are presented in Table 3.5. On average, a annualized return of 4.99% and an annualized volatility of 20.15% results to a Sharpe-ratio of 0.1933. As before, the low turnover leads to neglectable transaction costs.

Table 3.5: Results for the Equally Weighted Standard & Poor's 500 Subindices Portfolio
Results for the equally weighted portfolio consisting of the ten Standard & Poor's 500 subindices. Interval refers to the frequency at which a rebalancing is conducted. Values in parentheses refer to Sharpe-ratios and returns adjusted by transaction costs.

Interval	Sharpe Ratio		Return p.a.		Volatility p.a.	Turnover (R)	Turnover (A)
21	0.1916	(0.1912)	4.99%	(4.98%)	20.29%	0.03	4.75
63	0.1953	(0.1950)	5.04%	(5.03%)	20.16%	0.05	2.89
252	0.1929	(0.1928)	4.96%	(4.96%)	20.01%	0.11	1.37
Average	0.1933	(0.1930)	4.99%	(4.99%)	20.15%	0.06	3.00

As before, the application of the plain optimization improves the performance measures significantly (Tables 3.6 and 3.7, Panel A). Compared to the equally weighted portfolio, the volatility of the long-portfolio decreases by 4.83% whereas the return increases by 1.03% on average. Transaction costs vary between 0.007% and 0.035%. In contrast to the European indices portfolio, the allowance for short selling for the S&P500 portfolio leads to considerable improvements on the long portfolio with respect to volatility, return, and Sharpe-ratio. This goes along with a rise in averaged relative turnover from 0.21 to 0.56. The critical spreads reach considerably high values.

Table 3.6: Results for the Optimized Standard & Poor's 500 Subindices Portfolio and $0 < \omega_i < 1$

Results for the portfolio consisting of ten Standard & Poor's 500 subindices under the constraint $0 < \omega_i < 1$. For Panel A, # Data refers to the sample size used for the optimization. For Panel B and C, α refers to the significance level for the test for a constant covariance matrix. The interval refers to the frequency at which optimizations and tests are conducted. Values in parentheses refer to Sharpe-ratios and returns adjusted by transaction costs.

# Data / α	Interval	Sharpe Ratio		Return p.a.		Volatility p.a.	Turnover (R)	Turnover (A)	Critical Spread
Panel A: Plain Optimizations									
	21	0.3037	(0.3013)	5.63%	(5.60%)	14.93%	0.12	19.11	2.66%
250	63	0.3219	(0.3204)	5.93%	(5.91%)	15.00%	0.22	11.71	5.53%
	252	0.3694	(0.3686)	6.71%	(6.70%)	15.19%	0.55	7.15	14.87%
	21	0.3082	(0.3069)	5.75%	(5.73%)	15.09%	0.07	11.42	5.15%
500	63	0.3138	(0.3128)	5.87%	(5.86%)	15.20%	0.15	7.89	8.85%
	252	0.3459	(0.3452)	6.46%	(6.45%)	15.49%	0.40	5.22	22.09%
	21	0.2935	(0.2927)	5.65%	(5.64%)	15.51%	0.04	6.65	9.75%
1000	63	0.3050	(0.3044)	5.86%	(5.85%)	15.61%	0.09	4.84	18.81%
	252	0.3299	(0.3295)	6.34%	(6.33%)	15.88%	0.29	3.75	42.64%
	Average	0.3213	(0.3202)	6.02%	(6.01%)	15.32%	0.21	8.64	14.48%
Panel B: Optimization + Test for a Constant Covariance Matrix									
	21	0.3027	(0.3003)	5.62%	(5.58%)	14.93%	0.12	19.13	2.63%
5%	63	0.3349	(0.3336)	6.12%	(6.10%)	15.00%	0.21	11.13	6.50%
	252	0.3696	(0.3687)	6.71%	(6.70%)	15.19%	0.55	7.10	15.06%
	21	0.3088	(0.3066)	5.71%	(5.68%)	14.93%	0.11	17.89	2.99%
1%	63	0.3262	(0.3249)	5.99%	(5.97%)	14.99%	0.20	10.93	6.23%
	252	0.3655	(0.3647)	6.64%	(6.63%)	15.16%	0.51	6.69	16.01%
	Average	0.3346	(0.3331)	6.13%	(6.11%)	15.03%	0.28	12.14	8.24%
Panel C: Optimization + Daily Test for a Constant Covariance Matrix									
5%	1	0.3519	(0.3506)	6.33%	(6.32%)	14.88%	0.24	10.16	8.08%
1%	1	0.3667	(0.3657)	6.63%	(6.61%)	15.07%	0.30	8.75	10.93%
	Average	0.3593	(0.3581)	6.48%	(6.46%)	14.98%	0.27	9.45	9.51%

As presented in Tables 3.6 and 3.7 (Panel B), the application of the test for a constant covariance matrix yields to superior results on average. The long portfolio shows only slight improvements of the return whereas the return of the short portfolio increases by 0.52% on average. Moreover, the volatility decreases by 0.29% for the long and 0.25% for the short portfolio. Although the average trading volume rises by more than 40% com-

Table 3.7: Results for the Optimized Standard & Poor's 500 Subindices Portfolio and $|\omega_i| < 1$

Results for the portfolio consisting of ten Standard & Poor's 500 subindices under the constraint $|\omega_i| < 1$. For Panel A, # Data refers to the sample size used for the optimization. For Panel B and C, α refers to the significance level for the test for a constant covariance matrix. The interval refers to the frequency at which optimizations and tests are conducted. Values in parentheses refer to Sharpe-ratios and returns adjusted by transaction costs.

# Data / α	Interval	Sharpe Ratio		Return p.a.		Volatility p.a.	Turnover (R)	Turnover (A)	Critical Spread
Panel A: Plain Optimizations									
	21	0.4034	(0.3967)	6.83%	(6.73%)	14.20%	0.32	51.84	1.63%
250	63	0.4186	(0.4145)	7.15%	(7.09%)	14.45%	0.60	32.32	2.94%
	252	0.4960	(0.4935)	8.44%	(8.40%)	14.79%	1.53	19.95	6.86%
	21	0.3952	(0.3911)	6.75%	(6.70%)	14.31%	0.20	31.84	2.64%
500	63	0.3996	(0.3969)	6.92%	(6.88%)	14.56%	0.39	21.13	4.31%
	252	0.4569	(0.4552)	8.01%	(7.99%)	15.13%	1.06	13.75	9.44%
	21	0.2944	(0.2921)	5.44%	(5.41%)	14.74%	0.11	18.37	2.51%
1000	63	0.3228	(0.3213)	5.92%	(5.90%)	14.93%	0.22	11.85	5.46%
	252	0.3614	(0.3603)	6.67%	(6.66%)	15.42%	0.65	8.45	11.44%
	Average	0.3942	(0.3913)	6.90%	(6.86%)	14.73%	0.56	23.28	5.25%
Panel B: Optimization + Test for a Constant Covariance Matrix									
	21	0.4045	(0.3978)	6.84%	(6.75%)	14.20%	0.31	51.26	1.66%
5%	63	0.4169	(0.4130)	7.12%	(7.07%)	14.45%	0.57	30.68	3.08%
	252	0.4953	(0.4929)	8.43%	(8.40%)	14.80%	1.54	19.96	6.85%
	21	0.3968	(0.3906)	6.74%	(6.65%)	14.20%	0.30	48.26	1.70%
1%	63	0.3989	(0.3951)	6.87%	(6.81%)	14.46%	0.56	30.16	2.89%
	252	0.5013	(0.4989)	8.51%	(8.48%)	14.79%	1.47	19.06	7.35%
	Average	0.4356	(0.4314)	7.42%	(7.36%)	14.48%	0.79	33.23	3.92%
Panel C: Optimization + Daily Test for a Constant Covariance Matrix									
5%	1	0.4763	(0.4727)	7.83%	(7.78%)	14.14%	0.67	28.19	4.17%
1%	1	0.4580	(0.4547)	7.70%	(7.65%)	14.40%	0.88	25.44	4.45%
	Average	0.4672	(0.4637)	7.76%	(7.72%)	14.27%	0.77	26.82	4.31%

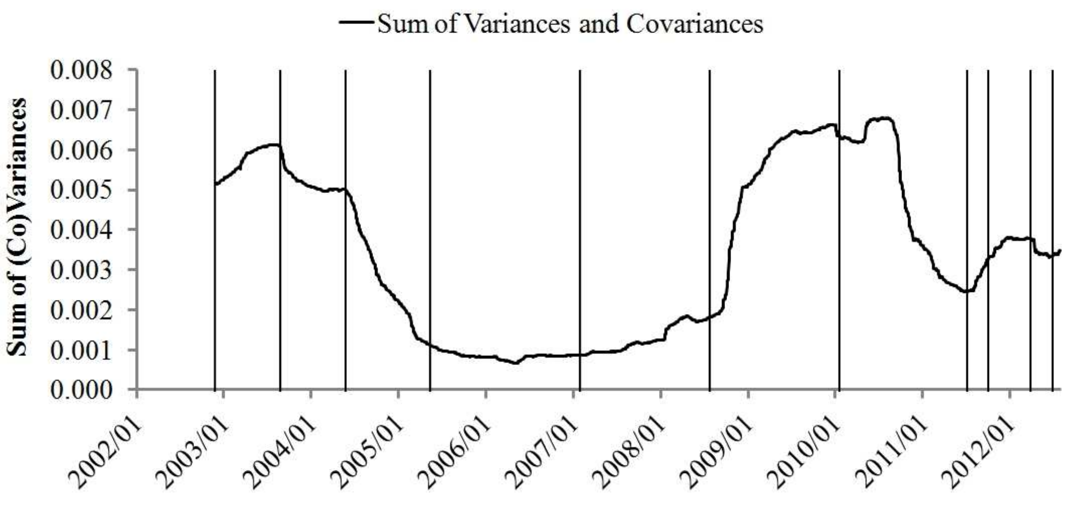
pared to the plain optimizations, the improvements of the results are not offset by a loss of return due to transaction costs. However, the critical spreads are somewhat lower compared to the plain optimizations. The choice of the significance level has no substantial impact to both return and volatility.

Panel C of the Tables 3.6 and 3.7 show the results for the covariance-test optimizations where the test is performed on a daily basis and the optimization is conducted only if the test rejects the null. On average, the results of this approach improve even on the covariance-test optimizations with a fixed test and reoptimization interval. Furthermore, the turnover is reduced considerably. In contrast to the first sample, the significance level has a minor impact on the results. Nevertheless, a level of 5% results in slightly superior results.

3.5.3 Rejection Dates

In this section we have a closer look at the rejection dates of the null. Considering the European indices dataset as an example, Figure 3.1 presents the dates at which the test for a constant covariance matrix rejects the null (63 days test interval / 1%-level) in connection with a trend of variances and covariances.

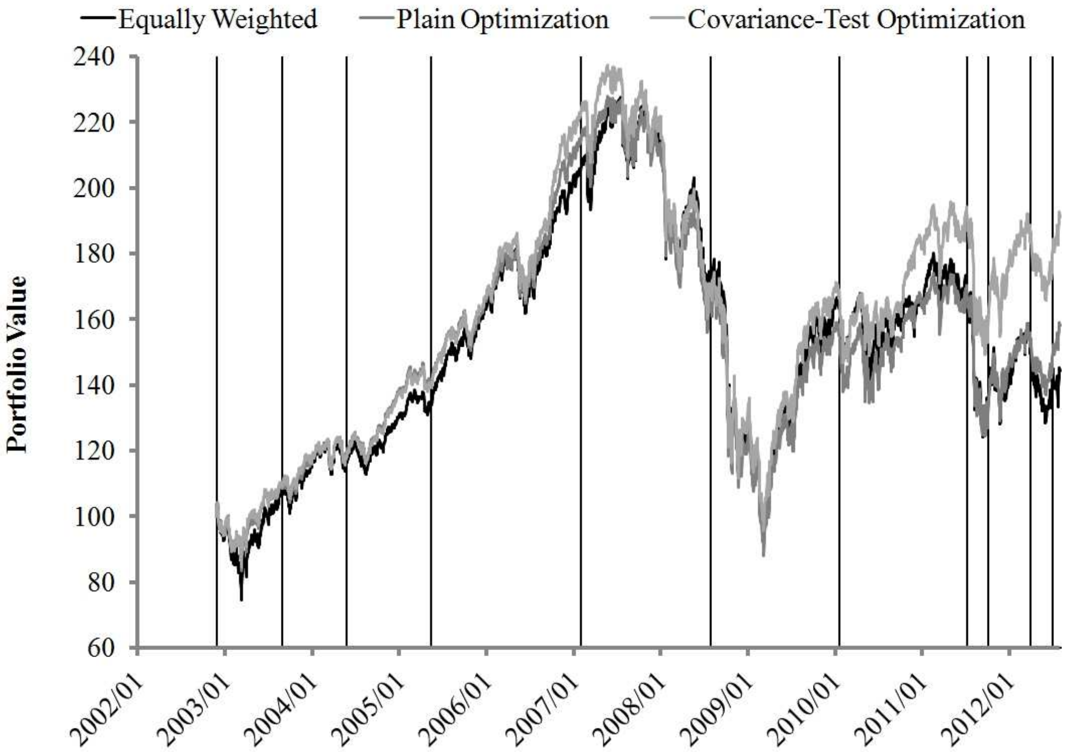
Figure 3.1: Trend of Variances and Covariances and Dates of Structural Breaks
The Figure shows the trend of the sum of variances and covariances for the European indices dataset over the time span November 26, 2002 to July 31, 2012 (2481 trading days). For each trading day, the sum results by adding up the entries on and below the diagonal of a covariance matrix. The matrix is computed on the basis of a rolling 500 trading day time-window. In addition, the points in time at which the test for a constant covariance matrix rejects the null (structural break) are marked by vertical bars. The tests are conducted under a setup of a 63 trading days test interval and a 1% significance level.



The chart illustrates that significant changes of variances and covariances are due to points in time at which the test rejects the null. Consequently, this procedure leads to considerably improved results with respect to volatility, return, and Sharpe-ratio compared to the optimizations with a fixed historical time-window. Figure 3.2 compares exemplary the performance of an equally weighted portfolio, a plain optimization portfolio, and a covariance-test optimization portfolio in connection with the dates at which the test for a constant covariance matrix rejects the null.

Figure 3.2: Portfolio Values

The Figure shows the portfolio values for the European indices dataset over the time span November 26, 2002 to July 31, 2012 (2481 trading days). The portfolio values are based on a rebalancing, reoptimization, and test interval of 63 trading days and a 500 trading day time-window with respect to the plain optimization. In addition, the points in time at which the test for a constant covariance matrix rejects the null are marked by vertical bars. The tests are conducted under a setup of a 63 trading days test interval and a 1% significance level.



The chart reveals that the covariance-test optimization outperforms the equally weighted portfolio and/or the plain optimization throughout most of the time. In particular during the late phase of the bull market 2006/2007 and the European sovereign-debt crisis beginning in the fall 2009, this new method outperforms the remaining portfolio selection approaches.

The results of the covariance-test optimization indicate that they are quite sensitive to the choice of the test and reoptimization interval, whereas the selected significance level plays only a minor role. This finding leads to a strategy, where we apply the test on a daily basis and conduct a reoptimization only if the test rejects the null. However, this strategy does not improve upon the covariance-test optimizations for fixed intervals in most settings. Moreover, the results are even worse for the European indices.

This behaviour is explained by the unreliable high number of detected structural breaks. For the S&P indices there are 29 (1%-level) and 42 (5%-level) rejections, respectively. The same holds true for the European indices where 17 (1%-level) and 26 (5%-level) rejections occured. This phenomenon can plausible be explained with the effect of sequential testing. Wied et al. [2013a] investigated this issue for a test of constant correlations. Hence, additional adjustments have to be carried out in order to make this strategy applicable for practice. However, this modifications are not in the scope of the present paper.

3.6 Conclusion

Our empirical study shows that minimum-variance optimization significantly improves return, volatility, and Sharpe-ratio compared to equally weighted portfolios. Although the optimizations lead to considerably increased trading volumes, the turnover in connection with relatively low bid-ask spreads for heavily traded blue chips causes modest transaction costs. Furthermore, the computation of critical relative bid-ask spreads suggests that an optimization is preferable even under much higher transaction costs. However, the study also reveals the sore point of the optimization setup: The results are very sensitive to the chosen historical time-window and to the reoptimization interval.

To overcome the issue of determining appropriate time-windows, we use the test of Aue et al. [2009] for a constant covariance matrix to detect structural breaks which set the starting point of a sample. We implement a consistent and essential advancement of the promising approach introduced by Wied et al. [2013b] and apply the optimizations in combination with the test in two different ways. First, we conduct the test and the optimization after a fixed interval where the rejection of the null sets a new beginning point for the time-window. Second, we apply the test on a daily basis and conduct a reoptimization only if the test rejects the null. That means, the procedure determines the length of the time-windows as well as the point in time where the portfolio is reoptimized.

Finally, we can conclude that minimum-variance optimizations in combination with the test for a constant covariance matrix provides a usable approach to replace an arbitrary sample selection for parameter estimation by a procedure which is statistically justified. Therefore, it can be used as an automated investment system for strategic asset allocations. Besides, there are some more remarkable benefits. First, the system is completely automated and no expensive funds managers and analysts are required. Hence, costs could be decreased significantly. Moreover, the out-of-sample study shows that there is a good chance to outperform an equally distributed portfolio over longer periods of time. Consequently, the approach seems to be an appropriate alternative for an usage in practice and in order to overcome the already mentioned weak points of actively managed portfolios. Nevertheless, the new approach is not suited so resolve the timing issue yet. To this end, some modifications considering sequential testing have to be performed. We will use the results achieved so far as a starting point and take up this topic in our future research.

Kapitel 4

A New Set of Improved Value-at-Risk Backtests

Revise and resubmit:

Journal of Banking and Finance (zusammen mit Daniel Ziggel, Gregor N.F. Weiß und Dominik Wied)

4.1 Introduction

Over the last two decades, Value-at-Risk (VaR) has become the de facto standard tool for measuring and managing risk in the financial services industry. Defined as the p-quantile of a relevant profit and loss (P/L) distribution where p is regularly set to 1% or 5%, it is now widely used by commercial banks and insurers as well as firms outside the financial industry to assess the risk exposure of single investments and portfolios.[1] A simple reason for this importance of VaR for the financial industry is given by the fact that under the 1996 Market Risk Amendment to the first Basel Accord, banks were allowed to employ internal VaR-models to calculate capital charges for their risky investments. Despite its popularity with practicioners, however, VaR has also received criticism from academia due to its lack of subadditivity [and thus coherence, see Artzner et al., 1999] in case of non-gaussian P/L distributions.[2] Even more importantly, commentators have blamed VaR in part for the severity of the recent financial crisis as the industry-wide use of VaR capital constraints enabled externalities to spread in financial markets through the pricing of risk

[1] Extensive discussions of the properties of VaR and its use in practice are given, e.g., by Dowd [1998], Jorion [2006], and Alexander [2008].

[2] Note, however, that evidence by Daníelsson et al. [2005] points out the subadditivity of VaR for most practical applications.

[see Shin, 2010].[3] Consequently, both regulators and financial risk managers have recently taken an increased interest in model validation and backtests of VaR-forecasts.

Despite its importance for bank regulation, VaR-backtesting has received relatively little attention in the financial econometrics literature compared to the numerous studies on the estimation and forecasting of VaR. One of the first formal statistical backtests for VaR was proposed by Kupiec [1995] who tests the sequence of VaR-violations for the correct number of violations (i.e., unconditional coverage). Christoffersen [1998] and Christoffersen and Pelletier [2004] extend these first tests of unconditional coverage by additionally testing for the independence of the sequence of VaR-violations yielding a combined test of conditional coverage. Recently, an integrated framework for VaR-backtesting that includes the previously mentioned tests was proposed by Berkowitz et al. [2011]. Further examples of the few backtests for VaR that are available to regulators are due to Berkowitz [2001], Engle and Manganelli [2004], Haas [2005] and Candelon et al. [2011], although the test of unconditional coverage continues to be the industry standard mostly due to the fact that it is implicitly incorporated in the framework for backtesting internal models proposed by the Basel Committee on Banking Supervision [1996].[4]

In this paper, we propose a new set of backtests for VaR-forecasts that significantly improve upon existing formal VaR-backtests like, e.g., the benchmark models proposed by Christoffersen and Pelletier [2004]. We first restate the definitions of the unconditional coverage property and propose a new test of the correct number of VaR-exceedances. Extending the current state-of-the-art, our new test can be used for both one-sided and two-sided testing and is thus able to test separately whether a VaR-model is too conservative or underestimates the actual risk exposure. Second, we stress the importance of testing both for the property of independent as well as the property of identically distributed VaR-exceedances and propose a simple approach for testing for both properties. While it has been noted in previous studies that VaR-violations should ideally be i.i.d.,

[3] Similar arguments in favor of a destabilizing effect of bank regulation based on VaR on the economy are stated by Leippold et al. [2006] and Basak and Shapiro [2001].

[4] A review of backtesting procedures that have been proposed in the literature is given by Campbell [2007].

standard backtests focus solely on the independence of the violations.[5] In this paper, we argue that the property of identically distributed VaR-exceedances is of vital importance to regulators and risk managers. In particular, we show that traditional VaR-backtests that center around first-order autocorrelation in violation processes are often not able to detect misspecified VaR-models during calm boom and highly volatile bust cycles. The new test of the i.i.d. property of VaR-violations explicitly tests for the presence of clusters in VaR-violation processes. This new feature is highly economically relevant as our test for violation clusters can identify VaR-models that yield inaccurate risk forecasts when they are most undesirable: during economic busts and financial crises when extreme losses on investments cluster due to a persistent increase in the volatility level. Finally, we also propose a weighted backtest of conditional coverage that simultaneously tests for a correct number and the i.i.d. property of VaR-violations. Our proposed weighted backtest is in the spirit of the original backtest of conditional coverage by Christoffersen and Pelletier [2004], but generalizes it by allowing the user to choose the weight with which the test of unconditional coverage enters the joint test of conditional coverage.[6] Our newly proposed set of backtests is simply based on i.i.d. Bernoulli random variables making them very intuitive and easy to implement. By construction, these tests automatically keep their level, even for very small sample sizes as they are often found in VaR-backtesting.

We employ our proposed backtests in a simulation study using several sets of simulated data that mimic real-life settings in which the simulated data violate the unconditional coverage, i.i.d., and conditional coverage properties to different degrees. The performance of the new tests is compared to classical tests frequently used in theory and practice as well as to a recently proposed powerful test. The results indicate that our tests significantly outperform the competing backtests in several distinct settings. In addition, we present an empirical application of the new tests using a unique data set consisting of the asset returns of an asset manager's portfolios.

[5] In fact, previous Markov- and duration-based tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011] only consider autocorrelation in VaR-violations as one possible reason why VaR-violations could be clustered.

[6] The approach of weighting the test statistics could also be pursued using classical uc and ind tests instead of our new uc and iid test. However, we believe this paper to be the first to explicitly point out the possibility to generate new tests by means of weighting uc and iid tests.

The paper is structured in a similar fashion as the one of Berkowitz et al. [2011] and is organized as follows. Section 4.2 introduces the notation, defines the properties of VaR-violations, and describes our new set of backtests. Section 4.3 evaluates the performance of the newly proposed backtests as well as several benchmark procedures for backtesting VaR-forecasts in a simulation study. Section 4.4 presents results from our empirical application study. Section 4.5 concludes the paper.

4.2 Methodology

In this section, we introduce the notation used throughout the paper, redefine the desirable properties of VaR-violations that are frequently discussed in the literature and present our new backtests.

4.2.1 Notation and VaR-Violation Properties

Let $\{y_t\}_{t=1}^{n}$ be a sample of a time series y_t corresponding to daily observations of the returns on an asset or a portfolio. We are interested in the accuracy of VaR-forecasts, i.e., an estimation of confidence intervals. Following Dumitrescu et al. [2012], the ex-ante VaR $VaR_{t|t-1}(p)$ (conditionally on an information set $\mathbb{F}_{t-1}$) is implicitly defined by $Pr(y_t < -VaR_{t|t-1}(p)) = p$, where p is the VaR coverage probability. Note that we follow the actuarial convention of a positive sign for a loss. In practice, the coverage probability p is typically chosen to be either 1% or 5% (see Christoffersen, 1998). This notation implies that information up to time $t-1$ is used to obtain a forecast for time t. Moreover, we define the ex-post indicator variable $I_t(p)$ for a given VaR-forecast $VaR_{t|t-1}(p)$ as

$$I_t(p) = \begin{cases} 0, \text{ if } y_t \geq -VaR_{t|t-1}(p); \\ 1, \text{ if } y_t < -VaR_{t|t-1}(p). \end{cases} \tag{4.1}$$

If this indicator variable is equal to 1, we will call it a VaR-violation.

To backtest a given sequence of VaR-violations, Christoffersen [1998] state three desirable properties that the VaR-violation process should possess. First, the VaR-violations are said to have unconditional coverage (uc thereafter) if the probability of a VaR-violation is equal to p, i.e.,

$$\mathbb{P}[I_t(p) = 1] = \mathbb{E}[I_t(p)] = p. \tag{4.2}$$

Second, the independence (ind thereafter) property requires that the variable $I_t(p)$ has to be independent of $I_{t-k}(p), \forall k \neq 0$. Finally, the uc and ind properties are combined via $\mathbb{E}[I_t(p) - p|\Omega_{t-1}] = 0$ to the property of conditional coverage (cc thereafter). In detail, a sequence of VaR-forecasts is defined to have correct cc if

$$\{I_t(p)\} \overset{i.i.d.}{\sim} \text{Bern}(p), \forall t. \tag{4.3}$$

While we agree with the formulation of the cc property, we point out that the uc and the ind properties as defined above suffer from some serious restrictions. The uc property requires a test whether the expected coverage is p for each day t individually. To be precise, the equation $\mathbb{P}[I_t(p) = 1] = \mathbb{E}[I_t(p)] = p$ holds only true if $\mathbb{P}[I_t(p) = 1] = p$ holds for all t. However, it is not feasible to verify if this assumption holds true for all t individually by means of a statistical test of uc. Moreover, it is quite likely that the sequence of VaR-violations is not stationary and that the actual p varies across different market phases even if $\frac{1}{n}\sum_{t=1}^{n} I_t$ equals p for the total sequence. Evidence for this conjecture is found by Escanciano and Pei [2012]. The practical relevance of this feature is demonstrated in our empirical study (see Section 4.4). Consequently, we redefine the uc property simply as

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] = p. \tag{4.4}$$

With respect to the ind property, it is interesting to note that the current state-of-the-art backtests in the financial econometrics literature do not focus on testing the property of VaR-violations being identically distributed. In fact, the sequence $\{I_t(p)\}$ could exhibit clusters of violations while still possessing the property of independence as defined above. In fact, unexpected temporal occurrences of clustered VaR-violations may have several potential reasons. On the one hand, $\{I_t(p)\}$ may be not identically distributed and p could vary over time. On the other hand, $I_t(p)$ may not be independent of $I_{t-k}(p), \forall k \neq 0$. We therefore reformulate the ind property as the i.i.d. property (i.i.d. thereafter). The hypothesis of i.i.d. VaR-violations holds true if

$$\{I_t(p)\} \overset{i.i.d.}{\sim} \text{Bern}(\tilde{p}), \forall t, \tag{4.5}$$

where $\tilde{p}$ is an arbitrary probability. Note that the i.i.d. hypothesis does not deal with the relative amount of VaR-violations. Hence, if appropriate, $\tilde{p}$ will be replaced by its empirical counterpart $\bar{p}$ (the estimated violation rate) within the respective test statistic, while it is specified to its desired value p (which is tested later on) within the cc property.

In the following, we describe our new set of backtests that includes separate tests for all mentioned properties of VaR-violation processes. Pseudocode for all new tests is provided in Chapter A.

4.2.2 A New Test of Unconditional Coverage

At this point, we are interested in testing the null hypothesis $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] = p$ against the alternative $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] \neq p$. In fact, as we will see later, our new test statistic also allows us to separately test against the alternatives $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] > p$ and $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] < p$. The most intuitive and commonly used test statistic for the test of uc is given by (see Christoffersen, 1998):

$$LR_{uc}^{kup} = -2 \log[L(p; I_1, I_2, ..., I_n)/L(\bar{p}; I_1, I_2, ..., I_n)] \overset{asy}{\sim} \chi^2(1), \tag{4.6}$$

where $\bar{p} = \frac{n_1}{n_1+n_0}$, n_1 is the number of violations and $n_0 = n - n_1$. Moreover, we have

$$L(p; I_1, I_2, ..., I_n) = p^{n_1}(1-p)^{n_0} \tag{4.7}$$

and

$$L(\bar{p}; I_1, I_2, ..., I_n) = \bar{p}^{n_1}(1-\bar{p})^{n_0}. \tag{4.8}$$

Candelon et al. [2011] recently introduced an alternative test for the uc hypothesis using orthonormal polynomials and the GMM test framework proposed by Bontemps [2006], Bontemps and Meddahi [2005] and Bontemps and Meddahi [2012]. Their test statistic is given by

$$J_{uc} = J_{cc}(1) = \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M_1(d_i; p) \right)^2 \overset{asy}{\sim} \chi^2(1), \tag{4.9}$$

where M_1 is an orthonormal polynomial associated with a geometric distribution with a success probability p and d_i denotes the duration between two consecutive violations [see Candelon et al., 2011, for more details].

However, both tests suffer from significant drawbacks. First, without modifications, it is not possible to construct one-sided confidence intervals. Such an additional feature, on the other hand, would be of particular interest to bank regulators and risk-averse investors who are primarily interested in limiting downside risk. While it is trivial to check whether a rejection was due to a model being too conservative or not conservative enough, none of the existing tests yields one-sided critical values. In this context, results from our simulation study illustrate that the power of one-sided tests is significantly higher. The second drawback is concerned with the behaviour of the tests in finite samples. As we deal with tail forecasts based on binary sequences, the number of violations is comparatively small and discrete. Hence, ties between the sample test value and those obtained from Monte Carlo simulation under the null hypothesis need to be broken. That means that we have to ensure that the probability for two equal values of the test statistic for two different data sets is zero. Christoffersen and Pelletier [2004] propose to use the Dufour [2006] Monte Carlo testing technique to break ties between test values. As their approach, however, is computationally demanding and unnecessarily complex, we propose a different tie breaking procedure.

We address the latter problem by exploiting an idea used, among others, by

Podolskij and Ziggel [2009] and propose to use the test statistic

$$MCS_{uc} = \sum_{t=1}^{n} I_t(p) + \epsilon, \tag{4.10}$$

where ϵ is a continuously distributed random variable with small variance that serves to break ties between test values.[7] Critical values of the test statistic are computed via Monte Carlo simulations (MCS) as is done for all other backtests throughout this paper. For fixed n and p, the distribution of the test statistic is known. We then simulate a large number of realizations of the test statistic under the respective null hypothesis and use the resulting quantile for testing the uc hypothesis. Adding the random variable ϵ guarantees that the test exactly keeps its size if the number of Monte Carlo simulations for obtaining the critical value tends to infinity.[8] Note that without the addition of the random variable ϵ, the test statistic would have a discrete distribution and not all possible levels could be attained. Additionally, note that the choice of ϵ is not crucial for testing the uc hypothesis. We noticed in robustness checks that the finite sample performances of the tests are not substantially affected by changes in the distribution of ϵ as long as it remains continuous with a small, non-zero variance. Consequently, it is intuitive to use normally distributed random variables for ϵ. Nevertheless, one needs to assure that the test statistic for $v - 1$ violations is smaller then the test statistic for v violations. Followingly, we set $\epsilon \sim 0.001 \cdot N(0, 1)$ in our simulation study. Finally, it is instructive to see that our new approach allows for one-sided and two-sided testing for every desired test level.

Critical values for all our tests are then computed via MCS instead of, e.g., making use of explicit expressions of the exact or asymptotic distributions. Basically, all test statistics we consider are given as the sum of a discrete random variable (determined by Bernoulli distributed random variables) and a continuous random variable with known distribution that is independent from the discrete random variable. Thus, on the one hand, the distributions of the test statistics are uniquely determined for fixed n and p and additionally it

[7] Podolskij and Ziggel [2009] employ the idea of adding a small random variable to a test statistic to construct a new class of tests for jumps in semimartigale models.

[8] The theoretical foundation of our approach is given by Dufour [2006] who considers a more general context and solves this problem by introducing randomized ranks according to a uniform distribution.

is basically useful to consider MCS. On the other hand, due to the continuous part, the test statistics are also continuously distributed. This follows from the general fact that, for a discrete random variable X with support M_X and a continuous random variable Y such that X and Y are independent,

$$P(X+Y \leq a) = \sum_{x \in M_X} P(x+Y \leq a | X = x) P(X = x) = \sum_{x \in M_X} P(Y \leq a-x) P(X = x).$$

Thus, the cumulative distribution function of $X + Y$ can be written as a countable sum of continuous functions so that it is continuous as well. Using a result from Dufour [2006], the empirical critical values then yield a test that exactly keeps its size if the number of MCS tends to infinity.

Instead of using MCS, one could basically also derive the exact distribution functions of the test statistics, although this would indubitably be a cumbersome task. It would also be possible to derive asymptotic results if the test statistics are appropriately standardized and if one imposes additional moment assumptions on the continuous random variable. For example, a suitably standardized uc test statistic might be $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (I_t(p) - p) + \frac{1}{\sqrt{n}} \epsilon$. However, we believe that, although of some interest, such an asymptotic analysis is not necessary in our setting. In practice, n and p are fixed and by an increasing number of Monte Carlo repetitions we can get arbitrarily exact critical values of the test statistics in reasonable time. Since one typically deals with a low number of VaR violations, one could moreover expect the asymptotic approximation to be highly inaccurate, which is confirmed by several studies [see, e.g., Berkowitz et al., 2011].

Basically, the one-sided version of our new uc test can be regarded as a generalization of the Basel traffic light approach as described in Campbell [2007]. The Basel approach provides a method which can be easily applied. Here, the 1% VaR violations in the last 250 days are counted. The traffic light is green whenever the number of violations is less than 5, yellow whenever the number lies between 5 and 9 and red otherwise. With the decision rule “Reject the null hypothesis of a valid VaR model whenever the traffic light is red” the procedure can be interpreted as a significance test. In fact, then the Basel test statistic is a special case (with $n = 250$, $p = 0.01$, $\alpha < 0.001$ and $\epsilon = 0$) of our uc test statistic. Information concerning the size and power of the Basel test can be found in Basel Committee on Banking Supervision [1996]. However, an apllication of this test is not possible as soon as the input parameters change. In contrast to that, our new approach allows, e.g., to increase the sample size or to vary the significance level.

4.2.3 A New Test of I.I.D. VaR-Violations

As stated in Christoffersen [1998], testing solely for correct uc of a VaR-model neglects the possibility that violations might cluster over time. Consequently, Christoffersen [1998] propose a test of the violations being independent against an explicit first-order Markov alternative. The resulting test statisic is given by:

$$LR_{iid}^{mar} = -2\,\log[L(\tilde{\Pi}_2; I_1, I_2, ..., I_n)/L(\tilde{\Pi}_1; I_1, I_2, ..., I_n)] \overset{asy}{\sim} \chi^2(1). \tag{4.11}$$

Here, the likelihood functions are given by:

$$\begin{aligned} L(\tilde{\Pi}_1; I_1, I_2, ..., I_n) = & \left(1 - \frac{n_{01}}{n_{00}+n_{01}}\right)^{n_{00}} \left(\frac{n_{01}}{n_{00}+n_{01}}\right)^{n_{01}} \\ & \left(1 - \frac{n_{11}}{n_{10}+n_{11}}\right)^{n_{10}} \left(\frac{n_{11}}{n_{10}+n_{11}}\right)^{n_{11}} \end{aligned} \tag{4.12}$$

and

$$\begin{aligned} L(\tilde{\Pi}_2; I_1, I_2, ..., I_n) = & \left(1 - \frac{n_{01}+n_{11}}{n_{00}+n_{10}+n_{01}+n_{11}}\right)^{n_{00}+n_{10}} \\ & \left(\frac{n_{01}+n_{11}}{n_{00}+n_{10}+n_{01}+n_{11}}\right)^{n_{01}+n_{11}}, \end{aligned} \tag{4.13}$$

where n_{ij} is the number of observations with value i followed by j. Note that this first-order Markov alternative has only limited power against general forms of clustering. Moreover, as shown in Christoffersen and Pelletier [2004], this test is not suited for several settings and has a poor behaviour in finite samples. The test can then be combined with the test of uc presented in the previous subsection to yield a full test of cc. Despite the aforementioned shortcomings, however, it is still one of the most frequently used backtests in practice [see Candelon et al., 2011].

In a subsequent work, Christoffersen and Pelletier [2004] introduce more flexible tests which are based on durations between the violations. The intuition behind these tests is that the clustering of violations will induce an excessive number of relatively short and long no-hit durations. Under the null hypothesis, the no-hit durations D should then be exponentially distributed with

$$f_{exp}(D; p) = pe^{-pD}, \tag{4.14}$$

where D is the no-hit duration. In their work, Christoffersen and Pelletier [2004] employ the Weibull and the gamma distribution to test for an exponential distribution of the no-hit durations. Nevertheless, we will only consider the Weibull test in our simulation study as it yields considerably better results than the gamma test [see Haas, 2005]. In addition to the mentioned tests, the literature on VaR-backtesting also includes the standard Ljung-Box test, the CAViaR test of Engle and Manganelli [2004], the regression based dynamic quantile test by Hurlin and Tokpavi [2006] and spectral density tests. However, the level of most of these tests is poor for finite samples and therefore critical values need to be calculated based on the Dufour Monte Carlo testing technique (see Berkowitz et al., 2011).

Recently, Candelon et al. [2011] introduced a new test for the i.i.d. hypothesis. As described above, this test is based on orthonormal polynomials and the GMM test framework. The test statistic is given by

$$J_{iid}(q) = \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M(d_i; \bar{p}) \right)^T \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M(d_i; \bar{p}) \right) \overset{asy}{\sim} \chi^2(q), \tag{4.15}$$

where $M(d_i; \bar{p})$ denotes a $(q, 1)$ vector whose components are the orthonormal polynomials $M_j(d_i; \bar{p})$, for $j = 1, ..., q$, evaluated for the true violation rate $\bar{p}$.

To introduce our new test statistic, we first define the set of points in time on which a

VaR-violation occurs via

$$V = \{t | I_t = 1\} = (t_1, ..., t_m). \tag{4.16}$$

The test statistic for our new i.i.d. hypothesis is then given by

$$MCS_{iid,m} = t_1^2 + (n - t_m)^2 + \sum_{i=2}^{m} (t_i - t_{i-1})^2 + \epsilon. \tag{4.17}$$

This sum essentially consists of the squared durations between two violations. Basically, the idea behind this test statistic follows the principle of the Run-Test proposed by Wald and Wolfowitz [1940]. To be precise, the sum of the squared durations between two violations is minimal if the violations are exactly equally spread across the whole sample period. If the violations are clustered and occur heaped, this sum increases. Just like in the Run-Test, both systematic and heaped occurences of violations could be undesirable in a risk management setting. For example, the process of VaR-violations could exhibit an undesirable cyclical or seasonal behaviour that is detected by our new test of the i.i.d. property as the test statistic tends to its minimum.[9] At the same time, too large values of $MCS_{iid,m}$ could indicate a clustering of violations indicating a significantly bad fit of the VaR-model in a particular time period. For the purposes of this study we concentrate on testing for clustered VaR-violations noting that two-tailed testing for both clusters and cyclical patterns in VaR-violations is straightforward.

Empirically, clustered VaR violations most often occur in a time of financial crisis with high volatility which follows an economically quiet time and vice versa. In the former case, an initially suitable VaR model becomes inadequate in times of market turmoil and increasing volatility. Assuming this, one could use our new i.i.d. test for detecting times of crises or volatility clusters. Note that such a test will work as long as the VaR model is not completely correctly specified. On the other hand, it is also possible that the VaR model is suitable for both quiet and volatile times leading to a failure of the test. Due to this fact, it would be interesting to investigate such a kind of test in more detail and useful

[9]This feature is of particular interest, e.g., in commodity and weather risk management.

to compare or combine an analysis based on the new i.i.d. test with e.g. a test for constant variances as presented in Wied et al. [2012a]. However, this issue is not in the scope of the present paper.

As before, we waive a formal derivation of the distribution of our test statistic. Instead, we obtain the critical values of the test statistic by means of a Monte Carlo simulation (thus inspiring the abbreviation $MCS_{iid,m}$). The simulation is straightforward as only n and p have to be adapted to the specific situation. Note that the critical values need to be simulated separately for each value of m as we are solely interested in the durations between the violations and not in the absolute number of it. We use the same continuously distributed random variable ϵ as before to break ties. Again, the choice of ϵ ensures the MCS to yield a valid test. Moreover, the computational complexity of the test is negligible.

4.2.4 A New Test of Conditional Coverage

We now describe our new test of cc that combines the two new tests for the uc and the i.i.d. property. Starting point is again the standard test of cc as proposed by Christoffersen [1998] which utilizes the test statistic

$$LR_{cc}^{mar} = -2\log[L(p; I_1, I_2, ..., I_n)/L(\tilde{\Pi}_1; I_1, I_2, ..., I_n)] \overset{asy}{\sim} \chi^2(2), \tag{4.18}$$

and which is based on the first-order Markov alternative described above. In a related study, Berkowitz et al. [2011] extend their Weibull test for the i.i.d. property and derive an alternative test of cc. They postulate a Weibull distribution for the duration variable D with distribution

$$h(D; a, b) = a^b b D^{b-1} e^{-(aD)^b}, \tag{4.19}$$

with $\mathbb{E}[D] = 1/p$. Then, the null hypothesis of their test of cc is given by

$$H_{0,cc} : b = 1, a = p. \tag{4.20}$$

Using orthonormal polynomials and the GMM test framework, Candelon et al. [2011]

propose a competing test of the cc hypothesis. Their test statistic is given by

$$J_{cc}(q) = \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M(d_i; p) \right)^T \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M(d_i; p) \right) \overset{asy}{\sim} \chi^2(q). \tag{4.21}$$

Again, $M(d_i; p)$ denotes a $(q, 1)$ vector whose entries are the orthonormal polynomials $M_j(d_i; p)$, for $j = 1, ..., q$.

To the best of our knowledge, the literature provides no modification of the mentioned tests in a way that they allow for a weighted influence of the uc and i.i.d. components in the combined test of cc. From the perspective of a risk manager, however, such a feature could be highly desirable as more weight could be assigned to one of the components of the test of cc. Hence, we are interested in a test of the form

$$MCS_{cc,m} = a \cdot f(MCS_{uc}) + (1-a) \cdot g(MCS_{iid,m}), 0 \leq a \leq 1, \tag{4.22}$$

where a is the weight of the test of uc in the combined cc test. The first component of our new cc test is then given by

$$f(MCS_{uc}) = \left| \frac{(MCS_{uc})/n - p}{p} \right| = \left| \frac{(\epsilon + \sum_{t=1}^{n} I_t)/n - p}{p} \right|. \tag{4.23}$$

This term measures (in percent) the deviation between the expected and observed proportion of violations. As the general sizes of MCS_{uc} and $MCS_{iid,m}$ are not the same the quantities would not be suitably comparable without a standardization. Moreover, the difference in size varies depending on the setting (i.e. n and p). As the quantities will appear in one sum, it is necessary to be able to compare them suitably.

To allow for a one-sided testing within the uc component, the above term is multiplied by $1_{\{\sum_{t=1}^{n} I_t/n \geq p\}}$ or $1_{\{\sum_{t=1}^{n} I_t/n \leq p\}}$, respectively.[10] The intuition behind this is that the weight of the uc part should be zero if the observed quantity is "on the opposite side" of the null hypothesis such that it is very unlikely that the alternative is true.

[10] A one-sided test seems to be useful as it can be considered as a generalization of the Basel traffic light approach and is of particular interest to risk-averse investors who are primarily interested in limiting downside risk.

The second component in the cc test in (4.22) is defined as

$$g(MCS_{iid,m}) = \frac{MCS_{iid,m} - \hat{r}}{\hat{r}} \cdot 1_{\{MCS_{iid,m} \geq \hat{r}\}}, \tag{4.24}$$

where $\hat{r}$ is an estimator of the expected value of the test statistic $MCS_{iid,m}$ under the null hypothesis (4.5), i.e., for $E(MCS_{iid,m}|H_0) =: r$ (see below and chapter A). The second component measures the deviation (in percent) between the expected and observed sum of squared durations. Again, we use random variables ϵ to break ties. In line with the new uc and i.i.d. tests, we abstain from a formal derivation of the distribution of our test statistic and obtain the critical values by means of a Monte Carlo simulation for each combination of sample size n and weighting factor a.

Note that the estimator $\hat{r}$ is calculated in a prior step before calculating the actual test statistics and deriving critical values (cf. the pseudocode in). Thus, for $MCS_{cc,m}$, the arguments regarding the correctness of the MCS from the end of Section 2.2 are also applicable.

As the weighting factor a can be chosen arbitrarily, a natural question to ask is how a should be chosen. On the one hand, small test samples (e.g., 250 days) and small values of p (e.g. $p = 1\%$) lead to a small expected number of VaR-violations. In these cases, a risk manager (or regulator) might be more interested in backtesting the VaR-violation frequency rather than the i.i.d. property of, for instance, only two or three violations. On the other hand, large test samples (e.g., 1,000 days) may include calm bull and volatile bear markets. A VaR-model which is not flexible enough to adapt to these changes may lead to non-identically distributed VaR-violations while at the same time yielding a correct uc. Therefore, risk managers could be inclined to select a lower level of a to shift the sensitivity of the cc test to the test of the i.i.d. property. Note, as both components of the test are strictly positive it is ruled out that one criteria could compensate the failing of the other. Therefore, the choice of a affects solely the sensitivity of the cc test to one of the components. Nevertheless, the selection of the optimal weighting factor a is an interesting task. Regarded as a mathematical optimization problem, one could basically find the optimal a which minimizes a suitably weighted sum of the type-1 and type-2 error for a given alternative. However, this mainly technical issue is not in the scope of the present paper.

4.3 Simulation Study

To examine the performance of our newly proposed backtests in finite samples, we perform a comprehensive simulation study in which we compare our new backtests to several different benchmarks. These include the classical tests proposed by Christoffersen [1998] and Christoffersen and Pelletier [2004] because these approaches are still very frequently used in theory (e.g. by Weiß and Supper, 2013) and in practice (see Basel Committee on Banking Supervision, 2011). In addition, we employ the tests recently proposed by Candelon et al. [2011] as a benchmark showing robust properties and a high power. The relevance of the benchmark tests is emphasized by the fact that in recent studies these procedures are applied in parallel (see, e.g., Asai et al., 2012 and Brechmann and Claudia, 2013).

Before starting with the uc tests, we want to point out that the time required to compute the critical values is quite short for all applied tests. The average calculation times for $p = 0.05$ and different values of n are presented in Table 4.1.

With the exception of the Weibull tests, all average calculation times lie within a corridor of 0.07 to 4.4 seconds. The longer calculation time of the Weibull tests, which lies between 25.79 to 27.95 seconds, is due to the required maximum likelihood estimates of the parameters of the Weibull distribution. However, none of the calculation times are critical for applications.

