

Essays on Measuring (Credit-)Risk in Banks, Financial Accounting and Auditing

– Theory and Practice –

Inaugural-Dissertation

zur

Erlangung der wirtschaftswissenschaftlichen Doktorwürde
des Fachbereichs Wirtschaftswissenschaften
der Philipps-Universität Marburg

eingereicht von:

Tobias Filusch

Diplom Wirtschaftsmathematiker aus Bad Hersfeld

Erstgutachter:	Prof. Dr. Sascha H. Mölls
Zweitgutachter	Prof. Dr. Bernhard Nietert
Einreichungstermin:	06. Juni 2019
Prüfungstermin:	23. Januar 2020
Erscheinungsort:	Marburg
Hochschulkennziffer:	1180

Inhaltsverzeichnis

Zusammenfassung	1
Beitrag #1	16
Quantile Score Estimation to Improve Value at Risk Forecasts (Unveröffentlichtes Manuskript)	
Beitrag #2	44
„(Lifetime) Expected Credit Losses“ als Mechanismus der Verlustantizipation nach IFRS 9 – Operationalisierung der Schätzparameter in kritischer Betrachtung , in: Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung IFRS 17(6), S. 249-255, 2017 (mit Mölls, S.H.).	
Beitrag #3	61
Konjunkturabhängige PD-Bestimmung nach IFRS 9 – Zum Informationsgehalt öffentlich verfügbarer Ratings , in: Crasselt, N., Lukas, E., Mölls, S.H., Timmreck, C. (Hrsg.): Handbuch Kapitalmarktorientierte Unternehmensbewertung – Grundlagen, Methoden, Regulierung und Branchentrends, S. 429-447, Stuttgart, 2018 (mit Mölls, S.H. und Ratz, C.).	
Beitrag #4	82
Calibrating and Forecasting Point in Time Probability of Default Term Structures (Unveröffentlichtes Manuskript)	
Beitrag #5	112
The Bargaining Game of the Audit – Regulatory Provisions for High-Quality Financial Statements and an Independent Audit (Unveröffentlichtes Manuskript)	
Eidesstattliche Versicherung	134

**Essays on Measuring (Credit-)Risk in Banks,
Financial Accounting and Auditing
– Theory and Practice –**

Zusammenfassung

vorgelegt von

Dipl.-Math. oec. Tobias Filusch

aus

Bad Hersfeld

Betreuer:

Prof. Dr. Sascha H. Mölls
Philipps-Universität Marburg
Fachbereich Wirtschaftswissenschaften
Professur für Rechnungslegung
Barfüßertor 2
35032 Marburg

Juni 2019

1 Beschreibung der übergeordneten Problemstellung

Die gezielte Steuerung des finanziellen Risikos bestimmt maßgeblich den wirtschaftlichen Erfolg von Unternehmen im Finanzsektor (vgl. u.a. Rolfes (2008)). Allerdings erhöhen risikoreiche Investmentstrategien auch die Gefahr von Schieflagen und Insolvenzen von Finanzinstituten, welche in der Folge sogar zu einer (globalen) Finanzkrise führen können (vgl. u.a. Friedman (2011)). Um das Insolvenzrisiko prinzipiell zu reduzieren, etablieren viele Legislativbehörden Mindestquoten für das vorzuhaltende Eigenkapital.¹ Der Ausgangspunkt solcher Regulierungsmaßnahmen ist häufig die fortlaufend aktualisierte Empfehlung von Organisationen wie dem Basel Committee of Banking Supervision (BCBS), die insbesondere für die Richtlinien Basel I, II und III verantwortlich sind (BCBS, 2019). In diesen Richtlinien werden vereinheitlichte Methoden präsentiert, welche die Berechnung von Risikokennzahlen sowie die davon abhängige Bestimmung der Untergrenze des Eigenkapitals umfassen.

Das nötige Eigenkapital beschaffen sich vor allem große Finanzunternehmen häufig an Kapitalmärkten (vgl. u.a. Hartmann-Wendels et al. (2010)). Unternehmen, die als Kapitalnehmer auf dem Kapitalmarkt aktiv werden möchten, sind verpflichtet, der Öffentlichkeit regelmäßig über die eigene wirtschaftliche Lage Auskunft zu geben. Eine wirklich verlässliche Informationsgrundlage liegt allerdings nur dann vor, wenn die bereitgestellten Informationen korrekt und über die Gesamtheit der am Kapitalmarkt aktiven Unternehmen vergleichbar sind (Wagenhofer und Ewert, 2015). Diese Informationsfunktion kann grundsätzlich durch Jahresabschlüsse gemäß anerkannter Rechnungslegungsvorschriften, wie den International Financial Reporting Standards (IFRS), erfüllt werden (Schildbach et al., 2013). Solche Standards umfassen dabei auch qualitative Berechnungsanforderungen für sämtliche auszuweisende Kennzahlen und fokussieren insgesamt auf die Finanzlage der Unternehmen, was folglich auch Risikomessungen mit einschließt (Zülch und Hendl, 2019). Neben den oben beschriebenen Richtlinien zur Bestimmung der Mindestquote des Eigenkapitals führt dies damit zu einem zweiten, parallel entstandenen, Regulierungsrahmen im Hinblick auf die Messung von unternehmensinternen Risiken, die darüber Einzug in die Jahresabschlüsse der Unternehmen erhalten.

Ein entscheidendes Kriterium für die Nützlichkeit von Jahresabschlüssen ist, dass diese die tatsächlichen Werte der gehaltenen Vermögenswerte des entsprechenden Unternehmens gemäß eines gegebenen Rechnungslegungsstandards widerspiegeln. Um die Verlässlichkeit dieses Anspruches für potenzielle Investoren zu stärken, finden regelmäßige Prüfungen von Jahresabschlüssen durch Dritte statt (Krommes, 2015). Der Berufsstand der Wirtschaftsprüfer ist in diesem Zusammenhang damit beauftragt, Unternehmen und deren Berichterstattung zu kontrollieren. Die Wirtschaftsprüfung soll die Jahresabschlüsse prüfen, die Korrektheit bestätigen und gefundene Fehler aufdecken. Damit dies auch tatsächlich so stattfindet, unterliegt auch der Wirtschaftsprüfermarkt einer behördlichen Regulierung, welche u.a. die Unabhängigkeit der Prüfer und damit die tatsächliche Meldung von gefundenen Fehlern oder Manipulationen in Jahresabschlüssen als ein zentrales Ziel aufweist (Wagenhofer und Ewert, 2015). Zusammenfassend liegen also drei Regulierungsquellen vor, welche das interne Risikomanagement und damit die interne Unternehmenssteuerung, die Berichterstattung im Umfeld der Kapitalmärkte und schließlich die Prüfung dieser Berichterstattung betreffen. Diese Regulierungsfelder beinhalten jeweils offene Forschungsfragen, welche nun enger eingegrenzt werden, um darüber

¹Vgl. z.B. die EU Kapitaladäquanzverordnung, welche in EU Verordnung Nr. 575/2013 formalisiert ist.

zu den in dieser Dissertation bearbeiteten Detailfragen zu gelangen.

Für das Risikomanagement in Banken ist das Risikomaß Value at Risk (VaR) von zentraler Bedeutung (vgl. u.a. Reichling (2007)). Der VaR gibt für Finanzpositionen (z.B. Wertpapierportfolios) definitorisch an, welche Verlusthöhe innerhalb eines gegebenen Zeitraums mit einer gegebenen Wahrscheinlichkeit nicht überschritten wird (Jorion, 2006). Die konkrete Ausgestaltung und Fortentwicklung der zur VaR-Messung herangezogenen Modelle wird seit den 1990er Jahren durch die Vorgaben des BCBS abgesteckt, die für die Praxisumsetzung – neben Konkretisierungen zu den anzuwendenden quantitativen Methoden und der Berechnungslogik zur Eigenkapitalunterlegung – auf operativer Ebene auch Anforderungen an die einzubeziehenden Daten sowie die Betrachtungsintervalle formulieren (vgl. u.a. Zhang (2017)). Die in der Praxis verwendeten VaR-Modelle stehen dabei fortlaufend auf dem Prüfstand, um Schwächen zu identifizieren, Verbesserungen zu initiieren und infolgedessen die Messung sowie Steuerung von finanziellen Risiken – vor allem zur Vermeidung von finanziellen Krisen – zu optimieren (Friedman, 2011; Neisen und Röth, 2018). Bei der Ausgestaltung der statistischen Methoden werden meistens weder von den gesetzgebenden Behörden, noch von den Standardsetzern Präzisierungen im Detail vorgeschrieben (vgl. u.a. BCBS (2019)). Insbesondere fällt im Bereich der verwendeten Parameterschätzung, welche für den VaR im Rahmen eines parametrischen Modells unerlässlich ist, auf, dass die Schätzmethodik bei Anwendern typischerweise nicht vom Maximum Likelihood Estimator (MLE) abweicht (vgl. u.a. McNeil et al. (2015)). Ob das Abweichen von diesem statistischen Standardverfahren tatsächlich einen positiven Effekt auf die Qualität der Risikomessung haben kann, ist bislang unklar und grenzt damit den ersten Untersuchungsschwerpunkt dieser Dissertation ab.

Die Fehleinschätzung von Risiken und die daraus resultierenden Schieflagen sowie Insolvenzen namhafter Banken – wie etwa Lehman Brothers (McDonald, 2015) – im Zuge der globalen Finanzmarktkrise um das Jahr 2008 haben nicht nur Forderungen nach einer weiteren Verbesserung der Risikomodelle von Finanzinstituten in den Fokus gerückt, sondern auch die Ausgestaltung der flankierenden kapitalmarktorientierten Rechnungslegung nach den IFRS tangiert (Bartmann et al., 2009). Die Entwicklung des IFRS 9, der den International Accounting Standard (IAS) 39 zum Jahr 2018 abgelöst hat, wurde dadurch maßgeblich beeinflusst, indem u.a. eine Lösung für die Kritik einer zu späten Bilanzierung von Verlusten integriert wurde (vgl. u.a. Henselmann et al. (2014)). Gemäß der im IFRS 9 verankerten Verlustantizipation werden für die Bestimmung von Wertminderungsbeträgen bereits solche Verluste einbezogen, die am Bilanzstichtag mit Blick auf die gesamte (Rest-)Laufzeit der Position zu erwarten sind (IASB, 2014). Diese Expected Credit Losses (ECLs) sind im Gegensatz zu den zuvor im Rahmen des IAS 39 maßgeblichen Incurred Losses zukunftsgerichtet und verschieben die Methoden der Rechnungslegung damit konzeptionell deutlich in Richtung der Ansätze des Risikomanagements (Schröder, 2015). Diese Vorgaben des IFRS 9 werden in den Richtlinien lediglich qualitativ verbalisiert und führen damit zu keiner starken Eingrenzung eines konkreten quantitativen Rahmens. Für die Umsetzung der Regelungen des IFRS 9 liegt es allerdings nahe, auf die bereits bestehenden und angewendeten Konzepte des Risikomanagements zurückzugreifen und diese zugunsten der abweichenden Ansprüche des IFRS 9 anzupassen (Grünberger, 2013). Entsprechend ergeben sich offene quantitative Umsetzungsfragen für die Bestimmung von ECLs vor dem Hintergrund des IFRS 9. Die Eingrenzung möglicher Umsetzungskonzepte, die Validierung vorliegender Daten für die Verwendung innerhalb dieser Konzepte und die Entwicklung konkreter Modelle für die im Rahmen des IFRS 9 vorgesehene

Verlustantizipation bilden daher den zweiten Schwerpunktbereich dieser Dissertation.

Die Regulierung der Jahresabschlussprüfung zielt häufig primär auf die Realisierung der tatsächlich unabhängigen Stellung des Wirtschaftsprüfers ab (vgl. u.a. Ewert (2003)). Wird die Jahresabschlussprüfung durch einen unabhängigen und hinreichend sorgfältigen Prüfer geführt, so wird erwartet, dass dieser Mängel im Zuge der Prüfung aufdeckt und der Öffentlichkeit wahrheitsgemäß darüber berichtet. Der zentrale Einflussfaktor, der dieser unabhängigen Stellung entgegenwirkt, ist dabei der (finanzielle) Druck, den das Management, als die für die Abschlüsse im geprüften Unternehmen Verantwortlichen, auf den Prüfer ausüben kann (DeAngelo, 1981a). Um dies zu verdeutlichen, muss der Blick auf die genaue (geschäftliche) Beziehung zwischen den beiden Parteien gerichtet werden, in der die Prüfungsgesellschaft durch das Unternehmen selbst für i.d.R. lediglich eine Einzelprüfung beauftragt wird und somit nicht zwangsläufig auch in Folgeperioden wiederbestellt werden müsste. Der Verlust von Mandaten ist für Prüfungsgesellschaften grundsätzlich unattraktiv, woraus Anreize für eine ggf. zu enge Zusammenarbeit mit dem Geprüften und ein im beidseitigen Interesse stehendes Prüfungsurteil entstehen können. Ein solches Urteil wird dabei häufig als das Resultat einer Auditor-Client Management Negotiation (ACMN) beschrieben, welche eine grundsätzlich kooperative Haltung der Akteure voraussetzt (vgl. u.a. Gibbins et al. (2001)). Inwiefern dieses mögliche Kooperationsverhalten die Qualität von Jahresabschlüssen mindert und wie ein derartiges Szenario durch gezielte Regulierung vermieden werden kann, sind die zentralen Forschungsfragen im dritten Schwerpunktbereich dieser Dissertation.

2 Darstellung der Forschungsobjekte

Für die Schätzung von parametrischen Risikomodellen, bzw. den zugehörigen Modellparametern, hat sich bei Zugrundelegung von Zeitreihendaten der MLE als geeigneter Schätzer bewährt (Greene, 2012). Wird der VaR basierend auf einer geeigneten Parameterschätzung prognostiziert, so kann diese Vorhersage im Rahmen von Validierungsverfahren, wie etwa dem Backtesting, hinsichtlich ihrer Güte bewertet werden (Campbell, 2005). Diese Bewertungsmethoden basieren im Kern auf der Verwendung von „Scoring-Funktionen“, die einem (Werte-)Paar – bestehend aus Vorhersage und realisierter Beobachtung – eine reelle Zahl zuordnen (Gneiting und Raftery, 2007) und damit implizit auch eine Vergleichsmöglichkeit zwischen unterschiedlichen Vorhersagemodellen bieten (McNeil et al., 2015). Statt eines solchen nachträglichen Vergleichs könnten solche Funktion allerdings auch bereits vor der Vorhersage eingesetzt werden, also die „Likelihood-Funktion“ als zentrales Schätzkriterium ablösen. Die Verwendung von an dem Vorhersageziel orientierten „Scoring-Funktionen“ zur Schätzung von Modellparametern ist in der Literatur bereits diskutiert und deren Relevanz herausgestellt worden (vgl. u.a. Weiss (1996)). Innerhalb dem gesamtheitlichen Konzept der M-Schätzung können „Scoring-Funktionen“ eine Verallgemeinerung des ML-Schätzers bilden (Huber, 1964). Insbesondere in den Fällen, in denen das verwendete Prognosemodell, für das die Parameter geschätzt werden sollen, nicht angemessen auf die zeitliche Dynamik der Daten passt (d.h. bei einem falsch spezifizierten Modell), entsteht eine Abhängigkeit des Schätzwertes von der im Schätzverfahren eingebetteten „Scoring-Funktion“ (Van der Vaart, 1998). Während dieser Fall bei Modellschätzungen, basierend auf beobachteten (Finanz-)Daten, regelmäßig mit unbe-

kanntem Ausmaß vorliegt (Box, 1976)², existiert eine derartige Abhängigkeit in einem korrekt spezifizierten Modell nicht, sodass dort die Verwendung alternativer „Scoring-Funktionen“ innerhalb des M-Schätzers asymptotisch (d.h. in hinreichend großer Datenmenge) im selben, nämlich dem wahren Modellparameter mündet (Van der Vaart, 1998). Die Verbesserung der Modellspezifizierung geht meist mit einer „teuren“ Komplexitätserhöhung des Modells einher (z.B. über die Hinzunahme von Modellparametern), weswegen ein grundsätzliches Beibehalten des gegebenen Modells ein anstrebenwertes Ziel darstellt. Inwiefern dann aber speziell für den VaR als Vorhersageziel gewählte Schätzverfahren basierend auf „Quantile Scoring-Funktionen“ (Gneiting, 2011) für ein einfaches Prognosemodell wie das GARCH(1,1) (Bollerslev, 1986) verwendet werden und dabei den MLE herausfordern oder sogar übertrumpfen kann, wird in Beitrag #1 untersucht und beantwortet.