Table 4.1: Comparison of the Backtests' Calculation Times

The table presents average calculation times (in seconds) for the different backtests used in the paper for $p = 0.05$, 10,000 simulations and different values of n based on 10 repetitions. All calculations are performed with Matlab2012a on a standard notebook. Note, the results of MCS_{iid} are taken over to MCS_{cc}. Hence, the upper bound for a direct calculation of MCS_{cc} is the sum of both single times.

	UC Tests			
n	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}	
252	0.08	1.48	0.07	
500	0.12	1.60	0.11	
1,000	0.20	1.84	0.20	
1,500	0.29	2.06	0.28	
2,500	0.45	2.57	0.45	
	I.I.D. Tests			
n	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
252	0.61	25.79	3.70	1.54
500	0.71	26.31	3.75	1.64
1,000	0.92	26.48	3.89	1.85
1,500	1.10	27.09	4.06	2.06
2,500	1.52	27.93	4.40	2.28
	CC-Tests			
n	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}
252	0.68	26.26	1.99	1.58
500	0.78	26.48	2.18	1.63
1,000	1.01	27.13	2.31	1.84
1,500	1.23	27.43	2.43	1.95
2,500	1.66	27.95	2.65	2.29

4.3.1 Tests of Unconditional Coverage

We analyze the performance of the different tests of uc by simulating $10,000$ samples[11] and using different parameter combinations for p, γ, and n to analyze the size and power of the backtests in more detail. In constrast to obtaining violations from a parametric VaR model, we simulate sequences of VaR-violations using the data generating process (DGP)

$$I_t \sim \text{Bern}(\gamma \cdot p),\ t = 1, ..., n. \tag{4.25}$$

Here, γ is a coverage parameter which allows for distinguishing between null hypothesis and alternatives. To determine the size of the tests, we set the coverage parameter $\gamma = 1.0$. For the analysis of the tests' power, we increase the violation probability and

[11] With this number of repetitions, the standard error of the simulated rejection probabilites is equal to $\frac{1}{100}\sqrt{p(1-p)}$, where p is the true rejection probability. That means, the standard error is of order $\frac{1}{100}$. A similar result holds for the accuracy of the simulated critical values, see below.

set $\gamma = 1.1,\ 1.25,\ 1.50$ and 2.00.[12] Each sequence I_t of simulated VaR-violations is then backtested using the new upper-tail MCS_{uc}^{ut} and the two-tailed MCS_{uc}^{tt} backtest as described in Section 4.2.2. To evaluate each test's power, we compute the fraction of simulations in which the test is rejected (hereafter referred to as rejection rate). Critical values of the test statistics for different parameters p and n are computed using $10,000$ MC simulations. Complementing our new backtests, we also apply the LR_{uc}^{kup} test of Christoffersen [1998] and the GMM_{uc} test of Candelon et al. [2011] to the simulated violation sequences and compare the results of the tests. The results of the simulation study on the performance of the tests of uc are presented in Table 4.2 and Table 4.3. Not surprisingly, due to the fact that the critical values for each of the tests are determined via simulation, the rejection frequencies for the setting $\gamma = 1.0$ are close to the nominal size of the tests. With respect to the power of the uc tests, the results of the LR_{uc}^{kup} test, the GMM_{uc} test, and the two-tailed MCS_{uc}^{tt} test are very similar. Only in a few cases do the results of the GMM_{uc} test deviate from the rejection rates of the LR_{uc}^{kup} test and the two-tailed MCS_{uc}^{tt} test in a positive or negative direction. However, all of the three analyzed two-tailed tests are outperformed by the one-sided MCS_{uc}^{ut} test in the vast majority of settings. Consequently, in addition to being of high practical relevance to regulators, our new one-tailed test of uc offers an increased test power compared to standard VaR-backtests from the literature.

[12]We calculate but do not report results for the setting $\gamma < 1$ and concentrate on the more practically relevant scenario of a VaR-model underestimating risk.

Table 4.2: Unconditional Coverage - Size and Power of Tests - 5% VaR

The table presents rejection rates obtained by applying unconditional coverage tests to 10,000 samples of Bernoulli simulated VaR-violation sequences. The VaR level p is set to 5%. Results are presented for various sets of sample sizes n and γ-factors which multiplies the probability of a VaR-violation by 1, 1.1, 1.25, 1.5, and 2. The results for $\gamma = 1p$ correspond to the evaluation of the size of the test. LR_{uc}^{kup} and GMM_{uc} refers to the unconditional coverage tests of Kupiec [1995] and Candelon et al. [2011]. MCS_{uc}^{tt} and MCS_{uc}^{ut} refer to the new two-tailed and upper-tail Monte Carlo simulation based tests. Top results are highlighted in bold type.

		Significance level: 1%				Significance level: 5%				Significance level: 10%			
$y \cdot p$	n	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}
	252	0.010	0.010	0.009	0.009	0.049	0.049	0.049	0.049	0.100	0.099	0.100	0.100
	500	0.011	0.011	0.010	0.010	0.049	0.049	0.050	0.047	0.099	0.103	0.099	0.097
5.00%	1,000	0.010	0.010	0.012	0.012	0.054	0.050	0.055	0.053	0.106	0.099	0.105	0.102
	1,500	0.010	0.012	0.009	0.009	0.047	0.052	0.049	0.048	0.098	0.099	0.097	0.101
	2,500	0.009	0.009	0.010	0.012	0.048	0.048	0.050	0.051	0.106	0.101	0.102	0.102
	252	0.015	0.005	0.015	**0.024**	0.062	0.059	0.064	**0.102**	0.111	0.128	0.124	**0.178**
	500	0.022	0.010	0.023	**0.036**	0.075	0.068	0.080	**0.128**	0.144	0.133	0.147	**0.223**
5.50%	1,000	0.033	0.020	0.034	**0.059**	0.105	0.099	0.118	**0.180**	0.195	0.190	0.191	**0.289**
	1,500	0.047	0.030	0.045	**0.076**	0.134	0.127	0.140	**0.215**	0.227	0.216	0.221	**0.345**
	2,500	0.083	0.055	0.082	**0.126**	0.201	0.186	0.204	**0.306**	0.336	0.296	0.310	**0.445**
	252	0.047	0.011	0.045	**0.072**	0.137	0.120	0.146	**0.223**	0.203	0.223	0.230	**0.338**
	500	0.089	0.048	0.095	**0.143**	0.211	0.215	0.240	**0.343**	0.331	0.331	0.346	**0.487**
6.25%	1,000	0.197	0.142	0.195	**0.281**	0.386	0.385	0.408	**0.530**	0.540	0.535	0.530	**0.667**
	1,500	0.342	0.268	0.328	**0.423**	0.549	0.542	0.560	**0.679**	0.672	0.666	0.679	**0.796**
	2,500	0.571	0.515	0.569	**0.661**	0.769	0.762	0.779	**0.859**	0.873	0.853	0.859	**0.922**
	252	0.196	0.061	0.192	**0.269**	0.377	0.349	0.396	**0.518**	0.481	0.510	0.519	**0.651**
	500	0.418	0.282	0.422	**0.516**	0.620	0.614	0.643	**0.754**	0.746	0.740	0.754	**0.852**
7.50%	1,000	0.761	0.700	0.769	**0.840**	0.894	0.898	0.907	**0.948**	0.951	0.950	0.948	**0.975**
	1,500	0.933	0.898	0.931	**0.958**	0.978	0.976	0.981	**0.992**	0.991	0.989	0.992	**0.997**
	2,500	0.996	0.993	0.996	**0.998**	0.999	**1.000**	0.999	0.999	**1.000**	**1.000**	0.999	**1.000**
	252	0.709	0.447	0.698	**0.777**	0.859	0.845	0.869	**0.922**	0.910	0.920	0.922	**0.960**
	500	0.961	0.924	0.961	**0.975**	0.988	0.988	0.988	**0.995**	0.996	0.996	0.995	**0.998**
10.00%	1,000	1.000	1.000	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	1,500	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	2,500	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Table 4.3: Unconditional Coverage - Size and Power of Tests - 1% VaR

The table presents rejection rates obtained by applying unconditional coverage tests to 10,000 samples of Bernoulli simulated VaR-violation sequences. The VaR level p is set to 1%. Results are presented for various sets of sample sizes n and γ-factors which multiplies the probability of a VaR-violation by 1, 1.1, 1.25, 1.5, and 2. The results for $\gamma = 1p$ correspond to the evaluation of the size of the test. LR_{uc}^{kup} and GMM_{uc} refers to the unconditional coverage tests of Kupiec [1995] and Candelon et al. [2011]. MCS_{uc}^{tt} and MCS_{uc}^{ut} refer to the new two-tailed and upper-tail Monte Carlo simulation based tests. Top results are highlighted in bold type.

		Significance level: 1%				Significance level: 5%				Significance level: 10%			
$y \cdot p$	n	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}
	252	0.010	0.012	0.009	0.010	0.051	0.050	0.049	0.050	0.101	0.103	0.100	0.104
	500	0.009	0.012	0.010	0.009	0.049	0.052	0.048	0.048	0.073	0.101	0.099	0.096
1.00%	1,000	0.014	0.009	0.012	0.011	0.048	0.050	0.053	0.051	0.105	0.102	0.103	0.107
	1,500	0.011	0.009	0.009	0.010	0.050	0.050	0.048	0.051	0.095	0.103	0.101	0.101
	2,500	0.010	0.008	0.010	0.011	0.054	0.047	0.052	0.051	0.106	0.099	0.100	0.100
	252	0.013	**0.017**	0.014	0.016	0.049	**0.074**	0.057	0.066	0.089	**0.138**	0.109	0.127
	500	0.010	0.014	0.015	**0.019**	0.046	0.054	0.062	**0.080**	0.082	0.135	0.115	**0.148**
1.10%	1,000	0.014	0.006	0.013	**0.023**	0.061	0.058	0.065	**0.089**	0.097	0.117	0.120	**0.166**
	1,500	0.015	0.010	0.017	**0.028**	0.069	0.058	0.070	**0.102**	0.136	0.132	0.127	**0.184**
	2,500	0.016	0.012	0.018	**0.036**	0.072	0.078	0.083	**0.130**	0.147	0.151	0.146	**0.221**
	252	0.026	**0.029**	0.020	**0.029**	0.058	0.108	0.076	**0.111**	0.095	0.187	0.134	**0.192**
	500	0.018	0.026	0.027	**0.039**	0.066	0.072	0.086	**0.136**	0.115	0.189	0.153	**0.234**
1.25%	1,000	0.032	0.003	0.039	**0.063**	0.112	0.119	0.131	**0.198**	0.164	0.207	0.207	**0.310**
	1,500	0.044	0.027	0.057	**0.091**	0.141	0.139	0.166	**0.253**	0.268	0.260	0.260	**0.371**
	2,500	0.082	0.050	0.087	**0.134**	0.220	0.219	0.232	**0.342**	0.334	0.335	0.344	**0.476**
	252	0.059	0.060	0.045	**0.069**	0.094	0.181	0.131	**0.192**	0.134	0.281	0.206	**0.305**
	500	0.054	0.081	0.072	**0.103**	0.137	0.160	0.186	**0.276**	0.220	0.339	0.282	**0.406**
1.50%	1,000	0.132	0.020	0.159	**0.220**	0.304	0.297	0.341	**0.447**	0.377	0.435	0.448	**0.580**
	1,500	0.194	0.140	0.227	**0.315**	0.401	0.401	0.439	**0.562**	0.573	0.569	0.563	**0.686**
	2,500	0.374	0.296	0.404	**0.506**	0.617	0.613	0.641	**0.747**	0.739	0.737	0.747	**0.848**
	252	0.182	**0.194**	0.143	0.194	0.238	**0.405**	0.291	0.401	0.281	0.518	0.405	**0.538**
	500	0.239	0.292	0.292	**0.358**	0.419	0.437	0.490	**0.605**	0.542	0.667	0.605	**0.721**
2.00%	1,000	0.533	0.213	0.583	**0.662**	0.747	0.749	0.778	**0.852**	0.810	0.845	0.852	**0.914**
	1,500	0.736	0.665	0.768	**0.831**	0.888	0.887	0.900	**0.941**	0.951	0.946	0.941	**0.969**
	2,500	0.944	0.911	0.947	**0.969**	0.988	0.984	0.987	**0.994**	0.992	0.993	0.994	**0.998**

4.3.2 Tests of the I.I.D. Property

As discussed in Section 4.2.1, a correctly specified VaR-model should yield i.i.d. violations. In this part of the simulation study, we analyze the power of the new backtests of i.i.d. VaR-violations using two data generating processes. First, we investigate the power of our new backtests and competing benchmark tests using dependent violations. Second, we repeat this analysis for non-identically distributed violation processes. In both settings, we perform the MCS_{iid} test and compare its finite sample behavior to that of the LR_{iid}^{mar} test of Christoffersen [1998], the LR_{iid}^{wei} test of Christoffersen and Pelletier [2004] and the GMM_{iid} test of Candelon et al. [2011].[13] Because clustering implies the occurance of at least two VaR-violations, the i.i.d. tests are not performed on samples where this minimum number is not achieved. To be more precise, $\sum_{t=1}^{n} I_t \geq 2$ holds true for each of the samples simulated by the procedures below, where I_t denotes a simulated VaR-violation sequence. Basically, each of the utilized tests are feasible under this condition. Only the LR_{iid}^{wei} test statistic cannot be computed for some simulated samples containing two violations (for more details see Candelon et al., 2011). We classify these cases as *not rejected*.

4.3.2.1 Independent VaR-Violations

In the first setting, we generate sequences of dependent VaR-violations with the degree of dependence inherent in the violation processes varying over time. For each λ and each n, we draw $10,000$ simulations of

$$y_t = \sigma_t z_t, \text{ with } \sigma_1 = 1 \tag{4.26}$$

and

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)z_{t-1}^2, 0 \leq \lambda \leq 1, t > 1. \tag{4.27}$$

[13] As suggested in Candelon et al. (2011) we set $q = 3$ for $p = 5\%$ and $q = 5$ for $p = 1\%$ throughout the simulation study. Critical values for the MCS_{iid} test are obtained as outlined in Section 4.2.3 using $10,000$ MC simulations.

Besides, $z_t \sim N(0,1), \forall t$. Note, this proceeding requires no pre-phasis in order to calculate σ. The distribution of y_t is based on the well-known exponentially weighted moving average (EWMA) type process. This approach allows for an easy regulation of the degree of dependence by determining λ as the single decay factor. To be more precise, λ controls the half-life interval of the observation weights (i.e., the interval in which the weight of an observed σ^2 decreases to half its original value) by $log(0.5)/log(\lambda)$. We apply the backtests to several different levels of λ representing half-life intervals of 5, 10, 20, 40, 60, and 80 days of data. This range of half-life intervals covers typical volatility persistence of asset return series.[14] Table 4.4 shows the half-life intervals and the corresponding λ level used to compute the power of the backtests.

Table 4.4: Half-Life Interval and λ-Level

The half-life interval is computed by $log(0.5)/log(\lambda)$ and refers to the time interval over which the weight of an observation decrease to one-half its original value. The corresponding λ refers to the decay factor of the EWMA type process of computing σ_t.

Half-Life Interval	5	10	20	40	60	80
λ	0.8706	0.9330	0.9659	0.9828	0.9885	0.9914

Dependent VaR-violations are ensured by setting a constant VaR for all $i = 1, \ldots, n$. For each decay factor λ, the VaR is determined separately by the empirical p-quantile of $10,000$ random values simulated by Equation (4.26). The simulated VaR-violations I_t are computed as defined by Equation (4.1).

Tables 4.5 and 4.6 show the results of the power study concerning the independence property of VaR-violations. We apply each test to 60 different combinations of coverage probability p, decay factor λ and sample size n. Together with the three significance levels of 1%, 5%, and 10%, we thus obtain 180 different settings in our simulation study. In total, the MCS_{iid} test outperforms the remaining tests in 104 out of the 180 test settings. Compared to the other test methods, this test possesses a high statistical power

[14]The EWMA approach can be used for VaR-forecasting purposes (RiskMetrics) whereas λ is typically set to 0.94 for one-day and 0.97 for one-month forecasts (see Mina et al., 2001). This corresponds to half-life intervals of 11 and 23 days. Furthermore, Berkowitz et al. [2011] estimated variance persistences for actual desk-level daily P/Ls from several business lines from a large international bank. The determined values are 0.9140, 0.9230, 0.9882 and 0.9941 which correspond to half-life intervals of 8, 9, 58, and 117 days.

in settings in which the half-life interval is relatively large. Furthermore, the superiority of the MCS_{iid} test increases with the significance level. The GMM test shows the best statistical power in almost one third of the considered settings. Compared to the remaining tests, the test performs well particularly for half-life intervals up to 20 days and for small significance levels. For significance level and coverage probability 1%, its power is almost always superior. The LR_{iid}^{mar} test yields the best statistical power in 21 out of 150 settings, this is especially true for small samples as well as for a half-life interval of five days. This result should be interpreted somewhat cautiously due to the fact that the vast majority of the top results are concentrated at the very short half-life interval of five days. It is to be expected that the LR_{iid}^{mar} test performs well in such circumstances, because short decay intervals lead to frequent occurrences of successive VaR-violations. Consequently, the power of this test deteriorates as the decay interval increases. Besides, the LR_{iid}^{mar} test performs surprisingly well for some settings with $n = 252$. However, in these cases the power decreases if n increases indicating asymptotic disturbances. A similar phenomenon was observed in Berkowitz et al. [2011]. For none of the 180 different settings does the LR_{iid}^{wei} test lead to the best statistical power of all analyzed test methods. Furthermore, for $p = 5\%$ and a half-life interval larger than 10 days, the test yields a statistical power below its nominal size and shows the undesired behavior of decreasing rejection rates as the sample size increases.

Table 4.5: I.I.D. VaR-Violations - Setting 1: Independence - Power of Tests - 5% VaR

The table presents rejection rates obtained by applying tests for i.i.d. VaR-violations to 10,000 samples of non-independent VaR-violation sequences simulated by Equation (4.26). The VaR level p is set to 5%. Results are presented for various sets of sample sizes n and half-life intervals which serve as a proxy for the degree of dependence. LR_{iid}^{mar}, LR_{iid}^{wei} and GMM_{iid} refers to the independence tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011]. MCS_{iid} refers to the new Monte Carlo simulation based test. Top results are highlighted in bold type.

Half-Life		Significance level: 1%				Significance level: 5%				Significance level: 10%			
Interval	n	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
	252	0.067	0.005	**0.108**	0.072	0.146	0.033	0.213	**0.220**	0.195	0.075	0.270	**0.339**
	500	0.093	0.016	**0.186**	0.142	0.170	0.081	**0.362**	0.354	0.224	0.153	0.451	**0.499**
5	1,000	0.126	0.047	**0.308**	0.264	0.217	0.160	**0.591**	0.552	0.308	0.260	0.689	**0.695**
	1,500	0.155	0.077	**0.423**	0.393	0.325	0.233	**0.741**	0.684	0.451	0.358	**0.823**	0.807
	2,500	0.308	0.170	**0.614**	0.611	0.515	0.396	**0.905**	0.858	0.631	0.535	**0.948**	0.933
	252	0.037	0.005	**0.086**	0.063	0.104	0.026	0.173	**0.188**	0.153	0.064	0.227	**0.296**
	500	0.047	0.006	**0.143**	0.120	0.098	0.038	0.281	**0.293**	0.145	0.080	0.357	**0.423**
10	1,000	0.049	0.014	**0.214**	0.211	0.104	0.065	0.454	**0.469**	0.168	0.122	0.556	**0.612**
	1,500	0.051	0.021	0.295	**0.315**	0.151	0.085	0.593	**0.600**	0.246	0.158	0.695	**0.732**
	2,500	0.096	0.033	0.425	**0.503**	0.234	0.134	**0.775**	0.774	0.338	0.223	0.860	**0.872**
	252	0.026	0.005	**0.061**	0.054	0.084	0.029	0.129	**0.149**	0.131	0.066	0.176	**0.236**
	500	0.029	0.005	**0.095**	0.092	0.073	0.029	0.195	**0.231**	0.112	0.062	0.262	**0.340**
20	1,000	0.025	0.004	0.135	**0.142**	0.067	0.027	0.300	**0.332**	0.119	0.058	0.392	**0.460**
	1,500	0.018	0.005	0.169	**0.202**	0.077	0.029	0.392	**0.438**	0.151	0.058	0.494	**0.578**
	2,500	0.034	0.005	0.228	**0.327**	0.107	0.027	0.536	**0.591**	0.181	0.055	0.645	**0.727**
	252	0.022	0.005	**0.052**	0.042	0.077	0.031	0.115	**0.128**	0.117	0.069	0.162	**0.210**
	500	0.022	0.008	**0.079**	0.077	0.064	0.030	0.163	**0.196**	0.103	0.068	0.226	**0.297**
40	1,000	0.018	0.003	0.095	**0.099**	0.052	0.024	0.219	**0.251**	0.103	0.051	0.293	**0.363**
	1,500	0.012	0.002	0.107	**0.129**	0.060	0.014	0.265	**0.307**	0.117	0.037	0.354	**0.430**
	2,500	0.017	0.002	0.128	**0.180**	0.073	0.010	0.324	**0.397**	0.132	0.025	0.424	**0.531**
	252	0.020	0.008	0.041	**0.042**	0.071	0.037	0.099	**0.130**	0.107	0.082	0.141	**0.211**
	500	0.023	0.005	**0.085**	0.080	0.059	0.032	0.164	**0.198**	0.095	0.070	0.224	**0.297**
60	1,000	0.016	0.005	0.093	**0.100**	0.049	0.024	0.204	**0.246**	0.098	0.049	0.275	**0.350**
	1,500	0.012	0.003	0.106	**0.119**	0.063	0.017	0.234	**0.280**	0.120	0.040	0.314	**0.396**
	2,500	0.016	0.001	0.110	**0.146**	0.065	0.009	0.269	**0.331**	0.122	0.026	0.363	**0.459**
	252	0.022	0009	0.032	**0.036**	0.072	0.041	0.089	**0.117**	0.107	0.086	0.130	**0.200**
	500	0.020	0.006	**0.085**	0.083	0.051	0.035	0.167	**0.206**	0.085	0.073	0.224	**0.305**
80	1,000	0.016	0.003	0.113	**0.119**	0.047	0.026	0.224	**0.263**	0.093	0.057	0.297	**0.371**
	1,500	0.014	0.002	0.113	**0.128**	0.065	0.021	0.250	**0.289**	0.122	0.045	0.323	**0.400**
	2,500	0.015	0.003	0.108	**0.150**	0.065	0.013	0.267	**0.323**	0.118	0.028	0.350	**0.436**

Table 4.6: I.I.D. VaR-Violations - Setting 1: Independence - Power of Tests - 1% VaR

The table presents rejection rates obtained by applying tests for i.i.d. VaR-violations to 10,000 samples of non-independent VaR-violation sequences simulated by Equation (4.26). The VaR level p is set to 1%. Results are presented for various sets of sample sizes n and half-life intervals which serve as a proxy for the degree of dependence. LR_{iid}^{mar}, LR_{iid}^{wei} and GMM_{iid} refers to the independence tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011]. MCS_{iid} refers to the new Monte Carlo simulation based test. Top results are highlighted in bold type.

Half-Life		Significance level: 1%				Significance level: 5%				Significance level: 10%			
Interval	n	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
	252	0.055	0.004	**0.068**	0.048	**0.181**	0.035	0.136	0.141	**0.237**	0.095	0.186	0.226
	500	0.065	0.010	**0.073**	0.047	**0.198**	0.065	0.152	0.148	**0.252**	0.132	0.212	0.241
5	1,000	**0.114**	0.038	0.099	0.055	**0.230**	0.137	0.211	0.182	**0.346**	0.224	0.285	0.296
	1,500	**0.141**	0.087	0.116	0.064	**0.283**	0.219	0.265	0.212	**0.388**	0.322	0.361	0.328
	2,500	**0.193**	0.179	0.149	0.083	**0.384**	0.362	0.363	0.255	**0.482**	0.475	0.470	0.393
	252	0.037	0.005	**0.076**	0.059	**0.156**	0.034	0.141	0.147	0.217	0.080	0.192	**0.227**
	500	0.039	0.009	**0.078**	0.051	0.151	0.051	**0.156**	0.150	0.225	0.104	0.211	**0.239**
10	1,000	0.064	0.026	**0.100**	0.058	0.152	0.100	**0.205**	0.187	0.265	0.173	0.281	**0.297**
	1,500	0.072	0.055	**0.111**	0.067	0.174	0.161	**0.250**	0.212	0.266	0.254	**0.343**	0.327
	2,500	0.094	0.117	**0.140**	0.098	0.236	0.275	**0.340**	0.273	0.324	0.384	**0.453**	0.404
	252	0.026	0.005	**0.084**	0.066	**0.158**	0.031	0.147	0.156	0.227	0.075	0.192	**0.237**
	500	0.028	0.008	**0.076**	0.052	0.114	0.049	0.144	**0.147**	0.198	0.099	0.194	**0.235**
20	1,000	0.040	0.020	**0.083**	0.067	0.103	0.078	0.173	**0.187**	0.209	0.137	0.244	**0.287**
	1,500	0.042	0.035	**0.098**	0.069	0.124	0.113	**0.216**	0.202	0.192	0.189	0.296	**0.320**
	2,500	0.048	0.071	**0.114**	0.084	0.149	0.181	**0.283**	0.258	0.225	0.271	0.380	**0.388**
	252	0.020	0.004	**0.079**	0.065	**0.199**	0.027	0.142	0.155	**0.266**	0.063	0.193	0.238
	500	0.023	0.010	**0.078**	0.070	0.107	0.048	0.135	**0.151**	0.204	0.093	0.187	**0.222**
40	1,000	0.031	0.026	**0.089**	0.068	0.083	0.077	0.154	**0.176**	0.181	0.136	0.216	**0.265**
	1,500	0.032	0.035	**0.087**	0.072	0.099	0.099	0.182	**0.195**	0.156	0.158	0.253	**0.295**
	2,500	0.031	0.050	**0.097**	0.088	0.119	0.126	0.223	**0.238**	0.180	0.195	0.308	**0.348**
	252	0.017	0.005	**0.077**	0.052	**0.257**	0.026	0.136	0.149	**0.330**	0.062	0.188	0.230
	500	0.024	0.010	**0.088**	0.074	0.116	0.045	0.142	**0.157**	0.212	0.095	0.189	**0.229**
60	1,000	0.031	0.030	**0.089**	0.073	0.081	0.084	0.155	**0.170**	0.174	0.135	0.213	**0.251**
	1,500	0.031	0.039	**0.092**	0.082	0.092	0.095	0.174	**0.189**	0.143	0.155	0.241	**0.280**
	2,500	0.029	0.052	**0.093**	0.091	0.109	0.118	0.199	**0.218**	0.162	0.183	0.277	**0.327**
	252	0.014	0.004	**0.064**	0.037	**0.302**	0.025	0.131	0.127	**0.374**	0.054	0.181	0.204
	500	0.023	0.006	**0.081**	0.071	0.112	0.039	0.135	**0.159**	0.211	0.084	0.182	**0.231**
80	1,000	0.030	0.031	**0.096**	0.083	0.083	0.085	0.157	**0.181**	0.171	0.135	0.211	**0.262**
	1,500	0.027	0.046	**0.090**	0.088	0.083	0.103	0.163	**0.193**	0.133	0.159	0.224	**0.279**
	2,500	0.033	0.054	0.097	**0.102**	0.116	0.118	0.194	**0.220**	0.175	0.177	0.265	**0.315**

4.3.2.2 Identically Distributed VaR-Violations

The data generating process for the second part of the simulation study is given by:

$$I_t = \begin{cases} \overset{i.i.d.}{\sim} \text{Bern}(p - 2\delta), 1 \leq t \leq \frac{n}{4}; \\ \overset{i.i.d.}{\sim} \text{Bern}(p + \delta), \frac{n}{4} < t \leq \frac{n}{2}; \\ \overset{i.i.d.}{\sim} \text{Bern}(p - \delta), \frac{n}{2} < t \leq \frac{3n}{4}; \\ \overset{i.i.d.}{\sim} \text{Bern}(p + 2\delta), \frac{3n}{4} < t \leq n. \end{cases} \tag{4.28}$$

Here, we choose $\delta = 0p$ to analyze the size of a test and $\delta = 0.1p,\ 0.2p,\ 0.3p,\ 0.4p$ and $0.5p$ for the power study. This setting leads to variations in the probability of obtaining a VaR-violation between the four equal-sized subsamples. Consequently, the violations will occur unequally distributed. Note that the probability variations are determined in a way which ensures $\mathbb{E}\left(\sum_{t=1}^{n} I_t\right) = n \cdot p$. The setup of this part of the simulation study covers a realistic scenario in which a VaR-model does not, or not fully, incorporate changes from calm market phases to highly volatile bear markets or financial crises and vice versa. This in turn leads to clustered VaR-violations regardless of the question whether the data might show signs of autocorrelation.

Alternatively, non-stationary VaR-violations could be identified by splitting a sample into several subsamples and applying the test for uc to each subsample. However, this approach suffers from two main drawbacks. First, for small subsamples the power of uc tests is relatively low (see Table 4.2). Second, it remains unclear at which points real data samples have to be split into two or more subsamples.

Tables 4.7 and 4.8 show the results of the power study concerning the property of identically distributed VaR-violations. We apply each test to 50 different combinations of coverage probability p, probability variation factor δ, and sample size n. Furthermore, we compute rejection rates for significance levels of 1%, 5%, and 10% which leads to a total of 150 different test settings.

Table 4.7: I.I.D. VaR-Violations - Setting 2: Identical Distribution - Size and Power of Tests - 5% VaR

The table presents rejection rates obtained by applying tests for i.i.d. VaR-violations to 10,000 samples of non-identically distributed VaR-violation sequences simulated by Equation (4.28). The VaR level p is set to 5%. Results are presented for various sets of sample sizes n and probability variation factors δ. Results for $\delta = 0p$ correspond to the evaluation of the size of the test. LR_{iid}^{mar}, LR_{iid}^{wei} and GMM_{iid} refers to the independence tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011]. MCS_{iid} refers to the new simulation based i.i.d. test. Top results are highlighted in bold type.

		Significance level: 1%				Significance level: 5%				Significance level: 10%			
δ	n	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
	252	0.010	0.010	0.011	0.010	0.048	0.053	0.049	0.053	0.095	0.104	0.101	0.101
	500	0.011	0.010	0.013	0.011	0.050	0.048	0.052	0.048	0.101	0.095	0.102	0.102
$0p$	1,000	0.009	0.010	0.010	0.008	0.046	0.046	0.046	0.050	0.097	0.096	0.097	0.097
	1,500	0.011	0.010	0.010	0.009	0.048	0.045	0.048	0.049	0.099	0.094	0.099	0.099
	2,500	0.010	0.009	0.009	0.010	0.051	0.049	0.049	0.051	0.101	0.102	0.101	0.101
	252	0.011	0.009	**0.014**	0.009	0.052	0.048	0.058	**0.060**	0.101	0.094	0.105	**0.111**
	500	0.011	0.009	**0.015**	**0.015**	0.050	0.044	0.054	**0.068**	0.100	0.087	0.102	**0.128**
$0.1p$	1,000	0.011	0.006	**0.019**	0.018	0.048	0.032	0.066	**0.074**	0.099	0.073	0.116	**0.136**
	1,500	0.009	0.007	**0.021**	0.020	0.047	0.036	0.071	**0.082**	0.094	0.076	0.124	**0.146**
	2,500	0.009	0.008	0.021	**0.023**	0.049	0.037	0.078	**0.093**	0.100	0.072	0.131	**0.170**
	252	0.015	0.008	**0.019**	0.015	0.060	0.037	0.068	**0.074**	0.111	0.075	0.117	**0.144**
	500	0.014	0.005	0.033	**0.035**	0.057	0.025	0.094	**0.124**	0.106	0.058	0.147	**0.208**
$0.2p$	1,000	0.011	0.002	0.055	**0.065**	0.049	0.020	0.140	**0.190**	0.094	0.044	0.204	**0.291**
	1,500	0.011	0.002	0.072	**0.090**	0.051	0.014	0.177	**0.238**	0.106	0.033	0.250	**0.344**
	2,500	0.012	0.001	0.096	**0.140**	0.057	0.008	0.243	**0.326**	0.111	0.019	0.329	**0.452**
	252	0.015	0.004	**0.037**	0.030	0.061	0.023	0.105	**0.130**	0.112	0.053	0.156	**0.227**
	500	0.020	0.003	0.094	**0.097**	0.061	0.018	0.202	**0.258**	0.106	0.050	0.275	**0.377**
$0.3p$	1,000	0.016	0.003	0.212	**0.241**	0.054	0.024	0.386	**0.456**	0.106	0.058	0.471	**0.579**
	1,500	0.015	0.005	0.297	**0.358**	0.063	0.028	0.504	**0.591**	0.130	0.068	0.593	**0.704**
	2,500	0.022	0.008	0.450	**0.549**	0.085	0.038	0.697	**0.771**	0.148	0.075	0.783	**0.856**
	252	0.027	0.001	**0.079**	0.053	0.080	0.017	0.181	**0.209**	0.131	0.043	0.240	**0.346**
	500	0.033	0.006	0.273	**0.283**	0.078	0.049	0.452	**0.540**	0.125	0.112	0.535	**0.664**
$0.4p$	1,000	0.032	0.043	0.613	**0.638**	0.079	0.164	0.783	**0.828**	0.140	0.275	0.838	**0.894**
	1,500	0.029	0.114	0.781	**0.838**	0.105	0.284	0.908	**0.943**	0.181	0.410	0.940	**0.971**
	2,500	0.053	0.248	0.942	**0.971**	0.158	0.482	0.987	**0.993**	0.250	0.616	0.993	**0.997**
	252	0.041	0.002	**0.158**	0.113	0.104	0.028	0.317	**0.378**	0.148	0.074	0.400	**0.552**
	500	0.053	0.040	0.688	**0.729**	0.109	0.213	0.863	**0.944**	0.157	0.376	0.915	**0.982**
$0.5p$	1,000	0.057	0.436	**1.000**	**1.000**	0.124	0.794	**1.000**	**1.000**	0.201	0.910	**1.000**	**1.000**
	1,500	0.061	0.892	**1.000**	**1.000**	0.186	0.986	**1.000**	**1.000**	0.299	0.998	**1.000**	**1.000**
	2,500	0.138	**1.000**	**1.000**	**1.000**	0.311	**1.000**	**1.000**	**1.000**	0.425	**1.000**	**1.000**	**1.000**

Table 4.8: I.I.D. VaR-Violations - Setting 2: Identical Distribution - Size and Power of Tests - 1% VaR

The table presents rejection rates obtained by applying tests for i.i.d. VaR-violations to 10,000 samples of non-identically distributed VaR-violation sequences simulated by Equation (4.28). The VaR level p is set to 1%. Results are presented for various sets of sample sizes n and probability variation factors δ. Results for $\delta = 0p$ correspond to the evaluation of the size of the test. LR_{iid}^{mar}, LR_{iid}^{wei} and GMM_{iid} refers to the independence tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011]. MCS_{iid} refers to the new simulation based i.i.d. test. Top results are highlighted in bold type.

		Significance level: 1%				Significance level: 5%				Significance level: 10%			
δ	n	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
	252	0.010	0.007	0.010	0.012	0.056	0.042	0.052	0.050	0.108	0.089	0.102	0.103
	500	0.009	0.009	0.009	0.011	0.050	0.050	0.048	0.053	0.101	0.097	0.099	0.101
$0p$	1,000	0.010	0.010	0.009	0.011	0.048	0.046	0.049	0.051	0.100	0.096	0.102	0.101
	1,500	0.010	0.010	0.010	0.010	0.047	0.047	0.050	0.052	0.098	0.095	0.100	0.096
	2,500	0.009	0.010	0.012	0.011	0.049	0.047	0.050	0.053	0.099	0.098	0.099	0.105
	252	**0.011**	0.008	0.009	0.009	**0.054**	0.042	0.049	0.050	**0.104**	0.087	0.099	0.098
	500	0.009	0.009	0.009	**0.012**	0.048	0.047	0.047	**0.054**	0.097	0.093	0.096	**0.103**
$0.1p$	1,000	0.011	0.011	**0.012**	**0.012**	0.053	0.049	0.054	**0.056**	0.102	0.099	0.107	**0.113**
	1,500	**0.013**	0.009	0.012	0.011	0.053	0.048	0.052	**0.059**	0.104	0.095	0.102	**0.113**
	2,500	**0.013**	0.008	0.012	**0.013**	0.055	0.042	0.056	**0.064**	0.104	0.088	0.111	**0.121**
	252	0.010	0.007	0.011	**0.014**	0.052	0.038	0.048	**0.056**	0.102	0.079	0.097	**0.106**
	500	0.010	0.009	0.011	**0.012**	**0.058**	0.042	0.052	0.053	**0.105**	0.086	0.098	**0.105**
$0.2p$	1,000	0.012	0.008	**0.015**	0.013	0.053	0.041	0.060	**0.065**	0.107	0.087	0.107	**0.126**
	1,500	0.012	0.007	**0.016**	**0.016**	0.056	0.039	0.064	**0.082**	0.114	0.085	0.122	**0.151**
	2,500	0.013	0.008	0.027	**0.031**	0.058	0.042	0.094	**0.120**	0.111	0.087	0.152	**0.204**
	252	0.013	0.006	0.014	**0.015**	0.057	0.033	0.054	**0.060**	0.105	0.073	0.102	**0.115**
	500	0.013	0.006	**0.017**	**0.017**	0.062	0.033	0.055	**0.067**	0.110	0.066	0.101	**0.124**
$0.3p$	1,000	0.015	0.005	**0.022**	0.020	0.064	0.034	0.076	**0.091**	0.123	0.078	0.132	**0.173**
	1,500	0.014	0.008	0.033	**0.034**	0.063	0.041	0.101	**0.143**	0.121	0.088	0.168	**0.250**
	2,500	0.017	0.011	0.077	**0.090**	0.070	0.058	0.193	**0.242**	0.125	0.119	0.278	**0.360**
	252	0.015	0.005	**0.017**	**0.017**	0.063	0.026	0.055	**0.070**	0.111	0.061	0.104	**0.121**
	500	0.016	0.003	**0.023**	0.022	0.069	0.023	0.066	**0.075**	0.114	0.057	0.111	**0.145**
$0.4p$	1,000	0.018	0.005	**0.038**	0.028	0.076	0.035	0.114	**0.138**	0.139	0.079	0.181	**0.253**
	1,500	0.020	0.014	**0.074**	0.073	0.074	0.065	0.191	**0.257**	0.139	0.129	0.280	**0.407**
	2,500	0.021	0.040	0.226	**0.259**	0.081	0.150	0.424	**0.522**	0.146	0.251	0.518	**0.645**
	252	0.018	0.001	0.022	**0.024**	0.069	0.011	0.066	**0.081**	0.114	0.039	0.108	**0.131**
	500	0.017	0.001	**0.031**	0.030	0.082	0.014	**0.091**	**0.091**	0.132	0.045	0.140	**0.164**
$0.5p$	1,000	0.025	0.007	**0.079**	0.053	0.087	0.051	0.197	**0.225**	0.157	0.113	0.277	**0.377**
	1,500	0.024	0.032	**0.174**	0.163	0.085	0.142	0.354	**0.487**	0.164	0.249	0.467	**0.670**
	2,500	0.027	0.167	0.597	**0.694**	0.099	0.437	0.822	**0.926**	0.172	0.602	0.893	**0.975**

In total, the MCS_{iid} test possesses a high statistical power regarding non-identically distributed VaR-violations and its test results are comparable to or better than the performance of the remaining three approaches for 130 out of the 150 settings. Particularly for significance levels of 5% and 10%, it outperforms the competing tests in almost all cases, irrespective of the degree of probability variation or sample size. The GMM test yields rejection rates which are equal or better than the results of the competing models for 30 of the 150 simulation settings. The test particularly achieves its top results for a significance level of 1%. The LR_{iid}^{mar} test is able to match the results of the competing tests in only seven cases which are restricted to settings in which $p = 1\%$ and $\delta = 0.1p$. The results of the LR_{iid}^{wei} test falls short of the performance of the remaining tests in almost all settings. Finally, it is striking that the power of the LR_{iid}^{mar} test and the LR_{iid}^{wei} test significantly exceed the nominal size only for large shifts in the VaR-violation probability, i.e. $\delta \geq 0.4p$.

4.3.3 Conditional Coverage

Table 4.9 illustrates the behavior of the MCS_{cc} test considering different levels of the weighting parameter a. For reasons of space we present results only for a single parameter combination for each of the two settings. This includes $n = 1000$, a half-life interval of 20 days, and $\gamma = 1.25$ for setting 1 and $n = 1000$, $\delta = 0.3p$, $\gamma = 1.25$ for setting 2. Depending on the setting, the VaR probability p, and the significance level, the test yields the highest rejection rates for values of a between 0.5 and 0.8. This is consistent with our expectation that the maximum of the statistical power is achieved when $0 < a < 1$, i.e., when the cc test addresses both the uc as well as the i.i.d. property of the violations. In the following, we only present the results for $a = 0.5$.

We continue with a comparison of the size and the power of the cc test MCS_{cc} to the LR_{cc}^{mar} test of Christoffersen [1998], the LR_{cc}^{wei} test of Christoffersen and Pelletier [2004] and the GMM_{cc} test of Candelon et al. [2011]. For this purpose, we combine each of the two settings described in Section 4.3.2 with increased probabilities of a VaR-violation outlined in Section 4.3.1. Note that we use the two-tailed uc component. For the

Table 4.9: Conditional Coverage - Power of the MCS_{cc} Test under Different Level of a

The table presents rejection rates obtained by applying the MCS_{cc} test to 10,000 samples of non-i.i.d. distributed VaR-violation sequences. Panel A and B contain rejection rates for sequences simulated by Equation (4.26) and Equation (4.28) with an increased violation probability. The parameter combinations used for the simulations are described at the top of each panel. The top result for each combination of a, VaR level, and significance level is highlighted in bold type.

	5% VaR			1% VaR		
	Significance level:			Significance level:		
a	1%	5%	10%	1%	5%	10%
Panel A: $n = 1,000$ / $\gamma = 1.25$ / half-life interval = 20 days						
0	0.107	0.294	0.440	0.056	0.171	0.283
0.1	0.123	0.329	0.482	0.053	0.184	0.295
0.2	0.149	0.376	0.535	0.068	0.219	0.337
0.3	0.169	0.449	0.607	0.082	0.232	0.356
0.4	0.231	0.511	0.649	0.106	0.265	0.378
0.5	0.310	**0.550**	**0.664**	0.128	0.277	0.372
0.6	0.350	0.545	0.641	**0.150**	**0.289**	**0.379**
0.7	**0.366**	0.539	0.621	0.144	0.254	0.340
0.8	0.343	0.511	0.604	0.140	0.256	0.330
0.9	0.318	0.468	0.553	0.149	0.264	0.342
1	0.306	0.455	0.536	0.125	0.224	0.300
Panel B: $n = 1,000$ / $\gamma = 1.25$ / $\delta = 0.3p$						
0	0.105	0.264	0.393	0.014	0.074	0.151
0.1	0.108	0.290	0.433	0.013	0.081	0.164
0.2	0.124	0.336	0.479	0.015	0.093	0.183
0.3	0.146	0.383	0.548	0.019	0.098	0.192
0.4	0.188	0.453	0.604	0.023	0.121	0.221
0.5	0.232	0.509	0.636	0.036	0.140	0.234
0.6	0.294	**0.542**	**0.657**	0.053	0.153	0.236
0.7	**0.299**	0.519	0.631	0.059	0.158	0.233
0.8	0.285	0.505	0.617	**0.067**	**0.163**	**0.238**
0.9	0.256	0.463	0.570	0.064	0.161	0.236
1	0.239	0.441	0.553	0.064	0.159	0.234

determination of critical values we perform the procedure as explained in Section 4.2.4 using $10,000$ MC simulations. In line with the settings above, for each combination of γ, δ, volatility half-life, and n we repeat the simulation of VaR-violation sequences $10,000$ times. We present the results of the simulation study concerning an increased probability of a VaR-violation combined with non-independent occurrence of violations (setting 1) in Tables 4.10 and 4.11, and combined with non-identically distributed violations (setting 2) in Table 4.12 and 4.13. [15]

[15]To save space, we do not present the rejection rates of all parameter combinations. The complete results are available from the authors upon request.

Table 4.10: Conditional Coverage - Setting 1: Independence - Power of Tests - 5% VaR

The table presents rejection rates obtained by applying cc tests to 10,000 samples of non-independent VaR-violation sequences simulated by Equation (4.26) with an increased violation probability. The VaR level p is set to 5%. Results are presented for various sets of sample sizes n, γ-factors which increase the probability of a VaR-violation, and decay intervals which serve as a proxy for the degree of dependence. LR_{cc}^{mar}, LR_{cc}^{wei} and GMM_{cc} refers to the cc tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011]. MCS_{cc} refers to the new simulation based test. Top results are highlighted in bold type.

Decay			Significance level: 1%				Significance level: 5%				Significance level: 10%			
Interval	p	n	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}
		252	0.052	0.028	0.044	**0.093**	0.103	0.088	0.212	**0.237**	0.193	0.154	0.318	**0.344**
		500	0.059	0.033	0.063	**0.150**	0.128	0.108	0.287	**0.340**	0.208	0.177	0.415	**0.463**
10	5.50%	1,000	0.074	0.047	0.107	**0.251**	0.166	0.142	0.435	**0.493**	0.231	0.218	0.571	**0.613**
		1,500	0.104	0.061	0.183	**0.371**	0.199	0.168	0.558	**0.613**	0.280	0.256	0.686	**0.721**
		2,500	0.152	0.095	0.360	**0.565**	0.290	0.226	0.767	**0.783**	0.377	0.331	0.857	**0.860**
		252	0.204	0.109	0.060	**0.235**	0.302	0.222	0.364	**0.457**	0.433	0.307	0.488	**0.555**
		500	0.353	0.259	0.144	**0.417**	0.493	0.429	0.565	**0.661**	0.599	0.526	0.693	**0.762**
10	7.50%	1,000	0.591	0.524	0.387	**0.704**	0.747	0.693	0.825	**0.878**	0.804	0.770	0.893	**0.929**
		1,500	0.795	0.708	0.669	**0.886**	0.878	0.847	0.939	**0.967**	0.915	0.899	0.970	**0.984**
		2,500	0.946	0.909	0.932	**0.985**	0.979	0.961	0.994	**0.998**	0.988	0.980	0.997	**0.999**
		252	0.096	0.060	0.047	**0.128**	0.160	0.142	0.227	**0.285**	0.258	0.215	0.335	**0.382**
		500	0.127	0.083	0.052	**0.182**	0.218	0.178	0.285	**0.372**	0.306	0.253	0.418	**0.486**
20	6.25%	1,000	0.179	0.137	0.096	**0.299**	0.318	0.264	0.438	**0.539**	0.393	0.345	0.572	**0.651**
		1,500	0.272	0.191	0.175	**0.451**	0.403	0.343	0.575	**0.680**	0.486	0.437	0.699	**0.776**
		2,500	0.409	0.300	0.388	**0.678**	0.591	0.475	0.771	**0.853**	0.666	0.577	0.856	**0.907**
		252	0.142	0.119	0.094	**0.166**	0.201	0.212	0.280	**0.308**	0.289	0.290	0.385	**0.404**
		500	0.156	0.124	0.075	**0.189**	0.234	0.219	0.289	**0.366**	0.314	0.287	0.404	**0.471**
40	6.25%	1,000	0.200	0.174	0.098	**0.267**	0.329	0.292	0.399	**0.490**	0.399	0.367	0.525	**0.604**
		1,500	0.279	0.216	0.150	**0.372**	0.399	0.354	0.495	**0.597**	0.473	0.445	0.618	**0.700**
		2,500	0.397	0.301	0.289	**0.552**	0.571	0.460	0.669	**0.765**	0.643	0.552	0.775	**0.838**
		252	0.223	0.224	**0.256**	0.220	0.310	0.374	**0.458**	0.406	0.416	0.466	**0.546**	0.535
		500	0.173	0.175	0.193	**0.224**	0.252	0.288	0.391	**0.395**	0.335	0.369	0.486	**0.505**
80	5.50%	1,000	0.129	0.124	0.149	**0.217**	0.215	0.207	0.357	**0.394**	0.275	0.277	0.456	**0.502**
		1,500	0.122	0.104	0.139	**0.223**	0.194	0.183	0.343	**0.401**	0.253	0.248	0.446	**0.510**
		2,500	0.126	0.092	0.142	**0.250**	0.219	0.163	0.376	**0.454**	0.278	0.223	0.483	**0.557**
		252	0.278	0.249	0.218	**0.292**	0.336	0.348	0.423	**0.449**	0.413	0.417	0.513	**0.542**
		500	0.326	0.294	0.220	**0.362**	0.404	0.388	0.473	**0.540**	0.474	0.452	0.577	**0.633**
80	7.50%	1,000	0.491	0.477	0.313	**0.564**	0.625	0.614	0.676	**0.764**	0.685	0.681	0.770	**0.837**
		1,500	0.696	0.626	0.478	**0.713**	0.789	0.762	0.821	**0.874**	0.839	0.821	0.888	**0.919**
		2,500	0.908	0.874	0.807	**0.927**	0.957	0.937	0.966	**0.981**	0.970	0.960	0.982	**0.991**

Table 4.11: Conditional Coverage - Setting 1: Independence - Power of Tests - 1% VaR

The table presents rejection rates obtained by applying cc tests to 10,000 samples of non-independent VaR-violation sequences simulated by Equation (4.26) with an increased violation probability. The VaR level p is set to 1%. Results are presented for various sets of sample sizes n, γ-factors which increase the probability of a VaR-violation, and decay intervals which serve as a proxy for the degree of dependence. LR_{cc}^{mar}, LR_{cc}^{wei} and GMM_{cc} refers to the cc tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011]. MCS_{cc} refers to the new simulation based test. Top results are highlighted in bold type.

Decay			Significance level: 1%				Significance level: 5%				Significance level: 10%			
Interval	p	n	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}
		252	0.038	0.017	0.093	**0.094**	0.140	0.066	**0.198**	0.191	**0.335**	0.128	0.273	0.266
		500	0.047	0.023	**0.092**	0.091	0.174	0.081	**0.201**	0.191	**0.274**	0.144	0.267	**0.274**
10	1.10%	1,000	0.044	0.037	0.023	**0.088**	0.158	0.129	0.194	**0.227**	0.242	0.210	0.303	**0.313**
		1,500	0.051	0.066	0.025	**0.094**	0.180	0.167	0.220	**0.253**	0.275	0.264	0.343	**0.359**
		2,500	0.057	0.120	0.042	**0.125**	0.194	0.271	**0.304**	**0.304**	0.326	0.383	**0.457**	0.426
		252	0.072	0.031	0.154	**0.162**	0.216	0.109	0.291	**0.297**	**0.455**	0.186	0.377	0.380
		500	0.127	0.059	0.177	**0.202**	0.341	0.147	0.327	**0.343**	**0.466**	0.220	0.402	0.436
10	1.50%	1,000	0.167	0.113	0.034	**0.229**	0.367	0.244	0.314	**0.426**	0.467	0.340	0.441	**0.528**
		1,500	0.225	0.210	0.048	**0.288**	0.439	0.366	0.413	**0.518**	0.568	0.476	0.553	**0.619**
		2,500	0.350	0.418	0.116	**0.424**	0.606	0.600	0.575	**0.672**	0.728	0.694	0.712	**0.771**
		252	0.069	0.040	**0.141**	0.135	0.182	0.104	**0.245**	0.238	**0.380**	0.168	0.317	0.312
20		500	0.067	0.034	0.118	**0.130**	0.214	0.093	0.231	**0.233**	0.311	0.154	0.297	**0.316**
	1.25%	1,000	0.074	0.051	0.023	**0.128**	0.207	0.133	0.219	**0.282**	0.289	0.207	0.329	**0.377**
		1,500	0.080	0.078	0.023	**0.150**	0.212	0.178	0.247	**0.321**	0.327	0.259	0.375	**0.423**
		2,500	0.107	0.141	0.038	**0.194**	0.277	0.284	0.324	**0.403**	0.409	0.389	0.466	**0.526**
		252	0.129	0.085	**0.183**	**0.183**	0.227	0.158	**0.273**	0.271	**0.387**	0.213	0.336	0.335
		500	0.099	0.064	0.135	**0.144**	0.230	0.124	0.228	**0.233**	**0.307**	0.183	0.294	0.306
40	1.25%	1,000	0.091	0.072	0.041	**0.146**	0.212	0.146	0.209	**0.271**	0.285	0.213	0.311	**0.356**
		1,500	0.095	0.089	0.035	**0.148**	0.206	0.172	0.221	**0.302**	0.312	0.247	0.334	**0.397**
		2,500	0.111	0.126	0.044	**0.190**	0.273	0.248	0.273	**0.377**	0.380	0.341	0.397	**0.491**
		252	0.243	0.192	**0.296**	**0.296**	0.342	0.273	0.373	**0.374**	**0.470**	0.329	0.424	0.427
		500	0.139	0.105	0.167	**0.174**	0.226	0.172	0.244	**0.244**	0.277	0.243	**0.307**	0.307
80	1.10%	1,000	0.109	0.103	0.085	**0.135**	0.198	0.190	**0.242**	0.236	0.278	0.263	**0.330**	0.321
		1,500	0.088	0.101	0.074	**0.128**	0.178	0.178	0.233	**0.248**	0.277	0.250	0.330	**0.339**
		2,500	0.077	0.098	0.068	**0.138**	0.182	0.183	0.222	**0.266**	0.260	0.257	0.320	**0.364**
		252	0.263	0.209	0.302	**0.316**	0.355	0.289	0.385	**0.388**	**0.480**	0.344	0.442	0.441
		500	0.198	0.141	0.217	**0.234**	**0.316**	0.209	0.308	0.313	**0.386**	0.267	0.367	0.378
80	1.50%	1,000	0.195	0.151	0.095	**0.222**	0.318	0.243	0.285	**0.352**	0.384	0.304	0.378	**0.440**
		1,500	0.213	0.172	0.078	**0.242**	0.351	0.272	0.296	**0.417**	0.464	0.351	0.409	**0.507**
		2,500	0.313	0.288	0.102	**0.331**	0.515	0.432	0.407	**0.560**	0.621	0.528	0.538	**0.658**

Table 4.12: Conditional Coverage - Setting 2: Identical Distribution - Size and Power of Tests - 5% VaR

The table presents rejection rates obtained by applying cc tests to 10,000 samples of non-identically distributed VaR-violation sequences simulated by Equation (4.28) with an increased violation probability. The VaR level p is set to 5%. Results are presented for various sets of sample sizes n and γ-factors which increase the probability of a VaR-violation, and probability variation factors δ. The results for $\delta = 0p$ correspond to the evaluation of the size of the test. LR_{cc}^{mar}, LR_{cc}^{wei} and GMM_{cc} refers to the cc tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011]. MCS_{iid} refers to the new simulation based test. Top results are highlighted in bold.

			Significance level: 1%				Significance level: 5%				Significance level: 10%			
δ	p	n	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}
		252	0.010	0.010	0.009	0.011	0.049	0.049	0.051	0.051	0.093	0.099	0.103	0.100
		500	0.010	0.010	0.008	0.011	0.048	0.053	0.046	0.051	0.105	0.103	0.098	0.100
$0p$	5.00%	1,000	0.011	0.009	0.010	0.011	0.052	0.046	0.053	0.052	0.104	0.100	0.105	0.098
		1,500	0.011	0.009	0.009	0.011	0.049	0.047	0.052	0.053	0.098	0.101	0.102	0.102
		2,500	0.010	0.009	0.009	0.012	0.052	0.049	0.049	0.054	0.101	0.097	0.100	0.102
		252	**0.016**	0.008	0.004	0.011	0.061	0.044	0.046	**0.065**	0.115	0.086	0.096	**0.124**
		500	**0.019**	0.015	0.005	0.015	0.066	0.056	0.045	**0.078**	0.129	0.105	0.108	**0.148**
$0.1p$	5.50%	1,000	0.020	0.016	0.006	**0.021**	0.082	0.068	0.058	**0.103**	0.138	0.129	0.123	**0.186**
		1,500	0.030	0.021	0.008	**0.033**	0.092	0.084	0.071	**0.125**	0.154	0.148	0.151	**0.218**
		2,500	0.036	0.034	0.011	**0.048**	0.129	0.107	0.106	**0.174**	0.198	0.181	0.209	**0.281**
		252	**0.147**	0.073	0.008	0.103	0.280	0.193	0.220	**0.330**	0.431	0.296	0.372	**0.442**
		500	**0.309**	0.230	0.028	0.180	0.488	0.442	0.399	**0.501**	0.622	0.569	0.589	**0.659**
$0.1p$	7.50%	1,000	**0.609**	0.563	0.151	0.464	**0.802**	0.775	0.733	0.801	0.868	0.855	0.864	**0.902**
		1,500	**0.847**	0.808	0.407	0.733	0.932	0.928	0.896	**0.942**	0.962	0.961	0.961	**0.976**
		2,500	**0.979**	0.974	0.853	0.958	0.996	0.993	0.994	**0.997**	0.997	0.998	0.998	**0.999**
		252	0.038	0.010	0.003	**0.043**	0.100	0.048	0.095	**0.166**	0.193	0.096	0.187	**0.258**
		500	0.061	0.027	0.012	**0.101**	0.151	0.097	0.174	**0.291**	0.252	0.169	0.308	**0.419**
$0.3p$	6.25%	1,000	0.112	0.066	0.051	**0.237**	0.273	0.188	0.348	**0.492**	0.373	0.285	0.508	**0.634**
		1,500	0.199	0.113	0.114	**0.402**	0.367	0.281	0.515	**0.670**	0.477	0.402	0.670	**0.779**
		2,500	0.374	0.236	0.306	**0.667**	0.617	0.456	0.765	**0.873**	0.710	0.593	0.864	**0.929**
		252	0.017	0.002	0.014	**0.088**	0.045	0.023	0.177	**0.260**	0.105	0.060	0.298	**0.382**
		500	0.024	0.029	0.165	**0.477**	0.068	0.115	0.602	**0.733**	0.134	0.209	0.733	**0.824**
$0.5p$	5.50%	1,000	0.039	0.180	0.778	**0.892**	0.105	0.414	0.947	**0.963**	0.161	0.561	0.967	**0.981**
		1,500	0.063	0.429	0.951	**0.980**	0.148	0.682	0.992	**0.995**	0.230	0.791	0.997	**0.998**
		2,500	0.117	0.775	0.999	**1.000**	0.259	0.911	**1.000**	**1.000**	0.347	0.953	**1.000**	**1.000**
		252	0.137	0.044	0.022	**0.206**	0.256	0.148	0.320	**0.469**	0.418	0.240	0.478	**0.589**
		500	0.306	0.199	0.125	**0.491**	0.493	0.408	0.618	**0.759**	0.628	0.532	0.756	**0.853**
$0.5p$	7.50%	1,000	0.621	0.541	0.491	**0.849**	0.805	0.772	0.918	**0.961**	0.871	0.856	0.965	**0.983**
		1,500	0.864	0.794	0.820	**0.974**	0.939	0.924	0.987	**0.996**	0.965	0.959	0.996	**0.999**
		2,500	0.984	0.973	0.991	**1.000**	0.996	0.993	**1.000**	**1.000**	0.999	0.998	**1.000**	**1.000**

Table 4.13: Conditional Coverage - Setting 2: Identical Distribution - Size and Power of Tests - 1% VaR

The table presents rejection rates obtained by applying cc tests to 10,000 samples of non-identically distributed VaR-violation sequences simulated by Equation (4.28) with an increased violation probability. The VaR level p is set to 1%. Results are presented for various sets of sample sizes n and γ-factors which increase the probability of a VaR-violation, and probability variation factors δ. The results for $\delta = 0p$ correspond to the evaluation of the size of the test. LR_{cc}^{mar}, LR_{cc}^{wei} and GMM_{cc} refers to the cc tests of Christoffersen [1998], Christoffersen and Pelletier [2004] and Candelon et al. [2011]. MCS_{iid} refers to the new simulation based test. Top results are highlighted in bold.

			Significance level: 1%				Significance level: 5%				Significance level: 10%			
δ	p	n	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}
		252	0.009	0.008	0.010	0.009	0.046	0.041	0.048	0.051	0.175	0.083	0.101	0.102
		500	0.011	0.012	0.009	0.010	0.061	0.052	0.050	0.053	0.115	0.098	0.094	0.105
$0p$	1.00%	1,000	0.012	0.011	0.009	0.010	0.048	0.053	0.048	0.050	0.091	0.102	0.098	0.100
		1,500	0.010	0.009	0.011	0.008	0.042	0.046	0.048	0.052	0.101	0.096	0.098	0.101
		2,500	0.010	0.009	0.011	0.010	0.047	0.046	0.051	0.050	0.093	0.097	0.103	0.104
		252	0.012	0.007	0.014	**0.016**	0.056	0.043	0.064	**0.066**	**0.211**	0.090	0.122	0.124
		500	0.012	0.007	0.013	**0.018**	**0.082**	0.043	0.058	0.065	**0.153**	0.092	0.106	0.125
$0.1p$	1.10%	1,000	**0.015**	0.008	0.006	0.012	0.062	0.041	0.038	**0.066**	0.110	0.086	0.090	**0.128**
		1,500	0.013	0.009	0.006	**0.015**	0.059	0.043	0.039	**0.071**	0.125	0.090	0.095	**0.134**
		2,500	0.013	0.011	0.005	**0.015**	0.069	0.050	0.041	**0.078**	0.142	0.102	0.099	**0.145**
		252	0.029	0.014	**0.053**	**0.053**	0.124	0.074	0.152	**0.158**	**0.385**	0.140	0.247	0.247
		500	0.055	0.015	0.050	**0.073**	**0.257**	0.071	0.167	0.203	**0.394**	0.128	0.233	0.309
$0.1p$	1.50%	1,000	**0.095**	0.037	0.002	0.084	**0.283**	0.129	0.120	0.259	0.387	0.221	0.241	**0.408**
		1,500	**0.148**	0.081	0.001	0.097	**0.355**	0.217	0.155	0.343	**0.528**	0.335	0.317	0.488
		2,500	**0.251**	0.222	0.006	0.170	**0.563**	0.445	0.269	0.506	**0.708**	0.576	0.457	0.646
		252	0.013	0.007	**0.027**	0.026	0.073	0.048	**0.098**	0.097	**0.269**	0.095	0.171	0.168
		500	0.021	0.007	0.023	**0.036**	**0.129**	0.040	0.097	0.116	**0.225**	0.082	0.152	0.198
$0.3p$	1.25%	1,000	0.028	0.008	0.002	**0.039**	0.115	0.047	0.062	**0.146**	0.185	0.096	0.141	**0.238**
		1,500	0.031	0.015	0.002	**0.050**	0.124	0.065	0.078	**0.188**	0.234	0.125	0.171	**0.293**
		2,500	0.046	0.033	0.006	**0.093**	0.188	0.122	0.129	**0.285**	0.312	0.203	0.253	**0.413**
		252	0.007	0.003	**0.022**	0.019	0.054	0.022	**0.082**	0.077	**0.212**	0.055	0.141	0.133
		500	0.011	0.004	**0.029**	**0.029**	0.087	0.025	0.102	**0.105**	0.167	0.059	0.161	**0.176**
$0.5p$	1.10%	1,000	0.010	0.006	0.008	**0.060**	0.062	0.037	0.119	**0.183**	0.117	0.088	0.219	**0.272**
		1,500	0.012	0.023	0.012	**0.127**	0.068	0.087	0.190	**0.321**	0.133	0.163	0.327	**0.450**
		2,500	0.010	0.082	0.109	**0.439**	0.077	0.238	0.551	**0.700**	0.162	0.365	0.715	**0.807**
		252	0.025	0.009	0.055	**0.059**	0.125	0.051	0.162	**0.170**	**0.380**	0.105	0.256	0.262
		500	0.058	0.011	0.075	**0.094**	**0.258**	0.060	0.206	0.224	**0.394**	0.111	0.279	0.332
$0.5p$	1.50%	1,000	0.091	0.033	0.006	**0.150**	0.293	0.118	0.199	**0.366**	0.395	0.201	0.339	**0.493**
		1,500	0.149	0.092	0.011	**0.250**	0.352	0.234	0.290	**0.518**	0.521	0.352	0.457	**0.641**
		2,500	0.263	0.258	0.046	**0.458**	0.569	0.474	0.500	**0.735**	0.719	0.605	0.674	**0.829**

Regarding both settings, the MCS_{cc} test yields the best rejection rates for the vast majority of test settings. To be precise, the MCS_{cc} test shows similar or better results compared to the competing tests in 157 out of 180 parameter combinations for setting 1 and 116 out of 150 parameter combinations for setting 2. With respect to setting 1, the LR_{cc}^{mar} test and the GMM_{cc} test achieve or exceed the rejection rates of the MCS_{cc} test in some cases in which the nominal VaR-level is set to 1%. This is especially true for the LR_{cc}^{mar} test for small samples and significance level 10%. Nevertheless, as described above, the power mostly decreases if n increases indicating asymptotic disturbances. The LR_{cc}^{wei} test does not achieve top rejection rates for any of the parameter combinations. Regarding setting 2, and parameter combinations for which the VaR-violation probability variation parameter is set to $\delta = 0.1p$, the LR_{uc}^{mar} test shows some superior results. In many cases, the rejection rates of the GMM_{cc} test show evidence of a good performance, but only in very few cases does it yield top results. For none of the reported parameter combinations does the LR_{cc}^{wei} test lead to results above the rejection rates of the remaining tests.

4.4 Empirical Application

To investigate the behavior of the new set of backtests and to illustrate their usefulness in a realistic risk management setting, we perform an empirical study using actual returns on a set of managed portfolios.

4.4.1 Data and Forecasting Scheme

We apply the new tests to a unique data set provided by a German asset manager.[16] The data set consists of 5,740 daily log-returns for each of four portfolios and covers a time period of 22 years (January 1, 1991 to December 31, 2012). While we exclude weekend days from our sample, it is not possible to easily eliminate holidays as well, because the portfolio assets are invested internationally and non-business days differ widely across

[16]Due to confidentiality reasons, the asset manager wishes to remain anonymous.

the countries in our sample. To this end, we add the returns of these days (e.g., accrued interest) to the next trading day. Table 4.14 presents summary statistics for the portfolio log-returns we use in our empirical study.

Table 4.14: Summary Statistics

Summary statistics of the portfolio data set used for the empirical application of the MCS and GMM tests. The data set consists of 5,740 log-returns for each of the four portfolios covering a period from January 1, 1991 to December 31, 2012. Mean Return p.a. and Volatility p.a. are annualized with 250 days.

Portfolio	1	2	3	4
Minimum	-2.691%	-3.086%	-3.473%	-2.805%
5% quantile	-0.651%	-0.531%	-0.657%	-0.638%
Median Return	0.016%	0.011%	0.016%	0.016%
Mean Return	0.025%	0.020%	0.026%	0.027%
95% quantile	0.657%	0.564%	0.683%	0.648%
Maximum	3.705%	2.683%	3.621%	3.745%
Volatility	0.417%	0.369%	0.426%	0.425%
Skewness	-0.133	-0.467	-0.300	0.083
Kurtosis	6.67	8.94	6.85	7.80
Mean Return p.a.	6.24%	4.95%	6.43%	6.84%
Volatility p.a.	6.59%	5.84%	6.73%	6.71%
Maximum Drawdown	-23.46%	-24.51%	-23.80%	-24.62%

The summary statistics in Table 4.14 show evidence of the usual stylized facts of returns on financial assets. In addition to having negligible (daily) mean returns, the portfolio returns exhibit signs of typical properties like negative skewed and leptokurtic asset returns indicating fat tails particularly on the downside. Nevertheless, overall portfolio risk over the complete sample period appears to be only moderate as evidenced by the estimates of the (unconditional) return series volatility with all four portfolios having significant positive annualized returns.

We calculate the one-day VaRs for each portfolio by the use of two different VaR-models. First, we choose standard historical simulation as the most widely used model in practice (see Pérignon and Smith, 2010). This concept assumes no particular distribution of the returns. The VaR is rather estimated solely based on historical returns. For each VaR-estimation, we use the value of the 1% and 5% quantile of the last 250 data points as an estimate for the portfolio's VaR. Second, we employ a GARCH(1,1) process as a parametric model to forecast the VaR using the estimated conditional variance of the

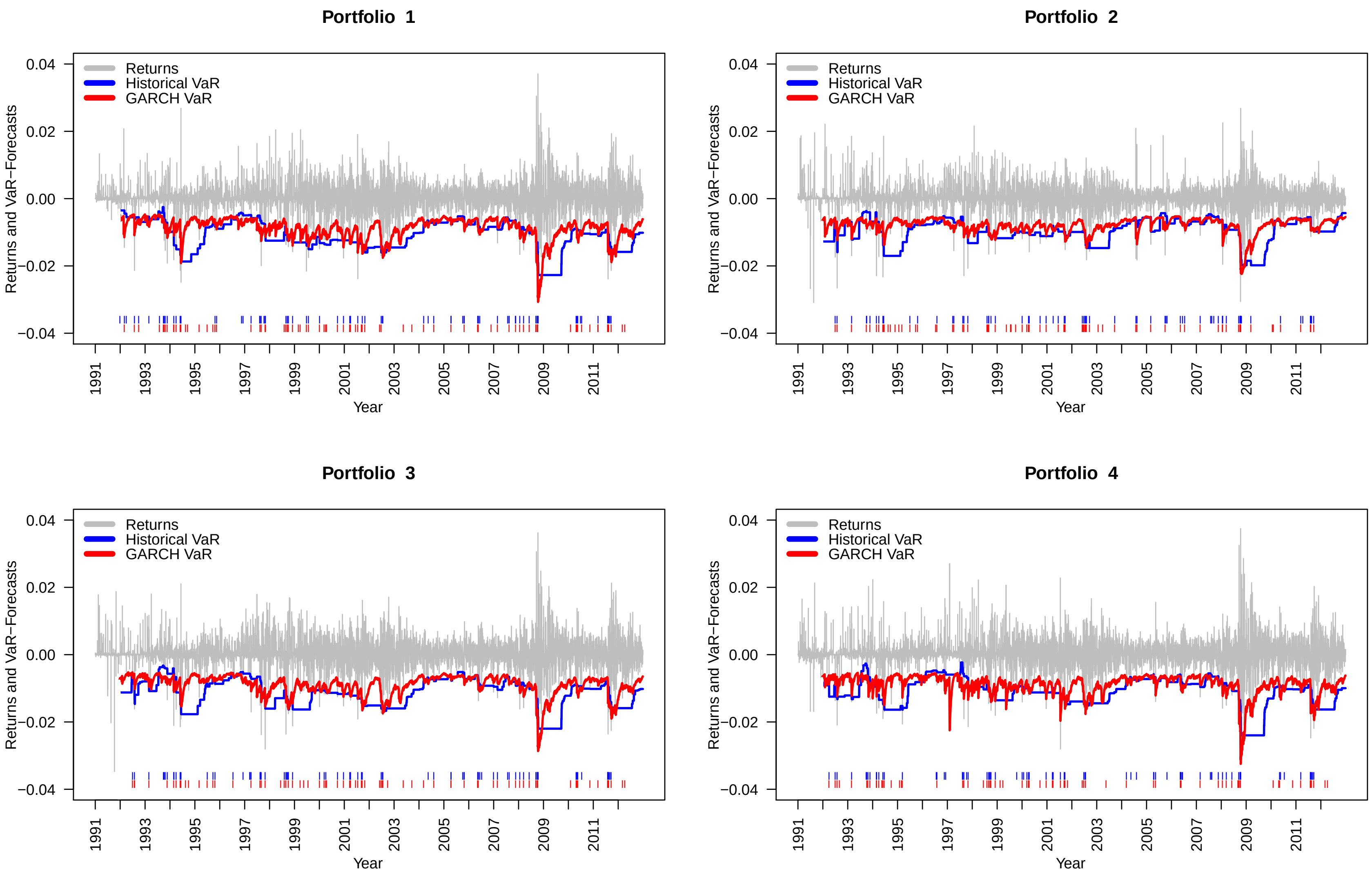
GARCH model. Compared to historical simulation, the GARCH model is more flexible because it accounts for autocorrelations in the return series' variances. We choose the simple GARCH(1,1) model rather than more sophisticated GARCH model specifications because Hansen and Lunde [2005] show that the GARCH(1,1) model is hard to beat in terms of volatility forecasting accuracy. For the sake of simplicity, we fit the GARCH parameter for each portfolio separately to the total sample of 5,740 log-returns.[17] The next-day VaR is then calculated simply by the quantile of a normal distribution with a zero mean and the standard deviation forecasted by the GARCH model on the basis of the last 250 log-returns. Figure 4.1 plots the daily portfolio returns together with the corresponding VaR-forecasts of the historical VaR and the GARCH model. In addition to the time-varying volatility of the returns, the charts illustrate the differences in the forecasts of the unconditional historical VaR approach and the conditional GARCH model. However, it can be seen for both models that the VaR-violations cluster to some degree during certain subsamples.

After calculating the VaR-violation sequence $I_t(p)$, we validate the VaR-estimation by making use of the new set of MCS tests to compute p-values and check the uc, i.i.d. and cc hypotheses separately. With respect to the MCS_{cc} test, we use the two-tailed uc component and opt for a weighting factor of $a = 0.5$. For comparison purposes, we additionally present p-values of the uc, i.i.d., and cc version of the GMM test as the results of our simulation study indicate that the set of GMM tests is a suitable benchmark. Moreover, we repeat our analysis for four separate time periods. For the first time period, we include 5,740 log-returns of the whole available time span (January 1, 1991 - December 31, 2012). We then focus on the volatility shift from the highly volatile bear market at the later stage of the dotcom-bubble burst (250 log-returns from April 16, 2002 to March 31, 2003) to the early stage of the subsequent calm bull market (250 log-returns from April 1, 2003 to March 15, 2004). Additionally, we apply the tests to the 500 log-returns of the combination of the latter two periods from April 16, 2002 to March 15, 2004.

[17] Of course, this procedure does not comply to the principle of out-of-sample forecasting. Nevertheless, as we focus on the performance of the backtests, the issue of optimally fitting the GARCH parameters to the data is not relevant for the purpose of this study.

Figure 4.1: Returns, VaR-Forecasts, and VaR-Violations

The figure presents returns, VaR-forecasts, and VaR-violations for the four portfolios considering a VaR-level of 1%. VaR-forecasts are plotted with lines whereas the dashes at the bottom of the charts mark the days on which a VaR-violation occurs.



4.4.2 Results

The results of applying the backtests to the total period data set are shown in Table 4.15.

Table 4.15: Empirical Application - Total Period

The table contains Violation Ratios (i.e., VaR-violation frequency divided by the number of VaR-forecasts) of the total period consisting of 5,490 VaR-forecasts for each portfolio (17.12.1991 to 31.12.2012). In addition, the table contains p-values for the unconditional coverage tests MCS_{uc}^{lt} (lower tail), MCS_{uc}^{ut} (upper tail), MCS_{uc}^{tt} (two tailed), and GMM_{uc}, for the i.i.d. tests MCS_{iid} and GMM_{iid}, and for the conditional coverage tests MCS_{cc} and GMM_{cc}. The extensions *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

VaR Model	Historical VaR				GARCH approach			
Portfolio	1	2	3	4	1	2	3	4
Panel A: 5% VaR								
$Viol. Ratio$	5.43%	5.37%	5.50%	5.66%	4.37%	4.12%	4.54%	4.12%
MCS_{uc}^{lt}	0.923	0.901	0.956	0.987	0.013**	0.001***	0.054*	0.001***
MCS_{uc}^{ut}	0.077*	0.099*	0.044**	0.013**	0.987	0.999	0.946	0.999
MCS_{uc}^{tt}	0.155	0.197	0.088*	0.025**	0.025**	0.002***	0.108	0.002***
GMM_{uc}	0.133	0.221	0.091*	0.035**	0.025**	0.002***	0.107	0.002***
MCS_{iid}	0.000***	0.000***	0.000***	0.000***	0.014**	0.000***	0.001***	0.001***
GMM_{iid}	0.000***	0.000***	0.000***	0.000***	0.005***	0.001***	0.005***	0.004***
MCS_{cc}	0.000***	0.000***	0.000***	0.000***	0.002***	0.000***	0.000***	0.000***
GMM_{cc}	0.001***	0.000***	0.001***	0.001***	0.002***	0.001***	0.004***	0.001***
Panel B: 1% VaR								
$Viol. Ratio$	1.20%	1.22%	1.35%	1.35%	1.53%	1.48%	1.46%	1.33%
MCS_{uc}^{lt}	0.924	0.949	0.994	0.992	1.000	1.000	0.999	0.994
MCS_{uc}^{ut}	0.076*	0.052*	0.006***	0.008***	0.000***	0.001***	0.001***	0.006***
MCS_{uc}^{tt}	0.151	0.103	0.012***	0.016***	0.000***	0.001***	0.001***	0.013**
GMM_{uc}	0.124	0.114	0.020**	0.020**	0.002***	0.003***	0.005***	0.026**
MCS_{iid}	0.022**	0.007***	0.003***	0.003***	0.130	0.204	0.578	0.057*
GMM_{iid}	0.022**	0.007***	0.004***	0.006***	0.439	0.012**	0.311	0.019**
MCS_{cc}	0.019**	0.006***	0.001***	0.001***	0.004***	0.009***	0.026**	0.011**
GMM_{cc}	0.034**	0.017**	0.008***	0.006***	0.019**	0.019**	0.051*	0.022**

First, we compute the VaR-violation ratios of each portfolio for each VaR-forecasting method and the nominal VaR levels of 5% and 1%. We define the VaR-violation ratio as the VaR-violation frequency divided by the number of VaR-forecasts. Both the historical VaR and the GARCH approach lead to VaR-violation ratios which deviate from the nominal VaR level of 5% and 1% to some degree. The p-values of the one-tailed MCS_{uc}^{lt} and MCS_{uc}^{ut} tests indicate that each of these deviations are statistically significant. However, some of the p-values yielded by the two-tailed MCS_{uc}^{tt} and the GMM_{uc} tests remain above the 10% significance level.

The MCS_{iid} test and the GMM_{iid} test reject the i.i.d. hypothesis for the violation sequences generated by the historical simulation VaR-model for the 5% and 1% VaR level. We expect a large sample like ours that consists of 22 years of data to suffer significantly from the stylized facts of financial returns (i.e., series of absolute or squared returns show profound serial correlation, volatility appears to vary over time, and extreme returns appear in clusters). Consequently, an inflexible and unconditional VaR-model like historical simulation should lead to non-i.i.d. VaR-violations. However, the p-values for the more flexible GARCH model suggest clustered VaR-violations only for the 5% VaR level. These findings are confirmed by significant p-values obtained for the MCS_{cc} and GMM_{cc} tests.

The test results for the bear and the bull market as well as for the combination of both market phases are reported in Table 4.16. We restrict the presentation of the results to the VaR level of 5%, because it vividly illustrates the effects of a shift from a bear to a bull market. The differences in the VaR-violation ratios between the bear and the bull market are significant. On average, for the bear market the historical VaR approach yields VaR-violations in 8.45% of the days whereas for the bull market the ratio amounts to 1.70%. Consequently, for both the bear and the bull market, the p-values of the relevant one-sided and the two-sided MCS_{uc} tests as well as the GMM_{uc} test are statistically significant in the vast majority of cases. With respect to the combined 500 trading days sample, the underestimated VaR of the bear market and the overestimated VaR of the bull market compensate each other and lead to an average VaR-violation ratio of 5.08%. Because this is very close to the nominal VaR level of 5%, all applied backtests imply a correct uc. This result underpins our redefinition of the uc property, because the backtests show no significant p-values although the probability for a VaR-violation is not equal to the nominal level p for all days t.

The i.i.d. tests show a remarkable behavior. Because the GARCH model accounts for autocorrelated volatility, it can be assumed that the VaR-violations are less dependent compared to the VaRs estimated with historical simulation. Consequently, the p-values regarding the GARCH model during the bear market and the bull market separately are statistically significant in only four out of twelve cases.

Table 4.16: Empirical Application - Bear, Bull, and Bear + Bull Market

For each portfolio, the table contains Violation Ratios (i.e., number of VaR-violations divided by VaR-forecasts) of the bear market period (250 VaR-forecasts from 16.04.2002 to 31.03.2003), the bull market period (250 VaR-forecasts from 01.04.2003 to 15.03.2004), and the combination of the bear and bull market period (500 VaR-forecasts from 16.04.2002 to 15.03.2004). The VaR level is set to 5%. In addition, the table contains the corresponding p-values for the unconditional coverage tests MCS_{uc}^{lt} (lower tail), MCS_{uc}^{ut} (upper tail), MCS_{uc}^{tt} (two tailed), and GMM_{uc}, for the i.i.d. tests MCS_{iid} and GMM_{iid}, and for the conditional coverage tests MCS_{cc} and GMM_{cc}. The extensions *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

VaR Model	Historical VaR				GARCH(1,1)			
Portfolio	1	2	3	4	1	2	3	4
Panel A: Bear Market / 5% VaR								
$Viol. Ratio$	7.60%	7.60%	8.40%	9.20%	8.80%	8.00%	9.20%	8.80%
MCS_{uc}^{lt}	0.967	0.964	0.988	0.998	0.995	0.982	0.998	0.995
MCS_{uc}^{ut}	0.033**	0.036**	0.012**	0.002***	0.005***	0.018**	0.002***	0.005***
MCS_{uc}^{tt}	0.065*	0.073*	0.023**	0.004***	0.011**	0.036**	0.003***	0.010**
GMM_{uc}	0.131	0.120	0.050*	0.019**	0.040**	0.064*	0.017**	0.045**
MCS_{iid}	0.000***	0.008***	0.025**	0.010**	0.033**	0.047**	0.207	0.051*
GMM_{iid}	0.000***	0.005***	0.042**	0.010**	0.078*	0.197	0.819	0.256
MCS_{cc}	0.001***	0.005***	0.007***	0.003***	0.006***	0.015**	0.015**	0.007***
GMM_{cc}	0.014**	0.037**	0.044**	0.020**	0.039**	0.090*	0.048**	0.048**
Panel B: Bull Market / 5% VaR								
$Viol. Ratio$	1.20%	2.00%	1.20%	1.60%	1.60%	2.80%	1.60%	1.60%
MCS_{uc}^{lt}	0.001***	0.007***	0.001***	0.004***	0.002***	0.046**	0.004***	0.004***
MCS_{uc}^{ut}	0.999	0.993	0.999	0.996	0.998	0.954	0.996	0.996
MCS_{uc}^{tt}	0.003***	0.013**	0.001***	0.007***	0.004***	0.093*	0.007***	0.008***
GMM_{uc}	0.001***	0.007***	0.001***	0.001***	0.004***	0.080*	0.003***	0.004***
MCS_{iid}	0.424	0.545	0.428	0.204	0.259	0.540	0.255	0.258
GMM_{iid}	0.657	0.634	0.659	0.787	0.770	0.643	0.757	0.761
MCS_{cc}	0.044**	0.095*	0.044**	0.025**	0.040**	0.237	0.036**	0.035**
GMM_{cc}	0.003***	0.013**	0.003***	0.003***	0.003***	0.193	0.005***	0.004***
Panel C: Bear + Bull Market / 5% VaR								
$Viol. Ratio$	4.40%	4.80%	4.80%	5.40%	5.20%	5.40%	5.40%	5.20%
MCS_{uc}^{lt}	0.269	0.404	0.457	0.687	0.580	0.666	0.684	0.591
MCS_{uc}^{ut}	0.731	0.596	0.543	0.313	0.420	0.334	0.316	0.409
MCS_{uc}^{tt}	0.538	0.807	0.914	0.627	0.841	0.668	0.633	0.818
GMM_{uc}	0.666	0.932	0.923	0.681	0.702	0.542	0.542	0.684
MCS_{iid}	0.003***	0.034**	0.004***	0.001***	0.003***	0.116	0.005***	0.003***
GMM_{iid}	0.001***	0.007***	0.003***	0.001***	0.003***	0.160	0.003***	0.005***
MCS_{cc}	0.003***	0.112	0.009***	0.004***	0.012**	0.306	0.011**	0.013**
GMM_{cc}	0.007***	0.030**	0.014**	0.010**	0.014**	0.374	0.024**	0.018**

These results are contrasted by the p-values for the sample where the bear and bull market are combined. Here, the i.i.d. tests attain p-values below the 1% level of significance in six out of eight cases. This result could be due to the large shift in the VaR-violation ratio. Only the p-values for portfolio two reveal no significance which can be explained by a smaller drop of the violation ratio from the bear to the bull market compared to the remaining portfolios. This outcome demonstrates the necessity of testing the independence as well as the identical distribution hypothesis using a powerful test. Finally, the results of the cc tests reflect the implications of the corresponding uc and i.i.d. tests.

4.5 Conclusion

Comparatively little attention has been paid in the literature to the development of proper tools for backtesting VaR-forecasts. This paper provides three main contributions to the issue of backtesting the performance of VaR-models. First, we extend the discussion of the desirable properties of violations originating from a correct VaR-model and restate the uc property of a VaR-violation process. Furthermore, we stress the need to require the VaR-violations to be identically distributed to adequately backtest models across different market phases. Second, we propose a new set of backtests that test VaR-violation processes for uc, the i.i.d. property as well as cc. Compared to existing standard approaches, these backtests contain new desirable features like one-tailed testing for uc and a test for cc that allows for different weightings of the uc and i.i.d. parts. The new backtesting procedures are based on i.i.d. Bernoulli random variables obtained by Monte Carlo simulation techniques and are very intuitive. Third, we perform a simulation study using generated VaR-violation samples that specifically violate the uc, i.i.d., and cc property to different controllable degrees. Compared to existing classical and state-of-the-art backtests, the new backtests outperform these benchmarks in several distinct settings. In addition, we use the new backtests in an empirical application study. We apply the backtests to return samples of calm boom and highly volatile bust cycles. The obtained results demonstrate the need for a backtest that accounts for non-identically distributed VaR-violations and, moreover, support the reformulation of the uc hypotheses.

As a natural extension of our work, one could think of multivariate versions of our newly proposed backtests which would need to take into account possible correlations in VaR-violations across assets and time. As this issue lies beyond the scope of the present work, we will address it in our future research.

Kapitel 5

Estimation Window Strategies for Value at Risk Forecasting

5.1 Introduction

Today, the Value at Risk (VaR) is the de facto standard tool for risk management in the financial services industry. Because of the high relevance of this risk measure, numerous parametric, semi-parametric, and non-parametric models for VaR-estimation and forecasting have been developed over the last two decades. All of these approaches estimate VaR-forecasts directly or indirectly on the basis of a sample of historical asset-returns. This leads to the question of how to determine the appropriate in-sample size for generating out-of-sample VaR-forecasts. This is not a trivial task, because it is well known that the volatility of asset returns varies over time and is subject to occasional structural breaks.[1] In particular, structural shifts may lead to significant biases in the parameter estimation of forecasting models. The aim of this study is to investigate whether different estimation window strategies lead to significant divergences in the VaR and ES forecasting accuracy. In addition, we analyze whether more sophisticated estimation window strategies outperform simple strategies like rolling or expanding windows. To this end, we analyze the characteristics of a comprehensive set of strategies in the context of various parametric, semi-parametric, and non-parametric VaR models.

Several studies deal with the issue of misspecified forecasting models. Among others, Hillebrand [2005] show that neglecting changes in financial time series can yield to overestimated persistence in conditional volatility parameters. Furthermore,

[1]For a comprehensive overview of structural breaks in financial time series see Andreou and Ghysels [2009]

Pesaran and Timmermann [2004] and Pesaran et al. [2006] find out that structural breaks can affect the accuracy of financial time series forecasts. Andreou and Ghysels [2003] states that failure to recognize the presence of structural breaks can have devastating and costly effects on financial risk management tools like VaR and Expected Shortfall (ES).

Estimation window strategies used in the finance literature are often limited to expanding windows, which include all available data preceding the forecasting time point, and rolling windows characterized by a moving interval of a more or less arbitrary fixed length. By nature, these strategies do not explicitly account for any changes in the in-sample data. To this end, Pesaran and Timmermann [2007] discuss the usage of the date for which a structural break is identified as a starting point for the determination of an estimation window. The appealing idea behind this strategy is to exclude historical data from the estimation sample which significantly differ from the more recent data. However, the employment of structural break tests for estimation window determination suffers from two serious drawbacks. First, the limited number of observations of a reduced sample size leads to an increased variance of the parameter estimates and hence to forecast errors. To this end, Pesaran and Timmermann [2007] analyze a trade-off between biased estimates and forecast error variance. They concluded that it can be useful to use pre-break data for parameter estimation. Second, Pesaran and Timmermann [2007] emphasize that the overall outcome crucially depends on how well the location of the break point is estimated by a statistical test. To overcome the difficulties of specifying an appropriate estimation window size, Pesaran and Timmermann [2007] and Clark and McCracken [2009] propose to combine forecasts from models estimated on different samples. The strategy of combining forecasts is based on the seminal paper of Bates and Granger [1969]. Timmermann [2006] provide a comprehensive overview over the forecast combination literature and discusses the pros and cons of this approach. In brief, he argues that pooling forecasts can lead to diversification gains which make it attractive to employ combinations rather than rely on a single model.

Compared to the vast amount of papers on volatility and VaR models, the finance literature provides relatively few comparisons of estimation window strategies in this context.

Focusing on volatility forecasting ability, Brownlees et al. [2012] compared several conditional volatility models estimated on different estimation window lengths. The results of this study indicate that the expanding window strategy gives a better forecasting accuracy compared to relatively long rolling windows of eight and four years. With respect to out-of-sample VaR-forecasting performance, Kuester et al. [2006] analyze different classical and sophisticated VaR models considering rolling windows of 1,000, 500, and 250 days. They stated that although less parameterized models seem to benefit from smaller sample sizes, general conclusions about model performance as the window length decreases cannot be made. Halbleib and Pohlmeier [2012] present empirical evidence from assessing the out-of-sample performance and robustness of VaR before and during the recent financial crisis with respect to the choice of sampling windows. They show that using a two year sampling window is sufficient before the crisis, while during the crisis the estimation window should also incorporate information on past extreme events. A deeper insight into estimation window strategies is provided by Rapach and Strauss [2008] and Rapach et al. [2008]. They investigate the forecasting accuracy of conditional volatility models accommodating structural breaks as well as forecast combinations of different volatility models or estimation windows. They find that combining forecasts often yields more accurate forecasts compared to a simple expanding window benchmark approach. The results for models estimated on window sizes determined by structural breaks are ambiguous. Accommodating structural breaks in exchange rate return variance often improves volatility forecasts while the same strategy for stock indices returns does not.

This paper is most related to the works of Rapach and Strauss [2008] and Rapach et al. [2008] but contributes to the literature by several important improvements and extensions. First, to the best of our knowledge, this is the only study that analyzes the impact of different estimation window strategies including structural breaks and forecast combinations explicitly on forecasting common risk measures like VaR and the ES rather than volatility. These strategies are investigated in a context of a comprehensive model set including different parametric as well as semi-parametric and non-parametric forecasting approaches. For parametric models, predicting VaR and ES is formally related to forecasting variance (see Section 5.2.1). Nevertheless, while the evaluation of variance forecasting is based on the entire loss distribution, the accuracy of VaR and ES forecasts depends on the specific shape characteristics of its lower left tail. Therefore, results from evaluating variance forecasting performance cannot generally be used to assess VaR and ES forecasting capability. In addition, semi- and non-parametric VaR and ES models are not or only indirectly related to variance forecasting. To measure VaR-forecasting accuracy we use powerful statistical tests of the unconditional coverage and i.i.d. properties of VaR-violations proposed by Ziggel et al. [2013], while the ES forecasts are evaluated by the test of McNeil and Frey [2000]. In addition, we compare the estimation window approaches by employing the conditional predictive ability (CPA) test of Giacomini and White [2006]. Second, the results of this study are very robust. The result evaluation of all related papers mentioned above suffer from the serious drawback that they are obtained by applying different approaches to just a few or in most cases one fixed out-of-sample period. But it is very likely that different out-of-sample periods lead to different results and conclusions. To this end, we perform the risk forecasts to a large number of different randomly selected samples of stock returns. Third, for individual forecasts and forecast combinations we use relatively small rolling estimation window sizes of 125, 250, 500, and 1,000 trading days which are very frequently used in theory and particularly in practice.[2] This is contrary to Rapach and Strauss [2008] and Rapach et al. [2008] who determine rolling estimation samples from about three to more than ten years.

The results of our empirical study reveal that the selection of the estimation window strategy leads to significant performance differences. Each of the evaluated estimation window strategies has its advantages and disadvantages. However, the usage of forecast combinations seems to be the preferable estimation window strategy, because it shows good results in most analyzed settings and backtests. The results reveal that applying combinations leads to more conservative VaR-forecasts and reduces the undesired occurrence of clustered VaR-violations on average.

The remainder of the paper proceeds as follows. In Section 5.2, we briefly review

[2] Basel Committee on Banking Supervision [2005] requires that the choice of the historical observation period for calculating VaR is constrained to a minimum length of one year.

the VaR-forecasting models and, additionally, present the estimation window strategies in detail. Section 5.3 describes the data set and the backtests used in this empirical study. The results are discussed in Section 5.4. Section 5.5 concludes.

5.2 VaR-Models and Estimation Window Strategies

We apply different estimation window strategies to a set of parametric, semi-parametric, and non-parametric VaR-forecasting models. In this section, we quickly describe the theoretical background of the VaR models and explain the application of the estimation window strategies.