Der Rechnungslegungsstandard IFRS 9 liefert eine Reihe von qualitativen Vorgaben für die Bestimmung von ECLs, bleibt im Hinblick auf die quantitative Umsetzung allerdings sehr unspezifisch (IASB, 2014). Entsprechend bedarf es mit Blick auf eine zielgerichtete Anwendung umfassender Präzisierungen, die bisher hauptsächlich von den großen Wirtschaftsprüfungs- und Beratungsgesellschaften (vgl. u.a. Grosskord et al. (2013)) sowie – mit etwas stärkerer quantitativer Ausrichtung – von Grünberger (2013) vorgeschlagen wurden. Generell bauen diese Arbeiten für die rechnungslegungsbezogenen Lösungen auf die bestehenden Kreditrisikomodelle von Banken auf. Eine derartige Übertragung der durch das BCBS vorgeschlagenen Ansätze verlangt allerdings weitreichende Anpassungen, da sich ihre Zielsetzung im Detail von denen des IASB z.T. deutlich unterscheidet. Eine mehrperiodische Zerlegung des ECL in Verlustquote (LGD – Loss Given Default), Forderungshöhe bei Verlust (EAD – Exposure at Default) und Ausfallwahrscheinlichkeit (PD – Probability of Default) ist im Zuge dessen möglich (Grünberger, 2013), allerdings gilt es, dann zeitpunkt-/ konjunkturabhängige, mehrjährige Laufzeitstrukturen dieser Kernparameter zu schätzen, was im Rahmen des Risikomanagements so nicht vorliegt. Während erwartete Verluste in der bankbetrieblichen Praxis unter Zugrundelegung eines einjährigen Betrachtungszeitraums konjunkturunabhängig quantifiziert werden (Bluhm, 2016), soll die „Lifetime-Betrachtung“ (Lifetime Expected Credit Loss – LECL) die gesamte Restlaufzeit von Krediten umfassen und damit die zeitpunktabhängige Zukunftsausrichtung explizit in den Fokus rücken.

Die Forschungsobjekte im Rahmen der Beiträge #2, #3 und #4 behandeln die Umsetzung einer LECL-Quantifizierung im Rahmen des IFRS 9 „Impairment-Modells“ schrittweise. In Beitrag #2 wird dabei die zentrale Zerlegung des LECL in die drei Kernparameter LGD, EAD und PD präzisiert, sowie systematische Operationalisierungsvorschläge dafür diskutiert. Unter anderem wird der Blick dabei genauer auf die Ausfallstruktur gerichtet, die sowohl das vorhandene (Kredit-)Risiko stark verkörpert, als auch sehr sensibel auf konjunkturelle Veränderungen reagiert. Diese Erkenntnis motiviert eine weitere Fokussierung auf genau diesen Parameter, der dabei – falls konjunkturbedingt modelliert – sogar auch in der Lage ist, Konjunktoreinflüsse direkt in den LECL zu integrieren. Die konjunkturelle Abhängigkeit von Ausfallwahrscheinlichkeiten wird in Beitrag #3 empirisch nachgewiesen, indem Ausfalldaten der Ratingagentur Standard & Poor’s analysiert werden. Damit ist die Voraussetzung für eine sinnvolle PD-Modellierung, basierend auf diesen Daten, grundsätzlich gegeben. Die-

²In diesem Zusammenhang ist das bekannte Zitat „all models are wrong, but some are useful“ zu nennen, welches regelmäßig dem Statistiker George Box zugesprochen wird.

ser entscheidende Schritt wird in Beitrag #3 bereits angerissen, eröffnet in der Folge aber weiterführende Verbesserungswünsche. Genau dort setzt Beitrag #4 an, der umfassendere Modellierungsansätze formal einführt, anwendet und diskutiert.

Im dritten Schwerpunktbereich dieser Dissertation wird die Regulierung der Jahresabschlussprüfung im Hinblick auf eine verbesserte Anreizstruktur zugunsten des regelkonformen Verhaltens von Prüfer und Management betrachtet. Die wenigen (mikro-)ökonomisch fundierten Analysen zur Unabhängigkeit des Wirtschaftsprüfers bauen regelmäßig auf Quasirentenansätzen (Milgrom und Roberts, 1992) auf, die die Anreize des Prüfers zur wahrheitsgemäßen Berichterstattung im Sinne eines Opportunitätskostenkalküls auf der Basis von Kapitalwerten messen. Sofern aus einem Mandat positive Wertbeiträge resultieren, also Quasirenten existieren, beeinträchtigen diese die Unabhängigkeit des Prüfers und sollten daher durch Regulierung möglichst gering gehalten werden (Wagenhofer und Ewert, 2015). In diesem Zusammenhang zeigt DeAngelo (1981 *a,b*), dass ausgehend von einer Situation des Low Balling³, der amtierende Prüfer bei den Folgeprüfungen aufgrund von Marktunvollkommenheiten, in Form von Transaktionskosten des Prüferwechsels, eine über seinen Kosten liegende Gebühr verlangen kann. Folglich ist von einer Einschränkung der Unabhängigkeit in einer derartigen Konstellation auszugehen.

Indem die anknüpfende Literatur zur Unabhängigkeit auf die qualitativen Befunde der ACMN-Forschung verzichtet, wird die Entscheidungssituation des Prüfers, innerhalb der engen Geschäftsbeziehung zum Mandanten, nur ungenügend berücksichtigt. Die dort beschriebenen Verhandlungen sind zwar kein offizieller Bestandteil von Abschlussprüfungen, werden in der Praxis aber nachweislich geführt. Sie beinhalten als Gegenstand dabei u.a. die genaue Ausgestaltung des Prüfberichts und determinieren damit auch das Prüfungsurteil (Gibbins et al., 2001). Indem das Management faktisch die Verantwortung für die zukünftige Wiederbestellung des Prüfers trägt (McCracken et al., 2008) und der Prüfer diese erreichen möchte, um damit die ggf. vorliegende Quasirente abschöpfen zu können, entsteht eine Drucksituation. Unter erhöhtem Druck des Managements könnte eine Koalition zwischen den beiden Parteien bzgl. möglicherweise sogar absichtlich verschönerter Bilanzkennzahlen einem beidseitigen ökonomischen Vorteil bedeuten (Bennett et al., 2015; Fearnley et al., 2005).⁴ Im Zuge dessen kommt es zur unausweichlichen Frage, wie genau diese Kooperationsrente aufgeteilt werden soll, was der zentrale Bestandteil der ACMN sein könnte und somit auch vor dem Hintergrund wirksamer Regulierungsmaßnahmen nicht ausgeblendet werden darf. Tritt der Prüfer dieser Verhandlung bei, so ist die Unabhängigkeit im Allgemeinen nicht gegeben. Je mehr sich das Management auf das Stattfinden einer solchen Verhandlung verlassen kann, desto attraktiver werden Manipulationen der Abschlüsse. Insbesondere gilt das, wenn die Verteilung der Verhandlungsmacht eher das Management begünstigt. Diese Gegebenheiten stehen im starken Konflikt zu der Qualitätssicherung von Jahresabschlüssen gegenüber den Adressaten, was das Handeln der Akteure im Rahmen der ACMN in direkte Beziehung zu der Wirksamkeit von Regulierungsmaßnahmen setzt.

Der (spiel-)theoretische Ansatz in Beitrag #5 modelliert das beschriebene Zusammenspiel der Prüfungsbeteiligten als Nash-Verhandlung (Nash, 1950), wobei insbesondere auch die Un-

³D.h. einer Gebührengestaltung, bei der die Kosten der Prüfung die entsprechenden Honorare übersteigen.

⁴Zu absichtlichen Verschönigungen im Rahmen des Earnings Management vgl. u.a. Healy und Wahlen (1999).

sicherheiten der Akteure über zukünftige Cash-Flows stochastische Berücksichtigung finden. Damit werden bestehende und auch aktuell diskutierte Maßnahmen – wie u.a. eine vorgeschriebene Prüferrotation oder ein erhöhtes Strafmaß bei Pflichtverletzungen – analysiert und bewertet. Die durch Regulierungsmaßnahmen determinierten Modellparameter beeinflussen maßgeblich die Entscheidungen der Beteiligten, welche im Sinne des öffentlichen Interesses zugunsten einer wahrheitsgemäßen Berichterstattung und/ oder einer unabhängigen Abschlussprüfung geführt werden sollten. Beitrag #5 zeigt auf, wie dieses Ziel durch eine geschickte Kombination von Regulierungsmaßnahmen erreicht werden kann.

Folgende Übersicht zeigt den Veröffentlichungsstand der einzelnen Beiträge zum Zeitpunkt der Abgabe der Dissertation sowie den prozentualen Anteil der beteiligten Autoren an der Erstellung der Beiträge:

- Beitrag #1: **„Quantile Score Estimation to Improve Value at Risk Forecasts“**
(Unveröffentlichtes Manuscript)
Autor: Tobias Filusch: 100%
- Beitrag #2: **„(Lifetime) Expected Credit Losses als Mechanismus der Verlustantizipation nach IFRS 9 – Operationalisierung der Schätzparameter in kritischer Betrachtung“**, in: Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung IFRS 17(6), S. 249-255, 2017 (mit Mölls, S.H.)
Hauptautor: Tobias Filusch: 50%
Koautor: Prof. Dr. Sascha H. Mölls: 50%
- Beitrag #3: **„Konjunkturabhängige PD-Bestimmung nach IFRS 9 – Zum Informationsgehalt öffentlich verfügbarer Ratings“**, in: Crasselt, N., Lukas, E., Mölls, S.H., Timmreck, C. (Hrsg.): Handbuch Kapitalmarktorientierte Unternehmensbewertung – Grundlagen, Methoden, Regulierung und Branchentrends, S. 429-447, Stuttgart, 2018 (mit Mölls, S.H. und Ratz, C.)
Hauptautor: Tobias Filusch: 33%
1. Koautor: Prof. Dr. Sascha H. Mölls: 33%
2. Koautor: Christopher Ratz: 33%
- Beitrag #4: **„Calibrating and Forecasting Point in Time Probability of Default Term Structures“**
(Unveröffentlichtes Manuscript)
Autor: Tobias Filusch: 100%
- Beitrag #5: **„The Bargaining Game of the Audit – Regulatory Provisions for High-Quality Financial Statements and an Independent Audit“**
(Unveröffentlichtes Manuscript)
Hauptautor: Tobias Filusch: 95%
Koautor: Prof. Dr. Sascha H. Mölls: 5%

3 Beitrag zur bestehenden Literatur

Innerhalb der behandelten Themengebiete existiert bereits eine Vielzahl an Literatur, welche in den einzelnen Beiträgen genauer vorgestellt wird. Die Detailanalysen dieser Arbeit erweitern die bestehende Literatur wie im Folgenden dargestellt:

- Beitrag #1 **„Quantile Score Estimation to Improve Value at Risk Forecasts“**:
Der Artikel bietet eine umfassende Anwendungsstudie nach formaler Einführung zweier quantilbasierter Schätzer für das GARCH-Modell. Die Studie zeigt die praktische Auswirkung und Relevanz eines expliziten Einbeziehens von „Quantile Scoring-Funktionen“ für gängige Risikomodelle.
- Beitrag #2 **„(Lifetime) Expected Credit Losses als Mechanismus der Verlustantizipation nach IFRS 9 – Operationalisierung der Schätzparameter in kritischer Betrachtung“**:
Die Analyse dieses Artikels präsentiert eine verallgemeinerte Zerlegung des LECL in die aus dem Risikomanagement bekannten Kernparameter LGD, EAD und PD. Es werden mögliche Berechnungsansätze für diese Parameter vorgestellt und im Hinblick auf die IFRS 9-Zielvorstellungen diskutiert.
- Beitrag #3 **„Konjunkturabhängige PD-Bestimmung nach IFRS 9 – Zum Informationsgehalt öffentlich verfügbarer Ratings“**:
Eine empirische Analyse von Ratingdaten führt in diesem Artikel zu Aussagen über deren Nützlichkeit im Hinblick auf eine Verwendung im Rahmen des IFRS 9. Darüber hinaus wird ein konjunkturbedingtes Migrationsmodell vorgestellt, welches zu konjunkturbedingten Ausfallstrukturkurven führt und somit einen essentiellen Baustein zur Bestimmung von LECLs liefert.
- Beitrag #4 **„Calibrating and Forecasting Point in Time Probability of Default Term Structures“**:
Im Artikel werden eigene Modellansätze für die konjunkturbedingte Modellierung von Ausfallstrukturkurven vorgestellt und im Rahmen einer praktisch relevanten Anwendungsstudie analysiert. Neben der Modellkalibrierung auf Daten wird insbesondere auch die Vorhersage konjunkturbedingter Ausfallstrukturkurven demonstriert.
- Beitrag #5 **„The Bargaining Game of the Audit – Regulatory Provisions for High-Quality Financial Statements and an Independent Audit“**:
Der Artikel entwickelt ein Modell für eine in der qualitativen Forschungsliteratur häufig diskutierte ACMN. Basierend darauf werden Wirksamkeitsanalysen von Regulierungsmaßnahmen der Jahresabschlussprüfung geführt und im Resultat wirksame Maßnahmen präsentiert.

4 Zusammenfassung der Ergebnisse

Die vorliegende Dissertation greift existierende Forschungsdesiderate zur Risikomessung mit dem VaR, zur Berechnung von LECLs innerhalb des IFRS 9 und zur modelltheoretischen Wirksamkeitsanalyse von Abschlussprüfungsregulierung auf. Die zuvor hergeleiteten Frage- und Problemstellungen aus diesen drei Schwerpunktbereichen werden zu den folgenden Ergebnissen geführt.