5.2.1 A Short Review of VaR-Forecasting Models

The VaR-forecasts provided by each of the models described in the following are estimated on the basis of a series of log-returns $r_t = \log(P_t/P_{t-1})$, where P_t denotes the quote of an asset at time $t = (0, 1, \ldots, T)$. Following the Market Risk Amendment proposed by the Basel Committee on Banking Supervision [2005], we estimate 1-day ahead ($k = 1$) and 10-day ahead ($k = 10$) VaR-forecasts. The $100(1-p)\%$ confidence level of the VaR is set to $p = 0.05$ and $p = 0.01$. As the Value-at-Risk does not fulfill the requirements of a coherent risk measure (see Artzner et al., 1999), we also estimate the ES which is generally defined as $ES_p(X) = E[X|X \leq VaR_p(X)]$.[3]

- *Normal Distribution*

  We start with the simplest parametric model based on the assumption of normal distributed returns. We include this model because Starica et al. [2005] and Rapach and Strauss [2008] find out that a simple approach based on the average of the squared returns often achieves good results compared to conditional models if this model is estimated on a relatively small moving window. They argue

[3]The terms ES and Conditional-Value-at-Risk (CVaR) are often used synonymously, although the latter is defined as $CVaR_p(X) = E[X|X < VaR_p(X)]$. Note that if the loss distribution is continuous, the ES is equal to the CVaR. The ES differs from the CVaR in the case of discontinuities in the loss distribution. In general, the relationship is defined by $CVaR \leq ES \leq VaR$. For further details see Acerbi and Tasche [2002].

that this moving average model is a convenient way of capturing a conditionally homoskedastic process with relatively frequent breaks. Let σ_t denote the standard deviation of the log-returns r_t. If the simplistic but popular assumption is made that asset returns follow a normal distribution, the 1-day VaR-forecasts are given by $VaR^p_{t+1} = \sigma_t z(p)$, where z is the quantile of the standard normal distribution. In line with the stylized facts of asset returns (see McNeil et al., 2005, p. 117), we assume the expected daily returns to be zero while the empirical standard deviation σ_t is estimated on the log-returns of different estimation windows. The k-day ahead forecasts for forecasting periods $k > 1$ are computed by $VaR_{t+k} = VaR_{t+1} \cdot \sqrt{k}$. The ES is calculated by $ES^p_{t+1} = \sigma_t \frac{\phi(z(p))}{p}$ for the 1-day ahead forecast and by $ES^p_{t+k} = ES^p_{t+1} \cdot \sqrt{k}$ for the k-day ahead forecast with $k > 1$, where ϕ denotes the density of the distribution.

- *Exponentially Weighted Moving Average*

  The very popular Exponential Weighted Moving Average (EWMA) approach, also used in the *RiskMetrics* framework, explicitly accounts for time-varying and auto-correlated volatility of asset returns. Considering an estimation window including m log-returns from time $t - m + 1$ to time t, the EWMA model estimates the next-day variance by $\hat{\sigma}^2_{t+1} = (1 - \lambda) \sum_{i=0}^{m+1} \lambda^i r^2_{t-i} = \lambda \hat{\sigma}^2_t + (1 - \lambda) r_t$. The decay factor λ is usually set to 0.94 and 0.97 for 1-day and one-month ahead volatility forecasts, respectively (see RiskMetrics Group, 1996). However, to investigate the impact of different estimation window strategies, we apply decay factors determined by $\lambda = e^{ln(0.001)/m}$. The 1-day VaR-forecasts are given by $VaR_{t+1} = \hat{\sigma}_{t+1} z(p)$ while the k-day ahead forecasts for forecasting periods $k > 1$ are computed by $VaR_{t+k} = VaR_{t+1} \cdot \sqrt{k}$. The ES is calculated by $ES^p_{t+1} = \hat{\sigma}_{t+1} \frac{\phi(z(p))}{p}$ for the 1-day ahead forecast and by $ES^p_{t+k} = ES^p_{t+1} \cdot \sqrt{k}$ for the k-day ahead forecast with $k > 1$, where ϕ denotes the density of the distribution.

- *GARCH(1,1)*

  The GARCH model of Bollerslev [1986] and its variants are very frequently used to forecast the volatility of asset returns. In contrast to the EWMA model, GARCH models accommodate the fact that the long-run average variance of asset returns tends to be relatively stable over time. Hansen and Lunde [2005] compare the volatility forecasting accuracy of a large range of different GARCH models. They find out that the simple GARCH(1,1) is not significantly outperformed by any of its more complex counterparts. Therefore in this empirical study we select the GARCH(1,1) model. The log-return of an asset is modelled by $r_t = \dot{\sigma}_t \epsilon_t$, where $\epsilon_t \sim i.i.d.\ N(0,1)$. Then the 1-day forecast of the asset return variance is denoted by $\dot{\sigma}^2_{t+1} = \omega + \alpha r^2_t + \beta \dot{\sigma}^2_t$, where $\alpha + \beta < 1$ and $\omega = \dot{\sigma}^2(1 - \alpha - \beta)$. The unconditional, or long-run average, variance is denoted by $\dot{\sigma}^2$. The k-day ahead forecasts of the variance for $k > 1$ are given by $\dot{\sigma}^2_{t+k} = k\dot{\sigma}^2 + \sum_{j=1}^{k}(\alpha + \beta)^{j-1}(\dot{\sigma}^2_{t+1} - \dot{\sigma}^2)$. The k-day ahead VaR-forecasts are determined by $VaR_{t+k} = \dot{\sigma}_{t+k} z(p)$. The k-day ahead ES forecast is given by $ES^p_{t+k} = \dot{\sigma}_{t+k}\frac{\phi(z(p))}{p}$, where ϕ denotes the density of the distribution.

- *GJR-GARCH(1,1)*

  The GARCH(1,1) model described above is symmetric in that sense that it does not distinguish between positive and negative return shocks. However, numerous studies evidenced that asset returns and conditional volatility are negatively correlated (for an overview see Bekaert and Wu, 2000). To accommodate this typical characteristic, we employ the GJR-GARCH(1,1) model of Glosten et al. [1993]. This variance forecasting approach modifies the classical GARCH(1,1) model in a way that not only the size but also the direction of a shock ϵ_t has an impact to the volatility forecast. The 1-day ahead volatility forecast provided by the GJR-GARCH(1,1) model is given by $\tilde{\sigma}^2_{t+1} = \omega + (\alpha + \gamma I_t) r^2_t + \beta \tilde{\sigma}^2_t$, where I_t is an indicator variable equaling one if $\epsilon_t < 0$. The computation of k-day ahead volatilities and forecasts of VaR and ES are identical to the GARCH(1,1) model.

- *Historical Simulation*

  As a non-parametric VaR model, the Historical Simulation (HS) approach assumes no particular distribution of the returns. Due to its simplicity it is easy to implement and very frequently used in practice. In a survey of international banks' VaR disclosures, Pérignon and Smith [2010] find out that 73 percent of these banks use HS for VaR-forecasting purposes. Therefore, the impact of different estimation window strategies to the VaR-forecasting accuracy of HS is of high interest particularly for practitioners. For each 1-day ahead VaR-forecast, we use the last m log-returns and arrange it in increasing order. Depending on the desired VaR confidence level, the corresponding percentile of the historical returns is used as VaR-forecast by $VaR^p_{t+1} = Percentile\left\{\{r_{t-i}\}^{m-1}_{i=0},\ 100p\right\}$. The k-day ahead VaR-forecasts for $k > 1$ are calculated by the same procedure. Here, VaR_{t+k} based on the percentile of $m-k+1$ returns calculated by $r_{t+1:k} = \sum^k_{j=1} r_{t+j}$. The k-day ahead ES forecast is determined by $ES^p_{t+k} = \frac{1}{p\cdot(m-k+1)} \cdot \sum^{m-k+1}_{i=1} r_{t+1:k} \cdot \mathbf{1}(r_{t+1:k} \leq VaR^p_{t+k})$, where $\mathbf{1}(\cdot)$ denotes the indicator function returning a 1 if a loss exceeds the VaR, and zero otherwise.

- *Filtered Historical Simulation*

  The non-parametric HS model has the advantage of not requiring any assumptions concerning the distribution of the asset returns but fails to accommodate conditional time-varying volatility. Exactly the opposite holds true for parametric GARCH models. The semi-parametric Filtered Historical Simulation (FHS) model of Barone-Adesi et al. [1999] combines the benefits of both models. We estimate the 1-day ahead volatility forecast $\bar{\sigma}_{t+1}$ using the GARCH(1,1) model as described above. For calculating VaR and ES forecasts, we follow Christoffersen [2009]. By multiplying the volatility forecast by the percentile of the standardized residuals we calculate the 1-day ahead VaR-forecast as $VaR_{t+1} = \bar{\sigma}_{t+1} Percentile\left\{\{\epsilon_{t-i}\}^{m-1}_{i=0},\ 100p\right\}$. To determine the k-step ahead VaR-forecast, we draw k random numbers from a discrete uniform distribution from 1 to m. Each drawn number from this distribution determines a historical standard residual $\epsilon_{i,k}$ obtained by estimating the GARCH(1,1) model on the log-returns of the estimation window. The return for the first day of the k-day holding period is calculated by $\bar{r}_{t+1} = \bar{\sigma}_{t+1}\epsilon_{1,k}$. Then, the return $\bar{r}_{t+1}$ is used to update the volatility for day two of the holding period $\bar{\sigma}_{t+2}$. The return of day two $\bar{r}_{t+2}$ is then again given by the multiplication of the updated volatility and the second drawn standardized residual. This process is repeated until a return sequence of length k is constructed. This return sequence sums up to a hypothetical future return $\bar{r}_{t+1:k} = \sum_{j=1}^{k} \bar{r}_{t+j}$. By repeating this procedure $L = 10000$ times, we obtain 10,000 hypothetical future returns $\bar{r}_{l,t+1:k}$. Again, the k-step ahead VaR-forecast is then given by $VaR^p_{t+k} = Percentile\left\{\{\bar{r}_{l,t+1:k}\}_{l=1}^{L},\ 100p\right\}$. The 1-day ahead ES forecast is determined by $ES^p_{t+1} = \bar{\sigma}_{t+1}\frac{1}{p\cdot m} \cdot \sum_{i=1}^{m} \epsilon_{t+1} \cdot \mathbf{1}(\epsilon_{t+1} \leq VaR^p_{t+1})$, where $\mathbf{1}(\cdot)$ denotes the indicator function returning a 1 if a loss exceeds the VaR, and zero otherwise. The k-day ahead ES forecast for $k > 1$ is calculated by $ES^p_{t+k} = \frac{1}{p\cdot L} \cdot \sum_{l=1}^{L} \bar{r}_{l,t+1:k} \cdot \mathbf{1}(\bar{r}_{l,t+1:k} \leq VaR^p_{t+k})$.

5.2.2 Estimation Window Strategies

In the following, we explain the different estimation window strategies used in this empirical study. As explained in detail in Section 5.3, each sample used in this empirical study for the 1-day ahead VaR-forecasts includes $T = 3000$ trading days, where the out-of-sample forecasting period comprises 2,000 trading days $t = [1001, \ldots, 3000]$. This in turn leads to a maximum in-sample size of 1,000 days for the first out-of-sample VaR-forecast for the day $t = 1001$.[4]

- Expanding Window

  The expanding window strategy includes the entire dataset of the sample available at time t. That means that the size m^{exp} of the expanding estimation window starts at 1,000 trading days for forecasting VaR for day $t = 1001$ and expands by one more observation per day. The last VaR-forecast estimated for $t = 3000$ is then based on the estimation window consisting of 2999 returns. Because the starting point of an

[4]The specifications of the total sample and the out-of-sample sizes relate to the 1-day ahead VaR-forecast period $k = 1$. Note that in case of the longer forecasting period $k > 1$, the total sample includes $3000 + k - 1$ trading days and we calculate 2,000 VaR-forecasts corresponding to the cumulated returns $\left\{\sum_{j=1}^{k} r_{1000+j}, \ldots, \sum_{j=1}^{k} r_{2999+j}\right\}$.

expanding estimation window is fixed to $t = 1$, this strategy neglects the occurrence of structural breaks which may lead to biased VaR estimations. However, using the longest available data history minimizes the forecast error variance. In line with Rapach and Strauss [2008] and Rapach et al. [2008] the expanding window strategy serves as a benchmark model in the context of this study.

- Rolling Window

  The rolling window strategy based on an estimation sample of a fixed size m^{rol}. For each new VaR-forecast estimation, the return of the day t is added to the sample while the return of the day $t - m^{rol}$ is excluded. This strategy is frequently used in finance research as well as by practitioners, because removing older observations from the sample reduces potential biases in the VaR estimation caused by structural breaks and, therefore, leads to flexible adjustments of the forecasting model to time-varying volatility. We employ rolling estimation windows of 125, 250, 500, and 1,000 trading days which covers a broad range of data history from approximately one half to four years.

- Structural Breaks

  The aim of the structural break strategy is to minimize the estimation biases resulting from significant changes in the volatility of an estimation sample. To detect structural breaks in the volatility, we perform the fluctuation test for constant variances of Wied et al. [2012a] to the daily log-returns. Basically, this test can be regarded as a one-dimensional special case of the test for a constant covariance structure of Aue et al. [2009]. Since this test is non-parametric, difficulties associated with parametric model selection, model fitting and parameter estimation are avoided. We test whether the variance $Var(X_t)$ of a sequence of random variables $(X_t,\ t = 1, \dots, T)$ is constant over time. In detail, we test the null hypothesis

  $H_0 :\ Var(X_t) = \sigma^2 \forall t \in \{1, \dots, T\}$

  against the alternative

  $H_1 :\ \exists t \in \{1, \dots, T-1\} : Var(X_t) \neq Var(X_t + 1)$

for a constant σ^2. The test statistic is given by

$$Q_T(X) = \underset{1\leq j\leq T}{max} \left| \hat{D}\frac{j}{\sqrt{T}} \left([Var\ X]_j - [Var\ X]_T\right)\right| \to \underset{z\in[0,1]}{sup} |B(z)|, \tag{5.1}$$

where $B(z)$ is a one-dimensional Brownian Bridge. The scalar $\hat{D}$ is needed for the asymptotic null distribution and mainly captures the long-run-dependence and the fluctuations resulting from estimating the expected value. The fluctuation of the empirical variances is measured by $\underset{1\leq j\leq T}{max} |([Var\ X]_j - [Var\ X]_T)|$. The weighting factor $\frac{j}{\sqrt{T}}$ scales down deviations for small j because the $[Var\ X]_j$ are more volatile at the beginning of the sample. For more formal details see Appendix B and Wied et al. [2012a]. To estimate the point of time where a change of the variance occurs, we employ a procedure based on Galeano and Wied [2014]. Within the total sample including all observations preceding the forecast day $t+1$ we identify the data point where the test statistic takes its maximum. If this maximum is equal to or above the critical value, the null of a constant variance is rejected.[5] The location of the maximum serves as a natural estimator of a so called dominating change point. At this point we split the sample into two parts and search for possible change points again in the latter part of the sample. The procedure stops if the test statistic remains below the critical value.

Basically, we use the latest date where a structural break is detected as the starting point for the estimation sample. However, Pesaran and Timmermann [2007] point out that using only the observations over the post-break period to estimate a VaR model need not be optimal in terms of forecasting performance. Although this approach yields unbiased forecast estimations, too small sample sizes could lead to increased forecast error variances. To overcome the trade-off between bias and forecast error variance, we limit the sample size to a minimum of 125 days which corresponds to the minimum rolling window size. Consequently, the size of the

[5] The limit distribution of $Q_T(X)$ is well known (see Billingsley, 2009) and its quantiles provide an asymptotic test. We follow Wied et al. [2012a] who find out that the test works well by using a critical value of 1.358 corresponding to a significance level of 5%.

estimation window using the structural break strategy can vary between 125 days and the length of the expanding window. The test for structural breaks and hence the adjustment of the length of the estimation sample is performed on a daily basis.

- Combinations

  Generally, a single dominant estimation window strategy which minimizes VaR-forecasting errors cannot be identified ex ante. Therefore, Pesaran and Timmermann [2007] and Clark and McCracken [2009] propose that it can be useful to combine VaR-forecast estimated on different sample sizes. Timmermann [2006] provide an overview of arguments for combining forecasts. One of the main arguments is that individual forecasts may be very differently affected by non-stationarities such as structural breaks. Some estimation window strategies may lead to quick adaptations while others do not, but lead to more precisely estimated parameters. As mentioned in Section 5.1, strategies based on the detection of structural breaks suffer from severe drawbacks, too. Consequently, it is possible that across periods with varying volatility, combinations of forecasts based on estimation window strategies representing different degrees of adaptability will outperform forecasts based on individual estimation window strategies. Each combination described below is computed by the simple equal-weighted average of the respective VaR-forecasts. Alternatively, Timmermann [2006] discuss several distinct techniques to determine optimal combination weights. However, he pointed out that simple combination schemes such as equal-weighting have widely been found to dominate estimations of optimal weights because optimization approaches tend to suffer from serious estimation errors. In the spirit of Rapach et al. [2008], we employ the following combinations:

  Mean All. This combination includes the total set of estimation window strategies consisting of the expanding window, the rolling windows of 125, 250, 500, and 1,000 days, and the windows determined by the structural break test.

  Mean All ex Structural Breaks. This combination excludes VaR-forecasts based on estimation windows determined by the structural break test. It is easier to implement because no statistical test for significant changes of the volatility has to be applied.

Mean Rolling Windows. This combination includes rolling windows of 125, 250, 500, and 1,000 days. Compared to the previous combinations it should be more flexible concerning estimations biased from structural breaks.

Mean Long Short. This combination averages the VaR-forecasts based on the long sized expanding windows and the short sized rolling windows of 125 days. These sample sizes correspond to very stable and very flexible VaR-forecast estimations, respectively.

Trimmed Mean All / Trimmed Mean All ex Structural Breaks / Trimmed Mean Rolling Windows. For calculating each of these combinations, the minimum and maximum individual VaR-forecast is excluded from the respective *mean all*, *mean all ex structural breaks*, and *mean rolling windows* combination.

5.3 Data and Backtesting

In the following we describe the data used in the empirical study. Furthermore, we explain the backtests which are used to evaluate and compare the performance of the distinct estimation window strategies.

5.3.1 Data

For the empirical study we compute daily log-returns by using total return quotes of companies listed on the German stock index DAX on June 30, 2013. We limit the selection to those 14 companies which provide a complete data history from January 1, 1973 to June 30, 2013 consisting of 10,370 log-returns. Zero returns caused by weekends and holidays are excluded from the data set. All quotes are obtained from *Thomson Reuters Financial Datastream*. Tabel 5.1 reports the summary statistics for each of the 14 time-series. Almost all stocks show significant positive annualized returns, but also an annua-

Table 5.1: Summary Statistics

Summary statistics of the data set used for the empirical study. The data set consists of 10,370 log-returns for each of the 14 stocks taken out of the DAX covering a period from January 1, 1973 to June 30, 2013. Mean p.a. and volatility p.a. are annualized with 250 days.

Company	Allianz	BASF	Bayer	Beiers -dorf	BMW	Commerz-Bank	Deutsche Bank
Minimum	-15.678%	-12.924%	-18.432%	-13.423%	-15.985%	-28.248%	-18.072%
5% Quantile	-2.773%	-2.352%	-2.449%	-2.292%	-2.787%	-3.120%	-2.695%
Median	0.008%	0.021%	0.021%	0.008%	0.014%	0.000%	0.011%
Mean	0.034%	0.044%	0.041%	0.043%	0.046%	-0.008%	0.021%
95% Quantile	2.758%	2.382%	2.456%	2.465%	2.924%	3.041%	2.675%
Maximum	19.273%	12.690%	32.305%	16.101%	13.518%	19.459%	22.304%
Volatility	1.892%	1.541%	1.631%	1.579%	1.842%	2.119%	1.863%
Skewness	0.245	-0.137	0.419	-0.041	-0.023	-0.174	0.135
Ex. Kurtosis	10.423	5.889	20.867	8.065	5.941	13.199	12.064
Mean p.a.	8.49%	10.98%	10.13%	10.63%	11.48%	-2.01%	5.29%
Volatility p.a.	29.92%	24.36%	25.78%	24.97%	29.12%	33.50%	29.46%
Company	**Deutsche Lufthansa**	**E.ON**	**Linde**	**Munich Re**	**RWE**	**Siemens**	**Thyssen-Krupp**
Minimum	-15.209%	-13.976%	-14.131%	-21.673%	-15.823%	-16.364%	-16.586%
5% Quantile	-3.106%	-2.349%	-2.276%	-2.835%	-2.305%	-2.543%	-2.996%
Median	0.005%	0.019%	0.010%	0.006%	0.017%	0.014%	0.017%
Mean	0.026%	0.033%	0.040%	0.044%	0.030%	0.032%	0.028%
95% Quantile	3.274%	2.362%	2.443%	3.047%	2.361%	2.546%	3.046%
Maximum	16.394%	15.886%	12.855%	16.528%	14.256%	16.601%	16.789%
Volatility	2.036%	1.567%	1.534%	1.931%	1.509%	1.688%	1.933%
Skewness	-0.033	-0.133	-0.094	-0.342	0.070	-0.135	-0.064
Ex. Kurtosis	4.332	7.353	5.925	9.959	7.666	8.768	5.234
Mean p.a.	6.47%	8.27%	10.03%	11.08%	7.43%	8.05%	7.08%
Volatility p.a.	32.19%	24.78%	24.26%	30.53%	23.86%	26.69%	30.57%

lized volatility well above 20%. In line with the stylized facts of financial time series (see McNeil et al., 2005), the excess kurtosis reveals fat tails of the log-return distributions. In addition, the majority of the time series are negatively skewed.

To obtain generalizable and robust results, we randomly select three different subsamples from each of the 14 samples under the condition that there are no more than 1,000 overlapping days with respect to two subsamples of the same series. In sum, this leads to 42 subsamples consisting of 3,000 trading days for the 1-day-ahead forecasts and 3,009 trading days for the 10-day-ahead forecasts. With respect to each of these subsamples, we generate 2,000 out-of-sample forecasts. Consequently, the results of our study are less biased than results obtained by restricting the samples to a particular period or market phase. Table 5.2 reports the subsample selection details. The table illustrates that the

Table 5.2: Subsample Selection

The table shows the 42 subsamples selected for the empirical study. From each of the 14 different samples, we select three different subsamples consisting of 3,000 trading days for the 1-day-ahead forecasts (3,009 trading days for the 10-day-ahead forecasts) under the condition that there are no more than 1,000 overlapping days. The dates refer to the 1-day-ahead forecasts. For the 10-day-ahead forecasts, nine days have to be added at the end of the subsample.

Company	Sample 1		Sample 2		Sample 3	
	from	to	from	to	from	to
Allianz	09/11/1979	04/08/1991	08/05/1986	06/17/1998	09/07/1994	07/18/2006
BASF	07/09/1976	01/07/1988	09/02/1982	05/13/1994	04/07/1997	02/04/2009
Bayer	04/27/1982	12/30/1993	10/24/1990	09/30/2002	07/09/1998	04/29/2010
Beiersdorf	05/19/1975	11/14/1986	08/27/1985	06/23/1997	01/21/1991	12/18/2002
BMW	10/01/1975	03/31/1987	11/06/1980	06/25/1992	03/10/1986	01/12/1998
Commerzbank	08/01/1991	07/02/2003	12/22/1995	10/29/2007	12/23/1999	10/06/2011
Deutsche Bank	01/14/1981	09/02/1992	01/06/1992	11/26/2003	01/22/2001	10/29/2012
Deutsche Lufthansa	01/08/1979	07/23/1990	10/29/1992	09/22/2004	02/02/1998	11/23/2009
E.ON	03/09/1979	09/21/1990	04/07/1988	03/06/2000	11/17/1994	09/25/2006
Linde	11/08/1985	09/04/1997	11/17/1994	09/25/2006	08/31/2001	06/14/2013
Munich Re	10/28/1975	04/27/1987	09/25/1981	05/26/1993	05/19/2000	02/27/2012
RWE	02/06/1979	08/21/1990	06/16/1988	05/18/2000	03/02/1998	12/21/2009
Siemens	05/18/1977	11/15/1988	09/26/1983	06/21/1995	03/19/1992	02/16/2004
ThyssenKrupp	04/25/1973	10/23/1984	03/21/1984	12/18/1995	10/30/1990	10/04/2002

subsamples cover a broad range of different market phases, for example the calm period in the mid-1990s or the highly volatile markets of the recent financial crisis.

5.3.2 Backtesting

We evaluate the VaR-forecasting accuracy depending on different estimation window strategies by applying the backtest framework recently proposed by Ziggel et al. [2013]. This new approach tests for the unconditional coverage property (uc thereafter) as well as for the property of i.i.d. VaR-violations. The set of backtests is directly based on i.i.d. Bernoulli random variables and uses Monte Carlo simulation techniques. Results from a simulation study indicate that these backtests significantly outperform competing backtests in several distinct settings. In addition to the evaluation of the VaR-forecasting accuracy for each estimation window strategy individually, we are interested in a comparison between distinct approaches. To this end, we employ the conditional predictive ability (CPA) test of Giacomini and White [2006]. This test of VaR-forecasting accuracy is derived under the assumption of data heterogeneity rather than stationarity and can be applied to the comparison between different estimation techniques and (finite) estimation windows. We follow Rapach et al. [2008] and choose the expanding window strategy as a benchmark because it uses all data available at time t and therefore minimizes the forecast error variance.

- Test of Unconditional Coverage

  We define the indicator variable $I_t(p)$ for a given VaR-forecast $VaR_{t|t-1}(p)$ as

$$I_t(p) = \begin{cases} 0, \text{ if } r_t \geq VaR_{t|t-1}(p); \\ 1, \text{ if } r_t < VaR_{t|t-1}(p), \end{cases} \tag{5.2}$$

  where $I_t(p) = 1$ indicates a VaR-violation. Considering the two-sided uc backtest of Ziggel et al. [2013], we test the null hypothesis $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} I_t(p)\right] = p$ against the alternative $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} I_t(p)\right] \neq p$. In addition, this test allows for directional testing, i.e. we can also test against the alternatives $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} I_t(p)\right] \geq p$ and $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} I_t(p)\right] \leq p$. The simple and intuitive test statistic is given by

$$MCS_{uc} = \epsilon + \sum_{t=1}^{T} I_t(p), \tag{5.3}$$

  where $\epsilon \sim 0.001 \cdot N(0,1)$ is a random variable that serves to break ties between test values and therefore guarantees that the test exactly keeps its size. For fixed T and p, critical values of the test statistic are computed via Monte Carlos simulation. We generate violation sequences by drawing $l = 10000$-times T random variables with distribution $\hat{I}_{j,t}(p) \sim$ Bern$(p),\ t = 1, ..., T,\ j = 1, ..., l$. Then, we calculate l test statistics by Equation (5.3) and compute the respective quantile of the test statistics corresponding to a significance level of 5%.

- Test of I.I.D. VaR-Violations

  If a VaR model is properly fitted, the resulting VaR-violations are independent and identically distributed (i.i.d.). The hypothesis of i.i.d. VaR-violations holds true if $\{I_t(\tilde{p})\} \overset{i.i.d.}{\sim}$ Bern$(\tilde{p}), \forall t$, where $\tilde{p}$ is an arbitrary probability. First, we define the set of points in time on which a VaR-violation occurs via $V = \{t | I_t = 1\} = (t_1, ..., t_m)$, where $m = \sum_{t=1}^{T} I_t(p)$. Following Ziggel et al. [2013], the test statistic for testing the i.i.d. hypothesis is then given by

$$MCS_{iid,m} = \epsilon + t_1^2 + (T - t_m)^2 + \sum_{i=2}^{m}(t_i - t_{i-1})^2, \quad (5.4)$$

where $\epsilon \sim 0.001 \cdot N(0, 1)$ is a random variable that serves to break ties. This backtest is based on the idea that the sum of the squared durations between two violations is minimal if the violations are exactly equally spread across the whole sample period. If the violations are clustered, this sum increases. Again, we obtain critical values by Monte Carlo simulations for fixed T and m. Similar to the uc test, we generate VaR-violation sequences by simulating $l = 10000$-times T random variables $\hat{I}_{j,t}(p)$ under the condition $\sum_{t=1}^{T} \hat{I}_t(p) = m$. Then, we calculate l test statistics by Equation (5.4) and set the 95% quantile of the test statistics as the critical value corresponding to a significance level of 5%. This critical value corresponds to the one-sided test for clustered VaR-violations.

- Expected Shortfall Backtesting

  To backtest the ES, we measure the average of the absolute deviations between the ES forecast and the realized shortfall in case of a VaR-violation. In addition, we apply the test of McNeil and Frey [2000] which evaluates the mean of the differences between the realized shortfall and the ES in the case of a VaR-violation. The average error should be zero. The backtest is a one-sided test against the alternative hypothesis that the residuals have mean greater than zero, i.e., the expected shortfall is underestimated on average.

- Conditional Predictive Ability

  For backtesting the 1-day ahead VaR-forecasts, we follow Santos et al. [2013] and assume an asymmetric linear (tick) loss function $\mathcal{L}$ of order p defined as $\mathcal{L}_p(e_{t+1}) = (p - \mathbf{1}(e_{t+1} < 0))e_{t+1}$, where $e_{t+1} = r_{t+1} - VaR_{t+1}^p$ and $\mathbf{1}(\cdot)$ is an indicator function equal to one if $e_{t+1} < 0$ and zero otherwise. The null hypothesis of equal conditional predictive ability claims that the out-of-sample loss difference between two models follows a martingale difference sequence. The test statistic is defined as

$$CPA = T\left(T^{-1}\sum_{t=1}^{T-1}\mathcal{I}_t\mathcal{LD}_{t+1}\right)'\hat{\Omega}^{-1}\left(T^{-1}\sum_{t=1}^{T-1}\mathcal{I}_t\mathcal{LD}_{t+1}\right) \tag{5.5}$$

where T is the sample size, $\mathcal{LD}$ is the loss difference between the two models, $\mathcal{I}_t = (1, \mathcal{LD}_t)$ is an instrument that helps predicting differences in forecast performance between the two models, and $\hat{\Omega}$ is a matrix that consistently estimate the variance of $\mathcal{I}_t\mathcal{LD}_{t+1}$. We reject the null hypothesis of equal unconditional predictive ability whenever $CPA > \chi^2_{T,1-p}$, where $\chi^2_{T,1-p}$ is the $(1-p)$-quantile of a χ^2_T distribution.[6]

5.4 Results

In the following, we present and discuss the results of our empirical study with respect to the uc and i.i.d. property, the ES forecast accuracy, and the conditional predictive ability.

5.4.1 Unconditional Coverage

We start with the evaluation of the uc properties of the different estimation window strategies. Tables 5.3 and 5.4 present the VaR-violation ratios, which are computed by dividing the actual number of VaR-violations by the total number of 2,000 VaR-forecasts. To evaluate the robustness of the results, the table additionally reports the standard deviation of the VaR-violation ratios across the 42 subsamples. Before analyzing the VaR-violation ratios of the different estimation window strategies in detail, it should be noted that obviously not only the application of different VaR models but also the selection of the estimation window strategy leads to significant differences in the results. Figure 5.1 illustrates the ranges between the minimum and the maximum VaR-violation ratio resulting from different estimation window strategies.

[6] For details of backtesting multi-day ahead VaR-forecasts, see Giacomini and White [2006].

Table 5.3: Unconditional Coverage - VaR-Violation Ratios - 1-Day Ahead Forecasts

For each VaR model and each estimation window strategy, the table reports the VaR-violation ratios averaged over the 42 subsamples. The VaR-violation ratio is calculated by dividing the actual number of VaR-violations by the total number of 2,000 1-day ahead VaR-forecasts. The values in brackets are the standard deviation of the VaR-violation ratio across the 42 subsamples. For each VaR model, VaR-violation ratios printed in bold are closest to the nominal VaR level.

Estimation Window Strategy	Normal Distr.		EWMA		GARCH		GJR-GARCH		HS		Filtered HS	
5%VaR												
Expanding Window	5.60%	(2.30%)	4.53%	(0.90%)	4.43%	(0.72%)	4.43%	(0.66%)	6.82%	(2.54%)	5.60%	(0.92%)
Rolling Window 125 days	4.74%	(0.55%)	**4.83%**	(0.51%)	**4.79%**	(0.46%)	5.23%	(0.38%)	6.06%	(0.38%)	6.12%	(0.28%)
Rolling Window 250 days	4.78%	(0.61%)	4.48%	(0.53%)	4.60%	(0.46%)	**4.80%**	(0.49%)	5.91%	(0.52%)	5.77%	(0.34%)
Rolling Window 500 days	**4.97%**	(0.85%)	4.34%	(0.53%)	4.55%	(0.56%)	4.63%	(0.57%)	6.10%	(0.73%)	5.58%	(0.45%)
Rolling Window 1,000 days	5.12%	(1.20%)	4.41%	(0.60%)	4.45%	(0.66%)	4.50%	(0.65%)	6.18%	(1.23%)	5.53%	(0.63%)
Structural Breaks Window	4.56%	(0.87%)	4.41%	(0.64%)	4.40%	(0.68%)	4.44%	(0.72%)	**5.54%**	(0.83%)	**5.40%**	(0.59%)
Comb. Mean All	4.58%	(0.89%)	4.27%	(0.53%)	4.39%	(0.59%)	4.48%	(0.58%)	5.59%	(0.95%)	5.41%	(0.46%)
Comb. Mean All ex SB	4.63%	(0.93%)	4.27%	(0.55%)	4.43%	(0.58%)	4.51%	(0.56%)	5.68%	(1.02%)	5.48%	(0.44%)
Comb. Mean Rolling Windows	4.58%	(0.75%)	4.34%	(0.49%)	4.45%	(0.55%)	4.58%	(0.54%)	5.60%	(0.73%)	5.51%	(0.38%)
Comb. Long Short	4.68%	(1.23%)	4.27%	(0.64%)	4.49%	(0.56%)	4.64%	(0.53%)	5.78%	(1.41%)	5.62%	(0.53%)
Comb. Trimmed Mean All	4.65%	(0.85%)	4.31%	(0.52%)	4.44%	(0.59%)	4.48%	(0.59%)	5.67%	(0.85%)	5.49%	(0.52%)
Comb. Trimmed Mean All ex SB	4.76%	(0.88%)	4.33%	(0.53%)	4.49%	(0.56%)	4.53%	(0.55%)	5.85%	(0.94%)	5.54%	(0.48%)
Comb. Trimmed Mean Rolling Windows	4.73%	(0.75%)	4.39%	(0.51%)	4.50%	(0.54%)	4.61%	(0.55%)	5.81%	(0.67%)	5.58%	(0.40%)
Average	4.80%		4.40%		4.49%		4.60%		5.89%		5.59%	
1%VaR												
Expanding Window	2.54%	(1.29%)	1.73%	(0.44%)	1.48%	(0.34%)	**1.44%**	(0.34%)	1.71%	(0.93%)	**1.14%**	(0.29%)
Rolling Window 125 days	1.83%	(0.29%)	1.66%	(0.27%)	1.70%	(0.29%)	1.98%	(0.30%)	1.99%	(0.27%)	2.02%	(0.24%)
Rolling Window 250 days	1.81%	(0.35%)	1.57%	(0.26%)	1.66%	(0.25%)	1.70%	(0.30%)	1.70%	(0.23%)	1.59%	(0.20%)
Rolling Window 500 days	1.99%	(0.41%)	1.63%	(0.26%)	1.58%	(0.29%)	1.60%	(0.29%)	1.62%	(0.29%)	1.38%	(0.19%)
Rolling Window 1,000 days	2.21%	(0.68%)	1.67%	(0.30%)	1.53%	(0.34%)	1.52%	(0.35%)	1.59%	(0.47%)	1.24%	(0.26%)
Structural Breaks Window	1.79%	(0.46%)	1.58%	(0.28%)	1.51%	(0.33%)	1.53%	(0.33%)	1.53%	(0.45%)	1.36%	(0.31%)
Comb. Mean All	1.78%	(0.46%)	1.53%	(0.27%)	**1.46%**	(0.24%)	1.47%	(0.28%)	**1.37%**	(0.39%)	1.25%	(0.20%)
Comb. Mean All ex SB	1.80%	(0.46%)	1.52%	(0.27%)	1.48%	(0.24%)	1.49%	(0.31%)	1.39%	(0.41%)	1.27%	(0.20%)
Comb. Mean Rolling Windows	**1.75%**	(0.37%)	1.52%	(0.23%)	1.52%	(0.23%)	1.54%	(0.28%)	1.44%	(0.32%)	1.34%	(0.18%)
Comb. Long Short	1.79%	(0.60%)	**1.46%**	(0.33%)	1.49%	(0.28%)	1.55%	(0.32%)	1.47%	(0.60%)	1.34%	(0.25%)
Comb. Trimmed Mean All	1.82%	(0.46%)	1.56%	(0.26%)	1.49%	(0.25%)	1.48%	(0.27%)	1.41%	(0.34%)	1.25%	(0.20%)
Comb. Trimmed Mean All ex SB	1.86%	(0.47%)	1.56%	(0.24%)	1.52%	(0.26%)	1.51%	(0.28%)	1.46%	(0.37%)	1.28%	(0.20%)
Comb. Trimmed Mean Rolling Windows	1.85%	(0.36%)	1.55%	(0.23%)	1.55%	(0.25%)	1.58%	(0.28%)	1.57%	(0.31%)	1.39%	(0.17%)
Average	1.91%		1.58%		1.54%		1.57%		1.56%		1.37%	

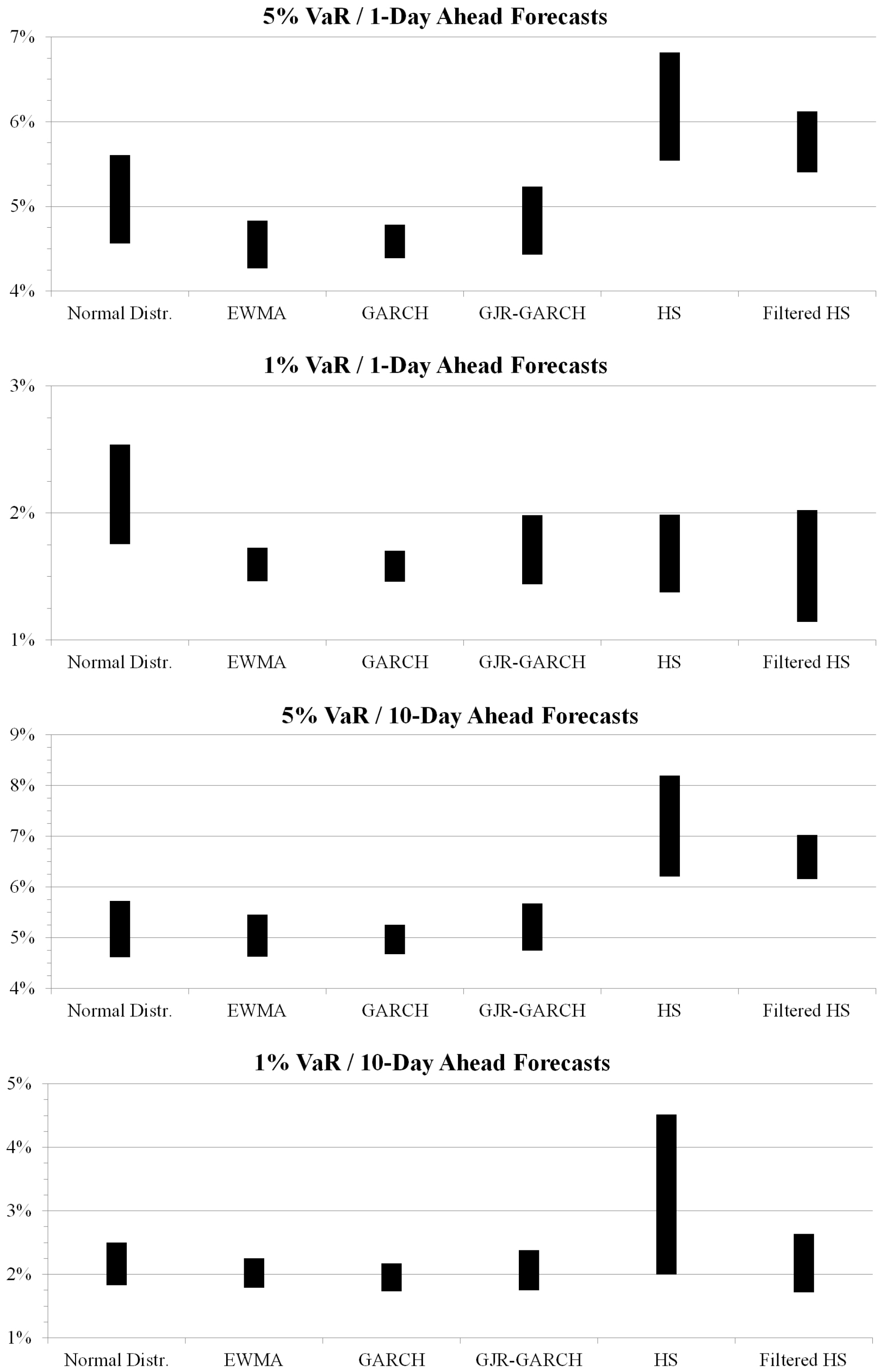
Table 5.4: Unconditional Coverage - VaR-Violation Ratios - 10-Day Ahead Forecasts

For each VaR model and each estimation window strategy, the table reports the VaR-violation ratios averaged over the 42 subsamples. The VaR-violation ratio is calculated by dividing the actual number of VaR-violations by the total number of 2,000 10-day ahead VaR-forecasts. The values in brackets are the standard deviation of the VaR-violation ratio across the 42 subsamples. For each VaR model, VaR-violation ratios printed in bold are closest to the nominal VaR level.

Estimation Window Strategy	Normal Distr.		EWMA		GARCH		GJR-GARCH		HS		Filtered HS	
5%VaR												
Expanding Window	5.72%	(2.37%)	4.72%	(1.13%)	4.90%	(1.40%)	4.86%	(1.45%)	7.16%	(2.91%)	6.23%	(1.77%)
Rolling Window 125 days	**5.06%**	(1.26%)	5.45%	(1.36%)	5.25%	(1.41%)	5.68%	(1.47%)	8.19%	(0.66%)	7.03%	(3.52%)
Rolling Window 250 days	4.80%	(1.00%)	**5.02%**	(1.30%)	**4.98%**	(1.17%)	5.19%	(1.20%)	6.94%	(0.80%)	6.50%	(2.54%)
Rolling Window 500 days	5.07%	(1.13%)	4.71%	(1.16%)	4.86%	(1.26%)	5.03%	(1.29%)	6.62%	(0.98%)	6.45%	(2.44%)
Rolling Window 1,000 days	5.24%	(1.41%)	4.62%	(1.04%)	4.79%	(1.20%)	4.84%	(1.25%)	6.53%	(1.49%)	6.38%	(1.99%)
Structural Breaks Window	4.62%	(1.18%)	4.76%	(1.09%)	4.68%	(1.16%)	4.74%	(1.27%)	6.34%	(1.29%)	6.28%	(1.95%)
Comb. Mean All	4.71%	(1.08%)	4.63%	(1.11%)	4.74%	(1.20%)	4.86%	(1.27%)	**6.20%**	(1.18%)	6.17%	(2.02%)
Comb. Mean All ex SB	4.80%	(1.14%)	4.65%	(1.13%)	4.78%	(1.22%)	4.90%	(1.29%)	6.27%	(1.24%)	**6.16%**	(2.05%)
Comb. Mean Rolling Windows	4.74%	(1.05%)	4.80%	(1.20%)	4.79%	(1.22%)	4.96%	(1.28%)	6.28%	(0.93%)	6.28%	(2.25%)
Comb. Long Short	4.86%	(1.43%)	4.69%	(1.15%)	4.94%	(1.32%)	5.06%	(1.37%)	6.70%	(1.74%)	6.29%	(2.22%)
Comb. Trimmed Mean All	4.75%	(1.07%)	4.65%	(1.10%)	4.78%	(1.23%)	4.87%	(1.30%)	6.28%	(1.09%)	6.28%	(2.03%)
Comb. Trimmed Mean All ex SB	4.90%	(1.13%)	4.67%	(1.13%)	4.85%	(1.24%)	4.95%	(1.31%)	6.45%	(1.22%)	6.31%	(2.07%)
Comb. Trimmed Mean Rolling Windows	4.85%	(1.05%)	4.86%	(1.24%)	4.88%	(1.25%)	**4.99%**	(1.29%)	6.53%	(0.87%)	6.38%	(2.29%)
Average	4.93%		4.79%		4.86%		5.00%		6.65%		6.36%	
1%VaR												
Expanding Window	2.50%	(1.41%)	1.87%	(0.64%)	1.91%	(0.75%)	1.83%	(0.75%)	2.04%	(1.34%)	**1.72%**	(0.74%)
Rolling Window 125 days	2.12%	(0.75%)	2.25%	(0.80%)	2.17%	(0.74%)	2.38%	(0.81%)	4.52%	(0.49%)	2.64%	(1.72%)
Rolling Window 250 days	1.91%	(0.69%)	2.04%	(0.72%)	1.99%	(0.61%)	2.07%	(0.65%)	2.97%	(0.62%)	2.10%	(1.22%)
Rolling Window 500 days	2.11%	(0.72%)	1.90%	(0.67%)	1.86%	(0.65%)	1.94%	(0.62%)	2.39%	(0.63%)	1.81%	(1.08%)
Rolling Window 1,000 days	2.20%	(0.81%)	1.84%	(0.66%)	1.86%	(0.68%)	1.84%	(0.67%)	2.03%	(0.83%)	1.74%	(0.94%)
Structural Breaks Window	**1.83%**	(0.76%)	1.84%	(0.66%)	**1.73%**	(0.62%)	**1.75%**	(0.68%)	2.24%	(0.88%)	1.79%	(0.93%)
Comb. Mean All	1.88%	(0.68%)	1.80%	(0.64%)	1.83%	(0.59%)	1.81%	(0.61%)	**2.00%**	(0.73%)	1.73%	(0.88%)
Comb. Mean All ex SB	1.93%	(0.69%)	1.82%	(0.66%)	1.86%	(0.61%)	1.87%	(0.62%)	2.06%	(0.77%)	1.77%	(0.90%)
Comb. Mean Rolling Windows	1.87%	(0.67%)	1.88%	(0.65%)	1.87%	(0.61%)	1.91%	(0.62%)	2.27%	(0.61%)	1.84%	(1.03%)
Comb. Long Short	1.99%	(0.82%)	**1.79%**	(0.64%)	1.95%	(0.67%)	1.95%	(0.67%)	2.31%	(1.09%)	1.92%	(0.95%)
Comb. Trimmed Mean All	1.88%	(0.70%)	1.83%	(0.66%)	1.83%	(0.59%)	1.81%	(0.63%)	2.09%	(0.68%)	1.80%	(0.95%)
Comb. Trimmed Mean All ex SB	1.97%	(0.70%)	1.86%	(0.66%)	1.88%	(0.61%)	1.90%	(0.64%)	2.17%	(0.73%)	1.84%	(0.97%)
Comb. Trimmed Mean Rolling Windows	1.94%	(0.69%)	1.92%	(0.66%)	1.94%	(0.61%)	1.99%	(0.66%)	2.50%	(0.61%)	1.91%	(1.09%)
Average	2.01%		1.90%		1.90%		1.93%		2.43%		1.89%	

Figure 5.1: Differences in VaR-violation Ratios Depending on the Estimation Window Strategy

For each forecast horizon, VaR level, and VaR model, the figure shows the range between the minimum and the maxiumum VaR-violation ratio resulting from different estimation window strategies. The individual VaR-violation ratios are averaged over the 42 subsamples of log-returns of stocks listed on the DAX as described in Table 5.2.



The dynamic VaR models EWMA, GARCH, GJR-GARCH and filtered historical simulation attribute a higher weighting to more recent returns. Therefore, differences in the VaR-violation ratio due to estimation window strategies tend to be somewhat lower compared to the static models employing a normal distribution and historical simulation. Nevertheless, even for such dynamic VaR models the results illustrate the importance of selecting a proper estimation window.

As an example, for each VaR model, Figures 5.2 and 5.3 show returns of the Allianz stock for the period 9 July 1998 to 18 July 2006 and the corresponding 1-day ahead VaR-forecasts at the 5% VaR level. The VaR-forecasts are estimated by using a selection of different estimation window strategies. The figures demonstrate that even for dynamic VaR models, the selection of an estimation window strategy can lead to differences of several percentage points for the next day VaR-forecast, particularly during volatile markets.

Tables 5.5 to 5.7 present the rejection rates for the two-sided and one-sided uc backtests. For each model and each estimation window strategy, the rejection rate is computed by the number of rejections divided by the total number of performed uc tests. The VaR-violation ratios of the expanding window strategy exceed the nominal VaR level and consequently indicate an underestimation of VaR, except for the dynamic parametric EWMA, GARCH, and GJR-GARCH models at the 5% VaR level. Additionally, the high level of the standard deviations of the VaR-violation ratios across the 42 subsamples indicates a lack of robustness in the results. The expanding windows show relatively high rejection rates of the two-tailed uc test for almost all models at the 5% VaR level in comparison to the competing estimation window strategies. For the 1% VaR level the rejection rates are more heterogeneous.

Figure 5.2: VaR-Forecasts and VaR-Violations (1/2)

The returns are computed by using total return quotes of Allianz for the period 9 July 1998 to 18 July 2006. VaR-forecasts are estimated by the normal distribution, the EWMA, and the GARCH model at the 5% VaR level. For each estimation window strategy, VaR-forecasts are shown with lines in different colors. The dashes at the bottom of the charts mark the data points where a VaR-violation occurs.

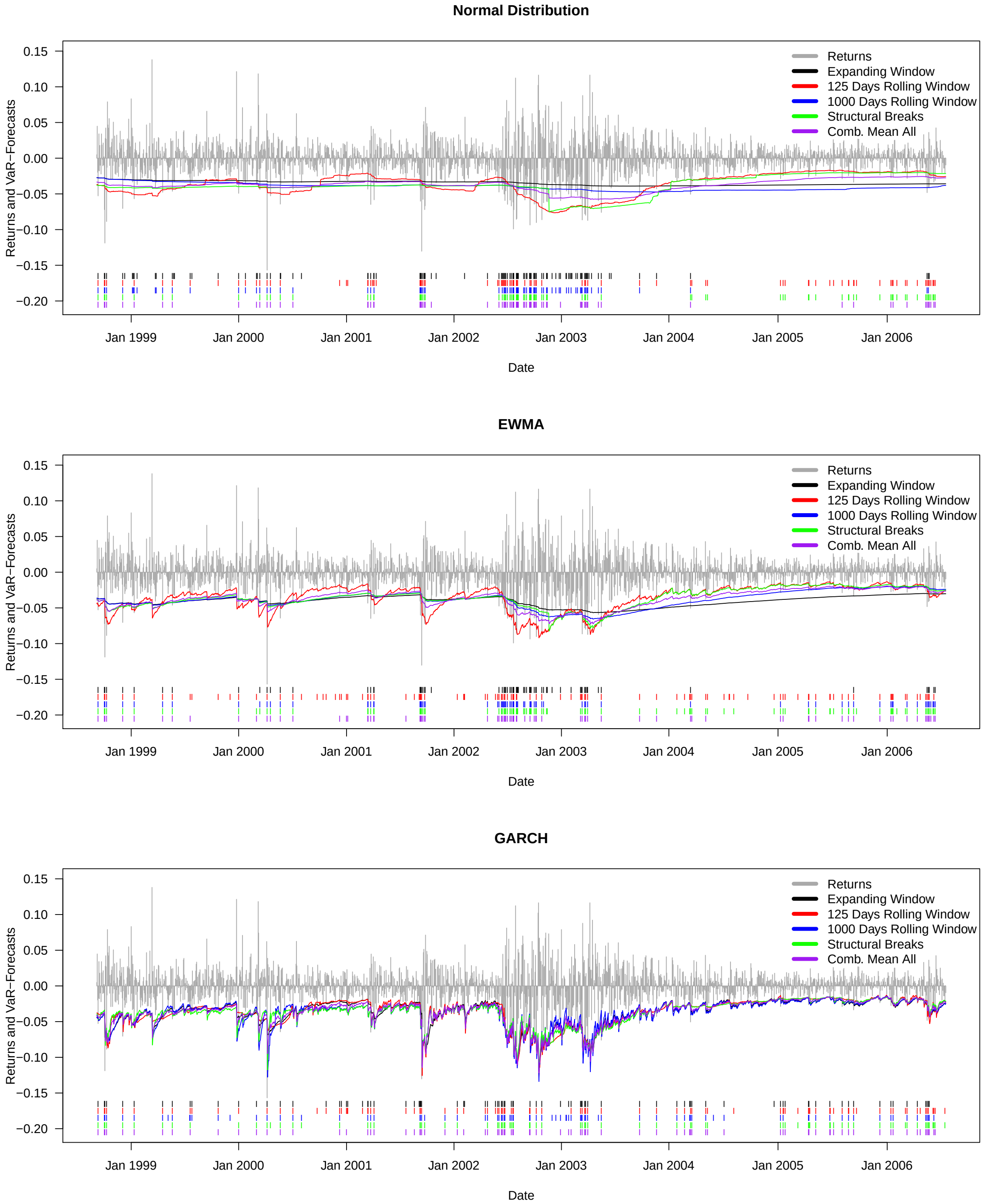


Figure 5.3: VaR-Forecasts and VaR-Violations (2/2)

The returns are computed by using total return quotes of Allianz for the period 9 July 1998 to 18 July 2006. VaR-forecasts are estimated by the GJR-GARCH, the historical simulation, and the filtered historical simulation model at the 5% VaR level. For each estimation window strategy, VaR-forecasts are shown with lines in different colors. The dashes at the bottom of the charts mark the data points where a VaR-violation occurs.

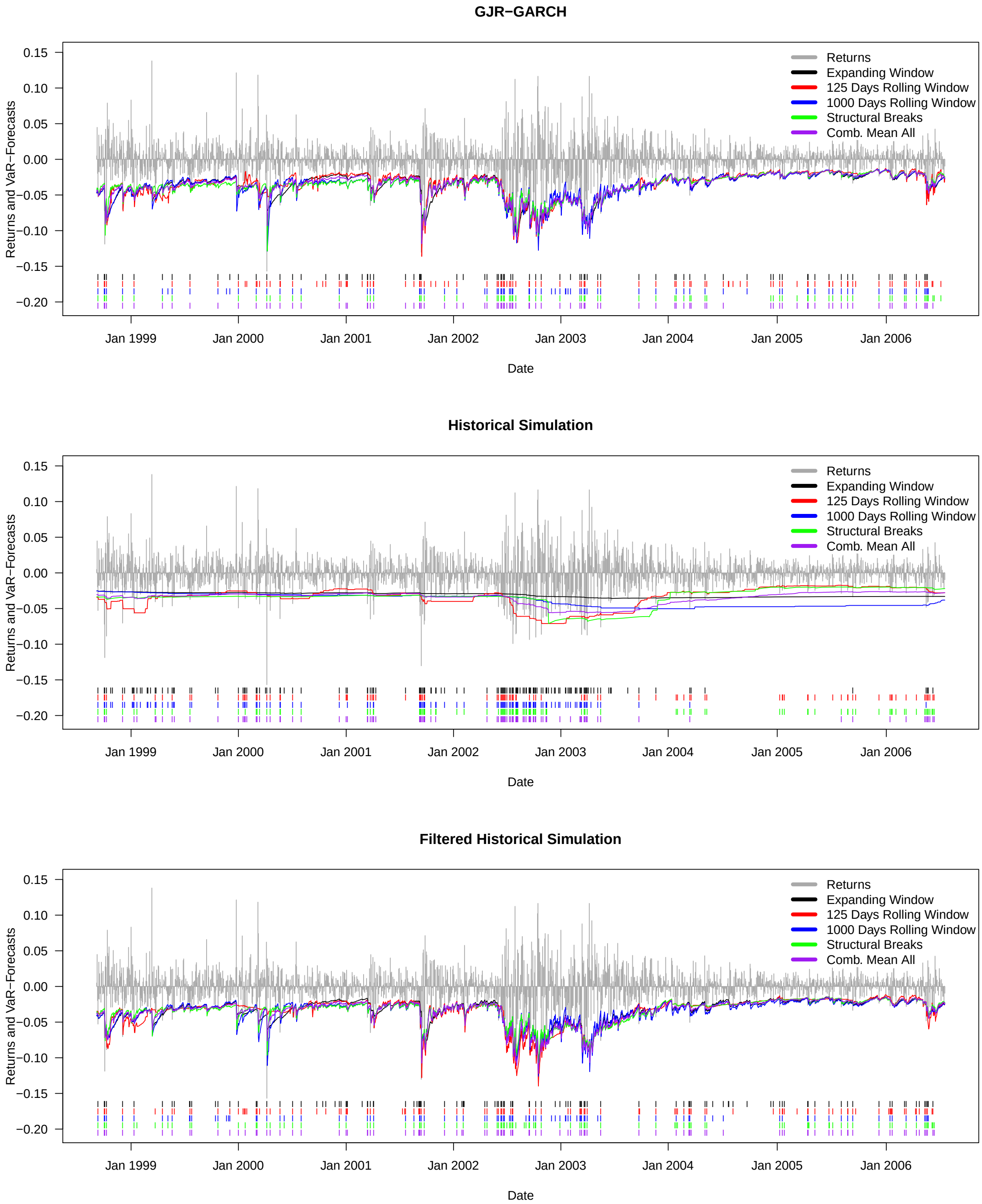


Table 5.5: Unconditional Coverage - Two-Sided Test

For each VaR model and each estimation window strategy, the table reports rejection rates of the two-sided uc tests. The rejection rate is computed by the number of rejections divided by the total number of performed uc tests. The rejection rates in bold are closest to the significance level of the test.

Estimation Window Strategy	Normal Distr.	EWMA	GARCH	GJR-GARCH	HS	Filtered HS	Average
5%VaR / 1-Day Ahead Forecasts							
Expanding Window	64.3%	45.2%	33.3%	31.0%	76.2%	45.2%	49.2%
Rolling Window 125 days	14.3%	**7.1%**	**7.1%**	2.4%	50.0%	66.7%	24.6%
Rolling Window 250 days	**9.5%**	23.8%	9.5%	**4.8%**	45.2%	31.0%	**20.6%**
Rolling Window 500 days	26.2%	31.0%	16.7%	21.4%	57.1%	19.0%	28.6%
Rolling Window 1,000 days	47.6%	35.7%	33.3%	26.2%	59.5%	26.2%	38.1%
Structural Breaks Window	35.7%	31.0%	28.6%	33.3%	38.1%	16.7%	30.6%
Comb. Mean All	45.2%	35.7%	31.0%	26.2%	42.9%	**9.5%**	31.7%
Comb. Mean All ex SB	50.0%	42.9%	31.0%	23.8%	42.9%	16.7%	34.5%
Comb. Mean Rolling Windows	33.3%	26.2%	28.6%	21.4%	**33.3%**	16.7%	26.6%
Comb. Long Short	52.4%	45.2%	28.6%	11.9%	47.6%	26.2%	35.3%
Comb. Trimmed Mean All	45.2%	28.6%	31.0%	26.2%	38.1%	16.7%	31.0%
Comb. Trimmed Mean All ex SB	38.1%	35.7%	28.6%	23.8%	47.6%	21.4%	32.5%
Comb. Trimmed Mean Rolling Windows	26.2%	28.6%	23.8%	23.8%	42.9%	21.4%	27.8%
Average	37.5%	32.1%	25.5%	**21.2%**	47.8%	25.6%	
1%VaR / 1-Day Ahead Forecasts							
Expanding Window	81.0%	78.6%	52.4%	**40.5%**	64.3%	**11.9%**	54.8%
Rolling Window 125 days	95.2%	69.0%	76.2%	100.0%	97.6%	97.6%	89.3%
Rolling Window 250 days	88.1%	73.8%	83.3%	81.0%	85.7%	81.0%	82.1%
Rolling Window 500 days	90.5%	73.8%	64.3%	66.7%	73.8%	28.6%	66.3%
Rolling Window 1,000 days	88.1%	73.8%	52.4%	45.2%	61.9%	16.7%	56.3%
Structural Breaks Window	81.0%	66.7%	59.5%	61.9%	54.8%	40.5%	60.7%
Comb. Mean All	78.6%	54.8%	**35.7%**	45.2%	**35.7%**	14.3%	**44.0%**
Comb. Mean All ex SB	78.6%	57.1%	47.6%	47.6%	38.1%	19.0%	48.0%
Comb. Mean Rolling Windows	81.0%	57.1%	54.8%	61.9%	45.2%	23.8%	54.0%
Comb. Long Short	**71.4%**	**40.5%**	47.6%	57.1%	54.8%	26.2%	49.6%
Comb. Trimmed Mean All	81.0%	59.5%	47.6%	45.2%	38.1%	16.7%	48.0%
Comb. Trimmed Mean All ex SB	83.3%	66.7%	52.4%	52.4%	50.0%	14.3%	53.2%
Comb. Trimmed Mean Rolling Windows	95.2%	61.9%	64.3%	59.5%	64.3%	33.3%	63.1%
Average	84.1%	64.1%	56.8%	58.8%	58.8%	**32.6%**	
5%VaR / 10-Day Ahead Forecasts							
Expanding Window	71.4%	40.5%	52.4%	57.1%	73.8%	**54.8%**	58.3%
Rolling Window 125 days	42.9%	52.4%	45.2%	54.8%	100.0%	85.7%	63.5%
Rolling Window 250 days	**28.6%**	45.2%	40.5%	50.0%	92.9%	73.8%	55.2%
Rolling Window 500 days	35.7%	35.7%	50.0%	47.6%	71.4%	73.8%	52.4%
Rolling Window 1,000 days	52.4%	**31.0%**	42.9%	**40.5%**	61.9%	59.5%	**48.0%**
Structural Breaks Window	54.8%	40.5%	**38.1%**	50.0%	66.7%	59.5%	51.6%
Comb. Mean All	45.2%	33.3%	45.2%	50.0%	**57.1%**	71.4%	50.4%
Comb. Mean All ex SB	47.6%	33.3%	42.9%	50.0%	**57.1%**	69.0%	50.0%
Comb. Mean Rolling Windows	33.3%	38.1%	42.9%	42.9%	59.5%	76.2%	48.8%
Comb. Long Short	57.1%	**31.0%**	45.2%	45.2%	73.8%	76.2%	54.8%
Comb. Trimmed Mean All	42.9%	33.3%	45.2%	45.2%	**57.1%**	69.0%	48.8%
Comb. Trimmed Mean All ex SB	42.9%	33.3%	42.9%	47.6%	59.5%	69.0%	49.2%
Comb. Trimmed Mean Rolling Windows	35.7%	47.6%	40.5%	47.6%	73.8%	73.8%	53.2%
Average	45.4%	**38.1%**	44.1%	48.4%	69.6%	70.1%	
1%VaR / 10-Day Ahead Forecasts							
Expanding Window	78.6%	76.2%	66.7%	66.7%	**71.4%**	66.7%	71.0%
Rolling Window 125 days	81.0%	78.6%	78.6%	83.3%	100.0%	76.2%	82.9%
Rolling Window 250 days	78.6%	76.2%	78.6%	76.2%	100.0%	71.4%	80.2%
Rolling Window 500 days	83.3%	78.6%	73.8%	73.8%	97.6%	**61.9%**	78.2%
Rolling Window 1,000 days	83.3%	81.0%	71.4%	66.7%	78.6%	66.7%	74.6%
Structural Breaks Window	**66.7%**	**69.0%**	**61.9%**	**61.9%**	92.9%	**61.9%**	**69.0%**
Comb. Mean All	76.2%	73.8%	73.8%	71.4%	76.2%	66.7%	73.0%
Comb. Mean All ex SB	73.8%	71.4%	76.2%	73.8%	76.2%	71.4%	73.8%
Comb. Mean Rolling Windows	83.3%	81.0%	73.8%	73.8%	88.1%	66.7%	77.8%
Comb. Long Short	**66.7%**	76.2%	76.2%	73.8%	81.0%	73.8%	74.6%
Comb. Trimmed Mean All	73.8%	76.2%	71.4%	66.7%	78.6%	66.7%	72.2%
Comb. Trimmed Mean All ex SB	73.8%	76.2%	71.4%	73.8%	83.3%	64.3%	73.8%
Comb. Trimmed Mean Rolling Windows	76.2%	81.0%	76.2%	73.8%	97.6%	69.0%	79.0%
Average	76.6%	76.6%	73.1%	72.0%	86.3%	**67.9%**	

Table 5.6: Unconditional Coverage - One-Sided Test / Lower Tail

For each VaR model and each estimation window strategy, the table reports rejection rates of the lower tail uc tests. The rejection rate is computed by the number of rejections divided by the total number of performed uc tests. The rejection rates in bold are closest to the tests' significance level of 5%.

Estimation Window Strategy	Normal Distr.	EWMA	GARCH	GJR-GARCH	HS	Filtered HS	Average
5%VaR / 1-Day Ahead Forecasts							
Expanding Window	26.2%	45.2%	45.2%	38.1%	19.0%	7.1%	30.2%
Rolling Window 125 days	16.7%	**11.9%**	**9.5%**	0.0%	0.0%	0.0%	**6.3%**
Rolling Window 250 days	**11.9%**	28.6%	21.4%	**9.5%**	0.0%	0.0%	11.9%
Rolling Window 500 days	23.8%	42.9%	33.3%	26.2%	0.0%	0.0%	21.0%
Rolling Window 1000 days	28.6%	42.9%	35.7%	35.7%	2.4%	0.0%	24.2%
Structural Breaks Window	35.7%	40.5%	38.1%	35.7%	**7.1%**	**4.8%**	27.0%
Comb. Mean All	40.5%	45.2%	35.7%	31.0%	**7.1%**	0.0%	26.6%
Comb. Mean All ex SB	40.5%	52.4%	35.7%	28.6%	**7.1%**	0.0%	27.4%
Comb. Mean Rolling Windows	40.5%	35.7%	35.7%	26.2%	2.4%	0.0%	23.4%
Comb. Long Short	40.5%	52.4%	33.3%	23.8%	9.5%	0.0%	26.6%
Comb. Trimmed Mean All	38.1%	42.9%	33.3%	31.0%	2.4%	0.0%	24.6%
Comb. Trimmed Mean All ex SB	33.3%	45.2%	33.3%	33.3%	2.4%	0.0%	24.6%
Comb. Trimmed Mean Rolling Windows	26.2%	33.3%	31.0%	26.2%	0.0%	0.0%	19.4%
Average	31.0%	39.9%	32.4%	26.6%	**4.6%**	0.9%	
1%VaR / 1-Day Ahead Forecasts							
Expanding Window	0.0%	0.0%	0.0%	0.0%	11.9%	0.0%	**2.0%**
Rolling Window 125 days	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Rolling Window 250 days	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Rolling Window 500 days	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Rolling Window 1000 days	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%	0.4%
Structural Breaks Window	0.0%	0.0%	0.0%	0.0%	**4.8%**	0.0%	0.8%
Comb. Mean All	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%	0.4%
Comb. Mean All ex SB	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%	0.4%
Comb. Mean Rolling Windows	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Comb. Long Short	0.0%	0.0%	0.0%	0.0%	9.5%	0.0%	1.6%
Comb. Trimmed Mean All	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%	0.4%
Comb. Trimmed Mean All ex SB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Comb. Trimmed Mean Rolling Windows	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Average	0.0%	0.0%	0.0%	0.0%	**2.7%**	0.0%	
5%VaR / 10-Day Ahead Forecasts							
Expanding Window	31.0%	28.6%	35.7%	38.1%	11.9%	**7.1%**	25.4%
Rolling Window 125 days	23.8%	**16.7%**	**21.4%**	**16.7%**	0.0%	28.6%	**17.9%**
Rolling Window 250 days	**21.4%**	21.4%	23.8%	23.8%	0.0%	19.0%	18.3%
Rolling Window 500 days	**21.4%**	28.6%	26.2%	26.2%	2.4%	16.7%	20.2%
Rolling Window 1000 days	28.6%	26.2%	31.0%	31.0%	**4.8%**	9.5%	21.8%
Structural Breaks Window	40.5%	31.0%	31.0%	35.7%	**4.8%**	9.5%	25.4%
Comb. Mean All	31.0%	26.2%	31.0%	31.0%	0.0%	19.0%	23.0%
Comb. Mean All ex SB	31.0%	26.2%	31.0%	28.6%	2.4%	16.7%	22.6%
Comb. Mean Rolling Windows	26.2%	26.2%	31.0%	28.6%	0.0%	19.0%	21.8%
Comb. Long Short	38.1%	26.2%	26.2%	21.4%	7.1%	16.7%	22.6%
Comb. Trimmed Mean All	33.3%	31.0%	31.0%	28.6%	0.0%	14.3%	23.0%
Comb. Trimmed Mean All ex SB	28.6%	28.6%	26.2%	28.6%	0.0%	14.3%	21.0%
Comb. Trimmed Mean Rolling Windows	23.8%	23.8%	31.0%	26.2%	0.0%	19.0%	20.6%
Average	29.1%	26.2%	28.9%	28.0%	2.6%	16.1%	
1%VaR / 10-Day Ahead Forecasts							
Expanding Window	**4.8%**	**4.8%**	0.0%	2.4%	16.7%	7.1%	6.0%
Rolling Window 125 days	2.4%	0.0%	0.0%	0.0%	0.0%	11.9%	2.4%
Rolling Window 250 days	2.4%	0.0%	0.0%	0.0%	0.0%	7.1%	1.6%
Rolling Window 500 days	2.4%	2.4%	**4.8%**	0.0%	0.0%	9.5%	3.2%
Rolling Window 1000 days	**4.8%**	2.4%	7.1%	0.0%	**4.8%**	7.1%	**4.4%**
Structural Breaks Window	**4.8%**	2.4%	2.4%	**4.8%**	9.5%	0.0%	4.0%
Comb. Mean All	2.4%	**4.8%**	2.4%	0.0%	0.0%	9.5%	3.2%
Comb. Mean All ex SB	2.4%	**4.8%**	2.4%	0.0%	2.4%	9.5%	3.6%
Comb. Mean Rolling Windows	**4.8%**	2.4%	2.4%	2.4%	0.0%	7.1%	3.2%
Comb. Long Short	**4.8%**	**4.8%**	0.0%	0.0%	4.8%	7.1%	3.6%
Comb. Trimmed Mean All	2.4%	2.4%	0.0%	**4.8%**	0.0%	7.1%	2.8%
Comb. Trimmed Mean All ex SB	2.4%	2.4%	0.0%	2.4%	0.0%	**4.8%**	2.0%
Comb. Trimmed Mean Rolling Windows	2.4%	2.4%	2.4%	2.4%	0.0%	7.1%	2.8%
Average	3.3%	2.7%	1.8%	1.5%	2.9%	7.3%	

Table 5.7: Unconditional Coverage - One-Sided Test / Upper Tail

For each VaR model and each estimation window strategy, the table reports rejection rates of the upper tail uc tests. The rejection rate is computed by the number of rejections divided by the total number of performed uc tests. The rejection rates in bold are closest to the tests' significance level of 5%.

Estimation Window Strategy	Normal Distr.	EWMA	GARCH	GJR-GARCH	HS	Filtered HS	Average
5%VaR / 1-Day Ahead Forecasts							
Expanding Window	40.5%	9.5%	**2.4%**	0.0%	61.9%	50.0%	27.4%
Rolling Window 125 days	2.4%	**2.4%**	0.0%	7.1%	76.2%	90.5%	29.8%
Rolling Window 250 days	**4.8%**	**2.4%**	0.0%	2.4%	54.8%	50.0%	19.0%
Rolling Window 500 days	14.3%	0.0%	0.0%	**4.8%**	69.0%	31.0%	19.8%
Rolling Window 1000 days	31.0%	0.0%	0.0%	0.0%	59.5%	38.1%	21.4%
Structural Breaks Window	7.1%	0.0%	0.0%	2.4%	**33.3%**	26.2%	**11.5%**
Comb. Mean All	11.9%	0.0%	**2.4%**	0.0%	38.1%	**21.4%**	12.3%
Comb. Mean All ex SB	14.3%	0.0%	0.0%	0.0%	45.2%	26.2%	14.3%
Comb. Mean Rolling Windows	7.1%	0.0%	0.0%	0.0%	40.5%	**21.4%**	**11.5%**
Comb. Long Short	21.4%	0.0%	0.0%	0.0%	42.9%	35.7%	16.7%
Comb. Trimmed Mean All	16.7%	0.0%	**2.4%**	0.0%	47.6%	26.2%	15.5%
Comb. Trimmed Mean All ex SB	14.3%	0.0%	**2.4%**	0.0%	47.6%	31.0%	15.9%
Comb. Trimmed Mean Rolling Windows	9.5%	0.0%	0.0%	0.0%	52.4%	31.0%	15.5%
Average	15.0%	1.1%	0.7%	1.3%	51.5%	36.8%	
1%VaR / 1-Day Ahead Forecasts							
Expanding Window	83.3%	83.3%	**57.1%**	**52.4%**	57.1%	**19.0%**	58.7%
Rolling Window 125 days	97.6%	88.1%	92.9%	100.0%	97.6%	100.0%	96.0%
Rolling Window 250 days	90.5%	78.6%	95.2%	90.5%	90.5%	83.3%	88.1%
Rolling Window 500 days	95.2%	88.1%	71.4%	76.2%	88.1%	47.6%	77.8%
Rolling Window 1000 days	88.1%	85.7%	61.9%	57.1%	61.9%	26.2%	63.5%
Structural Breaks Window	81.0%	83.3%	59.5%	69.0%	61.9%	45.2%	66.7%
Comb. Mean All	83.3%	64.3%	61.9%	57.1%	**45.2%**	21.4%	**55.6%**
Comb. Mean All ex SB	81.0%	71.4%	61.9%	**52.4%**	52.4%	21.4%	56.7%
Comb. Mean Rolling Windows	90.5%	66.7%	76.2%	69.0%	61.9%	35.7%	66.7%
Comb. Long Short	**73.8%**	**50.0%**	64.3%	71.4%	50.0%	40.5%	58.3%
Comb. Trimmed Mean All	83.3%	76.2%	69.0%	59.5%	47.6%	21.4%	59.5%
Comb. Trimmed Mean All ex SB	85.7%	78.6%	66.7%	64.3%	59.5%	26.2%	63.5%
Comb. Trimmed Mean Rolling Windows	95.2%	78.6%	73.8%	76.2%	69.0%	50.0%	73.8%
Average	86.8%	76.4%	70.1%	68.9%	64.8%	**41.4%**	
5%VaR / 10-Day Ahead Forecasts							
Expanding Window	45.2%	14.3%	26.2%	28.6%	61.9%	**52.4%**	38.1%
Rolling Window 125 days	23.8%	38.1%	28.6%	52.4%	100.0%	57.1%	50.0%
Rolling Window 250 days	**11.9%**	26.2%	21.4%	33.3%	95.2%	61.9%	41.7%
Rolling Window 500 days	19.0%	19.0%	23.8%	28.6%	76.2%	64.3%	38.5%
Rolling Window 1000 days	33.3%	9.5%	16.7%	19.0%	64.3%	**52.4%**	32.5%
Structural Breaks Window	19.0%	19.0%	**11.9%**	**16.7%**	66.7%	57.1%	31.7%
Comb. Mean All	16.7%	9.5%	16.7%	21.4%	**59.5%**	61.9%	31.0%
Comb. Mean All ex SB	19.0%	**7.1%**	14.3%	21.4%	**59.5%**	61.9%	**30.6%**
Comb. Mean Rolling Windows	14.3%	21.4%	16.7%	21.4%	71.4%	61.9%	34.5%
Comb. Long Short	26.2%	9.5%	23.8%	26.2%	66.7%	61.9%	35.7%
Comb. Trimmed Mean All	19.0%	14.3%	19.0%	21.4%	64.3%	59.5%	32.9%
Comb. Trimmed Mean All ex SB	19.0%	9.5%	16.7%	21.4%	64.3%	59.5%	31.7%
Comb. Trimmed Mean Rolling Windows	14.3%	26.2%	16.7%	26.2%	76.2%	61.9%	36.9%
Average	21.6%	**17.2%**	19.4%	26.0%	71.2%	59.5%	
1%VaR / 10-Day Ahead Forecasts							
Expanding Window	76.2%	76.2%	**66.7%**	**66.7%**	**59.5%**	64.3%	**68.3%**
Rolling Window 125 days	83.3%	81.0%	78.6%	83.3%	100.0%	71.4%	82.9%
Rolling Window 250 days	76.2%	78.6%	78.6%	76.2%	100.0%	66.7%	79.4%
Rolling Window 500 days	83.3%	78.6%	71.4%	78.6%	97.6%	**57.1%**	77.8%
Rolling Window 1000 days	88.1%	78.6%	69.0%	69.0%	76.2%	59.5%	73.4%
Structural Breaks Window	**66.7%**	76.2%	**66.7%**	**66.7%**	85.7%	64.3%	71.0%
Comb. Mean All	73.8%	76.2%	76.2%	73.8%	78.6%	61.9%	73.4%
Comb. Mean All ex SB	76.2%	76.2%	76.2%	73.8%	78.6%	64.3%	74.2%
Comb. Mean Rolling Windows	78.6%	78.6%	78.6%	73.8%	95.2%	61.9%	77.8%
Comb. Long Short	71.4%	**73.8%**	76.2%	76.2%	83.3%	71.4%	75.4%
Comb. Trimmed Mean All	71.4%	76.2%	76.2%	73.8%	78.6%	66.7%	73.8%
Comb. Trimmed Mean All ex SB	76.2%	78.6%	76.2%	73.8%	85.7%	61.9%	75.4%
Comb. Trimmed Mean Rolling Windows	81.0%	78.6%	78.6%	78.6%	100.0%	61.9%	79.8%
Average	77.1%	77.5%	74.5%	74.2%	86.1%	64.1%	

Regarding the rolling window strategy and the 5% VaR, the VaR-violation ratio of the historical simulation and the filtered historical simulation models exceed the nominal VaR level significantly while the ratios of the remaining models are around or below 5%.

For the 1% VaR level, the VaR-violation ratios of all models exceed the nominal level on average. For the 1-day ahead forecast the standard deviations of the results tend to increase with the length of the rolling window, whereas the 10-day ahead forecasts do not show a similar trend. With the exception of the normal distribution VaR model, larger moving samples tend to estimate VaR more conservatively compared to short windows. Consequently, the averaged rejection rates of the two-tailed and the upper-tail uc backtests decrease as the size of the rolling windows increases.

In most of the cases, the structural breaks strategy leads to more conservative VaR-forecasts compared to the rolling and the expanding window strategy, irrespective of the VaR level and the forecasting horizon. This result is partly reflected in higher lower-tail and lower upper-tail uc test rejection rates. However, the two-tailed uc test rejection rates are neither significantly better nor significantly worse than the competing strategies.

The averaged VaR-violation ratios of the combination strategies indicate that pooling forecasts leads to more conservatively estimated VaR-forecasts compared to the competing strategies. Interestingly, the VaR-violation ratios of the combinations are consistently lower than the mean of their respective component ratios. For example, the violation ratios considering 1-day ahead forecasts of the normal distribution VaR model and the 1%VaR for the 125-day rolling window and the expanding window are 4.74% and 5.60%, respectively, which amounts to an average of 5.17%. The violation ratio of the corresponding combination *mean long short* is significantly lower with 4.68%. On average, the combination *mean all* shows the lowest VaR-violation ratio of all combination strategies. However, the ratio tends to increase as the number of combination components decreases. This also applies to the trimmed combinations, where the ratios are slightly higher compared to their non-trimmed counterparts. These findings are largely confirmed by the results of the uc backtests and hold true for both VaR levels and both forecast horizons.

5.4.2 I.I.D. VaR-Violations

The results of the i.i.d. backtest are presented in Table 5.8. We start with an evaluation of the backtest results for the 1-day ahead forecasts. As to be expected, the rejection rates of

Table 5.8: I.I.D. VaR-Violations - Backtest Rejections

For each VaR model and each estimation window strategy, the table reports rejection rates of the i.i.d. backtest for the 1-day and 10-day ahead forecast horizons and the 5% and 1% VaR level. The rejection rate is computed by the number of rejections divided by the total number of performed i.i.d. backtests. The rejection rates in bold are closest to the significance level of the test.

Estimation Window Strategy	Normal Distr.	EWMA	GARCH	GJR-GARCH	HS	Filtered HS	Average
5%VaR / 1-Day Ahead Forecasts							
Expanding Window	95.2%	95.2%	33.3%	33.3%	97.6%	42.9%	66.3%
Rolling Window 125 days	**83.3%**	**31.0%**	38.1%	38.1%	95.2%	47.6%	55.6%
Rolling Window 250 days	85.7%	50.0%	40.5%	33.3%	92.9%	45.2%	57.9%
Rolling Window 500 days	95.2%	73.8%	42.9%	28.6%	97.6%	45.2%	63.9%
Rolling Window 1,000 days	97.6%	90.5%	40.5%	33.3%	95.2%	45.2%	67.1%
Structural Breaks Window	95.2%	81.0%	45.2%	47.6%	97.6%	59.5%	71.0%
Comb. Mean All	92.9%	71.4%	33.3%	26.2%	92.9%	40.5%	59.5%
Comb. Mean All ex SB	92.9%	71.4%	33.3%	**21.4%**	92.9%	28.6%	56.7%
Comb. Mean Rolling Windows	92.9%	54.8%	35.7%	26.2%	**90.5%**	35.7%	56.0%
Comb. Long Short	90.5%	66.7%	35.7%	23.8%	92.9%	**23.8%**	55.6%
Comb. Trimmed Mean All	95.2%	69.0%	33.3%	28.6%	92.9%	40.5%	59.9%
Comb. Trimmed Mean All ex SB	90.5%	69.0%	**31.0%**	26.2%	**90.5%**	28.6%	56.0%
Comb. Trimmed Mean Rolling Windows	92.9%	61.9%	**31.0%**	**21.4%**	92.9%	31.0%	**55.2%**
Average	92.3%	68.1%	36.4%	**29.9%**	94.0%	39.6%	
1%VaR / 1-Day Ahead Forecasts							
Expanding Window	88.1%	78.6%	19.0%	16.7%	88.1%	19.0%	51.6%
Rolling Window 125 days	**66.7%**	**4.8%**	**11.9%**	23.8%	**33.3%**	7.1%	**24.6%**
Rolling Window 250 days	76.2%	21.4%	28.6%	11.9%	76.2%	9.5%	37.3%
Rolling Window 500 days	78.6%	50.0%	21.4%	14.3%	83.3%	28.6%	46.0%
Rolling Window 1,000 days	90.5%	66.7%	19.0%	21.4%	83.3%	28.6%	51.6%
Structural Breaks Window	83.3%	52.4%	31.0%	19.0%	78.6%	35.7%	50.0%
Comb. Mean All	73.8%	47.6%	19.0%	**7.1%**	71.4%	9.5%	38.1%
Comb. Mean All ex SB	71.4%	42.9%	19.0%	**7.1%**	69.0%	7.1%	36.1%
Comb. Mean Rolling Windows	71.4%	33.3%	19.0%	9.5%	66.7%	9.5%	34.9%
Comb. Long Short	76.2%	40.5%	**11.9%**	9.5%	69.0%	**4.8%**	35.3%
Comb. Trimmed Mean All	78.6%	45.2%	21.4%	9.5%	69.0%	19.0%	40.5%
Comb. Trimmed Mean All ex SB	71.4%	47.6%	19.0%	**7.1%**	69.0%	9.5%	37.3%
Comb. Trimmed Mean Rolling Windows	73.8%	31.0%	19.0%	**7.1%**	73.8%	14.3%	36.5%
Average	76.9%	43.2%	20.0%	12.6%	71.6%	15.6%	
5%VaR / 10-Day Ahead Forecasts							
Expanding Window	100%	100%	100%	100%	100%	100%	100%
Rolling Window 125 days	100%	100%	100%	100%	100%	100%	100%
Rolling Window 250 days	100%	100%	100%	100%	100%	100%	100%
Rolling Window 500 days	100%	100%	100%	100%	100%	100%	100%
Rolling Window 1,000 days	100%	100%	100%	100%	100%	100%	100%
Structural Breaks Window	100%	100%	100%	100%	100%	100%	100%
Comb. Mean All	100%	100%	100%	100%	100%	100%	100%
Comb. Mean All ex SB	100%	100%	100%	100%	100%	100%	100%
Comb. Mean Rolling Windows	100%	100%	100%	100%	100%	100%	100%
Comb. Long Short	100%	100%	100%	100%	100%	100%	100%
Comb. Trimmed Mean All	100%	100%	100%	100%	100%	100%	100%
Comb. Trimmed Mean All ex SB	100%	100%	100%	100%	100%	100%	100%
Comb. Trimmed Mean Rolling Windows	100%	100%	100%	100%	100%	100%	100%
Average	100%	100%	100%	100%	100%	100%	
1%VaR / 10-Day Ahead Forecasts							
Expanding Window	100.0%	100.0%	100.0%	95.2%	100.0%	100.0%	99.2%
Rolling Window 125 days	100.0%	100.0%	100.0%	97.6%	100.0%	100.0%	99.6%
Rolling Window 250 days	100.0%	100.0%	100.0%	97.6%	100.0%	97.6%	99.2%
Rolling Window 500 days	100.0%	100.0%	97.6%	97.6%	100.0%	97.6%	98.8%
Rolling Window 1,000 days	100.0%	100.0%	97.6%	97.6%	100.0%	**97.6%**	98.8%
Structural Breaks Window	**97.6%**	**95.2%**	97.6%	**92.9%**	100.0%	100.0%	**97.2%**
Comb. Mean All	**97.6%**	97.6%	97.6%	**92.9%**	100.0%	97.6%	**97.2%**
Comb. Mean All ex SB	**97.6%**	97.6%	**95.2%**	95.2%	100.0%	100.0%	97.6%
Comb. Mean Rolling Windows	100.0%	100.0%	100.0%	95.2%	100.0%	100.0%	99.2%
Comb. Long Short	100.0%	97.6%	100.0%	97.6%	**97.6%**	100.0%	98.8%
Comb. Trimmed Mean All	100.0%	**95.2%**	100.0%	**92.9%**	100.0%	100.0%	98.0%
Comb. Trimmed Mean All ex SB	100.0%	97.6%	100.0%	**92.9%**	100.0%	97.6%	98.0%
Comb. Trimmed Mean Rolling Windows	100.0%	100.0%	100.0%	95.2%	100.0%	100.0%	99.2%
Average	99.5%	98.5%	98.9%	**95.4%**	99.8%	99.1%	

the dynamic models are significantly lower compared to the static normal distribution and historical simulation models. Depending on the VaR model, the impact of the different estimation window strategies to the rejection rates of the backtest is very specific. With respect to the normal distribution model, almost all estimation window strategies show rejection rates around 90%. Despite the fact that very short rolling windows of 125 and 250 days lead to slight improvements, our results contradict the findings of Starica et al. [2005] and Rapach and Strauss [2008] who state that a simple approach based on the average of the squared returns often achieves good results compared to conditional models if this model is estimated on a relatively small moving window. As mentioned in Section 5.2.1, the decay factor λ of the EWMA model is usually set to 0.94 for 1-day volatility forecasts which means that 99.9% of the information the model uses for the volatility estimation are contained in the last 112 days of historical data. Our results support this setting, because the smallest rolling window of 125 days yields the lowest rejection rates. Considering the GARCH and the GJR-GARCH models, the combination stragegies tend to outperform the remaining strategies. With respect to a preferable length of a rolling window, no clear conclusions can be made. The rejection rates for the historical simulation model are on a very high level compared to the dynamic VaR models, irrespective of the estimation window strategy. However, applying combination strategies leads to slight improvements. The results of the semi-parametric filtered historical simulation model are comparable to those of the GARCH-type models. Again, the rejection rates are improved by applying combinations, where the combination *mean long short* leads to the best performance.

Regarding the 10-day ahead forecasts, the i.i.d. hypothesis is almost consistently rejected for all of the combinations of VaR models and estimation window strategies. Multi-day ahead forecasts are generally vulnerable to dependent VaR-violations. For example, a large negative return shock of a single day has an impact on ten subsequent 10-day-period losses. It is very unlikely that a model is able to adjust its VaR-forecast quickly enough to accommodate this single shock. Consequently, the significant loss of one day can cause two or more subsequent VaR-violations.

5.4.3 Expected Shortfall

With respect to the ES, Tables 5.9 and 5.10 report the average absolute deviations and the rejection rates of the backtest of McNeil and Frey [2000]. We begin with the evaluation of the ES results corresponding to the 1-day ahead VaR-forecasts. Focusing the absolute deviations, the rolling window strategy outperforms the competing strategies on average. Dependent on the individual VaR model, for the 5% VaR the optimal rolling window size varies between 125 and 250 days. For the 1% VaR, rolling windows of 250 days consistently provide the smallest absolute deviations. The ES backtest of McNeil and Frey [2000] is a one-sided test concerning the simple mean of the deviations rather than absolute deviations. The differences in the rejection rates of the backtests due to the estimation window strategy are relatively small. However, for both analyzed VaR levels of 5% and 1%, the expanding window strategy and the combination *mean long short* strategy leads to slightly lower rejection rates compared to the competing strategies.

Considering the 10-day ahead forecasts, the evaluation of the estimation window strategies leads to conclusions which are fairly similar to the 1-day ahead forecasts, albeit the differences between the different settings are less significant. However, the selection of the proper VaR model has a larger impact to the forecasting accuracy than the chosen estimation window strategy, irrespective of the VaR level or the forecast horizon.

5.4.4 Conditional Predictive Ability

The results for the CPA test are reported in Tables 5.11 and 5.12.

The rolling window strategy significantly improve on the benchmark strategy of expanding windows only when less-parameterized approaches like the normal distribution, EWMA, and historical simulation models are applied. For 1-day ahead forecasts of the normal distribution and the historical simulation models, the proportion where the rolling window is preferred tends to increase as the length of the rolling windows decreases while for the GARCH, GJR-GARCH and the filtered historical simulation the opposite holds true. The 10-day ahead forecasts show no similar trend.

Table 5.9: Expected Shortfall - 1-Day Ahead Forecast

The table reports average absolute deviations of the ES and rejection rates of the ES backtest for the 1-day ahead forecasts and the 5% and 1% VaR level. The average absolute deviation is calculated as the average of the absolute deviations between the ES forecast and the realized shortfall in case of a VaR-violation. The rejection rate is computed by the number of rejections divided by the total number of performed ES backtests. For each VaR model, the smallest average absolute deviation is printed in bold. The rejection rates in bold are closest to the significance level of the test.

Estimation Window Strategy	Normal Distr.	EWMA	GARCH	GJR-GARCH	HS	Filtered HS	Average
5% VaR							
Average Absolute Deviation							
Expanding Window	0.651%	0.562%	0.455%	0.457%	0.173%	0.076%	0.396%
Rolling Window 125 days	0.785%	0.673%	**0.439%**	**0.410%**	0.235%	0.133%	0.446%
Rolling Window 250 days	**0.585%**	**0.433%**	0.465%	0.470%	0.089%	0.068%	**0.352%**
Rolling Window 500 days	0.594%	0.503%	0.466%	0.460%	**0.069%**	**0.061%**	0.359%
Rolling Window 1,000 days	0.688%	0.567%	0.451%	0.445%	0.137%	0.074%	0.394%
Structural Breaks Window	0.762%	0.609%	0.457%	0.431%	0.176%	0.097%	0.422%
Comb. Mean All	0.672%	0.540%	0.447%	0.423%	0.124%	0.072%	0.380%
Comb. Mean All ex SB	0.680%	0.537%	0.442%	0.423%	0.115%	0.071%	0.378%
Comb. Mean Rolling Windows	0.655%	0.515%	0.453%	0.434%	0.094%	0.072%	0.371%
Comb. Long Short	0.668%	0.516%	**0.439%**	0.420%	0.126%	0.070%	0.373%
Comb. Trimmed Mean All	0.669%	0.549%	0.444%	0.425%	0.131%	0.075%	0.382%
Comb. Trimmed Mean All ex SB	0.672%	0.545%	0.440%	0.425%	0.109%	0.071%	0.377%
Comb. Trimmed Mean Rolling Windows	0.649%	0.524%	0.454%	0.439%	0.103%	0.073%	0.374%
Average	0.672%	0.544%	0.450%	0.435%	0.129%	**0.078%**	
Backtest Rejections							
Expanding Window	**92.9%**	97.6%	**73.8%**	**66.7%**	16.7%	**0.0%**	**57.9%**
Rolling Window 125 days	97.6%	**90.5%**	95.2%	97.6%	**0.0%**	**0.0%**	63.5%
Rolling Window 250 days	97.6%	95.2%	95.2%	95.2%	**0.0%**	**0.0%**	63.9%
Rolling Window 500 days	97.6%	95.2%	88.1%	90.5%	2.4%	**0.0%**	62.3%
Rolling Window 1,000 days	95.2%	95.2%	83.3%	83.3%	4.8%	**0.0%**	60.3%
Structural Breaks Window	**92.9%**	92.9%	88.1%	90.5%	9.5%	**0.0%**	62.3%
Comb. Mean All	97.6%	**90.5%**	85.7%	88.1%	**0.0%**	**0.0%**	60.3%
Comb. Mean All ex SB	97.6%	92.9%	88.1%	78.6%	**0.0%**	**0.0%**	59.5%
Comb. Mean Rolling Windows	97.6%	**90.5%**	88.1%	90.5%	**0.0%**	**0.0%**	61.1%
Comb. Long Short	**92.9%**	**90.5%**	90.5%	78.6%	2.4%	**0.0%**	59.1%
Comb. Trimmed Mean All	97.6%	95.2%	85.7%	88.1%	**0.0%**	**0.0%**	61.1%
Comb. Trimmed Mean All ex SB	97.6%	92.9%	85.7%	83.3%	**0.0%**	**0.0%**	59.9%
Comb. Trimmed Mean Rolling Windows	97.6%	90.5%	85.7%	92.9%	0.0%	0.0%	61.1%
Average	96.3%	93.0%	87.2%	86.4%	2.7%	**0.0%**	
1% VaR							
Average Absolute Deviation							
Expanding Window	2.038%	1.980%	1.831%	1.804%	2.006%	1.774%	1.905%
Rolling Window 125 days	2.094%	2.146%	1.833%	1.811%	2.426%	1.957%	2.045%
Rolling Window 250 days	**1.902%**	**1.719%**	**1.714%**	**1.589%**	**1.503%**	**1.342%**	**1.628%**
Rolling Window 500 days	1.982%	1.871%	1.742%	1.722%	1.707%	1.554%	1.763%
Rolling Window 1,000 days	2.021%	1.923%	1.775%	1.755%	1.993%	1.735%	1.867%
Structural Breaks Window	2.063%	2.016%	1.805%	1.789%	2.251%	1.885%	1.968%
Comb. Mean All	2.114%	1.977%	1.833%	1.798%	2.187%	1.829%	1.956%
Comb. Mean All ex SB	2.121%	1.980%	1.813%	1.786%	2.190%	1.808%	1.950%
Comb. Mean Rolling Windows	2.083%	1.927%	1.793%	1.767%	2.053%	1.750%	1.896%
Comb. Long Short	2.148%	2.013%	1.809%	1.745%	2.160%	1.761%	1.939%
Comb. Trimmed Mean All	2.077%	1.956%	1.813%	1.790%	2.140%	1.834%	1.935%
Comb. Trimmed Mean All ex SB	2.092%	1.954%	1.799%	1.771%	2.111%	1.807%	1.922%
Comb. Trimmed Mean Rolling Windows	2.017%	1.927%	1.782%	1.745%	1.912%	1.716%	1.850%
Average	2.058%	1.953%	1.796%	1.759%	2.049%	1.750%	
Backtest Rejections							
Expanding Window	**92.9%**	97.6%	**73.8%**	**66.7%**	19.0%	**0.0%**	58.3%
Rolling Window 125 days	97.6%	**90.5%**	95.2%	97.6%	**0.0%**	**0.0%**	63.5%
Rolling Window 250 days	97.6%	95.2%	95.2%	95.2%	**0.0%**	**0.0%**	63.9%
Rolling Window 500 days	97.6%	95.2%	88.1%	90.5%	2.4%	**0.0%**	62.3%
Rolling Window 1,000 days	95.2%	95.2%	85.7%	83.3%	4.8%	**0.0%**	60.7%
Structural Breaks Window	**92.9%**	92.9%	88.1%	92.9%	7.1%	**0.0%**	62.3%
Comb. Mean All	97.6%	**90.5%**	85.7%	88.1%	**0.0%**	**0.0%**	60.3%
Comb. Mean All ex SB	97.6%	92.9%	88.1%	81.0%	**0.0%**	**0.0%**	59.9%
Comb. Mean Rolling Windows	97.6%	**90.5%**	88.1%	90.5%	**0.0%**	**0.0%**	61.1%
Comb. Long Short	**92.9%**	**90.5%**	85.7%	76.2%	2.4%	**0.0%**	**57.9%**
Comb. Trimmed Mean All	97.6%	95.2%	83.3%	88.1%	**0.0%**	**0.0%**	60.7%
Comb. Trimmed Mean All ex SB	97.6%	92.9%	85.7%	83.3%	**0.0%**	**0.0%**	59.9%
Comb. Trimmed Mean Rolling Windows	97.6%	90.5%	85.7%	90.5%	**0.0%**	**0.0%**	60.7%
Average	96.3%	93.0%	86.8%	86.4%	2.7%	**0.0%**	

Table 5.10: Expected Shortfall - 10-Day Ahead Forecast

The table reports average absolute deviations of the ES and rejection rates of the ES backtest for the 10-day ahead forecasts and the 5% and 1% VaR level. The average absolute deviation is calculated as the average of the absolute deviations between the ES forecast and the realized shortfall in case of a VaR-violation. The rejection rate is computed by the number of rejections divided by the total number of performed ES backtests. For each VaR model, the smallest average absolute deviation is printed in bold. The rejection rates in bold are closest to the significance level of the test.