Um die Qualität der Risikomessung in Finanzinstituten zu verbessern, wurde die Implementierung einer zielgerichteten Schätzroutine für die VaR-Prognose als aussichtsreicher Ansatzpunkt identifiziert. Daraus ergab sich die Notwendigkeit die Theorie genauer zu beleuchten, um darüber entsprechende Schätzer definieren und die Effektivität dieser in einer Anwendungsstudie überprüfen zu können. Dies führte zu der Erkenntnis, dass die Schätzung mit den quantilbasierten Schätzern für eine VaR-Vorhersage grundsätzlich möglich ist. Wie erwartet, kann die Verwendung dieser Schätzer bei einer in der Praxis kaum anzutreffenden, korrekten Modellspezifikation keinen Vorteil gegenüber dem MLE aufweisen. Ist dies allerdings nicht uneingeschränkt gegeben, führt der stark auf das explizite Vorhersageziel ausgerichtete Quantile Score Estimator (QSE) tatsächlich zu einer präziseren VaR-Vorhersage als es das durch den MLE geschätzte Vorhersagemodell vermag; er benötigt dafür allerdings ein sehr großes Datensample im Rahmen der Schätzung. Der zweite betrachtete Schätzer, der Continuous Ranked Probability Score Estimator (CRPSE), benötigt kein derart umfassendes Datensample, kann allerdings auch die Vorhersagegüte nicht wirksam verbessern. Diese Erkenntnisse resultieren aus der Simulationsstudie, in der eine Missspezifikation künstlich – aber realitätsnah – erzeugt wurde. Darüber hinaus konnte dies auch durch die weiterführende Anwendung der Verfahren auf Finanzdaten bestätigt werden. Die praktische Einsatzfähigkeit der quantilbasierten Schätzer hängt somit insgesamt von dem konkreten Einsatzgebiet und der dort verfügbaren Informations- und Datenbasis ab. Die Aussichten auf eine Verbesserung der VaR-Vorhersage werden darüber hinaus insbesondere durch einen höheren Missspezifikationsgrad begünstigt, d.h. einfache Modelle profitieren wesentlich stärker von der veränderten Schätzroutine als komplexe Modelle, die per se bereits eine hohe Flexibilität für die Anpassung an die Datendynamik besitzen. Basierend auf den gewonnenen Erkenntnissen könnten weiterführende Forschungsprojekte auf die Entwicklung eines kombinierten Schätzers abzielen, der die Stärken des QSE und des CRPSE bündelt und gleichzeitig deren Schwächen kaschiert. Dies könnte beispielsweise über eine gewichtete Betrachtung mehrerer Quantile, im für das Risikomanagement besonders wichtigen unteren Bereich der Gewinn- und Verlustverteilung, erreicht werden. Zudem könnte die geführte Vorhersageanalyse durch weitere Elemente, wie etwa eine „Out-of-Sample-Untersuchung“, ergänzt werden.

Im zweiten Schwerpunktbereich konnte ein grundlegender Mangel an durch den Rechnungslegungsstandard IFRS 9 vorgegebene quantitative Umsetzungsmodelle für die Bestimmung von LECLs identifiziert werden. Um diese Problematik zu überwinden, wurden vor diesem Hintergrund zunächst bereits vorhandene Modelle aus Literatur und Praxis beleuchtet. Dies führte zu der Erkenntnis, dass diese, meist aus dem Risikomanagementumfeld stammenden Ansätze, den Anforderungen des IFRS 9 nicht ohne umfassende Anpassungen standhalten können. Als besonders kritisch stach dabei die geforderte Einbeziehung von Konjunkturaussichten hervor, weswegen dieser Aspekt als das zentrale Forschungsobjekt für den weiteren Verlauf der Un-

tersuchungen festgelegt wurde. Erwartungsgemäß wurden konjunkturelle Abhängigkeiten vor allem bei der PD beobachtet und gemessen, die das Risiko des Portfolios zudem auch intuitiv sehr stark charakterisiert und daher als der zentrale Kernparameter für die weiteren Untersuchungen gewählt wurde.

Die zeitliche Dynamik von Ausfallwahrscheinlichkeiten und die sich darin widerspiegelnde Konjunkturabhängigkeit wurden im Rahmen dieser Dissertation primär auf Basis von Standard & Poor's Ratingdaten und deren Ratingmigrationen analysiert. Dabei konnten insbesondere für konjunkturelle Frühindikatoren ein hoher Erklärungsgehalt, für die im Folgejahr beobachteten Ausfallwahrscheinlichkeiten, beobachtet werden. Das daraufhin vorgestellte lineare Zeitreihenmodell wurde im weiteren Verlauf zu einem Migrationsmodell erweitert, welches die Möglichkeit zu einer konjunkturabhängigen – nämlich auf bereits beobachteten Indikatoren basierenden – Prognose von Ausfallstrukturkurven bot. Allerdings bezieht sich dieses Modell lediglich auf die Ausfallwahrscheinlichkeiten des direkt anschließenden Jahres, womit eine ausreichende Flexibilität dieses Ansatzes, zur angemessenen Darstellung sämtlicher in der Praxis auftretender Ausfallstrukturkurven, angezweifelt werden kann. Im weiteren Verlauf dieser Arbeit wurde diese Schwäche durch die zwei im vierten Beitrag vorgestellten Migrationsmodelle überwunden, welche auf historische Ausfallstrukturkurven innerhalb verschiedener zeitpunktabhängiger Betrachtungsfenster insofern kalibriert werden können, dass sie deren Verläufe widerspiegeln. Basierend darauf wurde zudem ein Vorhersagemechanismus präsentiert, der Ausfallstrukturkurven mit Hilfe von Frühindikatoren prognostizieren kann. Die Vorhersagequalität der Modelle ist im Ergebnis vergleichbar und übertrifft dabei jeweils die „Through-the-Cycle-Methodik“. Dies ist entscheidend, da eine zeitpunktsensitive Vorhersage andernfalls nicht vorliegen würde. Die in dieser Dissertation vorgeschlagenen Modellansätze befinden sich grundsätzlich – vor dem Hintergrund eines sehr herausfordernden Vorhersageziels ganzer Ausfallstrukturkurven – in einem frühen Entwicklungsstadium. Mehrere Teilaspekte der Modellausgestaltungen, wie etwa der lineare Vorhersagemechanismus, könnten durch anknüpfende Forschung aufgerüstet oder durch geeignetere Detailmethoden ersetzt werden. Darüber hinaus könnten weiterführende Untersuchungen die Aussagen über die Befähigung zur Vorhersage, insbesondere auch innerhalb anderer Datenuniversen und im Rahmen einer „Out-of-Sample-Studie“, präzisieren.

Die Literaturanalyse zu der Beziehung zwischen Jahresabschlussprüfern und dem Management des Mandanten ergab den qualitativen Nachweis von stattfindenden Verhandlungen (ACMNs) der beiden Parteien über die Ausgestaltung des Prüfberichtes. Der dritte Schwerpunktbereich dieser Dissertation verarbeitete diese Erkenntnisse zu einem eigens entwickelten (spiel-)theoretischen Modell für solche Verhandlungen. Hierbei konnten zwei Wirkungsgrade regulierender Maßnahmen auf die Verhandlungssituation unterschieden werden. Zum einen wirken eine Reihe von Maßnahmen, wie etwa eine vorgeschriebene Prüferrotation oder eine Verschärfung des Strafmaßes bei Fehlverhalten, auf alle möglichen Lösungen der Verhandlung (Einfluss A). Andererseits können alternative Regulierungsmaßnahmen eine isolierte Wirkung auf die konkret resultierende Verhandlungslösung erzielen, welche unter den Annahmen des Nash-Verhandlungsspiels eindeutig bestimmt ist (Einfluss B). Die Analysen anhand des Modells zeigen, wie die Verhandlung durch eine hinreichende Einflussnahme im Bereich A zwar theoretisch verhindert werden kann, die Möglichkeit einer der Umsetzung, abseits von Ausnahmefällen, allerdings fraglich ist. Dagegen kann aus der modelltheoretischen Untersuchung abgeleitet werden, dass eine gezielte Regulierung im Einflussbereich B, bzw. eine Kombinati-

on aus A und B, den Vorteil des Managements, resultierend aus verschönigten Abschlüssen, vollständig eliminieren kann. Damit kann die Einflussnahme verstärkt auf die Entscheidungssituation in erster Instanz ausgerichtet werden, die genau zu dem Zeitpunkt stattfindet, an dem das Management den Jahresabschluss im Hinblick auf die Prüfung erstellt, bzw. als verantwortliches Organ im Unternehmen erstellen lässt. Eine derart wirkende Regulierungsstrategie ist beispielsweise durch die Festlegung eines gewissen Rotationsturnusses (A), in Kombination mit einer darauf abgestimmten Mindestanzahl zugesicherter Jahresabschlussprüfungen für den betreffenden Prüfer (B), gegeben. Der erste Teil dieser Strategie legt über den Einflussbereich A die Rahmenbedingungen für die ACMN fest, während der zweite Teil die Verhandlungsposition des Prüfers gerade soviel stärkt, dass dieser den Löwenanteil der zur Disposition stehenden Rente, resultierend aus den verschönigten Abschlüssen, selbst abschöpfen kann. Wird dieses Resultat durch das Management antizipiert, so schwinden die Anreize für Manipulationen und es findet damit eine Qualitätssicherung von Jahresabschlüssen statt. Anknüpfende Forschung könnte die zentrale Ausgestaltung des Modellrahmens und die damit eng verbundene Abbildung der Realität verfeinern. Zudem könnte die Betrachtung auf mehrere Prüfperioden ausgeweitet werden, in denen sich die Entscheidungen der Akteure – z.B. vor dem Hintergrund neuer Sachlagen – womöglich ändern und somit die Wirksamkeit der Regulierung beeinträchtigen werden könnte.

Im Rahmen der drei Themenschwerpunkte können die Ergebnisse dieser Dissertation für Empfehlungen im Hinblick auf den Stand sowie die Weiterentwicklung des regulatorischen Rahmens genutzt werden. Quantitative Umsetzungsspielräume auf dem Weg hin zu einer angemessenen Bestimmung der i.d.R. qualitativ beschriebenen Risikokennzahlen liegen häufig vor. Im Risikomanagement ist dies nur für Details der Fall, was Modellverbesserungen (relativ zu den vordefinierten Validierungsmethoden) durch Anwender aus der Praxis und der Wissenschaft nur innerhalb dieses gesteckten Rahmens möglich macht. Die am Vorhersageziel ausgerichtete Parameterschätzung in VaR-Modellen ist zwar grundsätzlich in der Lage eine verbesserte Vorhersage zu liefern, kann andererseits aber nur in speziellen Fällen angewendet werden. Somit kann die Nützlichkeit dieses Ansatzes, beispielsweise für eine Risikomessung im Tagesgeschäft von Banken, nur durch Einzelbewertungen, in Anbetracht des konkret vorliegenden Modells und des für die Schätzung verlässlich zur Verfügung stehenden Datenumfanges, festgestellt werden; sie ist also nicht allgemein gegeben. Damit erscheint es nur schwer möglich, diesen Ansatz zu einem regulatorischen Marktstandard zu entwickeln.

Die Erkenntnisse der rechnungslegungsbezogenen Risikoquantifizierung sowie der Regulierung der Jahresabschlussprüfung können folgendermaßen zusammengefasst und verbunden werden. Im Standard IFRS 9 offenbaren sich große Spielräume für die Ausgestaltung von quantitativen Umsetzungskonzepten, was in diesem Kontext auch durch eine aktive Bilanzpolitik instrumentalisiert werden könnte. Da es der Wirtschaftsprüfung obliegt, ein derartiges Szenario einzuschränken, ist die Stellung von Wirtschaftsprüfungsgesellschaften in doppelter Hinsicht äußerst kritisch zu sehen. Einerseits beteiligen sich insbesondere die großen („Top Four“) Prüfer – ggf. im Rahmen ihrer Beratungssparte – selbst bei der Entwicklung von quantitativen Marktstandards, andererseits geraten diese im Zuge von ACMNs möglicherweise derart unter Druck, dass sie den Auslegungen oder Wünschen des eigentlich zu prüfenden Mandanten eher zustimmen. Basierend auf den Eindrücken der Untersuchungen der vorliegenden Arbeit, erscheinen die Freiheitsgrade bei der Berechnung von Wertminderungsbeträgen über LECLs derartig hoch, dass darüber prinzipiell eine Vielzahl von bilanzpolitischen Zielwerten bestimmt

werden könnte, die den Anforderungen des Standards dabei aber noch immer genügen. Davon kann die Forderung nach einer Entwicklung von konkreten und bindenden quantitativen Standards, direkt durch die Regulierungsbehörden, abgeleitet werden. Die angemessene Umsetzung solcher verfeinerter Vorgaben sollte für das Management des anwendenden Unternehmens dann entweder den Nutzen von geheimen Absprachen im Rahmen der ACMN übertreffen oder das Zuwiderhandeln müsste durch eine völlig unabhängige Prüfung wirksam an die Öffentlichkeit gelangen. Für das Erreichen eines dieser beiden Ziele ist es unerlässlich, die Beziehung zwischen Prüfern und Geprüften einzubeziehen und dabei insbesondere deren Anreize für inbegriffene Entscheidungen zu beachten. Damit kann die Signal- und Lenkungsfunction der Berichterstattung von Unternehmen für Investitionsentscheidungen auf Kapitalmärkten stabilisiert werden, was in der Folge die Märkte per se stärkt und damit insgesamt auch die volkswirtschaftliche Wohlfahrt fördert.

Literatur

- Bartmann, P., Buhl, H. U. und Hertel, M. (2009), Ursachen und Auswirkungen der Subprime-Krise, *Informatik Spektrum* **32**(2), S. 127–145.
- BCBS (2019), Minimum Capital Requirements for Market Risk, Basel Committee on Banking Supervision.
URL: <https://www.bis.org/bcbs/publ/d457.pdf>
- Bennett, G. B., Hatfield, R. C. und Stefaniak, C. (2015), The Effect of Deadline Pressure on Pre-Negotiation Positions: A Comparison of Auditors and Client Management, *Contemporary Accounting Research* **32**(4), S. 1507–1528.
- Bluhm, C. (2016), Introduction to Credit Risk Modeling, 2. Aufl., CRC Press, Boca Raton.
URL: <https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=4743029>
- Bollerslev, T. P. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **31**(3), S. 307–327.
- Box, G. E. P. (1976), Science and Statistics, *Journal of the American Statistical Association* **71**(356), S. 791–799.
- Campbell, S. D. (2005), A Review of Backtesting and Backtesting Procedures, *Finance and Economics Discussion Series* **21**.
- DeAngelo, L. E. (1981a), Auditor Independence, 'Low Balling', and Disclosure Regulation, *Journal of Accounting & Economics* **3**(2), S. 113–127.
- DeAngelo, L. E. (1981b), Auditor Size and Audit Quality, *Journal of Accounting & Economics* **3**(3), S. 183–199.
- Ewert, R. (2003), Prüfung, Beratung und externe Rotation: Ökonomische Forschungsergebnisse zur aktuellen Regulierungsdebatte im Bereich der Wirtschaftsprüfung, *Der Konzern* **1**, S. 528–539.
- Fearnley, S., Beattie, V. A. und Brandt, R. (2005), Auditor Independence and Audit Risk: A Reconceptualization, *Journal of International Accounting Research* **4**(1), S. 39–71.
- Friedman, J. (2011), What caused the Financial Crisis, University of Pennsylvania Press, Philadelphia.
URL: <http://www.jstor.org/stable/10.2307/j.ctt3fhz14>
- Gibbins, M., Salterio, S. und Webb, A. (2001), Evidence About Auditor-Client Management Negotiation Concerning Client's Financial Reporting, *Journal of Accounting Research* **39**(3), S. 535–563.
- Gneiting, T. (2011), Making and Evaluating Point Forecasts, *Journal of the American Statistical Association* **106**(494), S. 746–762.
- Gneiting, T. und Raftery, A. E. (2007), Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association* **102**(477), S. 359–378.