Estimation Window Strategy	Normal Distr.	EWMA	GARCH	GJR-GARCH	HS	Filtered HS	Average
5% VaR							
Average Absolute Deviation							
Expanding Window	2.004%	1.833%	1.684%	1.657%	2.515%	1.069%	1.794%
Rolling Window 125 days	2.315%	2.069%	1.690%	1.628%	**1.628%**	**0.968%**	**1.717%**
Rolling Window 250 days	2.055%	**1.762%**	1.756%	1.783%	2.464%	1.380%	1.867%
Rolling Window 500 days	**1.993%**	1.854%	1.700%	1.681%	2.439%	1.222%	1.815%
Rolling Window 1,000 days	2.104%	1.951%	**1.666%**	1.635%	2.563%	1.106%	1.838%
Structural Breaks Window	2.247%	1.982%	1.719%	1.684%	2.189%	1.047%	1.811%
Comb. Mean All	2.078%	1.877%	1.673%	**1.625%**	2.316%	1.028%	1.766%
Comb. Mean All ex SB	2.095%	1.887%	1.681%	1.645%	2.305%	1.038%	1.775%
Comb. Mean Rolling Windows	2.077%	1.846%	1.691%	1.664%	2.446%	1.066%	1.798%
Comb. Long Short	2.161%	1.847%	1.681%	1.677%	2.045%	1.003%	1.736%
Comb. Trimmed Mean All	2.065%	1.897%	1.676%	1.647%	2.327%	1.043%	1.776%
Comb. Trimmed Mean All ex SB	2.087%	1.918%	1.682%	1.668%	2.264%	1.030%	1.775%
Comb. Trimmed Mean Rolling Windows	2.050%	1.859%	1.696%	1.723%	2.395%	1.089%	1.802%
Average	1.819%	1.891%	1.692%	1.671%	1.724%	**1.084%**	
Backtest Rejections							
Expanding Window	**78.6%**	85.7%	85.7%	83.3%	61.9%	**45.2%**	**73.4%**
Rolling Window 125 days	85.7%	**83.3%**	**83.3%**	88.1%	71.4%	71.4%	80.6%
Rolling Window 250 days	85.7%	**83.3%**	**83.3%**	83.3%	69.0%	59.5%	77.4%
Rolling Window 500 days	85.7%	85.7%	85.7%	85.7%	69.0%	50.0%	77.0%
Rolling Window 1,000 days	88.1%	85.7%	85.7%	85.7%	**59.5%**	**45.2%**	75.0%
Structural Breaks Window	81.0%	**83.3%**	**83.3%**	**81.0%**	71.4%	**45.2%**	74.2%
Comb. Mean All	81.0%	85.7%	85.7%	83.3%	64.3%	47.6%	74.6%
Comb. Mean All ex SB	**78.6%**	85.7%	85.7%	83.3%	66.7%	52.4%	75.4%
Comb. Mean Rolling Windows	85.7%	85.7%	85.7%	83.3%	71.4%	54.8%	77.8%
Comb. Long Short	81.0%	85.7%	**83.3%**	**81.0%**	64.3%	50.0%	74.2%
Comb. Trimmed Mean All	83.3%	85.7%	**83.3%**	83.3%	66.7%	50.0%	75.4%
Comb. Trimmed Mean All ex SB	83.3%	85.7%	85.7%	83.3%	64.3%	52.4%	75.8%
Comb. Trimmed Mean Rolling Windows	85.7%	85.7%	83.3%	88.1%	73.8%	52.4%	78.2%
Average	83.3%	85.2%	84.6%	84.1%	67.2%	**52.0%**	
1% VaR							
Average Absolute Deviation							
Expanding Window	5.963%	5.834%	5.641%	5.703%	6.161%	5.689%	5.832%
Rolling Window 125 days	6.176%	6.243%	5.443%	5.524%	7.542%	5.749%	6.113%
Rolling Window 250 days	**5.780%**	**5.293%**	**5.254%**	**5.297%**	**4.068%**	**5.421%**	**5.186%**
Rolling Window 500 days	5.929%	5.620%	5.404%	5.389%	5.042%	5.697%	5.513%
Rolling Window 1,000 days	5.828%	5.871%	5.484%	5.438%	5.761%	5.850%	5.705%
Structural Breaks Window	6.118%	5.980%	5.434%	5.500%	6.743%	5.757%	5.922%
Comb. Mean All	6.125%	5.947%	5.461%	5.571%	6.896%	5.797%	5.966%
Comb. Mean All ex SB	6.140%	5.998%	5.424%	5.495%	6.925%	5.760%	5.957%
Comb. Mean Rolling Windows	6.123%	5.850%	5.430%	5.502%	6.331%	5.753%	5.831%
Comb. Long Short	6.200%	5.984%	5.377%	5.504%	6.645%	5.640%	5.892%
Comb. Trimmed Mean All	6.112%	5.923%	5.464%	5.567%	6.549%	5.780%	5.899%
Comb. Trimmed Mean All ex SB	6.086%	5.899%	5.431%	5.472%	6.536%	5.783%	5.868%
Comb. Trimmed Mean Rolling Windows	6.028%	5.811%	5.369%	5.415%	5.781%	5.773%	5.696%
Average	5.950%	5.865%	**5.432%**	5.490%	5.918%	5.727%	
Backtest Rejections							
Expanding Window	**78.6%**	85.7%	85.7%	83.3%	61.9%	**45.2%**	**73.4%**
Rolling Window 125 days	83.3%	**83.3%**	83.3%	88.1%	71.4%	69.0%	79.8%
Rolling Window 250 days	85.7%	**83.3%**	83.3%	83.3%	71.4%	57.1%	77.4%
Rolling Window 500 days	85.7%	85.7%	85.7%	85.7%	69.0%	50.0%	77.0%
Rolling Window 1,000 days	88.1%	85.7%	85.7%	85.7%	**59.5%**	**45.2%**	75.0%
Structural Breaks Window	**78.6%**	**83.3%**	83.3%	**81.0%**	71.4%	**45.2%**	73.8%
Comb. Mean All	81.0%	85.7%	85.7%	83.3%	64.3%	47.6%	74.6%
Comb. Mean All ex SB	81.0%	85.7%	85.7%	83.3%	66.7%	50.0%	75.4%
Comb. Mean Rolling Windows	85.7%	85.7%	85.7%	**81.0%**	73.8%	57.1%	78.2%
Comb. Long Short	83.3%	85.7%	**81.0%**	**81.0%**	66.7%	50.0%	74.6%
Comb. Trimmed Mean All	83.3%	85.7%	83.3%	83.3%	66.7%	50.0%	75.4%
Comb. Trimmed Mean All ex SB	83.3%	85.7%	85.7%	83.3%	64.3%	50.0%	75.4%
Comb. Trimmed Mean Rolling Windows	85.7%	85.7%	85.7%	85.7%	73.8%	52.4%	78.2%
Average	83.3%	85.2%	84.6%	83.7%	67.8%	**51.5%**	

Table 5.11: Conditional Predictive Ability - 1-Day Ahead Forecasts

For the 1-day ahead forecasts, the table reports the proportions of CPA tests where an alternative estimation strategy is preferred compared to the expanding window strategy which serves as the benchmark. The significance level is set to 5%. For each VaR model, the results of the best alternative estimation window strategy are printed in bold.

		5% VaR			1% VaR		
VaR-Model	Alternative Strategy	Expanding Window Preferred	Indifferent	Alternative Strategy Preferred	Expanding Window Preferred	Indifferent	Alternative Strategy Preferred
Normal Distribution	Rolling Window 125 days	0.0%	19.0%	81.0%	0.0%	28.6%	71.4%
	Rolling Window 250 days	0.0%	19.0%	81.0%	2.4%	26.2%	71.4%
	Rolling Window 500 days	7.1%	31.0%	61.9%	0.0%	47.6%	52.4%
	Rolling Window 1,000 days	31.0%	23.8%	45.2%	16.7%	38.1%	45.2%
	Structural Breaks Window	7.1%	40.5%	52.4%	4.8%	40.5%	54.8%
	Comb. Mean All	0.0%	11.9%	88.1%	0.0%	7.1%	92.9%
	Comb. Mean All ex SB	0.0%	11.9%	88.1%	**0.0%**	**4.8%**	**95.2%**
	Comb. Mean Rolling Windows	0.0%	14.3%	85.7%	2.4%	16.7%	81.0%
	Comb. Long Short	**0.0%**	**9.5%**	**90.5%**	0.0%	7.1%	92.9%
	Comb. Trimmed Mean All	0.0%	16.7%	83.3%	2.4%	14.3%	83.3%
	Comb. Trimmed Mean All ex SB	0.0%	11.9%	88.1%	2.4%	11.9%	85.7%
	Comb. Trimmed Mean Rolling Windows	0.0%	16.7%	83.3%	2.4%	19.0%	78.6%
EWMA	Rolling Window 125 days	0.0%	23.8%	76.2%	0.0%	45.2%	54.8%
	Rolling Window 250 days	0.0%	21.4%	78.6%	0.0%	40.5%	59.5%
	Rolling Window 500 days	0.0%	23.8%	76.2%	2.4%	42.9%	54.8%
	Rolling Window 1,000 days	0.0%	28.6%	71.4%	0.0%	40.5%	59.5%
	Structural Breaks Window	0.0%	40.5%	59.5%	2.4%	47.6%	50.0%
	Comb. Mean All	0.0%	11.9%	88.1%	0.0%	28.6%	71.4%
	Comb. Mean All ex SB	0.0%	11.9%	88.1%	0.0%	28.6%	71.4%
	Comb. Mean Rolling Windows	0.0%	16.7%	83.3%	0.0%	38.1%	61.9%
	Comb. Long Short	**0.0%**	**9.5%**	**90.5%**	**0.0%**	**21.4%**	**78.6%**
	Comb. Trimmed Mean All	0.0%	11.9%	88.1%	0.0%	33.3%	66.7%
	Comb. Trimmed Mean All ex SB	0.0%	14.3%	85.7%	0.0%	38.1%	61.9%
	Comb. Trimmed Mean Rolling Windows	0.0%	19.0%	81.0%	0.0%	38.1%	61.9%
GARCH	Rolling Window 125 days	26.2%	73.8%	0.0%	14.3%	83.3%	2.4%
	Rolling Window 250 days	28.6%	71.4%	0.0%	23.8%	73.8%	2.4%
	Rolling Window 500 days	19.0%	78.6%	2.4%	14.3%	78.6%	7.1%
	Rolling Window 1,000 days	26.2%	64.3%	9.5%	11.9%	78.6%	9.5%
	Structural Breaks Window	23.8%	71.4%	4.8%	19.0%	73.8%	7.1%
	Comb. Mean All	**4.8%**	**85.7%**	**9.5%**	**4.8%**	**81.0%**	**14.3%**
	Comb. Mean All ex SB	7.1%	85.7%	7.1%	7.1%	78.6%	14.3%
	Comb. Mean Rolling Windows	16.7%	76.2%	7.1%	11.9%	76.2%	11.9%
	Comb. Long Short	9.5%	85.7%	4.8%	2.4%	90.5%	7.1%
	Comb. Trimmed Mean All	7.1%	83.3%	9.5%	9.5%	78.6%	11.9%
	Comb. Trimmed Mean All ex SB	9.5%	83.3%	7.1%	7.1%	81.0%	11.9%
	Comb. Trimmed Mean Rolling Windows	11.9%	83.3%	4.8%	16.7%	73.8%	9.5%
GJR-GARCH	Rolling Window 125 days	38.1%	61.9%	0.0%	42.9%	57.1%	0.0%
	Rolling Window 250 days	19.0%	81.0%	0.0%	21.4%	73.8%	4.8%
	Rolling Window 500 days	21.4%	76.2%	2.4%	14.3%	76.2%	9.5%
	Rolling Window 1,000 days	11.9%	81.0%	7.1%	9.5%	76.2%	14.3%
	Structural Breaks Window	21.4%	78.6%	0.0%	14.3%	78.6%	7.1%
	Comb. Mean All	4.8%	83.3%	11.9%	0.0%	88.1%	11.9%
	Comb. Mean All ex SB	**4.8%**	**81.0%**	**14.3%**	**4.8%**	**78.6%**	**16.7%**
	Comb. Mean Rolling Windows	9.5%	76.2%	14.3%	7.1%	88.1%	4.8%
	Comb. Long Short	11.9%	78.6%	9.5%	9.5%	85.7%	4.8%
	Comb. Trimmed Mean All	4.8%	85.7%	9.5%	7.1%	83.3%	9.5%
	Comb. Trimmed Mean All ex SB	**4.8%**	**81.0%**	**14.3%**	4.8%	83.3%	11.9%
	Comb. Trimmed Mean Rolling Windows	4.8%	85.7%	9.5%	2.4%	90.5%	7.1%
Historical Simulation	Rolling Window 125 days	0.0%	26.2%	73.8%	0.0%	61.9%	38.1%
	Rolling Window 250 days	0.0%	19.0%	81.0%	0.0%	59.5%	40.5%
	Rolling Window 500 days	7.1%	31.0%	61.9%	9.5%	57.1%	33.3%
	Rolling Window 1,000 days	28.6%	14.3%	57.1%	11.9%	54.8%	33.3%
	Structural Breaks Window	4.8%	42.9%	52.4%	7.1%	52.4%	40.5%
	Comb. Mean All	**0.0%**	**4.8%**	**95.2%**	0.0%	21.4%	78.6%
	Comb. Mean All ex SB	**0.0%**	**4.8%**	**95.2%**	**0.0%**	**19.0%**	**81.0%**
	Comb. Mean Rolling Windows	0.0%	7.1%	92.9%	0.0%	31.0%	69.0%
	Comb. Long Short	**0.0%**	**4.8%**	**95.2%**	**0.0%**	**19.0%**	**81.0%**
	Comb. Trimmed Mean All	0.0%	7.1%	92.9%	0.0%	23.8%	76.2%
	Comb. Trimmed Mean All ex SB	0.0%	7.1%	92.9%	0.0%	23.8%	76.2%
	Comb. Trimmed Mean Rolling Windows	0.0%	7.1%	92.9%	0.0%	38.1%	61.9%
Filtered Historical Simulation	Rolling Window 125 days	26.2%	73.8%	0.0%	35.7%	64.3%	0.0%
	Rolling Window 250 days	26.2%	73.8%	0.0%	21.4%	73.8%	4.8%
	Rolling Window 500 days	16.7%	83.3%	0.0%	16.7%	78.6%	4.8%
	Rolling Window 1,000 days	19.0%	76.2%	4.8%	9.5%	88.1%	2.4%
	Structural Breaks Window	31.0%	69.0%	0.0%	28.6%	69.0%	2.4%
	Comb. Mean All	7.1%	85.7%	7.1%	9.5%	81.0%	9.5%
	Comb. Mean All ex SB	4.8%	85.7%	9.5%	2.4%	88.1%	9.5%
	Comb. Mean Rolling Windows	9.5%	88.1%	2.4%	7.1%	85.7%	7.1%
	Comb. Long Short	4.8%	85.7%	9.5%	9.5%	88.1%	2.4%
	Comb. Trimmed Mean All	4.8%	85.7%	9.5%	11.9%	76.2%	11.9%
	Comb. Trimmed Mean All ex SB	**7.1%**	**81.0%**	**11.9%**	**2.4%**	**85.7%**	**11.9%**
	Comb. Trimmed Mean Rolling Windows	9.5%	90.5%	0.0%	16.7%	78.6%	4.8%

Table 5.12: Conditional Predictive Ability - 10-Day Ahead Forecasts

For the 10-day ahead forecasts, the table reports the proportions of CPA tests where an alternative estimation strategy is preferred compared to the expanding window strategy which serves as the benchmark. The significance level is set to 5%. For each VaR model, the results of the best alternative estimation window strategy are printed in bold.

		5% VaR			1% VaR		
VaR-Model	Alternative Strategy	Expanding Window Preferred	Indifferent	Alternative Strategy Preferred	Expanding Window Preferred	Indifferent	Alternative Strategy Preferred
Normal Distribution	Rolling Window 125 days	14.3%	66.7%	19.0%	11.9%	81.0%	7.1%
	Rolling Window 250 days	7.1%	54.8%	38.1%	9.5%	66.7%	23.8%
	Rolling Window 500 days	16.7%	52.4%	31.0%	9.5%	69.0%	21.4%
	Rolling Window 1,000 days	14.3%	54.8%	31.0%	2.4%	76.2%	21.4%
	Structural Breaks Window	7.1%	61.9%	31.0%	4.8%	69.0%	26.2%
	Comb. Mean All	4.8%	54.8%	40.5%	**7.1%**	**59.5%**	**33.3%**
	Comb. Mean All ex SB	4.8%	57.1%	38.1%	11.9%	54.8%	33.3%
	Comb. Mean Rolling Windows	7.1%	61.9%	31.0%	11.9%	57.1%	31.0%
	Comb. Long Short	**4.8%**	**45.2%**	**50.0%**	2.4%	66.7%	31.0%
	Comb. Trimmed Mean All	9.5%	50.0%	40.5%	9.5%	64.3%	26.2%
	Comb. Trimmed Mean All ex SB	4.8%	54.8%	40.5%	9.5%	57.1%	33.3%
	Comb. Trimmed Mean Rolling Windows	7.1%	54.8%	38.1%	11.9%	57.1%	31.0%
EWMA	Rolling Window 125 days	19.0%	69.0%	11.9%	19.0%	69.0%	11.9%
	Rolling Window 250 days	19.0%	66.7%	14.3%	11.9%	73.8%	14.3%
	Rolling Window 500 days	9.5%	71.4%	19.0%	11.9%	76.2%	11.9%
	Rolling Window 1,000 days	9.5%	69.0%	21.4%	14.3%	71.4%	14.3%
	Structural Breaks Window	7.1%	73.8%	19.0%	11.9%	73.8%	14.3%
	Comb. Mean All	11.9%	61.9%	26.2%	7.1%	76.2%	16.7%
	Comb. Mean All ex SB	11.9%	59.5%	28.6%	7.1%	78.6%	14.3%
	Comb. Mean Rolling Windows	14.3%	61.9%	23.8%	9.5%	76.2%	14.3%
	Comb. Long Short	**9.5%**	**59.5%**	**31.0%**	**9.5%**	**71.4%**	**19.0%**
	Comb. Trimmed Mean All	11.9%	66.7%	21.4%	9.5%	73.8%	16.7%
	Comb. Trimmed Mean All ex SB	9.5%	64.3%	26.2%	9.5%	76.2%	14.3%
	Comb. Trimmed Mean Rolling Windows	16.7%	61.9%	21.4%	9.5%	76.2%	14.3%
GARCH	Rolling Window 125 days	16.7%	83.3%	0.0%	19.0%	81.0%	0.0%
	Rolling Window 250 days	14.3%	83.3%	2.4%	14.3%	83.3%	2.4%
	Rolling Window 500 days	7.1%	78.6%	14.3%	0.0%	90.5%	9.5%
	Rolling Window 1,000 days	11.9%	78.6%	9.5%	0.0%	95.2%	4.8%
	Structural Breaks Window	**2.4%**	**83.3%**	**14.3%**	2.4%	88.1%	9.5%
	Comb. Mean All	4.8%	85.7%	9.5%	0.0%	88.1%	11.9%
	Comb. Mean All ex SB	4.8%	85.7%	9.5%	**2.4%**	**83.3%**	**14.3%**
	Comb. Mean Rolling Windows	4.8%	85.7%	9.5%	**2.4%**	**83.3%**	**14.3%**
	Comb. Long Short	9.5%	83.3%	7.1%	9.5%	85.7%	4.8%
	Comb. Trimmed Mean All	7.1%	88.1%	4.8%	2.4%	92.9%	4.8%
	Comb. Trimmed Mean All ex SB	7.1%	85.7%	7.1%	0.0%	90.5%	9.5%
	Comb. Trimmed Mean Rolling Windows	7.1%	85.7%	7.1%	4.8%	85.7%	9.5%
GJR-GARCH	Rolling Window 125 days	19.0%	73.8%	7.1%	28.6%	69.0%	2.4%
	Rolling Window 250 days	14.3%	78.6%	7.1%	14.3%	83.3%	2.4%
	Rolling Window 500 days	9.5%	78.6%	11.9%	7.1%	85.7%	7.1%
	Rolling Window 1,000 days	9.5%	76.2%	14.3%	2.4%	92.9%	4.8%
	Structural Breaks Window	0.0%	90.5%	9.5%	2.4%	85.7%	11.9%
	Comb. Mean All	11.9%	73.8%	14.3%	**0.0%**	**85.7%**	**14.3%**
	Comb. Mean All ex SB	9.5%	81.0%	9.5%	4.8%	81.0%	14.3%
	Comb. Mean Rolling Windows	7.1%	81.0%	11.9%	7.1%	83.3%	9.5%
	Comb. Long Short	2.4%	88.1%	9.5%	16.7%	71.4%	11.9%
	Comb. Trimmed Mean All	7.1%	76.2%	16.7%	**0.0%**	**85.7%**	**14.3%**
	Comb. Trimmed Mean All ex SB	**7.1%**	**73.8%**	**19.0%**	0.0%	88.1%	11.9%
	Comb. Trimmed Mean Rolling Windows	11.9%	76.2%	11.9%	0.0%	90.5%	9.5%
Historical Simulation	Rolling Window 125 days	21.4%	71.4%	7.1%	40.5%	59.5%	0.0%
	Rolling Window 250 days	23.8%	66.7%	9.5%	14.3%	76.2%	9.5%
	Rolling Window 500 days	21.4%	61.9%	16.7%	14.3%	73.8%	11.9%
	Rolling Window 1,000 days	21.4%	64.3%	14.3%	16.7%	69.0%	14.3%
	Structural Breaks Window	16.7%	61.9%	21.4%	14.3%	61.9%	23.8%
	Comb. Mean All	11.9%	59.5%	28.6%	7.1%	69.0%	23.8%
	Comb. Mean All ex SB	9.5%	64.3%	26.2%	4.8%	71.4%	23.8%
	Comb. Mean Rolling Windows	11.9%	66.7%	21.4%	9.5%	76.2%	14.3%
	Comb. Long Short	**7.1%**	**61.9%**	**31.0%**	**7.1%**	**64.3%**	**28.6%**
	Comb. Trimmed Mean All	9.5%	64.3%	26.2%	9.5%	73.8%	16.7%
	Comb. Trimmed Mean All ex SB	7.1%	64.3%	28.6%	7.1%	73.8%	19.0%
	Comb. Trimmed Mean Rolling Windows	11.9%	69.0%	19.0%	9.5%	81.0%	9.5%
Filtered Historical Simulation	Rolling Window 125 days	38.1%	57.1%	4.8%	33.3%	64.3%	2.4%
	Rolling Window 250 days	16.7%	76.2%	7.1%	28.6%	69.0%	2.4%
	Rolling Window 500 days	14.3%	76.2%	9.5%	14.3%	78.6%	7.1%
	Rolling Window 1,000 days	**4.8%**	**83.3%**	**11.9%**	**7.1%**	**81.0%**	**11.9%**
	Structural Breaks Window	9.5%	83.3%	7.1%	14.3%	81.0%	4.8%
	Comb. Mean All	7.1%	81.0%	11.9%	9.5%	81.0%	9.5%
	Comb. Mean All ex SB	9.5%	81.0%	9.5%	7.1%	85.7%	7.1%
	Comb. Mean Rolling Windows	9.5%	83.3%	7.1%	7.1%	85.7%	7.1%
	Comb. Long Short	16.7%	76.2%	7.1%	23.8%	71.4%	4.8%
	Comb. Trimmed Mean All	4.8%	85.7%	9.5%	11.9%	81.0%	7.1%
	Comb. Trimmed Mean All ex SB	4.8%	85.7%	9.5%	7.1%	85.7%	7.1%
	Comb. Trimmed Mean Rolling Windows	9.5%	81.0%	9.5%	9.5%	83.3%	7.1%

Similar to the rolling windows, for the 1-day ahead forecasts the structural breaks strategy outperforms the expanding windows in the settings where the normal distribution, EWMA, and historical simulation models are used. Regarding the 10-day ahead forecasts, determining the estimation windows by structural break tests tend to be preferable for all VaR models, except for the filtered historical simulation approach.

The conditional predictive ability of the combination strategies is at least as good as, and in most cases better than, the expanding window strategy. Again, this applies in particular for the simple VaR models. The results of the CPA test do not reveal that one of the different combination approaches are clearly superior. However, the proportion where the expanding window strategy is outperformed by trimmed combinations tends to be slightly smaller compared to their plain counterparts.

5.5 Conclusion

Compared to the large number of VaR-forecasting models proposed in the literature, there are relatively little contributions to the question of which estimation window strategy is preferable to forecast common risk measures like VaR and ES. To this end, we perform an empirical study on the basis of returns of German blue chip stocks where thirteen different estimation window strategies are applied to a set of of seven different parametric, semi-parametric, and non-parametric VaR models. These strategies include simple approaches like expanding windows and rolling windows of different lengths as well as a more complex model that determines the length of an estimation window by using a test for detecting structural breaks in the variance of asset return series. In addition, we investigate combination strategies where the VaR-forecasts of several different models are pooled. We evaluate the VaR-forecasts of the different approaches by backtesting the uc and the i.i.d. properties of VaR-violations, the ES forecasting accuracy, and the conditional predictive ability.

The empirical study provides several interesting results. We demonstrate that not only the application of different VaR models but also the selection of the estimation window strategy leads to significant differences in the results. Considering the uc property of VaR-violations, the VaR-forecasts estimated by using the rolling window strategy become more conservative as the size of the rolling windows increases. Compared to the expanding and rolling window strategies, using structural break tests leads to a lower number of VaR-violations on average over all VaR models. Interestingly, the VaR-violation ratio of forecast combinations are lower than the mean of their individual component ratios. Considering the i.i.d. property of VaR-violations, short rolling windows are preferable for simple VaR models like normal distribution and EWMA. With respect to the remaining VaR models, forecast combinations show lower rejection rates than the competing strategies. Focusing on the average absolute deviations between the ES forecasts and the realized shortfalls in case of a VaR-violation, the rolling window strategy outperforms the competing strategies on average. However, the differences in the rejection rates of the statistical ES backtests caused by different estimation window strategies are relatively small. The comparison of the expanding windows as the benchmark strategy to the remaining strategies by the CPA test reveals that rolling windows as well as the structural break strategy are preferable when less-parameterized VaR models are applied. The combination strategies have an equal or better CPA compared to the expanding window benchmark in the vast majority of settings. In summary, although each estimation window strategy has its own strengths, the usage of forecast combinations seems to be the preferable estimation window strategy, because it shows convincing results in most settings and for most backtests and has less weaknesses compared to the remaining approaches.

Kapitel 6

Testing for Structural Breaks in Correlations: Does it Improve Value-at-Risk Forecasting?

Zur Veröffentlichung eingereicht in:

Journal of Empirical Finance (zusammen mit Gregor N.F. Weiß und Dominik Wied).

6.1 Introduction

It has become a stylized fact in the analysis of financial market data that correlations between asset returns are time-varying. Bollerslev et al. [1988] were among the first to stress the importance of accounting for dynamic covariances in international asset pricing. Further empirical evidence for time-varying asset correlations is found by Longin and Solnik [1995] and Ang and Bekaert [2002] who show that correlations between international equity markets increased over time and were higher in the high volatility regimes of bear markets.[1] In response to these findings, studies in the field of financial econometrics in recent years have tried to model the dynamics in asset correlations. Most notably, Engle [2002] proposed the Dynamic Conditional Correlation (DCC) model that combines the flexibility of univariate generalized autoregressive conditional heteroskedasticity (GARCH) models but at the same time circumvents the necessity to estimate a large number of parameters. Hillebrand [2005] showes that neglecting structural breaks in the data generating parameters of the conditional variance of GARCH models causes substanti-

[1] Evidence of correlations changing over time is also found by Pelletier [2006] and Colacito et al. [2011].

al estimation errors. As the finance literature still lacks a study dealing with the impact of structural parameter changes on the estimation of dynamic correlation models, it is an interesting question whether models accommodating for correlation shifts are able to outperform a standard dynamic correlation model.

In this paper, we investigate the question whether the constant conditional correlation (CCC) model of Bollerslev [1990] and the DCC model introduced by Engle [2002] and Engle and Sheppard [2001][2] are economically significantly outperformed with respect to their VaR and ES forecasting accuracy by CCC and DCC models combined with recently proposed tests for structural breaks in a) the pairwise correlations, b) the correlation matrix and c) the covariance matrix of asset returns to yield a set of seven candidate models with a diverse range of modeling flexibility.[3] More precisely, we modify the plain CCC and DCC benchmark models by combining them with the pairwise test for constant correlations of Wied et al. [2012b], the test for a constant correlation matrix of Wied [2012], and the test for a constant covariance matrix of Aue et al. [2009].[4] The motivation for choosing these three tests lies in the fact that they are nonparametric and do not impose restrictive assumptions on the structure of the time series. We conduct a horse race of these models and compare their out-of-sample forecasting accuracy by using ten-dimensional portfolios composed of international blue-chip stocks. Despite the benefits of the DCC model, the inclusion of the classical CCC model of Bollerslev [1990] in this study is useful for several reasons. First, contrary to the DCC model, the CCC model allows for a pairwise test for structural breaks in correlations.[5] Second, a simple CCC model that accounts for structural breaks in correlations could yield comparable accurate VaR-forecasts without imposing too strict assumptions on the dynamic behavior of correlations over time. Third, the empirical literature is ambiguous about the comparative performance of CCC and DCC models. For example, Santos et al. [2013] find in a comprehensive VaR

[2]See also Tse and Tsui [2002].

[3]As the focus of our paper lies on the modeling of the dynamics in the dependence structure between assets, we do not consider structural breaks in the assets' univariate volatilities. For a review of methods used for forecasting stock return volatility, see Poon and Granger [2003]. Structural breaks in volatility are examined, for example, by Rapach and Strauss [2008].

[4]As we will explain later, the test of constant pairwise correlations cannot be combined with the DCC model. Therefore, only seven instead of eight models are compared in our study.

[5]See Section 6.4 (3) for more details.

predictive ability comparison study that the performances of a CCC and a DCC model are comparable. Additionally, the results of their study indicate that the generalization of the DCC model proposed by Cappiello et al. [2006] does not lead to any significant improvements. Consequently, we abstain from implementing more sophisticated versions of the DCC model in this empirical study. The model performance is assessed by performing formal backtests of VaR- and Expected Shortfall (ES)- forecasts using the unconditional coverage test of Kupiec [1995], the CAViaR based test of Engle and Manganelli [2004] and Berkowitz et al. [2011], the ES backtest of McNeil and Frey [2000], the conditional predictive ability (CPA) test of Giacomini and White [2006] and a backtest procedure based on the Basel guidelines for backtesting internal models.

The contributions of our paper are numerous and important. First, we propose the use of tests for structural breaks in correlations and covariances together with static and dynamic correlation-based models for forecasting the VaR of asset portfolios. Second, to the best knowledge of the authors, this study presents the first empirical analysis of the question whether static and dynamic correlation-based VaR-models can be improved by additionally testing for structural breaks in correlations. Third, in a risk management context we empirically test which of the tests for structural breaks (pairwise correlations, correlation matrix and covariance matrix) is best suited for capturing significant changes in the correlations on financial assets.

The paper proceeds as follows. In Section 6.2, we quickly review the standard GARCH(1,1) model we use as marginal models in our study. In Section 6.3, we discuss the multivariate dependence models as well as the tests for structural breaks in correlations used in our empirical study. Section 6.4 presents the data and outlines the test procedure of our empirical study. The results of the empirical study are presented in Section 6.5. Section 6.6 concludes.

6.2 Univariate GARCH Model

GARCH-type models (see Bollerslev, 1986) have become the de-facto standard for describing the univariate behaviour of financial returns in a dynamic setting. The GARCH(1,1) model has been found to be the model of choice in the literature (see Hansen and Lunde, 2005). Consequently, in the empirical study we opt for the simple GARCH(1,1) as the standard model to forecast the volatility of the univariate marginals.

Let $r_{t,i}$ denote the log-return of an asset i $(i = 1, \ldots, n)$ at time t $(t = 0, 1, \ldots, T)$. Then the GARCH(1,1) process is defined by

$$r_{t,i} = \mu_i + \epsilon_{t,i} \tag{6.1}$$

$$\epsilon_{t,i} = \sigma_{t,i} z_{t,i} \tag{6.2}$$

$$\sigma^2_{t,i} = \alpha_{0,i} + \alpha_{1,i}\epsilon^2_{t-1,i} + \beta_{1,i}\sigma^2_{t-1,i} \tag{6.3}$$

where $\alpha_{0,i} > 0$ and $\alpha_{1,i} \geq 0$, $\beta_{1,i} \geq 0$ ensures a positive value of $\sigma^2_{t,i}$, and wide-sense stationarity requires $\alpha_{1,i} + \beta_{1,i} < 1$. Along the lines of Bollerslev and Wooldridge [1992], the innovations $z_{t,i}$ follow a strict white noise process from a Student's t distribution with mean 0, a scale parameter of 1, and $\nu > 2$ degrees of freedom. After estimating the parameters of the univariate GARCH models with, for example, maximum likelihood, one-step-ahead forecasts for the conditional variances are simulated from equation (6.3) for each of the n assets in a portfolio separately via plug-in estimation of

$$\sigma^2_{t+1,i} = \alpha_{0,i} + \alpha_{1,i}\epsilon^2_{t,i} + \beta_{1,i}\sigma^2_{t,i}. \tag{6.4}$$

6.3 Multivariate Dependence Models

In the following, the dependence models used in the empirical study are discussed. The selection includes five models employing statistical tests for the occurrence of structural breaks in the dependence structure and, for benchmarking purposes, the classical CCC- and DCC-GARCH models.

6.3.1 General Setup of Correlation-Based Dependence Models

The general definition of a multivariate GARCH model with linear dependence can be written as

$$r_t = \mu_t + \Sigma_t^{1/2} Z_t \qquad (6.5)$$

where r_t is a $(n \times 1)$ vector of log returns, μ_t is a $(n \times 1)$ vector of $\mathbb{E}(r_t)$ which we assume to be constant, and $\Sigma_t^{1/2}$ is the Cholesky factor of a positive definite conditional covariance matrix Σ_t which corresponds to the variance σ_t^2 in the univariate GARCH model. Furthermore, the innovations Z_t correspond to $z_{t,i}$ of the univariate GARCH process and are assumed to come from a Student's t distribution as described above. The conditional covariance matrix Σ_t can be expressed as

$$\Sigma_t = D_t P_t D_t \qquad (6.6)$$

where D_t is a $(n \times n)$ diagonal volatility matrix with the univariate conditional standard deviations $\sigma_{t,i}$ derived from (6.3) as its diagonal entries and $P_t = [\rho_{t,ij}]$ is a $(n \times n)$ positive definite correlation matrix where $\rho_{t,ii} = 1$ and $|\rho_{t,ij}| < 1$. From this it follows that the off-diagonal elements are defined as

$$[\Sigma_t]_{ij} = \sigma_{t,i}\sigma_{t,j}\rho_{t,ij}, \qquad i \neq j.$$

Our empirical study examines the one-step-ahead prediction of Value-at-Risk and Expected Shortfall. As we assume μ_t to be constant, the prediction solely depends on the forecast of the conditional covariance matrix $\Sigma_{t+1} = D_{t+1}P_{t+1}D_{t+1}$. Note that in our case, estimation of the univariate variances takes place before estimating the correlation matrices. For this reason and since the forecasts of univariate variances are identical for all examined dependence models, divergences in the performance of VaR- and ES-prediction thus depend only on the selected model to forecast the correlation matrix P_{t+1}.

6.3.2 Constant and Dynamic Conditional Correlation Models

The Constant Conditional Correlation GARCH model by Bollerslev [1990] constitutes a basic concept to specify the dependence structure of a given data set, since the conditional correlations are assumed to be constant over time. Let Σ_t be the conditional covariance matrix in a CCC-GARCH(1,1) process at time t. Corresponding to equations (6.5) and (6.6), the one-step-ahead forecast of the conditional covariance matrix can be obtained by a plug-in estimation of $\Sigma_{t+1} = D_{t+1} P_c D_{t+1}$. The correlation matrix P_c is assumed to be constant over time and its entries can be estimated with the arithmetic mean of products of the standardized residuals $\hat{z}_{t,i}$ [see Bollerslev, 1990, for details]. Here, $\hat{z}_{t,i} = \hat{\epsilon}_{t,i}\hat{\sigma}_{t,i}^{-1}$, where $\hat{\sigma}_{t,i}$ is the (plug-in-) estimated conditional standard deviation based on (6.3) and $\hat{\epsilon}_{t,i} = r_{t_i} - \hat{\mu}_i$. D_{t+1} is determined by the univariate conditional variances $\sigma^2_{t+1,i}$ obtained from (6.4) which are estimated by the plug-in method. The simplification of a constant dependence structure makes the model quite easy to estimate, in particular for high-dimensional portfolios. Due to its relatively simple design and its lasting popularity in the financial industry, we use the CCC-GARCH model as a useful benchmark. Furthermore, in contrast to the DCC model, the CCC model is combinable with the pairwise test for constant correlations of Wied et al. [2012b].

Several studies starting with the seminal work by Longin and Solnik [1995] show that correlations of asset returns are not constant over time. Therefore, as a generalization of the CCC model, Engle [2002] and Engle and Sheppard [2001] propose the Dynamic Conditional Correlation (DCC) GARCH model which allows the conditional correlation matrix to vary over time. The conditional covariance matrix is decomposed into conditional standard deviations and a correlation matrix via $\Sigma_t = D_t P_t D_t$. The correlation matrix P_t is assumed to be time-varying and is defined as

$$P_t = Q_t^{*-1} Q_t Q_t^{*-1}. \tag{6.7}$$

The time-varying character of the DCC-GARCH model is implemented by

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha(z_{t-1} z_{t-1}^T) + \beta Q_{t-1}. \tag{6.8}$$

Q_t^* is a diagonal matrix composed of the square root of the diagonal elements of Q_t and $\bar{Q}$ is the unconditional covariance matrix of the innovations $z_{t-1,i}$. The DCC parameters have to satisfy $\alpha \leq 1$, $\beta \leq 1$ and $\alpha + \beta < 1$. The one-step-ahead forecast of the conditional covariance matrix can then be obtained as a plug-in estimator of $\Sigma_{t+1} = D_{t+1} P_{t+1} D_{t+1}$. Here, D_{t+1} is determined by the univariate conditional variances $\sigma_{t+1,i}^2$ obtained from (6.4) and the conditional correlation matrix P_{t+1} is determined by $Q_{t+1} = (1 - \alpha - \beta)\bar{Q} + \alpha(z_t z_t^T) + \beta Q_t$ derived from (6.8). For details concerning the (maximum-likelihood) estimation of P_t, we refer to Engle [2002].

6.3.3 Tests for Structural Breaks in Correlations

In general, correlation based GARCH models can be extended by allowing for structural breaks in the dependence measure. We employ three recently proposed tests to detect structural breaks in P as well as in Σ and reestimate P after each change point. The basic motivation for using these tests is the fact that we want to know which data of the past we can use for estimating the correlation or covariance matrix. All three tests basically have the same structure: One compares the successively estimated quantities (bivariate correlations, correlation matrix, covariance matrix) with the corresponding quantities estimated from the whole sample and rejects the null of no-change if the difference becomes too large over time. All three tests work under mild conditions on the time series which makes them applicable to financial data. They are nonparametric in the sense that one does not need to assume a particular distribution such as a specific copula model or the normal distribution. Moreover, the tests allow for some serial dependence such that it is possible to apply the test on, for example, GARCH models. Principally, weak-sense stationarity is required for applying the fluctuation tests. While this is fulfilled in GARCH models under certain conditions, conditional heteroscedasticity might be a problem for the correlation tests as the tests might reject the null too often. To circumvent this problem, one can apply some kind of pre-filtering on the data. One potential drawback is the fact that it is a necessary condition to have finite fourth moments for deriving the asymptotic null distributions of the tests. While there is some evidence that second moments do exist in financial return data, the existence of finite fourth moments is doubtful. Nevertheless, we consider the fluctuation test to be applicable on returns as well. In the following, we will shortly present each test together with its respective null distributions.

6.3.3.1 Pairwise test for constant correlation

Wied et al. [2012b] propose a fluctuation test for constant bivariate correlations. The test compares the successively estimated bivariate correlation coefficients with the correlation coefficient from the whole sample. The test statistic is given by

$$\hat{D} \max_{2 \leq j \leq T} \frac{j}{\sqrt{T}} |\hat{\rho}_j - \hat{\rho}_T|, \tag{6.9}$$

where $\hat{D}$ is an estimator described in Wied et al. [2012b] that captures serial dependence and fluctuations of higher moments and serves for standardization. Also, the factor $\frac{j}{\sqrt{T}}$ serves for standardization, meaning that it compensates for the fact that correlations are in general better estimated for larger time series. The null hypothesis of constant correlation is rejected for too large values of the test statistic. Since the correlation test is designed for a bivariate vector, we control each entry of the population correlation matrix separately with this test. That means, we determine for each entry separately which data is used for its estimation. Under the null hypothesis of constant correlation, the test statistic converges to $\sup_{0 \leq z \leq 1} |B(z)|$, where B is a one-dimensional standard Brownian bridge. Under a sequence of local alternatives, the test statistic converges against $\sup_{0 \leq z \leq 1} |B(z) + C(z)|$, where C is a deterministic function.

6.3.3.2 Test for a constant multivariate correlation matrix

Wied [2012] proposes an extension of the bivariate correlation test to a d-dimensional correlation matrix. The test statistic in this case is rather similar to the former case with the difference that one does not just consider one deviation

$$|\hat{\rho}_j - \hat{\rho}_T|,$$

but the sum over all "bivariate deviations", that means,

$$\sum_{1 \leq i,j \leq p, i \neq j} \frac{k}{\sqrt{T}} \left| \hat{\rho}_k^{ij} - \hat{\rho}_T^{ij} \right| .$$

Also, the estimator $\hat{D}$ is calculated differently. While the bivariate test uses a kernel-based estimator, the multivariate test uses a block bootstrap estimator, see Wied [2012] for details. Under the null hypothesis of a constant correlation matrix, the test statistic converges to $\sup_{0 \leq z \leq 1} \sum_{i=1}^{d(d-1)/2} |B_i(z)|$, where $(B_i(z), z \in [0,1]), i = 1, \ldots, d(d-1)/2$ are independent standard Brownian bridges. Under local alternatives, we have convergence results that are similar to the ones with the former test.

6.3.3.3 Test for a constant multivariate covariance matrix

Aue et al. [2009] present a nonparametric fluctuation test for a constant d-dimensional covariance matrix of the random vectors $X_1, \ldots, X_T$ with $X_j = (X_{j,1}, \ldots, X_{j,d})$. Let $\mathsf{vech}(\cdot)$ denote the operator which stacks the columns on and below the diagonal of a $d \times d$ matrix into a vector and let A' be the transpose of a matrix A. At first, we consider the term

$$S_j = \frac{j}{\sqrt{T}} \left(\frac{1}{j} \sum_{l=1}^{j} \mathsf{vech}(X_l X_l') - \frac{1}{T} \sum_{l=1}^{T} \mathsf{vech}(X_l X_l') \right),$$

for $1 \leq j \leq T$, which measures the fluctuations of the estimated covariance matrices. Here, the factor $\frac{j}{\sqrt{T}}$ again serves for standardization for the same reasons as described above. The test statistic is then defined as $\max_{1 \leq j \leq T} S_j' \hat{E} S_j$, where $\hat{E}$ is an estimator which has the same structure as in the bivariate correlation test and is described in more detail in

Aue et al. [2009]. The limit distribution under the null hypothesis is the distribution of

$$\sup_{0 \leq z \leq 1} \sum_{i=1}^{d(d+1)/2} B_i^2(z),$$

where $(B_i(z), z \in [0,1]), i = 1, \ldots, d(d+1)/2$ are independent Brownian bridges.

Aue et al. [2009] show that the test is consistent against fixed alternatives. Note that the application of the test requires the assumption of constant first moments of the random vectors of the time series. The asymptotic result is derived under the assumption of zero expectation; if we had constant non-zero expectation, it would be necessary to subtract the arithmetic mean calculated from all observations from the original data which does not change the asymptotic distribution.