- Greene, W. H. (2012), *Econometric Analysis*, Pearson Series in Economics, 7. Aufl., Pearson, Boston [u.a.].
- Grosskord, M., Mach, P. und Reher, D. G. (2013), *Lifetime Expected Loss: Anwendungsfelder und Berechnungsmethoden*, Deloitte.
URL: <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/financial-services>
- Grünberger, D. (2013), *Kreditrisiko im IFRS-Abschluss: Handbuch für Bilanzersteller, Prüfer und Analysten*, Schäffer-Poeschel, Stuttgart.
- Hartmann-Wendels, T., Pfingsten, A. und Weber, M. (2010), *Bankbetriebslehre*, 5. Aufl., Springer, Berlin [u.a.].
- Healy, P. M. und Wahlen, J. M. (1999), A Review of the Earnings Management Literature and Its Implications for Standard Setting, *Accounting Horizons* **13**(4), S. 365–383.
- Henselmann, K., Ditter, D. und Holstein, T. (2014), Too little, too late? Ein empirisch gestützter Beitrag zur Bilanzierung der Kreditrisikovorsorge, *KoR: Internationale und Kapitalmarktorientierte Rechnungslegung IFRS* **14**(7/8), S. 355–362.
- Huber, P. J. (1964), Robust Estimation of a Location Parameter, *The Annals of Mathematical Statistics* **35**(1), S. 73–101.
- IASB (2014), *IFRS 9: Financial Instruments*, International Financial Accounting Standard Board.
URL: <http://www.ifrs.org>
- Jorion, P. (2006), *Value at Risk: The New Benchmark for Managing Financial Risk*, 3. Aufl., McGraw-Hill, New York.
- Krommes, W. (2015), *Handbuch Jahresabschlussprüfung: Ziele – Technik – Nachweise - Wegweiser zum sicheren Prüfungsurteil*, 4. Aufl., Springer Gabler, Wiesbaden.
- McCracken, S., Salterio, S. E. und Gibbins, M. (2008), Auditor-Client Management Relationships and Roles in Negotiating Financial Reporting, *Accounting, Organizations and Society* **33**(4-5), S. 362–383.
- McDonald, O. (2015), *Lehman's Valuation of Its Assets*, in: McDonald, O. (Hrsg.), *Lehman Brothers*, Manchester University Press, Manchester.
- McNeil, A. J., Frey, R. und Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton Series in Finance, Princeton University Press, Princeton, NJ.
- Milgrom, P. R. und Roberts, J. (1992), *Economics, Organization and Management*, Prentice Hall, Englewood Cliffs, NJ.
- Nash, J. F. (1950), The Bargaining Problem, *Econometrica* **18**(2), S. 155–162.
- Neisen, M. und Röth, S. (2018), *Basel IV: The Next Generation of Risk Weighted Assets*, 2. Aufl., Wiley-VCH, Weinheim.

- Reichling, P. (2007), *Praxishandbuch Risikomanagement und Rating*, Springer Fachmedien, Wiesbaden.
URL: <http://gbv.ebib.com/patron/FullRecord.aspx?p=750884>
- Rolfes, B. (2008), *Gesamtbanksteuerung: Risiken ertragsorientiert steuern*, Handelsblatt-Bücher, 2. Aufl., Schäffer-Poeschel, Stuttgart.
- Schildbach, T., Stobbe, T. und Brösel, G. (2013), *Der handelsrechtliche Jahresabschluss*, 10. Aufl., Verlag Wissenschaft & Praxis, Sternenfels.
- Schröder, T. (2015), Expected Loss im Kreditgeschäft nach IFRS 9 und Basel III: Kompatibilität von Rechnungslegung und Regulatorik?, *KoR: Internationale und Kapitalmarktorientierte Rechnungslegung IFRS* **15**(5), S. 245–251.
- Van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.
- Wagenhofer, A. und Ewert, R. (2015), *Externe Unternehmensrechnung*, 3. Aufl., Gabler, Berlin.
- Weiss, A. A. (1996), Estimating Time Series Models Using the Relevant Cost Function, *Journal of Applied Econometrics* **11**(5), S. 539–560.
- Zhang, H. (2017), Market Risk Modeling Framework Under Basel, in: Tian, W. (Hrsg.), *Commercial Banking Risk Management*, Palgrave Macmillan US, New York, S. 35–52.
- Zülch, H. und Hendler, M. (2019), *International Financial Reporting Standards (IFRS) 2019: Deutsch-Englische Textausgabe der von der EU gebilligten Standards. English & German edition of the official standards approved by the EU*, International Financial Reporting Standards (IFRS), 13. Aufl., Wiley-VCH, Weinheim.

Tobias Filusch*

Quantile Score Estimation to Improve Value at Risk Forecasts

Abstract

Risk assessment is an important part of the internal management of financial institutions. To improve market risk measurement, this paper introduces and examines the quantile score estimator (QSE) and the continuous ranked probability score estimator as alternatives to the maximum likelihood estimator in GARCH models. These estimators focus on the value at risk (VaR) as the forecast goal and can improve the forecast accuracy of the model in use. The broad application study demonstrates how a misspecified model can still lead to reasonable results when the QSE is used to estimate model parameters. It produces better VaR forecasts measured by scoring functions and VaR backtests without reconsidering the general model choice.

Contents

1	Introduction	17
2	Improving Value at Risk-Backtesting with New Estimation Routines	19
2.1	Measuring Risk and Evaluating Outcomes	19
2.2	A New Value at Risk Estimation Concept	21
2.3	Estimation under Model Misspecification	24
3	Quantile Score Estimation and Value at Risk Forecasting Application Study	25
3.1	Study Design and Financial Data Horizon	25
3.2	GARCH Parameter Estimation	26
3.3	Value at Risk Forecast Study	28
3.4	Application Findings	31
4	Conclusions	32

*Dipl.-Math. oec. Tobias Filusch, Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, ifG Marburg, Am Plan 2, D-35032 Marburg.

1 Introduction

The Basel framework – an international regulatory framework – requires banks to apply risk measurement to determine minimum capital requirements as part of a broadly defined capital market regulation (BCBS, 2019). At the cornerstone of the Basel framework is the popular quantitative risk measure value at risk (VaR), which was recently supplemented by expected shortfall (Jorion, 2006; Acerbi and Tasche, 2002; Tian, 2017). Both measures rely on the anticipated distribution of profits and losses (P&L), assuming an impending time horizon contingent upon current and relevant information (McNeil et al., 2015). VaR is represented by a particular quantile of that distribution, typically determined by a parametric family. However, an optimality criterion is still required to determine the best parameters for the forecast model in this case. Due to its desirable asymptotic properties such as consistency, normality, and efficiency, the maximum likelihood estimator (MLE) is widely considered the best or sometimes even only option; it is the status quo in financial time series estimation (Tsay, 2010) and, thus, forms a strong foundation for risk measurement. Due to effects of misspecification, different estimation criteria can produce different estimates given the same input data (Van der Vaart, 1998). Hence, selecting a proper estimation criterion is critical for eliciting optimal forecasts. Therefore, instead of comparing different VaR models like it is frequently done in the academic literature (Bao et al., 2006; Kuester et al., 2006; Bernardi and Catania, 2016), this paper considers only one basic parametric model throughout and instead focuses on the effect triggered by changing the estimation routine with regard to the forecast goal.

Practical applications require VaR models to be validated in an a posteriori procedure, which includes backtesting to evaluate the predictive quality of the model. In this process the original VaR forecasts are evaluated against actual P&L according to a presumed criterion (Auer, 2018). Backtesting criteria naturally use a hit variable, which counts the number of times portfolio realizations exceed the predicted VaR (Kupiec, 1995; Christoffersen, 1998; Christoffersen and Pelletier, 2004). This variable is also important for quantile scoring functions, a recent addition to traditional model validation techniques (Gneiting, 2011a). Despite its relevance, practitioners usually ignore the fact that these two evaluation concepts may lead to useful estimation criteria for optimal VaR forecasts under the relevant evaluation scale (Weiss, 1996). The design of our quantile-based estimators relies on three theoretical building blocks: first is the general concept of scoring rules and scoring functions, a popular evaluation concept in probabilistic forecasting (Gneiting and Katzfuss, 2014; Krüger, 2013). Second are the properties necessary to assess the feasibility of general scoring criteria and how to use them in parameter estimation. Any strictly consistent scoring rule can exploit the M-estimation framework (Huber, 1964) to create feasible estimators, which is the third and final building block. In this way, this paper presents two estimators specifically for the VaR.

Previous research has already treated similar estimators in general terms. Gneiting and Raftery (2007) outlined the concept of optimum score estimation with general scoring rules, referencing work by Pfanzagl (1969) and Birgé and Massart (1993) and their closely related concept of minimum contrast estimation. The continuous ranked probability score estimator (CRPSE) has previously been used for ensemble model output statistic coefficients in continuous weather variable prediction (Gneiting et al., 2005). With a greater focus on VaR, quantile

regression uses quantile scoring to directly fit a model for conditional quantiles (Koenker and Bassett, 1978). Closely related, Engle and Manganelli (2004) introduced the conditional autoregressive value at risk (CAViaR) – a direct autoregressive specification of the quantile evolution over time (see also Jeon and Taylor (2013) and Guler et al. (2017)). Xiao and Koenker (2009) has integrated the CAViaR model into a comparative study. Thereby, they propose to use quantile regression to estimate GARCH models, which they call QGARCH. The QGARCH is similar to our QSE-GARCH (quantile score estimator-GARCH) model, but with a different focus. This paper extends their estimation analysis by focusing on the actual improvement of VaR prediction and a comparison to the MLE as a competing method in the same VaR model. In addition, we also add the CRPSE to the analysis.

The core of this paper is the application study, which highlights the impact of the estimators as well as their weaknesses. Since conditional heteroscedasticity is a key feature of financial time series (White, 1980), we make predictions about the conditional distribution using the generalized autoregressive conditional heteroscedasticity (GARCH) model (Bollerslev, 1986). Hafner and Preminger (2010) found that GARCH models are often preferred by information criteria-driven model selection procedures even if the true data generating process is the stochastic volatility (SV) model (Heston, 1993), i.e. these criteria select a false rather than the correct model. This discrepancy is one frequent cause of misspecification that justifies our simulation study: while the simulated samples are generated by the SV model, we use the GARCH model as the forecast model to exactly replicate such a misspecification. Applying the GARCH model to over 3,000 stock price series complements the simulation study to provide even more practical importance.

For correctly specified models, all three estimators we apply are consistent. However, our study shows that under misspecification, estimates differ. Thus, the mean quantile score and the backtesting criteria allow to identify the preferred estimator with respect to the best VaR forecast. The essential observation of this study is that the QSE improves VaR forecasts compared to the MLE, which both the SV simulation and the stock data analysis show. This finding revives the use of simple models, which can be modified by replacing the estimator in such a way. If a forecast model does not perform well in practice, this could offer an inexpensive opportunity to remain the admission of the risk model by the authorities.

The paper is divided into two main parts. Section 2 provides the theoretical and practical background for risk measurement and model validation as well as presents scoring functions, scoring rules, their connection and necessary properties (2.1). Section 2.2 goes on to introduce the VaR-oriented estimators, before 2.3 provides a discussion on the implications of impending model misspecification in estimation. The second part, Section 3, begins with the description of the comprehensive application study (3.1). Furthermore, the proposed estimators are analyzed (3.2), after which the expected improvements of the resulting VaR forecasts are demonstrated (3.3). Section 3.4 summarizes the application findings and, finally, Section 4 concludes the paper.

2 Improving Value at Risk-Backtesting with New Estimation Routines

2.1 Measuring Risk and Evaluating Outcomes

Based on the uncertainty of future events, the risk of any financial instrument may be characterized as an upside or downside risk. Correspondingly this financial risk can be measured based on forecasted P&L distributions, which reflect possible future scenarios. Using statistical functionals allows to aggregate this comprehensive information into a single number. Concerning the downside risk of financial assets, this may be carried out by metrics such as expected shortfall or VaR (see Acerbi and Tasche (2002) and Jorion (2006)). The latter is especially relevant as the Basel framework has heavily relied upon VaR to calculate the required equity capital of financial institutions since the 1990s. VaR, as with other P&L-based measures, provides an aggregated forecast of market risk based on present information. Therefore, real market data is required to validate the VaR model. Statistical backtests use actual portfolio returns to evaluate the predictive quality of the model in use. The concept is reliable only if VaR models pass these tests. Regulatory provisions (i.e. Basel standards) also require backtesting; authorities use the outcome to determine whether the applied model is acceptable. A poor performance on coverage tests triggers extended monitoring by authorities and requires the model to be further developed – a potentially costly outlook. The procedure can also be scrutinized by using suitable scoring functions. These are only available for the class of so-called elicitable risk measures, to which the VaR belongs.

Determining the Value at Risk

VaR quantifies the loss in a certain future time frame, that will not be exceeded with a certain probability (confidence). A common choice is a confidence level of 95% over a period of ten business days (BCBS, 2009). The corresponding functional value derived from the P&L distribution forecasted for that 10-day horizon is the 5% quantile. Let F_t be the distribution of the only risk factor y_t in the portfolio conditional on the information set \mathcal{Y}_{t-1} at $t-1$. The VaR with level p (confidence $1-p$) for the horizon $[t-1, t]$ is then defined by

$$VaR^p(F_t) := -q_p(F_t), \quad (1)$$

where $q_p(F_t)$ is the p -quantile of distribution F_t .¹ To find a model for the risk factor distribution, the characteristics of financial time series must be considered – the so-called stylized facts (Tsay, 2010). Of the models capable of capturing those effects, the family of generalized auto-regressive conditional heteroscedastic (GARCH) processes is frequently used (Bollerslev, 1986). Let

$$y_t = \sigma_t \epsilon_t \quad (2)$$

be the risk factor return in period $[t-1, t]$ where the distribution of the innovation process is independent and identically distributed ($\epsilon_t \sim \mathcal{N}(0, 1)$ i.i.d.). The GARCH(1,1) focuses on an auto-regressive structure of conditional variance with three parameters: the time constant level of the conditional variance (ω), the influence of squaring a lagged observation (α) and the dependency from the lagged conditional variance (β). The GARCH(1,1) is hence given

¹Sometimes just VaR_t^p , if the conditional distribution is obvious.

by

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2. \quad (3)$$

The stochastic volatility (SV) model captures very similar features to the GARCH model, yet the representation is different as it is based on a randomly driven conditional variance (Taylor, 1986) given by

$$\begin{aligned} \log(\sigma_t^2) &= \gamma + \phi \log(\sigma_{t-1}^2) + \vartheta \xi_t, \text{ with} \\ \xi_t &\sim i.i.d. \mathcal{N}(0, 1). \end{aligned} \quad (4)$$

Both models induce a conditional risk factor distribution of $y_t \sim \mathcal{N}(0, \sigma_t^2)$, where the choice of another, perhaps more fat-tailed, distribution of ϵ_t can easily replace the normal distribution used here. The ability to reflect the stylized facts with only few parameters makes either of the two model classes a suitable choice for risk factor dynamics and, hence, reliable methods for VaR applications.

Backtesting the Value at Risk

For backtesting purposes, a sufficiently large sample of VaR forecasts VaR_t^p as well as corresponding portfolio realizations y_t must be available. Comparing a forecasts and realizations produces a hit sequence, which indicates whether the portfolio loss exceeded the VaR by turning to 1. It is given by

$$h_t(p) := \mathbb{1}_{\{y_t \leq -VaR_t^p\}}. \quad (5)$$

For a series $h_1(p), \dots, h_T(p)$ based on a certain VaR model, the number of hits should approach $T \cdot p$ simply based on the definition of VaR. If this is not the case, the model is unacceptable and should be rejected for failing to meet the unconditional coverage test (UCT), the first class of backtests for VaR. The most widely recognized UCT is the Kupiec test (KupT) (Kupiec, 1995).² The second crucial requirement for the series of hits $h_1(p), \dots, h_T(p)$ is the independence of hit occurrences. A hit should not affect the probability of further hits. The conditional coverage test (CCT) combines unconditional coverage and independence in a single test. A widely used test framework is the Christoffersen test (ChrT) (Christoffersen et al., 2001). Backtests are capable of rejecting poor VaR models. However, if two distinct models are both either rejected or not rejected, these tests cannot determine which of the models is better. This can significantly curtail their evaluative power. Model validation procedures needs additional information to identify the most accurate model for portfolio risk measurement.

Elicitable Risk Measures

Scoring functions play an important role in overcoming the limits of traditional backtests, especially when it comes to model comparison. Scoring functions map the pair consisting of a VaR forecast and the (a posteriori) portfolio observation to a real number score. The lowest (or highest) score will signify the best model, if a strictly consistent scoring function (SCSF) is used. General risk measures may not possess a SCSF, those that does are termed elicitable risk measures. A functional is called elicitable if and only if there exists at least one SCSF

²The probability

$$P(y_t \leq -VaR_t^p) = E(h_t(p)) \quad (6)$$

being equal to p is tested.

for it.³ Elicitable risk measures open up multiple possibilities for extending backtesting routines. VaR is an elicitable risk measure, so a corresponding SCSF can rank multiple series of VaR forecasts based on different forecast models that correspond to a single series of realized portfolio returns: the lower the mean score, the better the forecast model.⁴ Scoring functions present a particularly strong option for expanding the available toolbox for model validation. They define benchmarks that a VaR model must meet in order to be accepted and, in the case of multiple models, can use historical observations to identify the best one.

2.2 A New Value at Risk Estimation Concept

As explained in the previous section, the best VaR model is characterized by the lowest mean score and a certain consistency in passing the relevant backtests – both criteria are based on historical portfolio observations. Parameter estimates for parametric VaR forecast models like the GARCH model typically rely on the MLE, which considers the maximization of the (log-) likelihood to determine the optimal parameter (Francq and Zakoian, 2014) that describes the data best. A logical alternative for the likelihood concept is to use a scoring function-based approach instead, as this scoring is much more relevant for the VaR forecast, for which a low mean score is sought. Implicitly, we expect backtest results to be improved as well, as the hit variable (5) is part of both the quantile scoring functions and the coverage test statistics. Before we can introduce the estimators, we formally define and distinguish between the notions of a scoring function and a scoring rule, as well as present further properties necessary for an estimation purpose based on such more general criterion functions.

Quantile Scoring Rules and Scoring Functions

In general, a statistical *functional* \mathcal{T} is a map from a class of distributions \mathcal{F} into the set of real numbers \mathbb{R} . This makes every P&L-based risk measure a functional. A *scoring function* is a measurable mapping $S : \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$. With random variable $Y \sim F$, scoring function S is said to be *consistent for the functional \mathcal{T} relative to the class \mathcal{F}* , if

$$\forall x \in \mathbb{R}, F \in \mathcal{F} : E_F(S(\mathcal{T}(F), Y)) \leq E_F(S(x, Y)).^5 \quad (7)$$

Thus, the true functional value $\mathcal{T}(F)$ minimizes the expected score under F , the true distribution of Y . If the equality in (7) implies that $x = \mathcal{T}(F)$, then S is a *strictly consistent scoring function* (SCSF) for functional \mathcal{T} , relative to the class of distributions \mathcal{F} . Recall, a functional \mathcal{T} is considered *elicitable relative to a class of distributions \mathcal{F}* , if it possesses an SCSF relative to that class. Similar to a scoring function, the forecast of a distribution F can be evaluated by a measurable mapping $\mathbf{S} : \mathcal{F} \times \mathbb{R} \rightarrow \mathbb{R}$, which is called a *scoring rule*. The scoring rule is *proper relative to a class of distributions \mathcal{F}* , if for any $Y \sim F \in \mathcal{F}$,

$$E_F(\mathbf{S}(F, Y)) \leq E_F(\mathbf{S}(G, Y)) \quad \text{for all } G \in \mathcal{F} \quad (8)$$

³The expression elicitable originates from the social and behavioral sciences, where the problem of truthfully eliciting an expert's assessment of a distribution property – like a functional value – is treated (Lambert et al., 2008). Incentives to achieve that goal can only be implemented if a certain scoring function is able to distinguish between true and false statements.

⁴VaR is an elicitable functional relative to a quite general class of distributions. Other widely used measures, like expected shortfall, are not elicitable (Gneiting, 2011a). Ziegel (2016) showed that the so-called spectral risk measures are, in general, not elicitable. Fissler et al. (2016) showed that the combination of expected shortfall and VaR is elicitable.

⁵It is assumed that the relevant expected values exist and are finite.

holds. It is considered *strictly proper* relative to that class if equality in (8) holds if and only if $G = F$. Both scoring terms can be nested by a functional \mathcal{T} in such a way that a consistent scoring function S induces a proper scoring rule by

$$\mathbf{S}(\cdot, \cdot) = S(\mathcal{T}(\cdot), \cdot).^6 \quad (9)$$

Note that the bold letter denotes the scoring rule. This distinguishes it from the scoring function and allows the connection in composition (9) to be maintained. The result (9) shows that functional scoring based on given scoring functions can be used to evaluate the forecasted distribution with a proper scoring rule as well. If the distribution with the lowest score \mathbf{S} also possesses the functional value with the lowest functional score, and if the scoring function is an SCSF, then a functional-optimal distribution has been identified. Based on these considerations, we can determine a feasible class of distributions where first, the VaR is elicitable relative to, and second, the corresponding SCSF leads to a strictly proper scoring rule through (9). These two steps are necessary for any estimation.

Value at Risk Forecast Evaluation and Quantile Score Estimation

Let VaR_t^p be a VaR forecast and let y_t be the portfolio return observation for the same interval as this forecast. The *quantile score* (QS)

$$\begin{aligned} QS_p(VaR_t^p, y_t) &:= (\mathbb{1}_{\{y_t \leq -VaR_t^p\}} - p) (-VaR_t^p - y_t) \\ &= (h_t(p) - p) (-VaR_t^p - y_t) \end{aligned} \quad (10)$$

is an SCSF for VaR at nominal level p (Gneiting, 2011a).⁷ The following propositions provide the building blocks (a) and (b) pointed out in the previous section and delineate the possible application scope of the QS in parameter estimation.

Proposition 1. (*Class of centered normal distribution*)

Let the class of centered normal distributions be given by

$$\mathcal{F}^{cnorm} := \{\mathcal{N}(0, \sigma^2) \mid \sigma \in \mathbb{R}^+\}. \quad (11)$$

Then

(a) Functional $VaR^p(\cdot)$ is elicitable relative to class \mathcal{F}^{cnorm} with QS_p (see (10)) being an SCSF.

(b) Embedding (10) in (9) leads to a strictly proper scoring rule \mathbf{QS}_p relative to class \mathcal{F}^{cnorm} .

Proposition 2. (*Class of normal distributions*)

Consider the class of normal distributions

$$\mathcal{F}^{norm} := \{\mathcal{N}(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\} \quad (12)$$

Then

⁶See Theorem 2.4. in Gneiting (2011a)

⁷The QS is one specific member of the family of generalized piecewise linear scoring functions, which are given based on a non-decreasing function g as

$$QS_p(v, y) = (\mathbb{1}_{\{y \leq -v\}} - p) (g(-v) - g(y)), \quad v, y \in \mathbb{R}.$$

All members are strictly consistent for the VaR at nominal level p . See Gneiting (2011b) and the references therein for the emergence and evolution of quantile scoring functions. Notable are Koenker and Bassett (1978) for the quantile regression, as well as Thomson (1979) and Saerens (2000). QS (10) evolves with the identity function $g(x) := x$.

- (a) *Functional $VaR^p(\cdot)$ is elicitable relative to class \mathcal{F}^{norm} with QS_p (see (10)) being an SCSF.*
- (b) *The combination of QS_p and VaR^p does not produce a strictly proper scoring rule QS_p for this class \mathcal{F}^{norm} by applying (9).*

Proofs of both propositions can be found in Appendix 4. Proposition 1 has a strong positive impact on the applicability of the scoring rule

$$QS_p(F_t, y_t) := QS_p(VaR^p(F_t), y_t), \quad F_t \in \mathcal{F} \text{ and } y_t \in \mathbb{R}, \quad (13)$$

as it verifies that (13) is a strictly proper scoring rule relative to the class of centered normal distributions. Then again, Proposition 2 limits the utility of this concept, as the same implication based on (13) does not hold for the more general class of normal distributions. We restrict our study on class \mathcal{F}^{cnorm} , but it is worth to mention that there are further classes possible, in which a strictly proper scoring rule is induced by an elicitable functional and one of its SCSFs.

Based on scoring rule (13), VaR forecast models can be evaluated and compared in the same manner as the VaR forecast itself through scoring function (10). Given a series of portfolio return observations y_1, \dots, y_T and two competing VaR models $i = 1, 2$ resulting from two different conditional distribution forecast series F_t^i that implicitly represent forecast series $VaR^p(F_t^1)$ and $VaR^p(F_t^2)$, the lower mean score

$$\frac{1}{T} \sum_{t=1}^T QS_p(VaR^p(F_t^i), y_t), \quad i = 1, 2 \quad (14)$$

indicates the better VaR model at a fixed but arbitrary VaR level p . Directly minimizing this criterion in an estimation routine within a parametric family $\mathcal{F}_\Theta := \{F^\theta \mid \theta \in \Theta\}$ should produce a VaR-optimized conditional distribution series F_t^θ given the T observations.

Any strictly proper scoring rule \mathcal{S} defines an M-estimator

$$\hat{\theta}_T(\mathcal{S}) := \arg \min_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \mathcal{S}(F_t^\theta, y_t).^8 \quad (15)$$

The scoring rule QS_p is the first candidate to be used in (15), which leads to the *quantile score estimator* (QSE) denoted by $\hat{\theta}_T(QS_p)$.

The Continuous Ranked Probability Score

Although the QSE is the first feasible estimator that directly corresponds to VaR forecasting, Proposition 2 encourages the development of a framework that is also applicable to a more general class like \mathcal{F}^{norm} as well. The VaR forecast remains the focal point nevertheless, which requires the integration of quantile scoring into the scoring concept.

⁸This approach is motivated by the optimum score estimator, proposed by Gneiting and Raftery (2007). For a general introduction to M-estimators, see Huber (1964). For discussions of their properties and use, see Staudte and Sheather (1990) or Van der Vaart (1998).

The *continuous ranked probability score* (CRPS) is a VaR-oriented scoring rule that is indeed strictly proper relative to the class of distributions \mathcal{F}^{norm} (Gneiting and Raftery, 2007). The CRPS is even strictly proper relative to the general class of Borel probability measures that have a finite first moment (Gneiting and Raftery, 2007) and, consequently, a promising candidate. For $F_t^\theta \in \mathcal{F}_\Theta$ and $y_t \in \mathbb{R}$ for all t , it is given by

$$\begin{aligned} \mathbf{CRPS}(F_t^\theta, y_t) &:= \int_0^1 2 (\mathbb{1}_{\{y_t < -\text{VaR}^p(F_t^\theta)\}} - p) (\text{VaR}^p(F_t^\theta) - y_t) dp \\ &= \int_0^1 2 (h_t(p) - p) (\text{VaR}^p(F_t^\theta) - y_t) dp, \end{aligned} \quad (16)$$

which clearly reveals its close relationship to QS. The **CRPS** is the integral over another SCSF of the VaR functional. With the generalized piecewise linear function $g'(x) := 2x$, the SCSF in this case given by

$$\begin{aligned} QS'_p(v, y) &= (\mathbb{1}_{\{y \leq -v\}} - p) (g'(-v) - g'(y)) \\ &= 2 (\mathbb{1}_{\{y < -v\}} - p) (-v - y), \quad v, y \in \mathbb{R}. \end{aligned}$$

Conceptually, the (16) is still a straightforward generalization of \mathbf{QS}_p that includes all nominal quantile levels $p \in [0, 1]$. While leading to a broader application horizon, this characterization sacrifices the focus on a specific VaR level. Denoted by $\hat{\theta}_T(\mathbf{CRPS})$ the *continuous ranked probability score estimator* (CRPSE) is derived from combining (16) and (15). This completes the section on explicitly quantile-based estimators, leaving only some final thoughts on the maximum likelihood estimation as the main competitor prior to presenting the application study.

2.3 Estimation under Model Misspecification

Following Proposition 1, we can set-up an M-estimator to estimate within the class of centered normal distributions based on the strictly proper scoring rule (13), that implicitly uses the quantile scoring function. The estimation of VaR forecast model parameters is usually done with the *maximum likelihood estimator* (MLE). The connection to scoring rules and the fact that the MLE is a special case M-estimator as well can be easily illustrated by combining the negative log-likelihood scoring rule

$$\mathbf{nlogL}(F_t, y_t) := -\log f_{F_t}(y_t) \quad F_t^\theta \in \mathcal{F}_\Theta, \quad y_t \in \mathbb{R} \quad (17)$$

with the M-estimation representation (15) (White, 1982).⁹

A major consequence of model misspecification is that the asymptotic estimate $\theta^*(\mathbf{S}) \in \Theta$ of any M-estimator depends on which scoring rule is used (see White (1982) or Van der Vaart (1998)). In fact, the efficiency of the MLE is useless as different estimators result in different (asymptotic) estimates and, thus, in distinct VaR models and VaR forecasts. It is yet to show, which forecast is the best. In a nutshell, the quantile score makes VaR forecasts and VaR

⁹Note that the negative score representation of the log-likelihood with $f_{F_t} : \mathbb{R} \rightarrow [0, \infty)$ being the density of distribution F_t is used in order to uniformly refer to minimization when it comes to all included M-estimation. The existence of a density for the considered distributions is presumed.

forecast models directly comparable and backtests are able to reject wrong models. This establishes two criteria for the best VaR model to meet: frequently passing backtests and having a low mean quantile score. In order to fit financial data best, information criteria are commonly used to select time series models. In many cases, however, they skew the choice in favor of the GARCH model, even if the data contains stochastic volatility (Hafner and Preminger, 2010). Hence, using a GARCH forecast model in SV-originated samples is a reasonable and realistic scenario of model misspecification; it substantiates why our simulation-based study uses this set-up to intentionally produce misspecification.