6.4 Data and Test Procedure

Our empirical study is designed as follows:

(1) **Data and portfolio composition:** We compute log returns by using daily total return quotes of stocks listed on the indices AEX, DAX30, CAC40, FTSE100, IBEX35, and the S&P500. With respect to each of the six stock indices, we build a portfolio consisting of ten equal weighted assets which possess the highest market values on June 30, 2012 and meet the requirement of a complete data history. The data set for each of the portfolios contains log returns of $4,970$ trading days (we exclude non-trading days from our sample). The quotes cover a period from the autumn of 1992 to June 30, 2012. All quotes are obtained from *Thomson Reuters Financial Datastream*. Table 6.1 presents summary statistics for the log-returns of each portfolio.

    The annualized volatility of the (unconditional) portfolio log-returns ranges from 18.33% to 23.98% while all six portfolios show significant positive annualized returns above 12%. Furthermore, the summary statistics show evidence of leptokurtic portfolio returns indicating fat tails.

(2) **Univariate modeling:** To forecast the volatility of each asset in each portfolio at day

Table 6.1: Summary Statistics

Summary statistics of the data set used for the empirical study. The data set consists of 4,970 (unconditional) log-returns for each of the six portfolios covering a period from the autumn 1992 to June 30, 2012. Mean Return p.a. and Volatility p.a. are annualized with 250 days.

	Portfolio					
	AEX	CAC	DAX	FTSE	IBEX	S&P
Minimum	-8.914%	-9.342%	-9.650%	-8.514%	-8.929%	-8.780%
5% Quantile	-1.886%	-2.135%	-2.379%	-1.700%	-2.102%	-1.872%
Mean Return	0.049%	0.049%	0.051%	0.057%	0.052%	0.061%
Median Return	0.085%	0.086%	0.123%	0.063%	0.089%	0.080%
95% Quantile	1.853%	2.094%	2.211%	1.740%	2.136%	1.856%
Maximum	8.123%	11.285%	11.947%	9.392%	12.329%	10.990%
Volatility	1.250%	1.396%	1.517%	1.159%	1.381%	1.236%
Skewness	-0.186	0.077	-0.217	-0.066	-0.023	0.046
Excess Kurtosis	5.471	5.300	5.291	5.893	5.363	6.713
Mean Return p.a.	12.24%	12.20%	12.73%	14.25%	12.91%	15.21%
Volatility p.a.	19.76%	22.07%	23.98%	18.33%	21.84%	19.55%

$t+1$, GARCH(1,1) models are fitted to a moving time window consisting of the $1,000$ preceding log returns. The use of a moving time window of $1,000$ days is common in the literature and is in line with, e.g., McNeil et al. [2005] and Kuester et al. [2006]. Next, a one-step-ahead volatility forecast $\sigma_{t+1,i}$ is computed by the use of the estimated GARCH parameters α_0, α_1 and β_1 according to (6.4). Furthermore, degrees of freedom of the marginals are held to be constant at $\nu_c = 15$.

(3) **Testing for structural breaks and multivariate modeling:** The correlations P_c and P_t of the plain CCC and DCC models are fitted to a sample consisting of the standardized residuals obtained from the univariate GARCH estimation. Therefore, the sample includes a moving time-window of 1,000 trading days preceding the forecast day $t + 1$. We opt for a moving time window rather than for a fixed time-window, because a fixed time-window does not account for any changes in the correlation structure. As a second alternative, an expanding time-window could be used which is determined by a fixed starting point and a moving end. However, we do not use such a time-window, because the weighting of more recent data for the parameter fitting decreases when the time-window increases over time. In conclusion, the moving time-window approach allows the estimated parameter to change and therefore it is a

benchmark which is hard to beat.

The estimation of the CCC and DCC parameters in combination with each of the three different tests for structural breaks is designed as follows. Similar to Wied [2013], we apply the structural break tests to the standardized residuals $\hat{z}_{t,i}$ of a moving time-window of a constant length at each point in time t. Here, $\hat{z}_{t,i} = \hat{\epsilon}_{t,i}\hat{\sigma}_{t,i}^{-1}$, where $\hat{\sigma}_{t,i}$ is the (plug-in-) estimated conditional standard deviation based on (6.3) and $\hat{\epsilon}_{t,i} = r_{t_i} - \hat{\mu}_i$. For the purpose of this study, the time-window consists of $1,000$ trading days preceding the forecast day $t + 1$. In order to decide at which point in time a possible change occurs we use an algorithm based on Galeano and Wied [2014]. First, within the sample of $1,000$ trading days we identify the data point at which the test statistic takes its maximum. If this maximum is equal to or above the critical value, the null of a constant correlation/covariance is rejected.[6] In this case, the data point is a natural estimator of a so called dominating change point. Second, at this point we split the sample into two parts and search for possible change points again in the latter part of the sample. The procedure stops if no new change point is detected. Finally, the constant correlation coefficient P_c and the time-varying correlation coefficient P_t are estimated on the basis the standardized residuals of a subsample, which starts at the day of the latest detected change point and ends at day t. The sample size for estimating P is limited to $[100, \ldots, 1,000]$. Because we perform the tests on a daily basis, the nominal significance level might not be attained. Following Wied [2013], we do not address this topic within this study as we simply use the decisions of the tests in an explorative way. Note that in case of the application of the pairwise test for constant correlations this procedure is conducted for each of the off-diagonal elements of the correlation matrix. Because the resulting subsamples for each element are typically of different lengths, the estimation of DCC parameters is not feasible. Therefore, this test is only applied in combination with the CCC model.

[6]The critical values are computed for a significance level of 5% for each of the three structural break tests. We also tested a setup including a significance level of 1%. However, the forecasting results tend to be slightly worse. With respect to the test for a constant correlation matrix, we use a bootstrap approximation for a normalizing constant in order to approximate the asymptotic limit distribution of the test statistic. In line with Wied [2012], we chose 199 bootstrap replications.

Concerning the test for a constant covariance matrix, Aue et al. [2009] approximate asymptotic critical values by simulating Brownian bridges on a fine grid. Wied et al. [2013b] show that for a small sample size this approach leads to considerably overestimated critical values and hence to very infrequent rejections. To this end, based on Wied et al. [2013b], we simulate d-dimensional samples of standard normal distributed random variables representing $1,000$ trading days. This sample size corresponds to the size of the moving time-window as explained above. After that, we compute the test statistic for the sample. We repeat this procedure $10,000$ times. Finally, we determine the critical value by computing the 95%-quantile of the resulting test statistics. In addition, we verify whether the asymptotic critical values used for the pairwise test for constant correlation and the test for a constant correlation matrix are suitable for finite samples including $1,000$ trading days. To this end, we obtain critical values based on the procedure explained above and compare these to the corresponding asymptotic critical values. As shown in Table 6.2, in contrast to the differences for the test for a constant covariance matrix, the differences corresponding to the two tests for constant correlations are in an acceptable range.

Table 6.2: Critical Values

The table shows asymptotic and empirical critical values for the pairwise test for constant correlation, for the test for a constant correlation matrix, and for the test for a constant covariance matrix at the 5% significance level. Values in bold are used for the empirical study.

Test for	Constant Correlation (pairwise)	Constant Correlation (Matrix)	Constant Covariance (Matrix)
Asymptotic Critical Values	**1.358**	**23.124**	20.740
Empirical Critical Values	1.324	25.793	**14.265**

(4) **Simulations:** For calculating VaR and ES, we do not use analytical methods but simulations as it is done, e.g., by Giot and Laurent [2003] and Alexander and Sheedy [2008]. For each of the n assets in a portfolio and for each day t, $K = 100,000$ random simulations[7] of Student's t-distributed log returns $r_{t1}^{(k)}, \ldots, r_{tn}^{(k)}$ are generated

[7] Giot and Laurent [2003] state that the choice of 100,000 simulations provides accurate estimates of the quantile.

by use of the mean μ_t, the univariate volatility forecast $\sigma_{t+1,i}$, the correlation matrix P as estimated by the models described in Section 6.3, and the degrees of freedom $\nu_c = 15$.[8] Then, the simulated log returns for the individual portfolio assets are aggregated to 100,000 portfolio log returns.

(5) **Estimation of VaR and ES:** The daily VaR at the $100(1-\alpha)\%$ confidence level is given by the α-quantile of the simulated portfolio log returns. To analyze the effect of different levels of significance on the quality of our models' risk estimates, we set $\alpha = 0.05$ and $\alpha = 0.01$ and compare the results for the VaR-estimates with the realized portfolio losses in order to identify VaR-exceedances.

As the Value-at-Risk is not in general coherent, we also estimate the portfolios' Expected Shortfalls which are given by

$$ES_\alpha(X) = E[X|X \leq VaR_\alpha(X)]. \tag{6.10}$$

For day $t+1$, we determine the ES_α by computing the mean of the simulated log returns below the estimated VaR_α for that day.

(6) **Backtesting and performance measurement:** The performances of the different models are evaluated by applying appropriate backtests on the VaR- and ES- forecasts. Since the univariate volatility forecasts for each of the VaR models are equal. Hence, differences in VaR-forecasts and VaR-violations can only result from differences in the estimated correlations. We employ the commonly used test of Kupiec [1995] to evaluate whether the observed number of VaR-violations is consistent with the expected frequency (unconditional coverage). In addition, we take a look at the distribution of the VaR-violations. The day on which a VaR-violation occurs should be unpredictable, i.e., the violation-series should follow a martingale difference process. To this end, we perform the CAViaR-Test of Engle and Manganelli [2004] and

[8] We choose a fixed ν because the estimation of degrees of freedom leading to proper shapes particularly at the tails of a distribution is not a trivial task and is not in the focus of this paper. However, setting $\nu = 15$ corresponds to Santos et al. [2013] who estimate a range of 10 to 19 degrees of freedom for estimating the VaR of several multivariate portfolios by using a DCC-GARCH model.

Berkowitz et al. [2011]. The test is based on the idea that any transformation of the variables available when VaR is computed should not be correlated with the current violation. Consider the autoregression

$$I_t = \alpha + \sum_{k=1}^{n} \beta_{1k} I_{t-k} + \sum_{k=1}^{n} \beta_{2k} g(I_{t-k}, I_{t-k-1}, \cdots, R_{t-k}, R_{t-k-1}, \cdots) + u_t. \quad (6.11)$$

In line with Berkowitz et al. [2011], we set $g(I_{t-k}, I_{t-k-1}, \cdots, R_{t-k}, R_{t-k-1}, \cdots) = VaR_{t-k+1}$ and $n = 1$. The null hypothesis of a correctly specified model with $\beta_{1k} = \beta_{2k} = 0$ is tested with a likelihood ratio test. The test statistic is asymptotically χ^2 distributed with two degrees of freedom. Berkowitz et al. [2011] evaluate the finite-sample size and power properties of various different VaR-backtests by conducting a Monte Carlo study where the return generating processes are based on real life data. They find that the CAViaR-test shows a superior performance compared to competing models.

The Expected Shortfall is backtested with the test of McNeil and Frey [2000]. This test evaluates the mean of the shortfall violations, i.e., the deviation of the realized shortfall against the ES in the case of a VaR-violation. The average error should be zero. The backtest is a one-sided test against the alternative hypothesis that the residuals have mean greater than zero or, equivalently, that the expected shortfall is systematically underestimated.

The backtests described above are designed to evaluate the accuracy of a single model. Since we are also interested in a comparison between the performances of the correlation models we additionally employ the conditional predictive ability (CPA) test of Giacomini and White [2006]. This interval forecast test is derived under the assumption of data heterogeneity rather than stationarity and can be applied to the comparison between nested and non nested models as well as among different estimation techniques and (finite) estimation windows. We follow Santos et al. [2013]

and assume an asymmetric linear (tick) loss function $\mathcal{L}$ of order α defined as

$$\mathcal{L}_\alpha(e_{t+1}) = (\alpha - 1(e_{t+1} < 0))e_{t+1}, \tag{6.12}$$

where $e_{t+1} = r_{t+1} - VaR_{\alpha,t+1}$. The null hypothesis of equal conditional predictive ability claims that the out-of-sample loss difference between two models follows a martingale difference sequence.[9]

In addition to the statistical backtests, we assess the performance of the models from a practitioner's point of view. According to the framework for backtesting internal models proposed by the Basel Committee on Banking Supervision [1996], we measure the number of VaR-violations on a quarterly basis using the most recent twelve months of data. To be more precisely, we count the number of violations after every 60 trading days using the data of the most recent 250 trading days. We sum up the VaR-violations for each interval $[1, \ldots, 250]$, $[61, \ldots, 310]$, $\ldots$, $[3,721, \ldots, 3,970]$. This procedure leads to 63 results of one-year VaR-violation frequencies. Then, we follow McNeil et al. [2005] and compute the average absolute discrepancy between observed and expected numbers of VaR-violations. We abstain from using the calculation of capital requirements according to the Basel guidelines to evaluate the performance of the different models. Da Veiga et al. [2011] find that using models which underestimate the VaR lead to low capital charges because the current penalty structure for excessive violations is not severe enough. For this reason, we consider the capital requirement not to be an appropriate performance measure.

6.5 Results

In this section, the results of our empirical study are discussed focusing on the specified aspects mentioned in the introduction of this paper.

[9] For a detailed description of the test statistic, see Giacomini and White [2006]

6.5.1 Total Number of VaR Violations

We start the discussion of our results with the analysis of the total number of VaR-violations. A key requirement with regard to VaR-forecasting models is that the actual number of VaR-violations should match the expected number related to the selected α-quantile. For each of the different models, we compute the VaR-violation ratio by dividing the actual number of VaR-violations by the total number of 3,970 VaR-forecasts. Furthermore, we apply the unconditional coverage test of Kupiec [1995] to test the null hypothesis of a correctly specified model. The results are reported in Table 6.3.

With respect to Panel A, the average VaR-violation ratios for the $\alpha = 5\%$ and $\alpha = 1\%$ quantiles (hereinafter referred to as 5% VaR and 1% VaR) amount to 4.924% and 1.231%, respectively, which is close to the corresponding nominal VaR levels. In the vast majority of settings, the average VaR-violation ratio of the models including tests for structural breaks are closer to the nominal VaR levels than the corresponding ratios of the plain models. The p-values of the unconditional coverage test of Kupiec [1995] are reported in Panel B. For the 5% VaR, in only a very few cases the p-values are below the 10% threshold for statistical significance and, therefore, it is difficult to derive conclusions. For the 1% VaR, the models including the test for a constant correlation matrix show less significant p-values than the remaining approaches.

6.5.2 Distribution of VaR Violations

The total number of VaR-violations is not an exhaustive criterion to evaluate the fit of the analyzed dependence models, because it gives no indication about the distribution of the VaR-violations. Among others, Longin and Solnik [2001] as well as Campbell et al. [2002] show that in particular in volatile bear markets correlations tend to increase. Consequently, in times where an effective risk management is most needed, inflexible dependence models may not be able to adequately adapt to changes in the dependence structure. This could lead to the undesired occurrence of clustered VaR-violations which in turn could lead to disastrous losses. To this end, we perform the CAViaR-based backtest

Table 6.3: Results Value-at-Risk

For each portfolio and for the 5% and 1% VaR, the table shows the VaR-Violation Ratio (i.e., number of VaR-violations divided by VaR-forecasts) and the p-values for the unconditional coverage test of Kupiec [1995], and the CAViaR based test of Engle and Manganelli [2004] and Berkowitz et al. [2011]. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	Model	Test	AEX	CAC	DAX	FTSE	IBEX	S&P	Average
Panel A: VaR-Violation Ratio									
5% VaR	CCC	no Test	5.592%	5.189%	5.390%	4.509%	4.811%	5.264%	5.126%
		Correlation (pairwise)	5.239%	4.962%	5.063%	4.584%	4.811%	4.861%	4.920%
		Correlation (Matrix)	5.088%	4.786%	4.912%	4.383%	4.761%	4.559%	4.748%
		Covariance (Matrix)	5.340%	5.063%	5.315%	4.509%	4.786%	4.861%	4.979%
	DCC	no Test	5.315%	5.063%	5.038%	4.307%	4.912%	5.013%	4.941%
		Correlation (Matrix)	5.214%	4.811%	5.038%	4.232%	4.786%	4.710%	4.798%
		Covariance (Matrix)	5.315%	5.013%	5.365%	4.433%	4.811%	4.811%	4.958%
1% VaR	CCC	no Test	1.713%	1.209%	1.159%	1.385%	1.335%	1.259%	1.343%
		Correlation (pairwise)	1.310%	1.008%	1.259%	1.360%	1.335%	1.134%	1.234%
		Correlation (Matrix)	1.335%	0.957%	1.159%	1.234%	1.234%	1.033%	1.159%
		Covariance (Matrix)	1.486%	1.033%	1.184%	1.360%	1.335%	1.108%	1.251%
	DCC	no Test	1.461%	1.134%	1.108%	1.234%	1.310%	1.134%	1.230%
		Correlation (Matrix)	1.285%	0.982%	1.234%	1.259%	1.234%	1.083%	1.180%
		Covariance (Matrix)	1.360%	1.033%	1.259%	1.259%	1.310%	1.108%	1.222%
Panel B: p-Value UC Test									
5% VaR	CCC	no Test	1.000	1.000	1.000	0.149	1.000	1.000	
		Correlation (pairwise)	1.000	1.000	1.000	0.224	1.000	1.000	
		Correlation (Matrix)	1.000	1.000	1.000	0.069*	1.000	0.196	
		Covariance (Matrix)	1.000	1.000	1.000	0.149	1.000	1.000	
	DCC	no Test	1.000	1.000	1.000	0.040**	1.000	1.000	
		Correlation (Matrix)	1.000	1.000	1.000	0.023**	1.000	1.000	
		Covariance (Matrix)	1.000	1.000	1.000	0.095*	1.000	1.000	
1% VaR	CCC	no Test	0.000***	0.200	0.327	0.021**	0.044**	0.114	
		Correlation (pairwise)	0.061*	0.962	0.114	0.031**	0.044**	0.408	
		Correlation (Matrix)	0.044**	0.785	0.327	0.152	0.152	0.837	
		Covariance (Matrix)	0.004***	0.837	0.258	0.031**	0.044**	0.500	
	DCC	no Test	0.006***	0.408	0.500	0.152	0.061*	0.408	
		Correlation (Matrix)	0.084*	0.911	0.152	0.114	0.152	0.603	
		Covariance (Matrix)	0.031**	0.837	0.114	0.114	0.061*	0.500	
Panel C: p-Value CAViaR Test									
5% VaR	CCC	no Test	0.032**	0.070*	0.005***	0.011**	0.117	0.064*	
		Correlation (pairwise)	0.017**	0.124	0.164	0.090*	0.131	0.313	
		Correlation (Matrix)	0.341	0.208	0.225	0.156	0.186	0.794	
		Covariance (Matrix)	0.044**	0.076*	0.044**	0.012**	0.074*	0.143	
	DCC	no Test	0.088*	0.136	0.014**	0.035**	0.170	0.123	
		Correlation (Matrix)	0.018**	0.176	0.435	0.097*	0.146	0.397	
		Covariance (Matrix)	0.011**	0.075*	0.034**	0.045**	0.118	0.370	
1% VaR	CCC	no Test	0.466	0.527	0.232	0.213	0.565	0.664	
		Correlation (pairwise)	0.920	0.289	0.884	0.372	0.424	0.825	
		Correlation (Matrix)	0.441	0.181	0.833	0.849	0.239	0.625	
		Covariance (Matrix)	0.521	0.434	0.808	0.350	0.403	0.787	
	DCC	no Test	0.794	0.615	0.473	0.526	0.694	0.450	
		Correlation (Matrix)	0.891	0.216	0.542	0.360	0.262	0.789	
		Covariance (Matrix)	0.898	0.419	0.853	0.762	0.351	0.581	

of Engle and Manganelli [2004] and Berkowitz et al. [2011] to analyze the performance of the models used in our empirical study.

The results of the CAViaR test are presented in Panel C of Table 6.3. Considering the 5% VaR, the p-values for the plain models fall short of the 10% threshold for statistical significance in five (CCC) and three (DCC) out of six cases. Both, the CCC and DCC model are improved by the test for a constant correlation matrix leading to zero and two rejections of the null hypothesis, respectively. The pairwise test for constant correlations shows less rejections than the plain CCC model, too, while the test for a constant covariance matrix does not lead to any improvements. With respect to the 1% VaR, the CAViaR test does not lead to any statistically significant results.

In addition to the statistical tests, we evaluate the performance of the different VaR-forecasting models from a perspective which is more relevant in practical terms. As explained in section 6.4, we follow the Basel guidelines for backtesting internal models and count the number of VaR-violations after every 60 trading days using the data of the preceding 250 trading days. Based on the resulting 63 quarterly VaR-violation frequencies, we compute average absolute discrepancies of VaR-violations which are presented in Table 6.4.

Table 6.4: Average Absolute Discrepancy of VaR-Violations

Based on the framework for backtesting internal models proposed by the Basel Committee on Banking Supervision [1996], we count the number of violations on a quarterly basis (every 60 trading days) using the most recent year (250 trading days) of data. We then sum up the VaR-violations for each interval $[1, \ldots, 250]$, $[61, \ldots, 310]$, $\ldots$, $[3,721, \ldots, 3,970]$. This procedure leads to 63 results of one-year VaR-violation frequencies. The table shows the average absolute discrepancy between observed and expected numbers of VaR-violations.

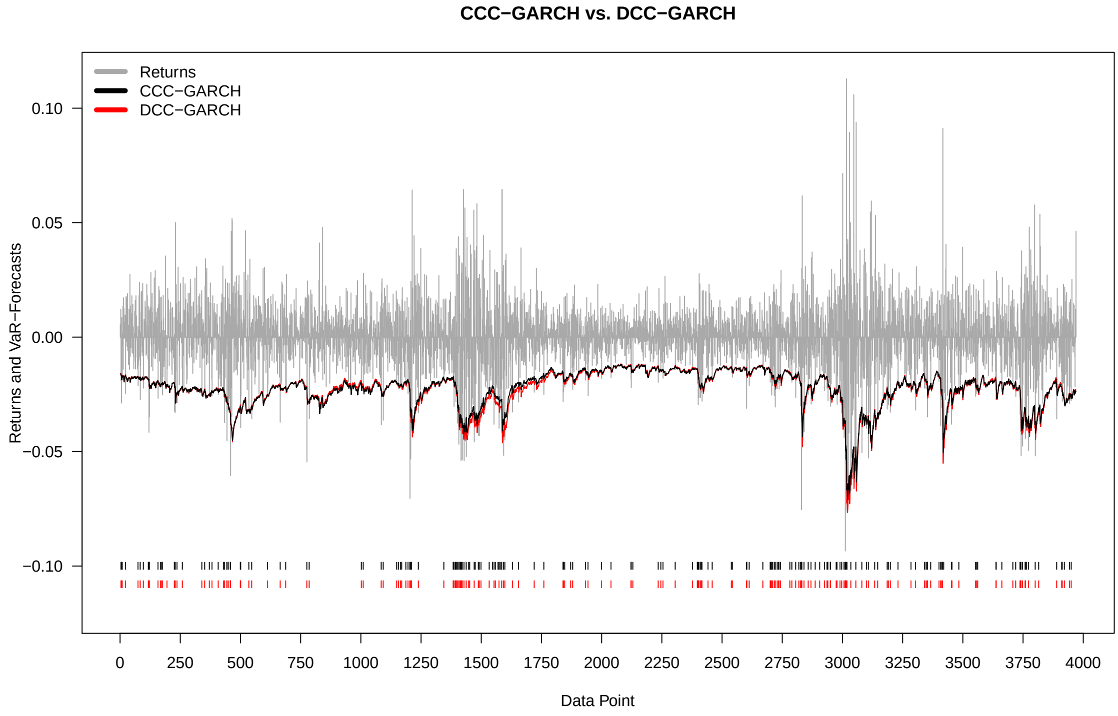
	Model	Test	AEX	CAC	DAX	FTSE	IBEX	S&P	Average
5% VaR	CCC	no Test	6.008	4.770	6.119	4.198	4.421	5.262	5.130
		Correlation (pairwise)	5.278	3.722	5.389	3.754	4.103	4.357	4.434
		Correlation (Matrix)	5.056	3.706	5.151	3.802	3.897	4.024	4.272
		Covariance (Matrix)	6.135	4.310	5.659	4.421	4.421	4.865	4.968
	DCC	no Test	5.738	4.563	5.421	4.119	4.468	4.754	4.844
		Correlation (Matrix)	5.563	3.992	5.389	4.135	3.992	4.421	4.582
		Covariance (Matrix)	6.056	4.357	5.548	4.087	4.405	4.675	4.854
1% VaR	CCC	no Test	2.437	1.278	1.611	2.198	2.119	1.897	1.923
		Correlation (pairwise)	1.817	1.214	1.881	2.087	1.865	1.643	1.751
		Correlation (Matrix)	2.008	1.262	1.706	1.770	1.817	1.437	1.667
		Covariance (Matrix)	2.183	1.151	1.786	2.151	2.024	1.659	1.825
	DCC	no Test	1.929	1.183	1.421	2.056	2.008	1.690	1.714
		Correlation (Matrix)	1.833	1.214	1.913	2.087	1.817	1.627	1.749
		Covariance (Matrix)	1.976	1.214	1.849	1.960	1.817	1.579	1.733

We start with the discrepancies for the 5% VaR. The CCC models in combination with structural break tests show lower average absolute discrepancies compared to their plain counterpart. In particular, the CCC model in combination with the test for a constant correlation matrix leads to less clustered VaR-violations. There are only small differences in the average absolut discrepancies of the DCC based models. However, the models which include the test for a constant correlation matrix improve slightly on the plain DCC. Continuing with the results for the 1% VaR, the average absolute discrepancies of the plain CCC and DCC models are outperformed by the models accounting for structural breaks. Again, the models including the test for a constant correlation matrix show the lowest discrepancies on average.

Because the (averaged) absolute discrepancies of quarterly VaR-violation frequency is a highly aggregated performance measure, we analyze the effects of the application of tests for structural breaks by taking a detailed view at the VaR-forecasts and VaR-violations for the 5% VaR using the CAC40 portfolio as an example.[10] To illustrate the differences in the behaviour of the analyzed approaches, Figures 6.1 to 6.3 show the portfolio returns and corresponding daily VaR-forecasts of the CCC and DCC based models. Comparing the VaR-forecasts of the plain CCC and the DCC model, there are just small differences in the VaR-forecasts observable. Particularly at the high volatility periods around the data points 1,500 and 3,000, the forecasts of the DCC model are slightly more conservative than the CCC forecasts. Conversely, the VaR-forecasts of the DCC model during the calm period around the data point 1,000 are slightly lower compared to the CCC model. The VaR-forecasts of the models accounting for structural breaks deviate from the plain CCC forecasts in the same direction as the DCC forecasts do, but to a markedly larger extent. This applies in particular to both models including the tests for constant correlations. The same pattern is observable for the comparison of the plain DCC model and its structural break counterparts, but the deviations between the VaR-forecasts of these models are smaller compared to the CCC based models.

[10] For the remaining portfolios we provide the same figures on request. However, these charts follow similar patterns.

Figure 6.1: VaR-Forecasts and VaR-Violations for the CCC Based Models (1/2)
The figure presents returns, VaR-forecasts, and VaR-violations for the CAC40 portfolio at $\alpha = 5\%$. VaR-forecasts are shown with lines and the dashes at the bottom of the charts mark the data points where a VaR-violation occurs.



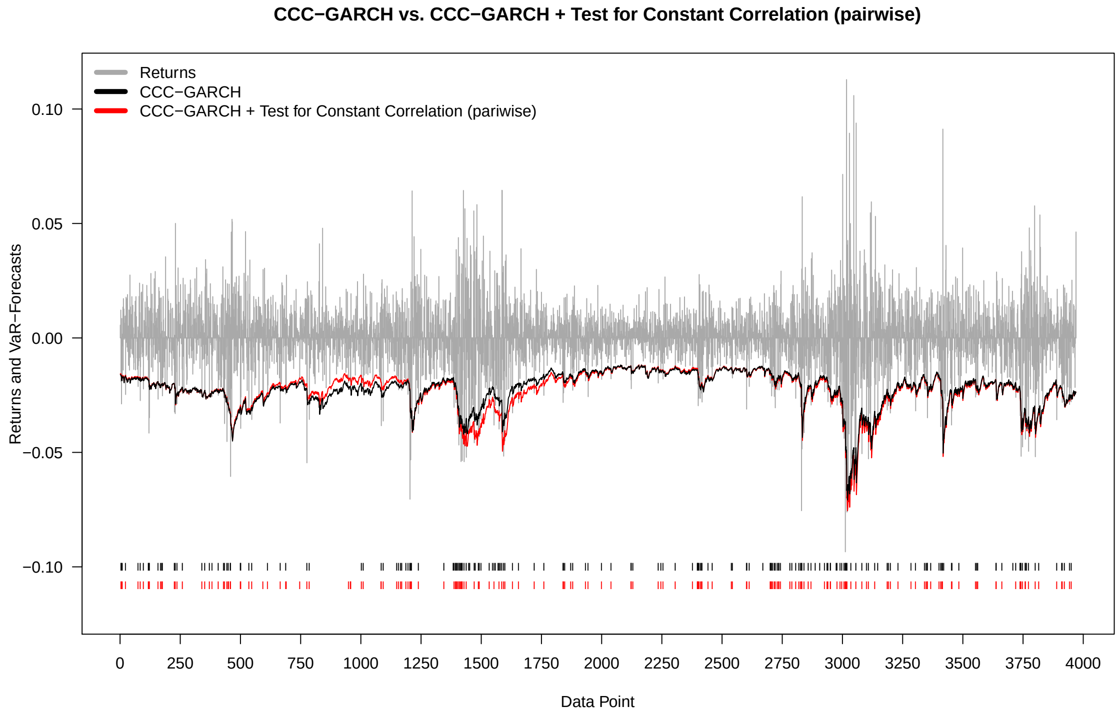
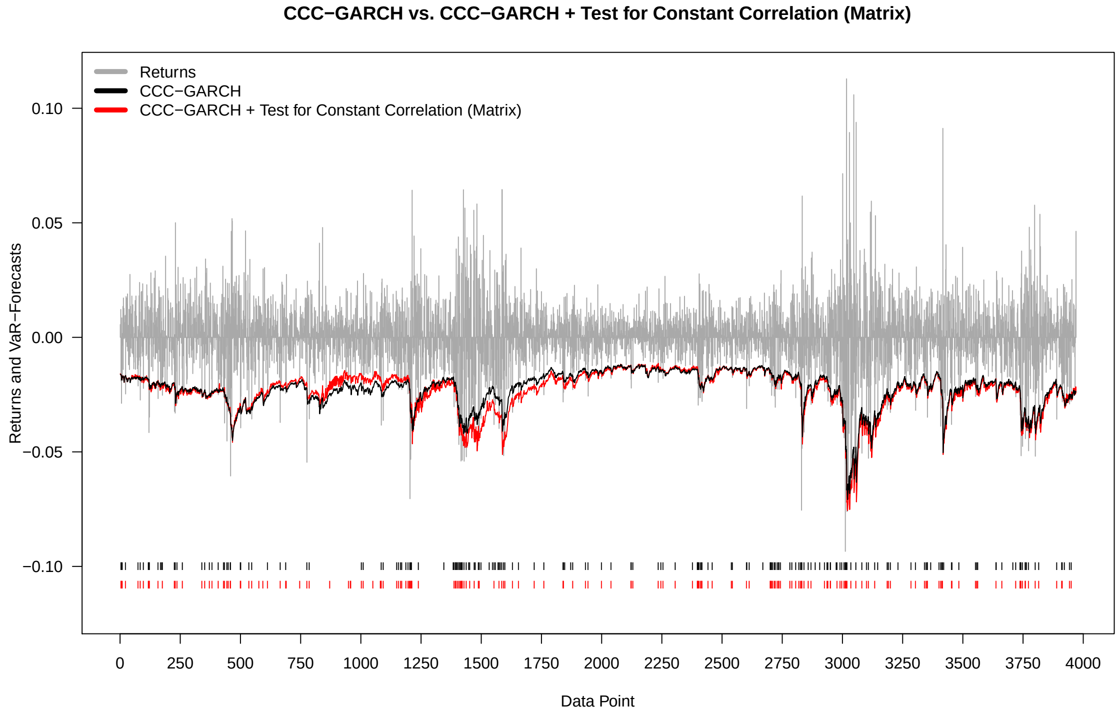


Figure 6.2: VaR-Forecasts and VaR-Violations for the CCC Based Models (2/2)
The figure presents returns, VaR-forecasts, and VaR-violations for the CAC40 portfolio at $\alpha = 5\%$. VaR-forecasts are shown with lines and the dashes at the bottom of the charts mark the data points where a VaR-violation occurs.



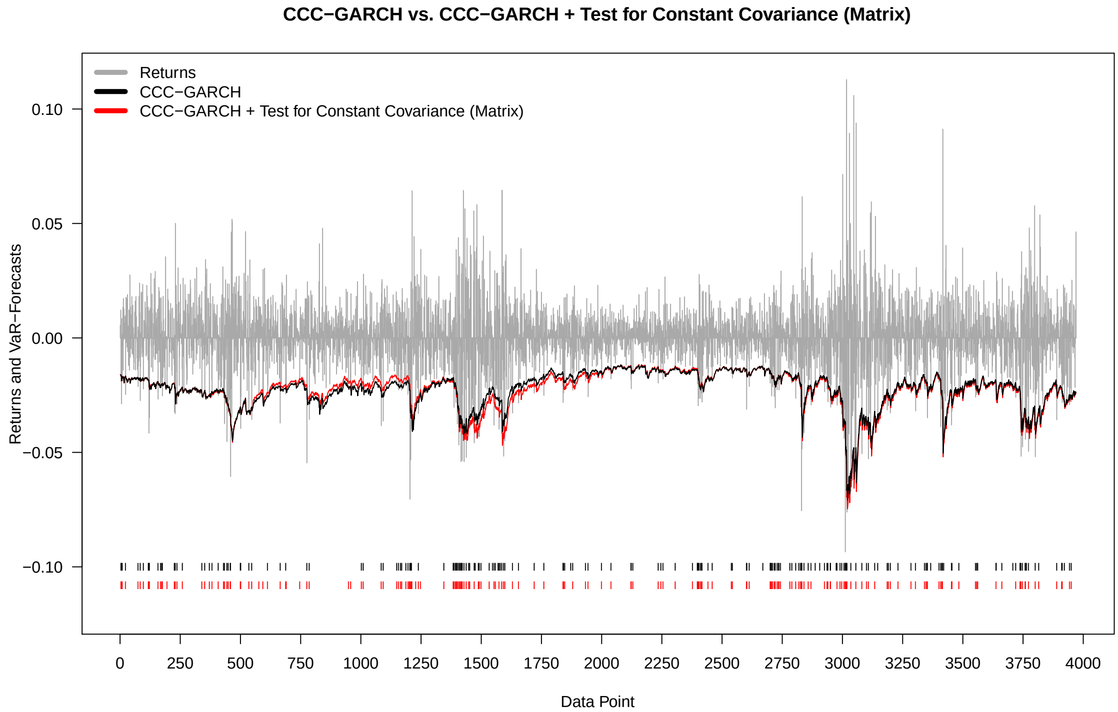
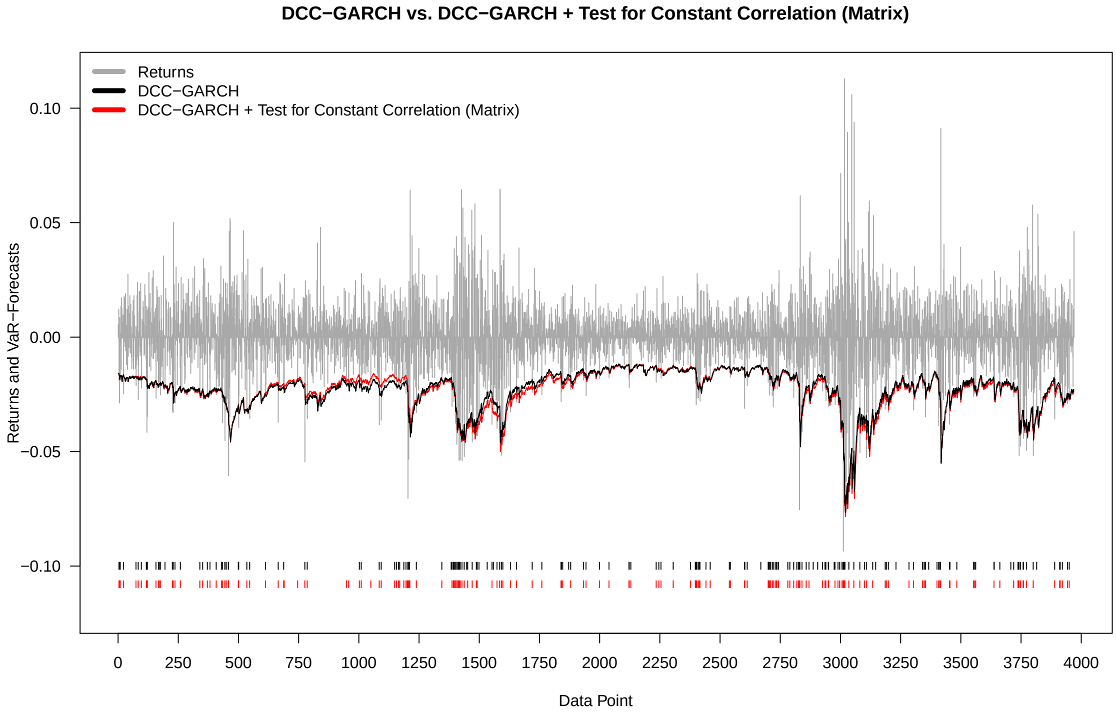
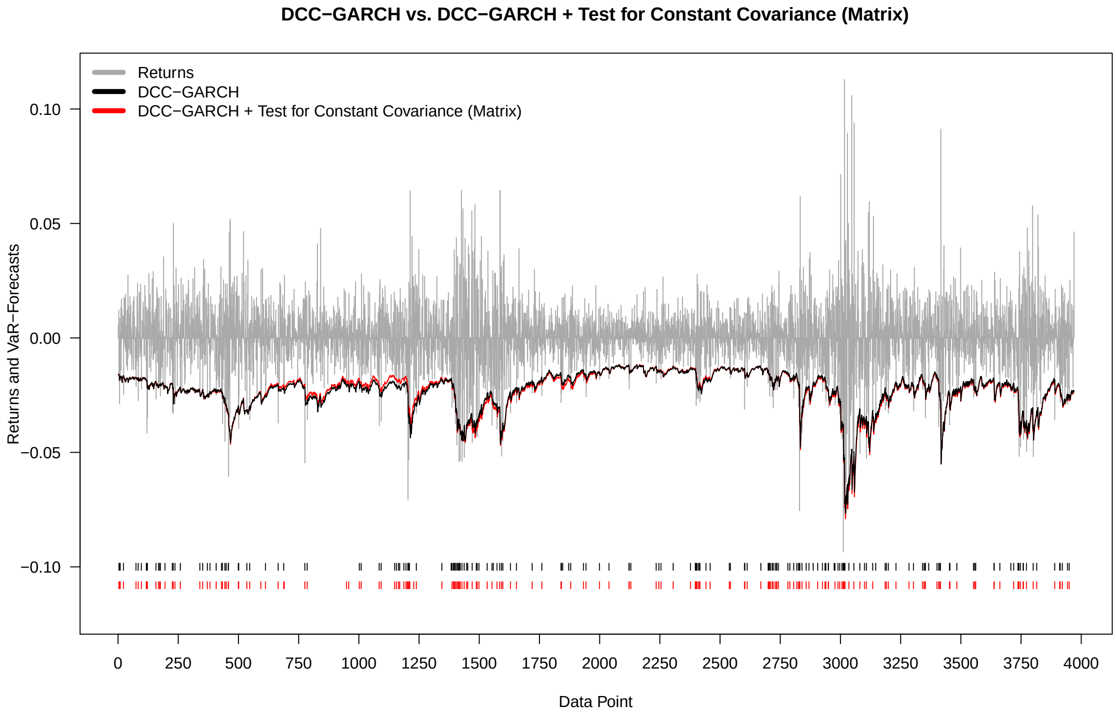


Figure 6.3: VaR-Forecasts and VaR-Violations for the DCC Based Models
The figure presents returns, VaR-forecasts, and VaR-violations for the CAC40 portfolio at $\alpha = 5\%$. VaR-forecasts are shown with lines and the dashes at the bottom of the charts mark the data points where a VaR-violation occurs.





We continue with the Basel based periodically computation of VaR-violation frequencies. Figure 6.4 illustrates the number of VaR-violations on a 60 trading day basis using the data of the most recent 250 trading days for the CAC40 portfolio. The VaR-violation frequency of the plain CCC and DCC models deviate from the required level in particular during four specific periods:

a) at the later stage of the dot-com bubble between the data points 730 (October 28, 1999) and 1,150 (June 27, 2001);

b) at the bear market after the burst of the dot-com bubble and the 9/11 attacks between the data points 1,450 (August 30, 2002) and 1,690 (August 11, 2003);

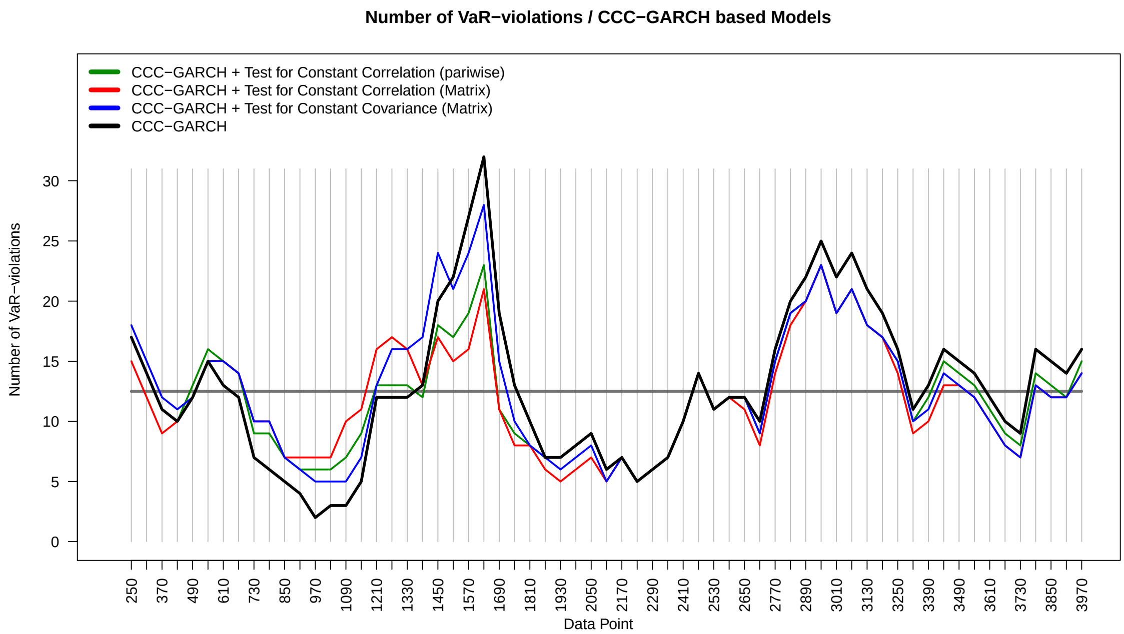
c) at the economic recovery between the data points 1,810 (January 29, 2004) and 2,410 (May 30, 2006);

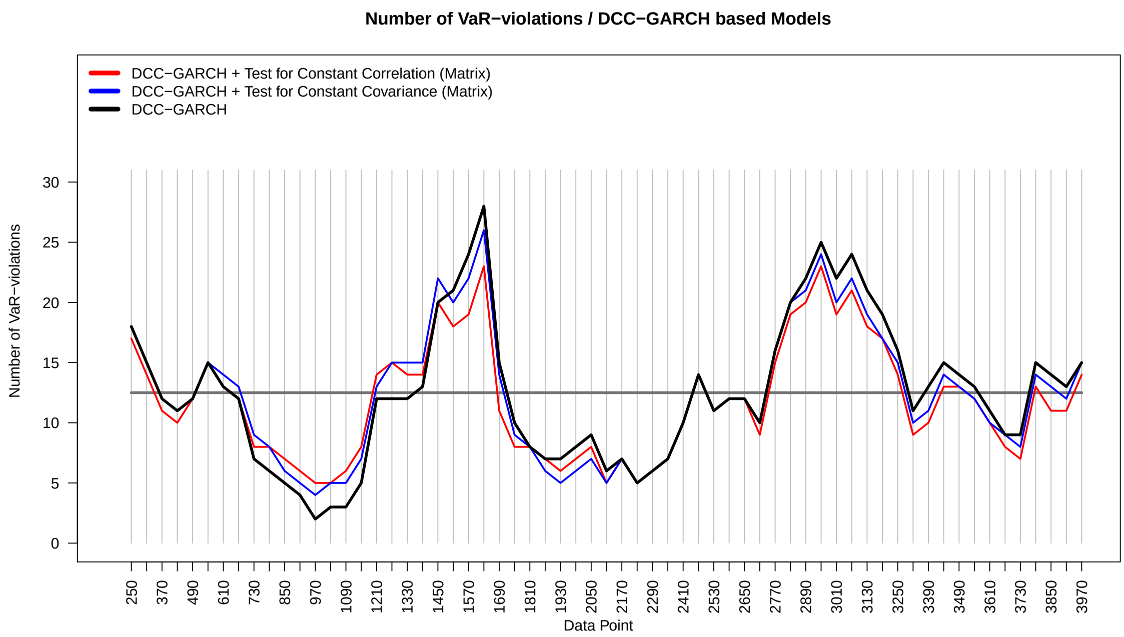
d) and at the financial crisis between the data points 2,770 (October 24, 2007) and 3,250 (September 11, 2009).

Turning to period a), the numbers of VaR-exceedances of the plain CCC and DCC models are far too low. The implementation of the structural break tests results in VaR-forecasts which are less conservative to some degree and therefore more accurate. Consequently, the additional use of these tests leads to a reduction of the extent by which the violation frequency falls short of the expected level. In contrast, during periods b) and d), the plain models show far too many exceedances, whereas the number of VaR-violations of the structural break test models are significantly lower. This applies particularly to the models in combination with the tests for constant correlations whose daily VaR-forecasts are distinctly more conservative. However, during the calm stock markets of period c), the daily VaR-forecasts of the different models show hardly any different results. Nevertheless, the plain CCC and DCC models show a slightly lower degree of risk-overestimation compared to the remaining approaches.

Figure 6.4: Quarterly VaR-Violations

The chart shows the number of VaR-violations for the CAC40 portfolio at $\alpha = 5\%$ on a quarterly basis (every 60 trading days) using the most recent year (250 trading days) of data. The chart at the top shows the results of the CCC based models. The chart at the bottom shows the results of the DCC based models. The horizontal grey lines mark the expected number of VaR-violations.





6.5.3 Expected Shortfall

In addition to the measurement of the Value-at-Risk, we evaluate the different risk-models with respect to their accuracy in forecasting Expected Shortfall. To this end, we compare the models on the basis of the deviation of the realized shortfall against the ES in the case of a VaR-violation. Furthermore, we apply the backtest of McNeil and Frey [2000]. The results are presented in Tables 6.5 and 6.6. Overall, the realized shortfall of the models show only small deviations from the ES which ranges from -0.20 to 0.02 percentage points for the 5% VaR and -0.34 to 0.01 percentage points for the 1% VaR, whereas a negative deviation indicates a risk underestimation. Concerning the 5% VaR, the average absolute deviation of the plain CCC model is undercut by the deviations of its counterparts accounting for structural breaks, in particular by the CCC model including the pairwise test for constant correlations. Regarding the DCC model, only the test for a constant covariance matrix outperforms the plain model. With respect to the 1% VaR, none of the structural break models yield lower average absolute deviations than the plain CCC and DCC approaches. The results of the one-sided ES backtest of McNeil and Frey [2000] does not lead to any further significant conclusions. The test leads to p-values rejecting the null hypothesis at a significance level of 10% for the majority of portfolios and therefore indicate that all models tend to underestimate the ES. Because there are only small differences in the number of rejections it is difficult to derive conclusions from this backtest.