3 Quantile Score Estimation and Value at Risk Forecasting Application Study

3.1 Study Design and Financial Data Horizon

This application study demonstrates the impact the QSE and the CRPSE has on VaR forecasts by underlining several advantages and identifying potential disadvantages as compared to the MLE – the status quo in estimating VaR models. Two simulation studies and the application to a set of stock returns form the group of potential and observed risk factor dynamics included. Each risk factor constitutes a single portfolio, which results in a straightforward and unbiased relationship between the model for the conditional risk factor distribution and the VaR forecast of that portfolio. For each data source of this study, the data can be analyzed in three ways. The first evaluation considers the estimated values and its (finite-sample) distribution. The second and third evaluations make use of the mean score and backtesting criteria as an in-sample forecast and evaluation assessment.

The correctly specified GARCH setting is the starting point, which makes the estimation concepts' performance more tangible and comparable with the chosen parameter being known for the generated samples. Note that using the VaR scoring or VaR backtesting criterion is somewhat inefficient in this case since every included estimator – and all M-estimators in general – estimates the true model parameter, respectively. Thus, there is no comparison possible. Based on its efficiency, the MLE is expected to perform well and serves as a benchmark; we will examine how close the QSE and the CRPSE gets to it. Step two presents a second simulation similar in design to the first one, but uses SV samples instead of GARCH samples. Following this misspecification, we expect different estimates and, ultimately, different forecasts. Finally, a large set of daily log-return data obtained from the Thomson Reuters Eikon database is used. Data from four stock exchanges was obtained for the period April 1, 1992 to March 31, 2017. These series are common risk factors that require sophisticated VaR modeling as they may be in a bank's portfolio. While each instrument of our data set is still actively traded at the end date, some may have been first listed later than the start date of this study. Table 1 outlines the size and basic characteristics of the stock data. For each instrument, we set a minimum number of observations needed to $T_{min} = 3,000$, which ensures each risk factor series provides enough information to get reasonable estimates and make appropriate forecasts. Instruments that fail to meet this criterion were excluded. In this case, the correctness of the GARCH(1,1) specification is unclear. It may fit well in some samples, but will not in others. We implemented all estimation routines in software R. To guarantee that the comparison is fair, the same optimization routine to solve (15) is used for

Stock Exchange	Number of Instruments	Average Number of Observations
Frankfurt	403	5810
London	787	5238
Hong Kong	903	5079
New York	1071	5294
Total	3164	5399

Table 1: Summary of the historical stock exchange data set. Each return series has at least 3,000 observations.

each estimator.¹⁰

3.2 GARCH Parameter Estimation

The GARCH(1,1) model is used to forecast in each of our studies. It requires an estimate for parameter $\theta = (\omega, \alpha, \beta)'$ (see (3)). The first simulation-based study generates samples based on $\theta_0 = (0.0001, 0.075, 0.9)'$ – a prototypical choice for an MLE-based GARCH(1,1) estimate as applied to financial data. The HafPremSV setting is the counterpart to the GARCH(1,1) model that reflects the behavior of financial data in the SV model (see (4)). HafPremSV is represented by $\phi = 0.95$, $\gamma = -0.411$, and $\vartheta = 0.484$ (Hafner and Preminger, 2010). A second SV setting with a smaller persistence ϕ is referred to as the HafHerwSV and given by $\phi = 0.9$, $\gamma = -0.821$, and $\vartheta = 0.675$ (Hafner and Herwartz, 2000). In both simulation studies the sample size is set to $T = 5,000$, which we replicate $R = 5,000$ times, respectively. Correspondingly, we create enough data for the focus to remain on the methods and eliminate small sample issues. Simultaneously, the study is feasible in a reasonable amount of time.¹¹

Estimation Performance Analysis

In the correctly specified GARCH setting the outcome of the estimation is compared with the mean estimate (ME), the standard error (SE) and the mean squared error (MSE). In addition to these descriptive statistics, which are presented in Table 6, Figure 2 visually depicts the corresponding density plots. This part of the study verifies that all three estimators are essentially unbiased towards the true parameter θ_0 . The QSE generally has a higher MSE and SE than the CRPSE, which is rather comparable to the MLE. An apparent disadvantage of the QSE is revealed by the column to $QSE_{0.01}$. Based on these metrics, the smallest quantile levels included display the worst precision in the entire group. One potential explanation for this poor performance is the substantially smaller number of returns in the lower tail of each sample’s distribution. The QSE, of course, solely relies on this observations within the estimation process. Even though $T = 5,000$ seems to be a sufficiently large sample size to begin with, this statement is only relative to the chosen p -level and, in this case, may still be too small.

Tables 2 and 7 present the GARCH(1,1) parameter estimation summary of the study for the HafPremSV and the HafHerwSV setting, respectively. The density plots are presented in Appendix 4. Both SV settings leave similar impressions. The different scoring rules in-

¹⁰Every performed optimization uses function `nlm` of R-package `stats`.

¹¹Note that parallel computing capacity for this study was necessary to use and provided by the University of Marburg.

	ω	α	β	$\alpha + \beta$	rate $\alpha + \beta < 1$
start	0.00012	0.05	0.8	0.85	
<i>MLE</i>	9.84e-06 (3.54e-06)	0.388 (0.037)	0.699 (0.022)	1.086 (0.021)	0%
<i>CRPSE</i>	1.203e-05 (3.2e-06)	0.255 (0.022)	0.678 (0.025)	0.933 (0.011)	100%
<i>QSE</i> _{0.01}	2.757e-05 (1.488e-05)	0.408 (0.105)	0.713 (0.056)	1.121 (0.067)	1.04%
<i>QSE</i> _{0.05}	1.676e-05 (6.36e-06)	0.301 (0.051)	0.693 (0.04)	0.994 (0.027)	60.5%
<i>QSE</i> _{0.075}	1.459e-05 (5.42e-06)	0.279 (0.045)	0.687 (0.039)	0.966 (0.024)	92.94%
<i>QSE</i> _{0.1}	1.328e-05 (4.89e-06)	0.264 (0.042)	0.682 (0.039)	0.947 (0.022)	99.44%

Table 2: Mean estimates of a misspecified GARCH model in the HafPremSV sample of size $T = 5,000$ with $R = 5,000$ replications. Standard errors are shown in parentheses below the mean estimates.

deed lead to different GARCH parameters. This confirms the influence the embedded scoring function has on the estimate and, subsequently, the presence of misspecification. Once again, the SE of the QSE is greater than either the SE of the MLE or the CRPSE. Interestingly, the MLE does not outperform the CRPSE as before. Instead, if we take a closer look at parameter α , for example, we can witness that the MLE has an SE of 0.037 (0.056), while the CRPSE has a lower SE of just 0.022 (0.025) in the HafPremSV (HafHerwSV) setting. Obviously, misspecification may cause an efficient estimator to lose its SE-dominance.

Handling the GARCH Persistence

The *persistence* of a stochastic process describes how long a shock in conditional variance persists. For the GARCH(1,1) model, the persistence is characterized by $\alpha + \beta$. If bigger than one, the GARCH(1,1) process does not return to an unconditional variance (Nelson, 1990), which usually is a problematic choice. Persistence is a relatable property of both the SV and the GARCH models, which makes it an additional feature for comparing estimates (GARCH) on the basis of the input data (SV). As the criterion, the ability to keep the persistence of the estimated process below one (stationary GARCH) and, moreover, close to the persistence of the SV data is considered. Tables 2 and 7 show the mean GARCH persistence ($\alpha + \beta$) of the estimates. It not only varies, but also frequently exceeds level one; with the MLE, for instance, this is true for every single replication (the rate of $\alpha + \beta < 1$ is zero percent). The input persistence of the two SV settings is chosen close to, yet still below, one (0.95 and 0.9). Originally, this choice reflected known dynamics of financial data. The MLE is never estimating a stationary GARCH in our study, which does not come by surprise. It is a documented phenomenon that, given the same data, GARCH persistence is estimated higher than SV persistence (Carnero, 2004). However, our study does not support this statement without further distinction. In contrast to the MLE, the CRPSE underestimates the persistence for both SV settings (mean persistence of 0.933 in HafPremSV and 0.859 in HafHerwSV) and also has a lower mean persistence within the stock data (see Table 8). In the latter case,

the true persistence may vary between instruments and is ultimately unknown. Additionally, the CRPSE keeps the persistence below one for every single replication in the HafPremSV, HafHerwSV, and stock data study without exception.

Besides $QSE_{0.01}$, all mean QSE-estimates present a stationary GARCH. The mean estimate of the $QSE_{0.01}$ does not and, in addition, the rates of estimated stationary processes are quite low (1% and 4%). This may be explained by the poor overall estimation performance of level $p = 0.01$, as previously mentioned. Yet, a closer look at the tables also reveals that a decrease in p generally leads to an increase in the mean GARCH persistence of the QSE_p and eventually to a non-stationary outcome. This holds true for the stock data as well: the lower the p , the higher the persistence. This rule of thumb is supposed to be a direct consequence of misspecification. A single optimal VaR model does not seem not exist for all VaR levels p , so the estimates vary to make up for that. Under the influence of misspecification, the estimation routine seems to cause the deviation and – in case of the MLE – leads to a clear overestimation of the persistence. This would explain the observations made by Carnero (2004), who, as in most application studies, used the MLE as the standard estimation procedure for GARCH models without considering alternatives. The CRPSE produces a different result and leads to a more suitable parameter. The QSE provides interesting impressions, but will not reliably solve the non-stationarity issue.

3.3 Value at Risk Forecast Study

After comparing the different estimates in detail, we are now turning the attention to the VaR forecast capabilities of the estimated models. Based on an estimator $E \in \{MLE, CRPSE, QSE_p\}$, we deduce the in-sample GARCH(1,1) forecast series $VaR_t^{p,E}$ at VaR-level p for any sample y_t and $p \in \{0.01, 0.05, 0.075, 0.1\}$. Note that we do not include the forecast $VaR_t^{p_1, QSE_{p_2}}$ where $p_1 \neq p_2$. Although this is a valid forecast series in general, we would drop the intended conformity between forecast goal and estimation criterion choice, which is what we originally built the QSE for.

Mean Score Evaluation

The mean score criterion is based on scoring QS_p (see (10)) – an SCSF for the VaR. Thus, we calculate the mean score for every VaR forecast series belonging to estimator E by

$$\frac{1}{T} \sum_{t=1}^T QS_p(VaR_t^{p,E}, y_t), \quad p \in \{0.01, 0.05, 0.075, 0.1\}, \quad (18)$$

where y_t is the portfolio observation for the corresponding VaR forecast – either an SV sample or a log-return series in this particular case. Table 3 summarizes this information; it shows the mean of the series of each mean score for every replication. For each of the three data sources shown in the left column, VaR series are calculated for four levels of p .

The QSE_p is the best estimator to predict the VaR, as it has the minimum mean score over all replications, respectively. In addition, the VaR_t^{p, QSE_p} series has the lowest mean score for both the HafPremSV and the HafHerwSV in 5,000 out of 5,000 replications for every level p , respectively. The bottom third of Table 3 displays the result for the 3,164 stock return series. Again, the QSE-GARCH(1,1)-VaR has the minimum mean score, proving it

Data Source	p	$VaR_t^{p,MLE}$	$VaR_t^{p,CRPSE}$	VaR_t^{p,QSE_p}
HafPremSV	1%	0.00091440	0.00103401	0.00089047
	5%	0.00292248	0.00292579	0.00288864
	7.5%	0.00385681	0.00380178	0.00378688
	10%	0.00465041	0.00455157	0.00454562
HafHerwSV	1%	0.00100130	0.00118482	0.00096610
	5%	0.00305503	0.00306793	0.00301161
	7.5%	0.00399898	0.00392625	0.00390419
	10%	0.00479483	0.00465625	0.00464848
Stock Data	1%	0.00124712	0.00179485	0.00116924
	5%	0.00346969	0.00360437	0.00334212
	7.5%	0.00449078	0.00443094	0.00427636
	10%	0.00535128	0.00513535	0.00504818

Table 3: Mean scores of all VaR models for each p and for each data source. Mean scores (based on function QS_p) of each sample of size $T = 5,000$ are aggregates over $R = 5,000$ replications. Bold numbers indicate the minimum of each row, which also indicates the best VaR model for a given p .

the best estimator here as well. An in-depth look at the return series shows that the QSE-GARCH(1,1)-VaR provides the lowest mean score in 3,040 out of 3,164 cases. Hence, the QSE can be considered the best VaR forecast estimator for the vast majority of instruments. The mean score criterion overwhelmingly shows that the QSE has a positive impact on VaR forecasts. Ultimately this finding is unsurprising, as the estimation routine was specifically designed to minimize the mean score criterion in this analysis. Nevertheless, the results leave an intriguing impression of the QSE concept.

Backtest Evaluation

To underline the practical implications of our framework, we draw on the KupT and the ChrT for unconditional and conditional coverage. Compared to the mean score criterion, the backtesting criterion indisputably presents a less-biased method for evaluating the VaR forecast quality induced by the estimators. Once again, the two SV simulation settings (HafPremSV and HafHerwSV) as well as the stock return data are considered. Every combination of forecast series $VaR_t^{p,E}$ and observation series y_t leads to a single backtest decision of either rejecting the tested null hypothesis H_0 (“the model is correct”) or not. If a test fails to reject H_0 , practitioners typically accept the model as appropriate. However, it is important to reiterate that “fails to reject H_0 ” is not the equivalent of saying the VaR model is entirely correct. Rather, it means that the model cannot be rejected based on the test subject and scope. Nevertheless, a model that these tests frequently do not reject can be considered superior to one that is frequently rejected. After all, regulators still require the model to pass the tests and, if this is not accomplished, further costs arise.

Tables 4 treats the HafPremSV and Table 10 the HafHerwSV simulation. Table 5 presents the results of backtesting within the stock data. To see the difference between a correct and a possibly incorrect specification, Table 9 shows the results for the correctly specified

GARCH simulation study. Presented in each of those tables are relative frequencies of the test decision of failing to reject the H_0 . As expected, the test of the correct specification is rarely rejected with the rate given in the table staying close to the chosen test confidence of 95%, respectively. The CCT causes some problems following all three included estimators for $p = 0.01$, which leads to the suspicion the sample size is too small to test the conditional coverage of the 99%-VaR forecast. As the SV model simulations apparently causes an

Forecast Objective	Estimator	ChrT fails to reject H_0	KupT fails to reject H_0
$VaR^{0.01}$	<i>MLE</i>	0.50	0.48
	<i>CRPSE</i>	0.00	0.00
	<i>QSE</i> _{0.01}	30.20	99.98
$VaR^{0.05}$	<i>MLE</i>	65.16	56.66
	<i>CRPSE</i>	0.02	0.00
	<i>QSE</i> _{0.05}	98.40	100.00
$VaR^{0.075}$	<i>MLE</i>	2.50	0.80
	<i>CRPSE</i>	7.12	3.44
	<i>QSE</i> _{0.075}	98.42	100.00
$VaR^{0.1}$	<i>MLE</i>	0.00	0.00
	<i>CRPSE</i>	61.26	50.70
	<i>QSE</i> _{0.1}	97.92	100.00

Table 4: Percentages of “fails to reject the H_0 ” (H_0 : correct VaR model) of $R = 5,000$ backtests with 95% confidence. HafPremSV samples are used with size $T = 5,000$. Forecast model is the GARCH(1,1), in which three competing estimators (the MLE, CRPSE, and QSE_p) are compared with four different VaR levels p . Note that for the $QSE_{0.01}$ -VaR, the ChrT returns a significant amount of NA’s (MLE: 2124, CRPSE: 148 and $QSE_{0.01}$: 3488), which distorts the results for this level.

incorrect model specification, the VaR forecast series of the MLE-GARCH(1,1) is rejected for nearly every replication at each level p (see the rows corresponding to MLE in Tables 10 and 4). This clearly shows the ability of these two tests to identify an incorrect model. The CRPSE-GARCH(1,1) is also rejected at a very high rate (in both simulations), from which misspecification can be concluded, as well. However, the VaR forecast series of the QSE-GARCH(1,1) presents a very different outcome. Both tests accept this VaR model at an exceptionally high rate. Although misspecification is still present, the estimation routine seemingly hides it with a suitable estimate within the given parameter space, effectively deceiving the test procedures. This remarkable result leads to two different conclusions. First, if a certain degree of misspecification is present or suspected similar to the GARCH-SV scenario in this study, then the QSE_p dramatically improves the outcome by estimating models that are accepted at much higher rates than those of the MLE.

Performing the same procedure on the stock data (Table 5) creates a similar effect, however, as the QSE indeed impressively affects the outcome. Yet again, the KupT fails to reject the QSE-GARCH(1,1) model for well over 90% of the samples. Furthermore, the ChrT fails to reject it for at least 56% of the samples. In contrast, the MLE and CRPSE estimate models are frequently rejected by both tests at a high rate of 80% or more. Each stock surely has a unique dynamic, which depends on various factors, i.e. the trading activity of the instrument.

Forecast Objective	Estimator	ChrT fails to reject H0	KupT fails to reject H0
$VaR^{0.01}$	<i>MLE</i>	17.23	24.68
	<i>CRPSE</i>	0.00	0.00
	<i>QSE</i> _{0.01}	62.86	99.68
$VaR^{0.05}$	<i>MLE</i>	10.11	13.94
	<i>CRPSE</i>	2.05	2.84
	<i>QSE</i> _{0.05}	66.72	97.91
$VaR^{0.075}$	<i>MLE</i>	2.65	3.79
	<i>CRPSE</i>	12.80	21.90
	<i>QSE</i> _{0.075}	59.96	96.24
$VaR^{0.1}$	<i>MLE</i>	1.39	2.02
	<i>CRPSE</i>	25.54	44.53
	<i>QSE</i> _{0.1}	56.16	93.87

Table 5: Percentages of “fails to reject the H0” (H0: correct VaR model) of $R = 3,164$ backtests with 95% confidence. Stock return data samples are used as in Table 1. Forecast model is the GARCH(1,1), in which three competing estimators (the MLE, CRPSE, and QSE_p) are compared based on four VaR levels p .

Nevertheless, when combined with the basic GARCH(1,1) model, the QSE provides definitely the best GARCH(1,1)-VaR forecasts.

3.4 Application Findings

Data Origin			
	GARCH(1,1) Samples (Correct Specified GARCH(1,1))	Stochastic Volatility Samples (Misspecified GARCH(1,1) Estimation)	Stock Return Samples
Evaluation	All estimators unbiased $SE(QSE) > SE(CRPSE) > SE(MLE)$ $MSE(QSE) > MSE(CRPSE) > MSE(MLE)$ → All estimators work/ MLE efficient	Estimates differ following used scoring $SE(QSE) > SE(MLE) > SE(CRPSE)$ MLE persistence overestimated (> 1) CRPSE persistence < 1 (stationary process) → CRPSE is useful estimation routine	Estimates slightly differ following used scoring MISCE persistence $<$ MLE persistence < 1 $SE(QSE) > SE(MLE) > SE(CRPSE)$ → QSE/ CRPSE useful estimation routines
VaR Scoring	no distinct estimates/ models evolve	QSE VaR has always the lowest mean score → QSE improves VaR forecast	QSE VaR has frequently the lowest mean score → QSE improves VaR forecast
VaR Backtesting	no distinct estimates/ models evolve	QSE VaR model has much higher acceptance rate → QSE improves VaR forecast	QSE VaR model has much higher acceptance rate → QSE improves VaR forecast

Figure 1: Overview of the examination, observations, and the implications of this study.

Figure 1 presents an aggregated overview of the application design. Clusters 1-7 represent the focus areas, which are enhanced by key observations as well as essential intermediate conclusions. Each of those seven clusters endorse the quantile score estimation approach. Similar to the MLE, the CRPSE is a solid estimator with only slightly larger SEs and MSEs. The QSE is incomparable in this regard due to its greater standard error; nevertheless, it remains an unbiased and, hence, reasonable estimator (Cluster 1). We show that the different embedded scoring functions in the SV sample and the stock data lead to distinct VaR models through

estimation, which allows the comparison to be extended. Given the same data and opposite to other statements in the academic literature, the GARCH persistence is not generally higher than the SV persistence, as the CRPSE-GARCH persistence is frequently smaller than the chosen SV persistence of the samples. Hence, the CRPSE presents a useful alternative to the MLE for estimating a stationary GARCH process, even if the model is suspected to be misspecified, which, combined with the estimator at hand, seems to be the reason for an overestimated persistence. Interestingly, the CRPSE has a smaller SE in this scenario compared to the MLE (Clusters 2 and 3), so efficiency is no reasonable argument in favor of the MLE in this particular case. The QSE persistence varies depending on the level p , which seems to be a compensation mechanism for handling misspecification. This flexibility reflects the strength of the QSE: different levels clearly demand different models, which cannot be supplied by either the MLE or the CRPSE. In this regard, the two latter concepts can only generate models independent from the targeted VaR level.

The QSE proves to be an useful estimation concept for VaR models due to its unparalleled minimal mean score. It far outperforms the MLE and the CRPSE (Clusters 4 and 5) and warrants further study. In our analysis, the backtest criterion responds to the QSE in a remarkable, albeit questionable, manner. Surprisingly, both backtests fail to reject the QSE-GARCH(1,1) model at a very high rate, concealing the deliberately incorrect model specification. Neither the MLE-GARCH(1,1) nor the CRPSE-GARCH(1,1) do so, which further substantiates the conclusion. Although controversial, this must be mentioned, as the estimation routine is clearly capable of undermining the validity of backtests (Cluster 6). Applying the backtest criterion to the stock data essentially produces similar results as with the SV samples. Slightly different conclusions may be drawn from the result, however, because the correctness of any model specification is unknown. Additionally the fit may differ over the wide range of risk factors included. The QSE provides the least rejected and, hence, the most appropriate GARCH(1,1)-VaR model under the backtesting criterion. While the QSE's overall impact on coverage tests has been assumed due to the apparent relationship between the quantile scoring and coverage test statistics, this study reveals its unexpectedly strong impact. Based on both criteria, the VaR-optimal parameter for the GARCH(1,1) is clearly the QSE estimate, when misspecification is present. Moreover, without misspecification it appears robust enough to maintain its usefulness, unless the sample size is too small to make reasonable estimates. In order for the QSE to compete with the MLE, a required minimum sample size is critical. Our study shows this to be particularly true for the $QSE_{0.01}$. A general rule of thumb indicates that as p decreases, the sample size must increase. This is especially unfavorable in risk management applications, as large sample sizes are often unavailable and the small quantiles are of particular interest. Moreover, the CRPSE proves itself irrelevant for VaR forecast applications (Cluster 7).

4 Conclusions

We introduce and discuss a new concept for estimation in VaR forecast models that implicitly focuses on the forecast goal and its evaluation. We propose two estimators for the GARCH model – the QSE and the CRPSE. Both are built based on a strictly consistent scoring function of the quantile functional plugged into the M-estimation framework. We present a feasible environment for an application of the QSE by two propositions, so we must conclude that the

QSE is – other than the MLE and the CRPSE – not applicable within general parametric families of distributions. We show how different choices of the estimation criterion leads to different M-estimates under model misspecification, which is important in risk management because intended VaR forecasts based on that estimate will differ as well. Our application underlines how VaR forecasts are improved significantly, which we verify based on scoring function and backtest validation.

We demonstrate how different GARCH(1,1) estimates in the SV simulation study result corresponding to the different estimators, which is a potential stumbling block, as both the GARCH and SV models are common alternative choices to model financial time series and, consequently, such misspecification may be a frequent side effect in risk management. Interestingly, we observe a similar estimate deviation in the stock return data set, which explains our suspicion of misspecification being present there as well. The QSE-GARCH(1,1)-VaR regularly produces the lowest mean (quantile) scores and the model is significantly less often rejected by the Kupiec and the Christoffersen backtest. Hence, an explicit focus on the targeted VaR at the estimation stage appears to trump misspecification with regard to a valid VaR forecast. Without specific requirements regarding estimation routines, misspecified models can be altered to perform well in common model validation procedures. This could support further applications of simple model classes with limited parameters, where the estimation routine covers up the models' shortcomings. In order to meaningfully use the QSE, extended data sets that particularly contain many observations within the targeted tail are necessary. Consequently, the smaller the VaR level of interest, the greater the volume of required information must be – an unfortunate conclusion for practical risk management applications.

Further research into the quantile score-based estimation of VaR models could be directed towards an improvement of the CRPSE. Despite its flexibility and wide-ranging applicability, it remained unconvincing for improving VaR forecasts within our studies. Only the fact that the CRPSE – unlike the MLE – keeps the estimated GARCH persistence low and, as a result, the estimated GARCH processes more often stationary can be pointed out. An improved weighting of the quantiles used for the CRPSE and, in particular with regard to the targeted VaR, may result in a more effective CRPSE towards VaR forecasting. Thereby, the QSE advantages could be integrated, while the more general CRPSE framework compensates its disadvantages. Furthermore, the family of generalized piecewise linear functions provides many candidates of strictly consistent scoring functions for VaR, so improvements of the quantile-based estimators may be found with different criterion choices or combinations out of this family as well.

References

- Acerbi, C. and Tasche, D. (2002), Expected Shortfall A Natural Coherent Alternative to Value at Risk, *Economic Notes* **31**(2), pp. 379–388.
- Auer, M. (ed.) (2018), *Hands-On Value-at-Risk and Expected Shortfall: A Practical Primer, Management for Professionals*, Springer International Publishing, Cham.
- Bao, Y., Lee, T. H. and Saltoglu, B. (2006), Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check, *Journal of Forecasting* **25**(2), pp. 101–128.
- BCBS (2009), *Revisions to the Basel II Market Risk Framework*, Basel Committee on Banking Supervision.
URL: <https://www.bis.org/publ/bcbs158.htm>
- BCBS (2019), *Minimum Capital Requirements for Market Risk*, Basel Committee on Banking Supervision.
URL: <https://www.bis.org/bcbs/publ/d457.pdf>
- Bernardi, M. and Catania, L. (2016), Comparison of Value-at-Risk Models Using the MCS Approach, *Computational Statistics* **31**(2), pp. 579–608.
- Birgé, L. and Massart, P. (1993), Rates of Convergence for Minimum Contrast Estimators, *Probability Theory and Related Fields* **97**(1-2), pp. 113–150.
- Bollerslev, T. P. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **31**(3), pp. 307–327.
- Carnero, M. A. (2004), Persistence and Kurtosis in GARCH and Stochastic Volatility Models, *Journal of Financial Econometrics* **2**(2), pp. 319–342.
- Christoffersen, P. (1998), Evaluating Interval Forecasts, *International Economic Review* **39**(4), pp. 841–862.
- Christoffersen, P., Hahn, J. and Inoue, A. (2001), Testing and Comparing Value-at-Risk Measures, *Journal of Empirical Finance* **8**(3), pp. 325–342.
- Christoffersen, P. and Pelletier, D. (2004), Backtesting Value-at-Risk: A Duration-Based Approach, *Journal of Financial Econometrics* **2**(1), pp. 84–108.
- Engle, R. F. and Manganelli, S. (2004), CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business and Economic Statistics* **22**(4), pp. 367–381.
- Fissler, T., Ziegel, J. F. and Gneiting, T. (2016), Expected Shortfall is Jointly Elicitable with Value-at-Risk Implications for Backtesting, *Risk* **29**(1), pp. 58–61.
- Francq, C. and Zakoian, J. M. (2014), Quasi-Maximum Likelihood Estimation of GARCH Models with Heavy-Tailed Likelihoods: Comment, *Journal of Business and Economic Statistics* **32**(2), pp. 198–201.

- Gneiting, T. (2011*a*), Making and Evaluating Point Forecasts, *Journal of the American Statistical Association* **106**(494), pp. 746–762.
- Gneiting, T. (2011*b*), Quantiles as Optimal Point Forecasts, *International Journal of Forecasting* **27**(2), pp. 197–207.
- Gneiting, T. and Katzfuss, M. (2014), Probabilistic Forecasting, *Annual Review of Statistics and Its Application* **1**, pp. 125–151.
- Gneiting, T. and Raftery, A. E. (2007), Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association* **102**(477), pp. 359–378.
- Gneiting, T., Raftery, A. E., Westveld, A. H. and Goldman, T. (2005), Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation, *Monthly Weather Review* **133**(5), pp. 1098–1118.
- Guler, K., Ng, P. T. and Xiao, Z. (2017), Mincer-Zarnowitz Quantile and Expectile Regressions for Forecast Evaluations under Asymmetric Loss Functions, *Journal of Forecasting* **36**(6), pp. 651–679.
- Hafner, C. M. and Herwartz, H. (2000), Testing for Linear Autoregressive Dynamics Under Heteroskedasticity, *Econometrics Journal* **3**(2), pp. 177–197.
- Hafner, C. M. and Preminger, A. (2010), Deciding between GARCH and Stochastic Volatility via Strong Decision Rules, *Journal of Statistical Planning and Inference* **140**(3), pp. 791–805.
- Heston, S. (1993), A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *The Review of Financial Studies* **6**(2), pp. 327–343.
- Huber, P. J. (1964), Robust Estimation of a Location Parameter, *The Annals of Mathematical Statistics* **35**(1), pp. 73–101.
- Jeon, J. and Taylor, J. W. (2013), Using CAViaR Models with Implied Volatility for Value-at-Risk Estimation, *Journal of Forecasting* **32**(1), pp. 62–74.
- Jorion, P. (2006), *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd edn., McGraw-Hill, New York.
- Koenker, R. W. and Bassett, G. (1978), Regression Quantiles, *Econometrica* **46**(1), pp. 33–50.
- Krüger, F. (2013), *Four Essays on Probabilistic Forecasting in Econometrics*, PhD thesis, University of Konstanz, Konstanz.
- Kuester, K., Mittnik, S. and Paolella, M. S. (2006), Value-at-Risk Prediction: A Comparison of Alternative Strategies, *Journal of Financial Econometrics* **4**(1), pp. 53–89.
- Kupiec, P. (1995), Techniques for Verifying the Accuracy of Risk Management Models, *Journal of Derivatives* **3**(2), pp. 73–84.
- Lambert, N. S., Pennock, D. M. and Shoham, Y. (2008), Eliciting Properties of Probability Distributions, in: Fortnow, L. (ed.), *Proceedings of the 9th ACM Conference on Electronic Commerce*, ACM Digital Library, ACM, New York, NY, pp. 129–138.