Table 6.5: Results Expected Shortfall - 5% VaR

For each portfolio and for the 5% VaR, the table shows the mean Expected Shortfall in case of a VaR-violation, the mean realized shortfall in the case of a VaR-violation, the difference between the mean ES and the mean realized shortfall, and the p-value of the backtest of McNeil and Frey [2000]. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	Model	Test	Exp. Shortfall	Actual Shortfall	Dev.	p-Value
AEX	CCC	no Test	-2.66%	-2.82%	-0.16%	0.002***
		Correlation (pairwise)	-2.69%	-2.81%	-0.12%	0.022**
		Correlation (Matrix)	-2.65%	-2.78%	-0.13%	0.011**
		Covariance (Matrix)	-2.68%	-2.81%	-0.14%	0.008***
	DCC	no Test	-2.71%	-2.85%	-0.14%	0.008***
		Correlation (Matrix)	-2.71%	-2.81%	-0.10%	0.036**
		Covariance (Matrix)	-2.69%	-2.81%	-0.12%	0.020**
CAC	CCC	no Test	-3.07%	-3.12%	-0.05%	0.192
		Correlation (pairwise)	-3.11%	-3.10%	0.01%	0.578
		Correlation (Matrix)	-3.15%	-3.13%	0.02%	0.635
		Covariance (Matrix)	-3.13%	-3.12%	0.01%	0.579
	DCC	no Test	-3.11%	-3.11%	0.00%	0.485
		Correlation (Matrix)	-3.16%	-3.13%	0.02%	0.648
		Covariance (Matrix)	-3.14%	-3.12%	0.02%	0.628
DAX	CCC	no Test	-3.45%	-3.53%	-0.09%	0.093*
		Correlation (pairwise)	-3.37%	-3.49%	-0.12%	0.047**
		Correlation (Matrix)	-3.41%	-3.51%	-0.10%	0.076*
		Covariance (Matrix)	-3.41%	-3.50%	-0.09%	0.094*
	DCC	no Test	-3.48%	-3.56%	-0.08%	0.123
		Correlation (Matrix)	-3.35%	-3.47%	-0.12%	0.047**
		Covariance (Matrix)	-3.43%	-3.49%	-0.06%	0.178
FTSE	CCC	no Test	-2.57%	-2.76%	-0.20%	0.000***
		Correlation (pairwise)	-2.54%	-2.70%	-0.16%	0.002***
		Correlation (Matrix)	-2.55%	-2.71%	-0.16%	0.002***
		Covariance (Matrix)	-2.55%	-2.74%	-0.19%	0.000***
	DCC	no Test	-2.60%	-2.77%	-0.18%	0.001***
		Correlation (Matrix)	-2.56%	-2.75%	-0.19%	0.000***
		Covariance (Matrix)	-2.56%	-2.73%	-0.17%	0.001***
IBEX	CCC	no Test	-2.99%	-3.13%	-0.15%	0.008***
		Correlation (pairwise)	-3.01%	-3.11%	-0.11%	0.032**
		Correlation (Matrix)	-3.01%	-3.12%	-0.11%	0.033**
		Covariance (Matrix)	-3.04%	-3.15%	-0.11%	0.038**
	DCC	no Test	-3.02%	-3.09%	-0.07%	0.105
		Correlation (Matrix)	-3.04%	-3.12%	-0.08%	0.088*
		Covariance (Matrix)	-3.04%	-3.12%	-0.08%	0.080*
S&P	CCC	no Test	-2.62%	-2.75%	-0.12%	0.012**
		Correlation (pairwise)	-2.68%	-2.77%	-0.08%	0.083*
		Correlation (Matrix)	-2.66%	-2.76%	-0.10%	0.055*
		Covariance (Matrix)	-2.70%	-2.79%	-0.09%	0.061*
	DCC	no Test	-2.65%	-2.76%	-0.11%	0.022**
		Correlation (Matrix)	-2.71%	-2.78%	-0.07%	0.110
		Covariance (Matrix)	-2.68%	-2.77%	-0.09%	0.066

Table 6.6: Results Expected Shortfall - 1% VaR

For each portfolio and for the 1% VaR, the table shows the mean Expected Shortfall in case of a VaR-violation, the mean realized shortfall in the case of a VaR-violation, the difference between the mean ES and the mean realized shortfall, and the p-value of the backtest of McNeil and Frey [2000]. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

	Model	Test	Exp. Shortfall	Actual Shortfall	Dev.	p-Value
AEX	CCC	no Test	-3.65%	-3.76%	-0.11%	0.153
		Correlation (pairwise)	-3.56%	-3.80%	-0.24%	0.019**
		Correlation (Matrix)	-3.58%	-3.79%	-0.21%	0.040**
		Covariance (Matrix)	-3.60%	-3.77%	-0.17%	0.074*
	DCC	no Test	-3.61%	-3.76%	-0.15%	0.084*
		Correlation (Matrix)	-3.51%	-3.76%	-0.25%	0.019**
		Covariance (Matrix)	-3.60%	-3.81%	-0.21%	0.040**
CAC	CCC	no Test	-4.00%	-3.99%	0.01%	0.544
		Correlation (pairwise)	-3.74%	-3.84%	-0.11%	0.254
		Correlation (Matrix)	-3.72%	-3.84%	-0.13%	0.223
		Covariance (Matrix)	-3.79%	-3.87%	-0.08%	0.315
	DCC	no Test	-3.86%	-3.86%	0.00%	0.487
		Correlation (Matrix)	-3.73%	-3.85%	-0.11%	0.251
		Covariance (Matrix)	-3.79%	-3.86%	-0.08%	0.322
DAX	CCC	no Test	-4.64%	-4.92%	-0.28%	0.043**
		Correlation (pairwise)	-4.44%	-4.68%	-0.24%	0.070*
		Correlation (Matrix)	-4.47%	-4.74%	-0.27%	0.059*
		Covariance (Matrix)	-4.45%	-4.78%	-0.34%	0.023**
	DCC	no Test	-4.55%	-4.82%	-0.27%	0.052*
		Correlation (Matrix)	-4.40%	-4.67%	-0.27%	0.049**
		Covariance (Matrix)	-4.45%	-4.68%	-0.23%	0.076*
FTSE	CCC	no Test	-3.52%	-3.68%	-0.15%	0.051*
		Correlation (pairwise)	-3.46%	-3.59%	-0.13%	0.082*
		Correlation (Matrix)	-3.47%	-3.60%	-0.13%	0.092*
		Covariance (Matrix)	-3.45%	-3.64%	-0.19%	0.017**
	DCC	no Test	-3.55%	-3.67%	-0.12%	0.114
		Correlation (Matrix)	-3.47%	-3.61%	-0.15%	0.050*
		Covariance (Matrix)	-3.44%	-3.62%	-0.17%	0.028**
IBEX	CCC	no Test	-3.62%	-3.83%	-0.20%	0.037**
		Correlation (pairwise)	-3.82%	-3.94%	-0.12%	0.155
		Correlation (Matrix)	-3.72%	-3.89%	-0.18%	0.062*
		Covariance (Matrix)	-3.79%	-3.92%	-0.13%	0.123
	DCC	no Test	-3.73%	-3.89%	-0.17%	0.106
		Correlation (Matrix)	-3.75%	-3.89%	-0.14%	0.106
		Covariance (Matrix)	-3.78%	-3.88%	-0.10%	0.199
S&P	CCC	no Test	-3.51%	-3.76%	-0.25%	0.023**
		Correlation (pairwise)	-3.52%	-3.75%	-0.23%	0.046**
		Correlation (Matrix)	-3.47%	-3.76%	-0.29%	0.023**
		Covariance (Matrix)	-3.43%	-3.70%	-0.28%	0.016**
	DCC	no Test	-3.55%	-3.83%	-0.28%	0.019**
		Correlation (Matrix)	-3.54%	-3.76%	-0.22%	0.055*
		Covariance (Matrix)	-3.42%	-3.67%	-0.25%	0.027**

6.5.4 Comparison of Conditional Predictive Ability

The statistical tests used above are designed to evaluate the appropriateness of an individual model rather than directly comparing the forecasting accuracy between different VaR models. To provide a ranking of the analyzed forecasting approaches, we additionally perform the CPA test of Giacomini and White [2006]. Tables 6.7 and 6.8 report the results of the CPA test. With respect to the results of the 5% VaR quantile shown in Panel A, the plain CCC model is clearly outperformed by each approach accounting for structural breaks. The comparison between the different tests for structural breaks indicate that the pairwise test for constant correlations of Wied et al. [2012b] is preferred in this setting. The results of the comparison of the plain DCC models with its counterparts including structural breaks are ambiguous. Only the DCC model in combination with the test for a constant covariance matrix slightly improves the conditional predictive ability. Furthermore, the CCC model in combination with the pairwise test for constant correlations outperforms the plain DCC-model. Considering the 1% VaR level in Panel B, the results of the plain CCC model and the CCC model including structural breaks are comparable while the plain DCC model outperforms the corresponding structural break approaches.

6.6 Conclusion

The aim of this paper was to examine the question whether the VaR- and ES- forecasting accuracy of plain CCC and DCC models can be improved by the implementation of recently proposed tests for structural breaks in covariances and correlations. To this end, we perform an empirical out-of-sample study by using ten-dimensional portfolios composed of international blue-chip stocks. In addition to the plain CCC and the DCC benchmarks, we modify these models by combining them with the pairwise test for constant correlations of Wied et al. [2012b], the test for a constant correlation matrix of Wied [2012], and the test for a constant covariance matrix of Aue et al. [2009].

Table 6.7: Results CPA-Test - 5% VaR

The table shows the results of a comparison between the CCC and DCC models and their counterparts including tests for structural breaks based on the CPA-test. The results indicate how often a forecasting model is preferred when applying it to each of the six different portfolios. The models printed in bold yield a statistically significant better performance.

				Statistically Significant Results		
Model 1	Model 2	Model 1 Preferred	Model 2 Preferred	Model 1 Preferred	Model 2 Preferred	Indifferent
CCC	**DCC**	0	6	0	3	3
CCC	**CCC + Correlation Test (pairwise)**	1	5	0	3	3
CCC	**CCC + Correlation Test (Matrix)**	1	5	0	1	5
CCC	**CCC + Covariance Test (Matrix)**	2	4	0	2	4
CCC	**DCC + Correlation Test (Matrix)**	2	4	0	2	4
CCC	**DCC + Covariance Test (Matrix)**	1	5	0	3	3
CCC + Correlation Test (Matrix)	**CCC + Correlation Test (pairwise)**	2	4	0	1	5
CCC + Covariance Test (Matrix)	CCC + Correlation Test (Matrix)	1	5	0	0	6
CCC + Covariance Test (Matrix)	**CCC + Correlation Test (pairwise)**	1	5	0	1	5
DCC	**CCC + Correlation Test (pairwise)**	3	3	1	2	3
DCC	CCC + Correlation Test (Matrix)	3	3	0	0	6
DCC	CCC + Covariance Test (Matrix)	5	1	0	0	6
DCC	DCC + Correlation Test (Matrix)	3	3	1	1	4
DCC	DCC + Covariance Test (Matrix)	2	4	1	1	4
DCC + Correlation Test (Matrix)	**DCC + Covariance Test (Matrix)**	2	4	0	2	4

Table 6.8: Results CPA-Test - 1% VaR

The table shows the results of a comparison between the CCC and DCC models and their counterparts including tests for structural breaks based on the CPA-test. The results indicate how often a forecasting model is preferred when applying it to each of the six different portfolios. The models printed in bold yield a statistically significant better performance.

				Statistically Significant Results		
Model 1	Model 2	Model 1 Preferred	Model 2 Preferred	Model 1 Preferred	Model 2 Preferred	Indifferent
CCC	**DCC**	0	6	0	4	2
CCC	CCC + Correlation Test (pairwise)	3	3	1	0	5
CCC	CCC + Correlation Test (Matrix)	3	3	1	0	5
CCC	**CCC + Covariance Test (Matrix)**	3	3	0	1	5
CCC	DCC + Correlation Test (Matrix)	3	3	1	0	5
CCC	DCC + Covariance Test (Matrix)	2	4	1	1	4
CCC + Correlation Test (Matrix)	CCC + Correlation Test (pairwise)	2	4	0	0	6
CCC + Correlation Test (Matrix)	**CCC + Covariance Test (Matrix)**	4	2	0	1	5
CCC + Covariance Test (Matrix)	CCC + Correlation Test (pairwise)	2	4	0	0	6
DCC	CCC + Correlation Test (pairwise)	5	1	1	0	5
DCC	CCC + Correlation Test (Matrix)	5	1	1	0	5
DCC	CCC + Covariance Test (Matrix)	5	1	3	0	3
DCC	DCC + Correlation Test (Matrix)	5	1	2	0	4
DCC	DCC + Covariance Test (Matrix)	5	1	2	0	4
DCC + Correlation Test (Matrix)	**DCC + Covariance Test (Matrix)**	2	4	0	2	4

In order to evaluate the accuracy of the VaR-forecasts, we conduct the unconditional coverage test of Kupiec [1995] and the CAViaR based test of Engle and Manganelli [2004] and Berkowitz et al. [2011]. The results of both backtests indicate that testing for constant correlations can lead to a more accurate conditional coverage and less dependent VaR-violations. Evaluating the accuracy of the ES by performing the backtest of McNeil and Frey [2000] leads to no clear conclusion whether applying tests for structural breaks are beneficial or not. Additionally, we compare the conditional predictive ability of the models using the test of Giacomini and White [2006]. The results show that the extension of a plain CCC model by tests for structural breaks leads to an equally or better unconditional predictive ability while it seems hard to beat plain DCC model.

To get a deeper insight into the characteristics of the different models, we change from the statistical backtest perspective towards a backtest which is of relevance in regulatory practice. To this end, we perform a backtest procedure based on the Basel guidelines for backtesting internal models. On a quarterly basis, we measure the number of VaR-violations within the most recent one-year period and evaluate the absolute discrepancies from the expected VaR-violation frequency. The plain CCC model is clearly outperformed by its counterparts modified by structural break tests, particularly by the test for a constant correlation matrix. However, the results with respect to the DCC models are more ambiguous.

Anhang A

Pseudocode

A.1 Test of Unconditional Coverage

(i) Generate the violation sequence resulting from the observed returns and the corresponding VaR forecasts by

$$I_i(p) = \begin{cases} 1, & \text{if } y_i < VaR_{i|i-1}(p); \\ 0, & \text{else.} \end{cases}$$

(ii) Draw $l+1$ random variables by

$$\epsilon_j \sim N(0,1) \cdot 0.001, \ j = 1, ..., l+1.$$

(iii) Calculate the test statistic for the observed violation sequence by

$$MCS_{uc} = \epsilon_{l+1} + \sum_{i=1}^{n} I_i.$$

(iv) Simulate violation sequences by drawing l-times n random variables with distribution

$$\hat{I}_{j,i}(p) \sim \text{Bern}(p), \ i = 1, ..., n, \ j = 1, ..., l.$$

(v) Calculate the test statistic for each simulated violation sequence by

$$M\hat{C}S_{uc,j} = \epsilon_j + \sum_{i=1}^{n} \hat{I}_{i,j}, \ j = 1, ..., l.$$

(vi) Sort the resulting values of the simulated statistic $M\hat{C}S_{uc,j}$ in descending order.

(vii) Compute the quantiles for the desired significance level and compare the test statistic for the observed violation sequence to the resulting critical values.

A.2 Test of the I.I.D. Property

(i) Generate the violation sequence resulting from the observed returns and the corresponding VaR forecasts by

$$I_i(p) = \begin{cases} 1, \text{ if } y_i < VaR_{i|i-1}(p); \\ 0, \text{ else}. \end{cases}$$

(ii) Calculate the sum of observed VaR violations by

$$m = \sum_{i=1}^{n} I_i.$$

(iii) Identify the time indexes where an observed VaR violation occurred by

$$V = \{i | I_i = 1\} = (t_1, ..., t_m).$$

(iv) Draw $l + 1$ random variables by

$$\epsilon_j \sim N(0, 1) \cdot 0.001, \; j = 1, ..., l + 1.$$

(v) Calculate the test statistic for the observed violation sequence by

$$MCS_{iid,m} = t_1^2 + (n - t_m)^2 + \sum_{s=2}^{m} (t_s - t_{s-1})^2 + \epsilon_{l+1}.$$

(vi) Simulate violation sequences by drawing l-times n random variables with distribution

$$\hat{I}_{i,j}(p) \sim \text{Bern}(p), \; i = 1, ..., n, \; j = 1, ..., l,$$

under the condition that $\sum_{i=1}^{n} \hat{I}_{i,j} = m, \; \forall \; j$.

(vii) For each simulated violation sequence, identify the set of time indexes of the violations by

$$\hat{V}_j = \{t_j | \hat{I}_{i,j} = 1\} = (t_{j,1}, ..., t_{j,m}).$$

(viii) Calculate the test statistic for the simulated violation sequences by

$$M\hat{C}S_{iid,m,j} = t_{j,1}^2 + (n - t_{j,m})^2 + \sum_{s=2}^{m} (t_{j,s} - t_{j,s-1})^2 + \epsilon_j.$$

(ix) Sort the resulting values of the simulated statistic $M\hat{C}S_{iid,m,j}$ in descending order.

(x) Compute the quantile for the desired significance level and compare the test statistic for the observed violation sequence to the resulting critical value.

A.3 Test of Conditional Coverage

(i) Simulate violation sequences by drawing l-times n random variables with distribution

$$\hat{I}_{i,j}(p) \sim \text{Bern}(p),\ i = 1, ..., n,\ j = 1, ..., l,$$

under the condition that $\sum_{i=1}^{n} \hat{I}_{i,j} > 1,\ \forall\, j.$

(ii) For each simulated violation sequence, identify the set of time indexes of the violations by

$$\hat{V}_j = \{\hat{t}_j | \hat{I}_{j,i} = 1\} = (\hat{t}_{j,1}, ..., \hat{t}_{j,m}).$$

(iii) Draw $l + 1$ random variables by

$$\epsilon_j \sim N(0, 1) \cdot 0.001,\ j = 1, ..., l + 1.$$

(iv) Calculate the violation frequency of each of the simulated sequences

$$\hat{m}_j = \sum_{i=1}^{n} \hat{I}_{i,j}.$$

(v) Define $\hat{m} = (\hat{m}_1, ..., \hat{m}_l)$ and set $\hat{m}_{min} = \max(2, \min(\hat{m}))$ and $\hat{m}_{max} = \max(\hat{m})$ for the lower and upper bound of possible VaR violation frequencies.

(vi) For each $k = \hat{m}_{min}, \hat{m}_{min+1}, \ldots, \hat{m}_{max}$, simulate violation sequences by drawing l^*-times n random variables with distribution

$$\tilde{I}_{i,j}(k/n) \sim \text{Bern}(k/n),\ i = 1, ..., n,\ j = 1, ..., l^*,$$

under the condition that $\sum_{i=1}^{n} \tilde{I}_{i,j}(k/n) = k,\ \forall\, j.$

(vii) For k and each simulated violation sequence, identify the set of time indexes of the violations by

$$\tilde{V}_{j,k} = \{\tilde{t}_{j,k} | \tilde{I}_{i,j,k} = 1\} = (\tilde{t}_{j,1}, ..., \tilde{t}_{j,k}).$$

(viii) For each k, calculate r_k, an estimator for $E(MCS_{iid,k}|H_0)$, by

$$r_k = \frac{1}{l^*} \cdot \sum_{j=1}^{l^*} \left(\tilde{t}_{j,1}^2 + (n - \tilde{t}_{j,k})^2 + \sum_{s=2}^{k} (\tilde{t}_{j,s} - \tilde{t}_{j,s-1})^2 \right).$$

(ix) Calculate the test statistic for each violation sequence simulated in step (i) by

$$M\hat{C}S_{cc,k,j} = af(M\hat{C}S_{uc,j}) + (1-a)g(M\hat{C}S_{iid,k,j}), 0 \leq a \leq 1,$$

where

$$f(M\hat{C}S_{uc,j}) = \left| \frac{\left(\epsilon_j + \sum_{i=1}^{n} \hat{I}_i\right)/n - p}{p} \right|,$$

and

$$g(M\hat{C}S_{iid,k,j}) = \frac{M\hat{C}S_{iid,k,j} - r_k}{r_k} \cdot 1_{\{M\hat{C}S_{iid,k,j} \geq r_k\}},\ k = \sum_{i=1}^{n} \hat{I}_{i,j}.$$

(x) Sort the resulting values of the simulated statistic $M\hat{C}S_{cc,k,j}$ in descending order.

(xi) Compute the quantile for the desired significance level.

(xii) Generate the violation sequence resulting from the observed returns and the corresponding VaR forecasts by

$$I_i(p) = \begin{cases} 1, \text{ if } y_i < VaR_{i|i-1}(p); \\ 0, \text{ else.} \end{cases}$$

(xiii) Calculate the sum of observed VaR violations by

$$m = \sum_{i=1}^{n} I_i.$$

(xiv) Identify the set of time indexes where an observed VaR violation occurred by

$$V = \{t|I_i = 1\} = (t_1, ..., t_m).$$

(xv) If $m \notin [\hat{m}_{min}, \hat{m}_{min+1}, \ldots, \hat{m}_{max}]$, determine r_m by repeating steps (vi) to (viii) where k is replaced by m.

(xvi) Calculate the test statistic for the observed violation sequence by

$$MCS_{cc,m} = af(MCS_{uc}) + (1-a)g(MCS_{iid,m}), 0 \leq a \leq 1,$$

where

$$f(MCS_{uc}) = \left| \frac{(\epsilon_{l+1} + \sum_{i=1}^{n} I_i)/n - p}{p} \right|,$$

and

$$g(MCS_{iid,m}) = \frac{MCS_{iid,m} - r_m}{r_m} \cdot 1_{\{MCS_{iid,m} \geq r_m\}}.$$

(xvii) Compare the test statistic for the observed violation sequence to the critical value.

Anhang B

Test for Constant Variances

The test statistic of the test for constant variances of Wied et al. [2012a] is given by

$$Q_T(X) = \max_{1 \leq j \leq T} \left| \hat{D} \frac{j}{\sqrt{T}} \left([Var\ X]_j - [Var\ X]_T \right) \right|,$$

where

$$[Var\ X]_l = \frac{1}{l} \sum_{i=1}^{l} X_i^2 - \left(\frac{1}{l} \sum_{i=1}^{l} X_i \right)^2 =: \overline{X}_l^2 - (\overline{X}_l)^2$$

is the empirical variance from the first l observations. Furthermore,

$$\hat{D} = (1 - 2\overline{X}_T)^{-1} (\hat{D}_1)^{-1/2}$$

is a scalar with

$$\hat{D}_1 = \frac{1}{T} \sum_{i=1}^{T} \hat{U}_i' \hat{U}_i + 2 \sum_{j=1}^{T} k\left(\frac{j}{\gamma T} \right) \frac{1}{T} \sum_{i=1}^{T-j} \hat{U}_i' \hat{U}_{i+j}$$

and

$$\hat{U}_l = \begin{pmatrix} X_l^2 - \overline{X}_T^2 \\ X_l - \overline{X}_T \end{pmatrix},$$

$$k(x) = \begin{cases} 1 - |x|, & |x| > 1 \\ 0, & \text{otherwise} \end{cases},$$

$$\gamma_n = \sqrt{T}.$$

For technical assumptions and proofs see Wied et al. [2012a].

Anhang C

Test for Constant Covariances

For $l = 0, \dots, [\log(T)]$, let $\sigma_{l,1}$ and $\sigma_{l,2}$ be matrices with $d(d+1)/2$ columns and $T-l$ rows such that the columns contain certain products (component by component) of the one-dimensional marginal time series. Concretely, if the entries on and below the diagonal of a $d \times d$ matrix are numbered from $c = 1, \dots, d(d+1)/2$ such that c corresponds to one pair $(i,j), 1 \leq i, j \leq d$, it holds that the c-th column of $\sigma_{l,1}$ is equal to the vector

$$(X_{l+1,i} \cdot X_{l+1,j}, \dots, X_{T,i} \cdot X_{T,j})$$

and that the c-th column of $\sigma_{l,2}$ is equal to the vector

$$(X_{1,i} \cdot X_{1,j}, \dots, X_{T-l,i} \cdot X_{T-l,j})\,.$$

Define $\hat{\Sigma}_l$ as the empirical covariance matrix of $\sigma_{l,1}$ and $\sigma_{l,2}$. Then, we introduce the quantity

$$\hat{\Sigma} = \hat{\Sigma}_0 + 2\sum_{l=1}^{[\log(T)]} \left(1 - \frac{l}{[\log(T)]}\right) \hat{\Sigma}_l$$

which is an estimator for the covariance matrix of S_k that captures fluctuations in higher moments and serial dependence and thus also serves for standardization. The test statistic is then the maximum over quadratic forms, i.e.

$$\Lambda_T = \max_{1 \leq k \leq T} S_k' \hat{\Sigma}^{-1} S_k.$$

Literaturverzeichnis

ACERBI, C. AND D. TASCHE (2002): “On the coherence of expected shortfall,” *Journal of Banking & Finance*, 26, 1487–1503.

ALEXANDER, C. (2008): *Market Risk Analysis, Value at Risk Models*, vol. 4, Chichester: John Wiley & Sons Ltd.

ALEXANDER, C. AND E. SHEEDY (2008): “Developing a stress testing framework based on market risk models,” *Journal of Banking & Finance*, 32, 2220–2236.

ANDREOU, E. AND E. GHYSELS (2003): “Quality control for financial risk management: monitoring disruptions in the distribution of risk exposure,” Tech. rep., Discussion paper, UNC.

——— (2009): “Structural breaks in financial time series,” in *Handbook of financial time series*, Springer, 839–870.

ANG, A. AND G. BEKAERT (2002): “International Asset Allocation with Time-Varying Correlations,” *Review of Financial Studies*, 15(4), 1137–1187.

ARNOLD, M., S. STAHLBERG, AND D. WIED (2013): “Modeling different kinds of spatial dependence in stock returns,” *Empirical Economics*, 44(2), 761–774.

ARTZNER, P., F. DELBAEN, J.-M. EBER, AND D. HEATH (1999): “Coherent Measures of Risk,” *Mathematical Finance*, 9, 203–228.

ASAI, M., M. MCALEER, AND M. C. MEDEIROS (2012): “Asymmetry and long memory in volatility modeling,” *Journal of Financial Econometrics*, 10, 495–512.

AUE, A., S. HÖRMANN, L. HORVÁTH, AND M. REIMHERR (2009): “Break detection in the covariance structure of multivariate time series models,” *The Annals of Statistics*, 37, 4046–4087.

BAI, Z., H. LIU, AND W. WONG (2009): "Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory," *Mathematical Finance*, 19(4), 639–667.

BALTUTIS, M. (2009): "Non-stationary stock returns and time to revise the optimal portfolio," *SSRN working paper, online: http://ssrn.com/abstract=1397184*.

BALTUTIS, M. AND E. DOCKNER (2007): "Do Conditional Covariance Estimates Generate Value?" *SSRN working paper, online: http://ssrn.com/abstract=1397188*.

BARONE-ADESI, G., K. GIANNOPOULOS, AND L. VOSPER (1999): "VaR without correlations for portfolios of derivative securities," *Journal of Futures Markets*, 19, 583–602.

BASAK, S. AND A. SHAPIRO (2001): "Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices," *Review of Financial Studies*, 14, 371–405.

BASEL COMMITTEE ON BANKING SUPERVISION (1996): "Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Model-Based Approach to Market Risk Capital Requirements," BIS.

——— (2005): "Amendment to the Capital Accord to incorporate market risks," `http://www.bis.org/publ/bcbs119.htm`.

——— (2011): "Messages from the academic literature on risk measurement for the trading book," `http://www.bis.org/publ/bcbs_wp19.htm`.

——— (2013): "Fundamental review of the trading book: A revised market risk framework," `https://www.bis.org/publ/bcbs265.pdf`.

BATES, J. M. AND C. W. GRANGER (1969): "The combination of forecasts," *OR*, 451–468.

BEHR, P., A. GUETTLER, AND F. MIEBS (2013): "On portfolio optimization: imposing the right constraints," *Journal of Banking & Finance*, 37, 1232–1242.

BEHR, P. AND F. MIEBS (2008): “Is minimum-variance investing really worth the while? An analysis with robust performance inference,” *EDHEC-Risk working paper*.

BEKAERT, G. AND G. WU (2000): “Asymmetric volatility and risk in equity markets,” *Review of Financial Studies*, 13, 1–42.

BERKOWITZ, J. (2001): “Testing Density Forecasts, with Applications to Risk Management,” *Journal of Business and Economic Statistics*, 19, 465–474.

BERKOWITZ, J., P. CHRISTOFFERSEN, AND D. PELLETIER (2011): “Evaluating Value-at-Risk Models with Desk-Level Data,” *Management Science*, 57, 2213–2227.

BEST, M. AND R. GRAUER (1991): “On the Sensitivity of Mean-Variance Portfolios to Changes in Asset Means: Some Analytical and Computational Results,” *The Review of Financial Studies*, 4(2), 315–342.

BILLINGSLEY, P. (2009): *Convergence of probability measures*, vol. 493, Wiley. com.

BISSANTZ, N., V. STEINORTH, AND D. ZIGGEL (2011a): “Stabilität von Diversifikationseffekten im Markowitz-Modell,” *AStA Wirtschafts- und Sozialstatistisches Archiv*, 5(2), 145–157.

BISSANTZ, N., D. ZIGGEL, AND K. BISSANTZ (2011b): “An empirical study of correlation and volatility changes of stock indices and their impact on risk figures,” *Acta Universitatis Danubius (Œconomica)*, 7(4), 127–141.

BLACK, F. AND R. LITTERMAN (1992): “Global portfolio optimization,” *Financial Analysts Journal*, 28–43.

BOLLERSLEV, T. (1986): “Generalized Autoregressive Conditional Heteroscedasticity,” *Journal of econometrics*, 31, 307–327.

——— (1990): “Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model,” *Review of Economics and Statistics*, 72, 498–505.

BOLLERSLEV, T., R. F. ENGLE, AND J. M. WOOLDRIDGE (1988): "A capital asset pricing model with time-varying covariances," *Journal of Political Economy*, 96, 116–131.

BOLLERSLEV, T. AND J. WOOLDRIDGE (1992): "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Reviews*, 11, 143–172.

BONTEMPS, C. (2006): "Moment-based tests for discrete distributions," *Working Paper*.

BONTEMPS, C. AND N. MEDDAHI (2005): "Testing normality: A GMM approach," *Journal of Econometrics*, 124, 149–186.

——— (2012): "Testing distributional assumptions: A GMM approach," *Journal of Applied Econometrics*, 27(6), 978–1012.

BRECHMANN, C. E. AND C. CLAUDIA (2013): "Risk management with high-dimensional vine copulas: An analysis of the Euro Stoxx 50," *Statistics & Risk Modeling*, 30, 307–342.

BROWNLEES, C., R. ENGLE, AND B. KELLY (2012): "A practical guide to volatility forecasting through calm and storm," *Journal of Risk*, 14, 3.

CAMPBELL, R., K. KOEDIJK, AND P. KOFMAN (2002): "Increased Correlation in Bear Markets," *Financial Analysts Journal*, 58, 87–94.

CAMPBELL, S. (2007): "A review of backtesting and backtesting procedures," *Journal of Risk*, 9(2), 1–18.

CANDELON, B., G. COLLETAZ, C. HURLIN, AND S. TOKPAVI (2011): "Backtesting Value-at-Risk: A GMM Duration-based Test," *Journal of Financial Econometrics*, 9(2), 314–343.

CAPPIELLO, L., R. F. ENGLE, AND K. SHEPPARD (2006): "Asymmetric dynamics in the correlations of global equity and bond returns," *Journal of Financial Econometrics*, 4, 537–572.

CHAN, L. K., J. KARCESKI, AND J. LAKONISHOK (1999): "On portfolio optimization: Forecasting covariances and choosing the risk model," *Review of Financial Studies*, 12(5), 937–974.

CHOPRA, V. AND W. ZIEMBA (1993): "The effect of errors in means, variances and covariances on optimal portfolio choice," *The Journal of Portfolio Management*, 19(2), 6–11.

CHOW, T., J. HSU, V. KALESNIK, AND B. LITTLE (2011): "A Survey of Alternative Equity Index Strategies," *Financial Analysts Journal*, 67(5), 37–57.

CHRISTOFFERSEN, P. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

——— (2009): "Value–at–Risk models," in *Handbook of Financial Time Series*, Springer, 753–766.

CHRISTOFFERSEN, P. AND D. PELLETIER (2004): "Backtesting Value-at-Risk: A Duration-Based Approach," *Journal of Financial Econometrics*, 2(1), 84–108.

CHU, C.-S. J., M. STINCHCOMBE, AND H. WHITE (1996): "Monitoring structural change," *Econometrica*, 64(5), 1045–1065.

CLARK, T. E. AND M. W. MCCRACKEN (2009): "Improving forecast accuracy by combining recursive and rolling forecasts," *International Economic Review*, 50, 363–395.

CLARKE, R., H. DE SILVA, AND S. THORLEY (2006): "Minimum-Variance Portfolios in the U.S. Equity Market," *Journal of Portfolio Management*, 33(1), 10–24.

CLARKE, R., H. DE SILVA, AND S. THORLEY (2013): "Minimum Variance, Maximum Diversification and Risk Parity: An Analytic Perspective," *Journal of Portfolio Management*, 39, 39–53.

COLACITO, R., R. F. ENGLE, AND E. GHYSELS (2011): "A component model for dynamic correlations," *Journal of Econometrics*, 164, 45–59.

DA VEIGA, B., F. CHAN, AND M. MCALEER (2011): "It pays to violate: how effective are the Basel accord penalties in encouraging risk management?" *Accounting & Finance*, 52, 95–116.

DANÍELSSON, J., B. N. JORGENSEN, M. SARMA, G. SAMORODNITSKY, AND C. G. DE VRIES (2005): "Subadditivity Re-Examined: the Case for Value-at-Risk," Discussion paper 549, LSE Financial Markets Group.

DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009): "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" *Review of Financial Studies*, 22(5), 1915–1953.

DEMIGUEL, V., Y. PLYAKHA, R. UPPAL, AND G. VILKOV (2012): "Improving portfolio selection using option-implied volatility and skewness," *Journal of Financial and Quantitative Analysis*, 1–57.

DONOHUE, C. AND K. YIP (2003): "Optimal portfolio rebalancing with transaction costs," *The Journal of Portfolio Management*, 29, 49–63.

DOWD, K. (1998): *Beyond Value at Risk: The New Science of Risk Management*, Wiley Frontiers in Finance, Chichester: John Wiley & Sons Ltd.

DUFOUR, J.-M. (2006): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," *Journal of Econometrics*, 133, 443–477.

DUMITRESCU, E.-I., C. HURLIN, AND V. PHAM (2012): "Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests," *Finance*, 33, 79–112.

ENGLE, R. (2002): "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business and Economic Statistics*, 20, 339–350.

ENGLE, R. AND K. SHEPPARD (2001): "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH," *Technical report, National Bureau of Economic Research Working paper 8554*.

ENGLE, R. F. AND S. MANGANELLI (2004): "CAViaR," *Journal of Business & Economic Statistics*, 22, 367–381.

ESCANCIANO, J. C. AND P. PEI (2012): "Pitfalls in backtesting Historical Simulation VaR models," *Journal of Banking & Finance*, 36, 2233–2244.

GALEANO, P. AND D. WIED (2014): "Multiple break detection in the correlation structure of random variables," *to appear in: Computational Statistics and Data Analysis*.

GARCIA, R. AND P. PERRON (1996): "An Analysis of the Real Interest Rate under Regime Shifts," *Review of Economics and Statistics*, 78(1), 111–125.

GIACOMINI, R. AND H. WHITE (2006): "Tests of conditional predictive ability," *Econometrica*, 74, 1545–1578.

GIOT, P. AND S. LAURENT (2003): "Value-At-Risk for long and short trading positions," *Journal of Applied Econometrics*, 18, 641–664.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. E. RUNKLE (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The journal of finance*, 48, 1779–1801.

GOHOUT, W. AND K. SPECHT (2007): "Mean-variance portfolios using Bayesian vector-autoregressive forecasts," *Statistical Papers*, 48(3), 403–418.

GOLOSNOY, V., S. RAGULIN, AND W. SCHMID (2011): "CUSUM control charts for monitoring optimal portfolio weights," *Computational Statistics and Data Analysis*, 55(11), 2991–3009.

GOLOSNOY, V. AND W. SCHMID (2007): "EWMA control charts for monitoring optimal portfolio weights," *Sequential Analysis*, 26, 195–224.

HAAS, M. (2005): “Improved duration-based backtesting of Value-at-Risk,” *Journal of Risk*, 8(2), 17–36.

HALBLEIB, R. AND W. POHLMEIER (2012): “Improving the value at risk forecasts: Theory and evidence from the financial crisis,” *Journal of Economic Dynamics and Control*, 36, 1212–1228.

HANSEN, P. R. AND A. LUNDE (2005): “A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?” *Journal of Applied Econometrics*, 20, 873–889.

HAUGEN, R. AND N. BAKER (1991): “The Efficient Market Inefficiency of Capitalization-Weighted Stock Portfolios,” *Journal of Portfolio Management*, 17(3), 35–40.

HILLEBRAND, E. (2005): “Neglecting parameter changes in GARCH models,” *Journal of Econometrics*, 129, 121–138.

HUI, T.-K. (2005): “Portfolio diversification: a factor analysis approach,” *Applied Financial Economics*, 15(12), 821–834.

HURLIN, C. AND S. TOKPAVI (2006): “Backtesting value-at-risk accuracy: a simple new test,” *Journal of Risk*, 9, 19.

JACOBSEN, B. (2011): “Does Active Management Provide Investor Surplus?” *Journal of Portfolio Management*, 38(1), 131–139.

JAGANNATHAN, R. AND T. MA (2003): “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constrains Helps,” *The Journal of Finance*, 58(4), 1651–1684.

JORION, P. (2006): *Value at Risk: The New Benchmark for Managing Financial Risk*, Chicago: McGraw-Hill Professional, 3 ed.

KEMPF, A. AND C. MEMMEL (2002): “Schätzrisiken in der Portfoliotheorie,” in *Handbuch Portfoliomanagement*, ed. by J. Kleeberg and H. Rehkugler, Uhlenbruch, 893–919.

KRISHAN, C., R. PETKOVA, AND P. RITCHKEN (2009): "Correlation risk," *Journal of Empirical Finance*, 16(3), 353–367.

KRITZMAN, M., S. MYRGREN, AND S. PAGE (2009): "Optimal rebalancing: a scalable solution," *Journal of investment management*, 9.

KRZANOWSKI, W. (2000): *Principles of multivariate analysis*, Oxford University Press, USA.

KUESTER, K., S. MITTNIK, AND M. S. PAOLELLA (2006): "Value-at-risk prediction: A comparison of alternative strategies," *Journal of Financial Econometrics*, 4, 53–89.

KUPIEC, P. (1995): "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 3, 73–84.

LEIPPOLD, M., F. TROJANI, AND P. VANINI (2006): "Equilibrium impact of value-at-risk regulation," *Journal of Economic Dynamics and Control*, 30, 1277–1313.

LENOIR, G. AND N. S. TUCHSCHMID (2001): "Investment time horizon and asset allocation models," *Financial Markets and Portfolio Management*, 15(1), 76–93.

LONGIN, F. AND B. SOLNIK (1995): "Is the correlation in international equity returns constant: 1960-1990?" *International Money and Finance*, 14(1), 3–26.

——— (2001): "Extreme correlation of international equity markets," *Journal of Finance*, 56(2), 649–676.

MANDELBROT, B. (1962): "The Variation of Certain Speculative Prices," *IBM Research Report NC-87*.

MARKOWITZ, H. (1952): "Portfolio selection," *Journal of Finance*, 7(1), 77–91.

MARKOWITZ, H. M. AND E. L. VAN DIJK (2003): "Single-Period Mean: Variance Analysis in a Changing World," *Financial Analysts Journal*, 30–44.

MASTERS, S. J. (2003): "Rebalancing," *The Journal of Portfolio Management*, 29, 52–57.

McNeil, A. J. and R. Frey (2000): “Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach,” *Journal of empirical finance*, 7, 271–300.

McNeil, A. J., R. Frey, and P. Embrechts (2005): *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press.

Merton, R. (1980): “On estimating the expected return on the market: an exploratory investigation,” *Journal of Financial Economics*, 8(4), 323–361.

Michaud, R. (1989): “The Markowitz-Optimization Enigma: Is ’Optimized’ Optimal?” *Financial Analysts Journal*, 45(1), 31–42.

Mina, J., J. Y. Xiao, et al. (2001): “Return to RiskMetrics: the evolution of a standard,” *RiskMetrics Group*.

Pelletier, D. (2006): “Regime switching for dynamic correlations,” *Journal of Econometrics*, 131(1-2), 445–473.

Pérignon, C. and D. R. Smith (2010): “The level and quality of Value-at-Risk disclosure by commercial banks,” *Journal of Banking & Finance*, 34, 362–377.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006): “Forecasting time series subject to multiple structural breaks,” *The Review of Economic Studies*, 73, 1057–1084.

Pesaran, M. H. and A. Timmermann (2004): “How costly is it to ignore breaks when forecasting the direction of a time series?” *International Journal of Forecasting*, 20, 411–425.

——— (2007): “Selection of estimation window in the presence of breaks,” *Journal of Econometrics*, 137, 134–161.

Ploberger, W., W. Krämer, and K. Kontrus (1989): “A new test for structural stability in the linear regression model,” *Journal of Econometrics*, 40(2), 307–318.

PODOLSKIJ, M. AND D. ZIGGEL (2009): “New tests for jumps in semimartigale models,” *Statistical Inference for Stochastic Processes*, 13(1), 15–41.

POJARLIEV, M. AND W. POLASEK (2003): “Portfolio construction by volatility forecasts: Does the covariance structure matter?” *Financial Markets and Portfolio Management*, 17(1), 103–116.

POON, S.-H. AND C. GRANGER (2003): “Forecasting volatility in financial markets: a review,” *Journal of Economic Literature*, 41, 478–539.

RAPACH, D. E. AND J. K. STRAUSS (2008): “Structural breaks and GARCH models of exchange rate volatility,” *Journal of Applied Econometrics*, 23, 65–90.

RAPACH, D. E., J. K. STRAUSS, AND M. E. WOHAR (2008): “Forecasting stock return volatility in the presence of structural breaks,” *Frontiers of Economics and Globalization*, 3, 381–416.

RISKMETRICS GROUP (1996): “RiskMetrics-Technical Document,” *New York-JP Morgan*.

SANTOS, A. A., F. J. NOGALES, AND E. RUIZ (2013): “Comparing univariate and multivariate models to forecast portfolio value-at-risk,” *Journal of Financial Econometrics*, 11, 400–441.

SHARPE, W. (1991): “The Arithmetic of Active Management,” *Financial Analysts Journal*, 47(1), 7–9.

SHIN, H. S. (2010): *Risk and Liquidity*, Clarendon Lectures in Finance, New York: Oxford University Press Inc.

STANDARD & POOR’S (2012): “S&P Indices Versus Active Funds Scorecard,” Mid-Year 2012.

STARICA, C., S. HERZEL, AND T. NORD (2005): “Why does the GARCH (1, 1) model fail to provide sensible longer-horizon volatility forecasts,” *Manuscript, Chalmers University of Technology*.

SUN, W., A. FAN, L.-W. CHEN, T. SCHOUWENAARS, AND M. A. ALBOTA (2006): "Optimal rebalancing for institutional portfolios," *The Journal of Portfolio Management*, 32, 33–43.

TIMMERMANN, A. (2006): "Forecast combinations," *Handbook of economic forecasting*, 1, 135–196.

TSE, Y. K. AND A. K. C. TSUI (2002): "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations," *Journal of Business & Economic Statistics*, 20, 351–362.

VOSTRIKOVA, L. (1981): "Detecting 'disorder' in multidimensional random processes," *Soviet Mathematics Doklady*, 24, 55–59.

WALD, A. AND J. WOLFOWITZ (1940): "On a Test Whether Two Samples are from the Same Population," *The Annals of Mathematical Statistics*, 11(2), 147–162.

WEISS, G. N. AND H. SUPPER (2013): "Forecasting Liquidity-Adjusted Intraday Value-at-Risk with Vine Copulas," *Journal of Banking & Finance*, 37, 3334–3350.

WERMERS, R. (2000): "Mutual Fund Performance:An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses," *The Journal of Finance*, 55(4), 1655–1695.

WIED, D. (2012): "A nonparametric test for a constant correlation matrix," *preprint, arXiv: http://arxiv.org/abs/1210.1412*.

——— (2013): "CUSUM-type testing for changing parameters in a spatial autoregressive model of stock returns," *Journal of Time Series Analysis*, 34(1), 221–229.

WIED, D., M. ARNOLD, N. BISSANTZ, AND D. ZIGGEL (2012a): "A new fluctuation test for constant variances with applications to finance," *Metrika*, 75, 1111–1127.

——— (2013a): "Über die Anwendbarkeit eines neuen Fluktuationstests für Korrelationen auf Finanzzeitreihen," *AStA Wirtschafts- und Sozialstatistisches Archiv*, 6(3), 87–103.

WIED, D. AND P. GALEANO (2013): “Monitoring correlation change in a sequence of random variables,” *Journal of Statistical Planning and Inference*, 143(1), 186–196.

WIED, D., W. KRÄMER, AND H. DEHLING (2012b): “Testing for a change in correlation at an unknown point in time using an extended functional delta method,” *Econometric Theory*, 28(3), 570–589.

WIED, D., D. ZIGGEL, AND T. BERENS (2013b): “On the application of new tests for structural changes on global minimum-variance portfolios,” *Statistical Papers*, 54, 955–975.

ZHOU, X. AND G. YIN (2003): “Markowitz’s mean-variance portfolio selection with regime switching: A Continuous-Time Model,” *SIAM Journal on Control and Optimization*, 42(4), 1466–1482.

ZIGGEL, D., T. BERENS, G. N. WEISS, AND D. WIED (2013): “A New Set of Improved Value-at-Risk Backtests,” *Available at SSRN 2316134*.

ZIGGEL, D. AND D. WIED (2012): “Trading strategies based on new fluctuation tests,” *IFTA-Journal (Journal of the International Federation of Technical Analysts)*, 2012 Edition, 17–20.

Eidesstattliche Versicherung

Hiermit versichere ich, dass ich diese Dissertation selbstständig verfasst habe. Bei der Erstellung der Arbeit habe ich mich ausschließlich der angegebenen Hilfsmittel bedient. Die Dissertation ist nicht bereits Gegenstand eines erfolgreich abgeschlossenen Promotions- oder sonstigen Prüfungsverfahrens gewesen.

Dortmund, 16.05.2014