- McNeil, A. J., Frey, R. and Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton Series in Finance, Princeton University Press, Princeton, NJ.
- Nelson, D. B. (1990), Stationarity and Persistence in the GARCH (1,1) Model, *Econometric Theory* **6**(3), pp. 318–334.
- Pfanzagl, J. (1969), On the Measurability and Consistency of Minimum Contrast Estimates, *Metrika* **14**(1), pp. 249–272.
- Saerens, M. (2000), Building Cost Functions Minimizing to Some Summary Statistics, *IEEE Transactions on Neural Networks* **11**(6), pp. 1263–1271.
- Staudte, R. G. and Sheather, S. J. (1990), *Robust Estimation & Testing*, Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ.
- Taylor, S. J. (1986), *Modelling Financial Time Series*, Wiley, Hoboken, NJ.
- Thomson, W. (1979), Eliciting Production Possibilities from a Well-Informed Manager, *Journal of Economic Theory* **20**(3), pp. 360–380.
- Tian, W. (ed.) (2017), *Commercial Banking Risk Management: Regulation in the Wake of the Financial Crisis*, Palgrave Macmillan US, New York.
- Tsay, R. S. (2010), *Analysis of Financial Time Series*, Wiley Series in Probability and Statistics, 3rd edn., Wiley, Hoboken, NJ.
- Van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.
- Weiss, A. A. (1996), Estimating Time Series Models Using the Relevant Cost Function, *Journal of Applied Econometrics* **11**(5), pp. 539–560.
- White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* **48**(4), pp. 817–838.
- White, H. (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica* **50**(1), pp. 1–25.
- Xiao, Z. and Koenker, R. (2009), Conditional Quantile Estimation for Generalized Autoregressive Conditional Heteroscedasticity Models, *Journal of the American Statistical Association* **104**(488), pp. 1696–1712.
- Ziegel, J. F. (2016), Coherence and Elicitability, *Mathematical Finance* **26**(4), pp. 901–918.

Appendix

Proofs of Propositions*

Proof of Proposition 1. For (a) see proof of Proposition 2, which also includes this class as it is $\mathcal{F}^{cnorm} \subset \mathcal{F}^{norm}$.

For (b), an arbitrary distribution G is chosen, that has the same expected value than the true distribution F , under F . It is then demonstrated that both distributions must be equal.

Let $p \in (0, 1)$ and $Y \sim F \in \mathcal{F}^{cnorm}$. Let $G \in \mathcal{F}^{cnorm}$ be arbitrary with

$$E_F \left[\mathbf{QS}_p(F, Y) \right] = E_F \left[\mathbf{QS}_p(G, Y) \right].$$

Both F and G are represented by standard deviations σ_F and σ_G , respectively following the definition of \mathcal{F}^{cnorm} . Then by definition of \mathbf{QS}_p , also

$$E_F \left[\mathbf{QS}_p(\text{VaR}^p(F), Y) \right] = E_F \left[\mathbf{QS}_p(\text{VaR}^p(G), Y) \right]$$

holds. \mathbf{QS}_p is a SCSF for VaR^p , so this equality holds if and only if

$$\text{VaR}^p(F) = \text{VaR}^p(G).$$

By definition of the VaR and basic relationships of quantiles of the normal distribution, thus

$$q_p(F) = \sigma_F \Phi^{-1}(p) = \sigma_G \Phi^{-1}(p) = q_p(G),$$

which holds if and only if $\sigma_F = \sigma_G$. Hence $F = G$ holds, too, which by definition proves \mathbf{QS}_p to be strictly proper relative to class \mathcal{F}^{cnorm} . \square

Proof of Proposition 2. For (a) see Gneiting (2011a).

The proof of (b) is designed as follows. It is shown, that different combinations of (μ, σ) and, hence, multiple members $F_1 \neq F_2 \in \mathcal{F}^{norm}$ could lead to the same functional value $\text{VaR}^p(F_1) = \text{VaR}^p(F_2)$ and by that also to the same expected score following scoring rule \mathbf{QS}_p . This rules out \mathbf{QS}_p for being strictly proper relative to \mathcal{F}^{norm} .

Let $F_i := \mathcal{N}(\mu_i, \sigma_i^2) \in \mathcal{F}^{norm}$ with $i \in \{1, 2\}$ be two normal distributions represented by parameter vectors

$$(\mu_1, \sigma_1)' := (0, 1)' \neq (-0.05, 0.9696021584)' =: (\mu_2, \sigma_2)'$$

Clearly, these are two different representatives $F_1 \neq F_2$ in \mathcal{F}^{norm} , but they possess the same VaR to level 0.05, that is

$$\text{VaR}^{0.05}(F_1) = 1.644853627 = \text{VaR}^{0.05}(F_2).$$

Hence, with $Y \sim F_1$ it follows that

$$E_{F_1} \left[\mathbf{QS}_{0.05}(\text{VaR}^{0.05}(F_1), Y) \right] = E_{F_1} \left[\mathbf{QS}_{0.05}(\text{VaR}^{0.05}(F_2), Y) \right],$$

and finally by definition of scoring rule \mathbf{QS}_p

$$E_{F_1} \left[\mathbf{QS}_{0.05}(F_1, Y) \right] = E_{F_1} \left[\mathbf{QS}_{0.05}(F_2, Y) \right].$$

But still $F_1 \neq F_2$, which means \mathbf{QS}_p is not strictly proper relative to class \mathcal{F}^{norm} . \square

CRPS Closed Form Special Case

In case of $F_t \in \mathcal{F}^{norm} \forall t$, the integral in (16) has the closed form representation

$$\text{CRPS}(F_t, y_t) = \sigma_t \left[\frac{1}{\sqrt{\pi}} - 2 \phi\left(\frac{y_t - \mu_t}{\sigma_t}\right) - \frac{y_t - \mu_t}{\sigma_t} \left(2 \Phi\left(\frac{y_t - \mu_t}{\sigma_t}\right) - 1 \right) \right], \quad (19)$$

with ϕ , Φ being the density and distribution function of the standard normal distribution, respectively (Gneiting and Raftery, 2007). Recall that $F_t \in \mathcal{F}^{norm}$ means that F_t is a normal distribution with mean μ_t and variance σ_t^2 .

Estimation Results

Figure 2 shows the densities of the correct specified GARCH estimation. All estimators turn out to be unbiased. As expected, the MLE provides the best precision. Figure 3 presents estimation densities of the GARCH model given the simulated HafPremSV samples; Figure 4 includes HafHerwSV samples. Due to the misspecification, the estimates differ in both cases and, therefore, lead to different GARCH models.

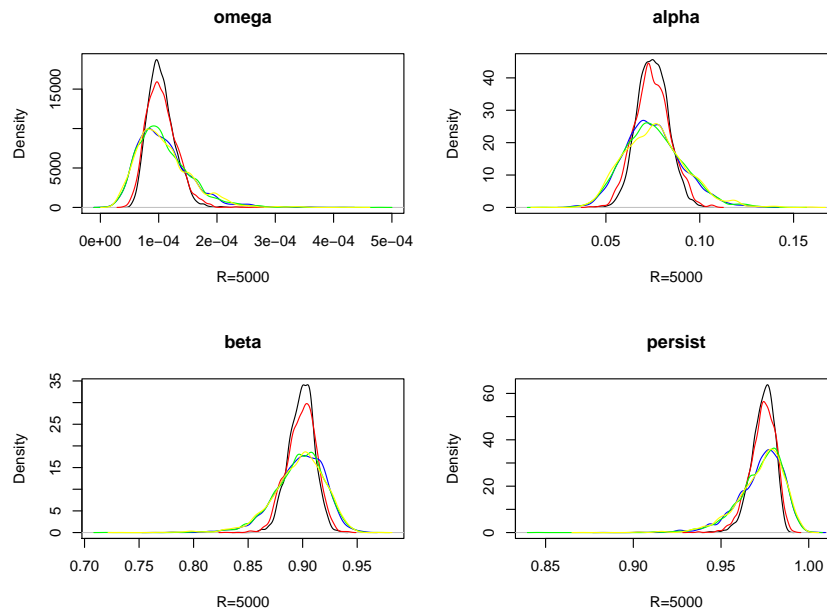


Figure 2: Densities of correct specified GARCH estimation in the GARCH sample of size $T = 5,000$ with $R = 5,000$ replications. The MLE density is black and the CRPSE is red. The QSE's are blue ($p = 0.05$), green ($p = 0.075$), and yellow ($p = 0.1$).

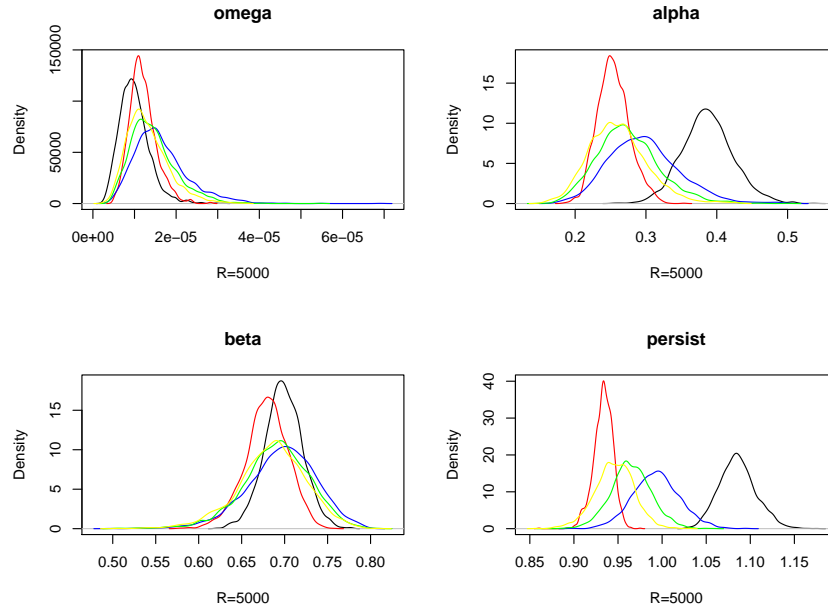


Figure 3: Densities of misspecified GARCH estimation in the HafPremSV sample of size $T = 5,000$ with $R = 5,000$ replications. The *MLE* density is black and the *CRPSE* is red. The *QSE*'s are blue ($p = 0.05$), green ($p = 0.075$), and yellow ($p = 0.1$).

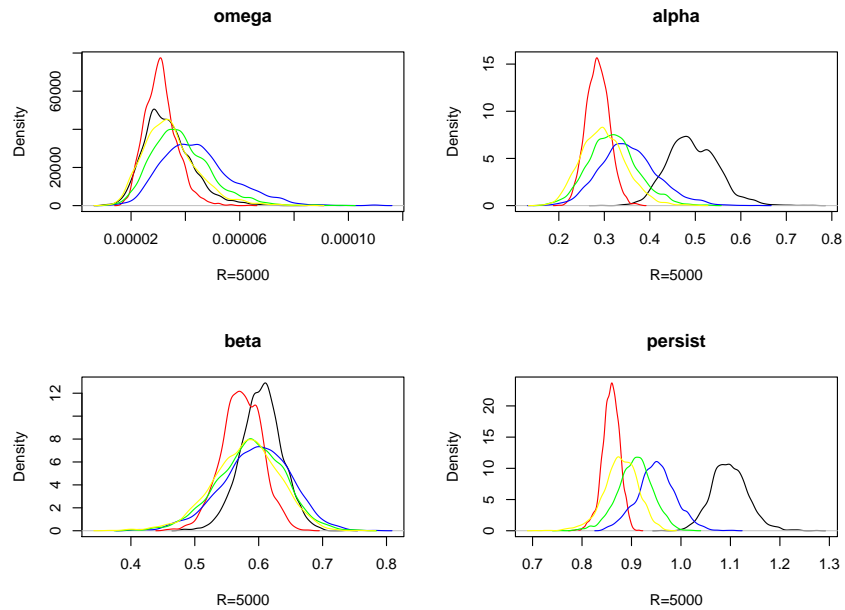


Figure 4: Densities of misspecified GARCH estimation in the HafHerwSV sample of size $T = 5,000$ with $R = 5,000$ replications. The *MLE* density is black and the *CRPSE* is red. The *QSE*'s are blue ($p = 0.05$), green ($p = 0.075$), and yellow ($p = 0.1$).

Table 6 provides mean estimates of the GARCH model estimation study using the GARCH-simulated data. The estimates stay close to the true parameter, however, the MLE closely followed by the CRPSE achieve the best precision. The situation is clearly changing under misspecification. Table 7 summarizes the estimation within the HafHerwSV sample. Like discussed earlier by means of Table 2, there are different estimates resulting corresponding to the different estimators and it is not possible to name a clear frontrunner – neither by considering the mean nor the SE. Interestingly, this does not change within the stock data analysis given in Table 8.

	ω	α	β	$\alpha + \beta$	rate $\alpha + \beta < 1$
true	0.0001	0.075	0.9	0.975	
start	0.00012	0.05	0.8	0.85	
<i>MLE</i>	0.00010501 (2.275e-05) [0]	0.075 (0.008) [0]	0.898 (0.012) [0]	0.974 (0.006) [0]	100
<i>CRPSE</i>	0.00010631 (2.627e-05) [0]	0.075 (0.01) [0]	0.898 (0.014) [0]	0.973 (0.007) [0]	100
<i>QSE</i> _{0.01}	0.00011342 (5.661e-05) [0]	0.077 (0.019) [0]	0.895 (0.028) [0.001]	0.971 (0.016) [0]	99.44
<i>QSE</i> _{0.05}	0.00010953 (4.53e-05) [0]	0.076 (0.016) [0]	0.896 (0.023) [0.001]	0.973 (0.013) [0]	99.98
<i>QSE</i> _{0.075}	0.0001095 (4.613e-05) [0]	0.076 (0.016) [0]	0.896 (0.023) [0.001]	0.973 (0.013) [0]	99.98
<i>QSE</i> _{0.1}	0.00010988 (4.798e-05) [0]	0.076 (0.017) [0]	0.896 (0.024) [0.001]	0.972 (0.014) [0]	99.88

Table 6: Mean estimates of a correct specified GARCH model in the GARCH sample of size $T = 5,000$ based on $R = 5,000$ replications. Standard errors are given in parentheses and mean squared errors are given in brackets below the mean estimates.

	ω	α	β	$\alpha + \beta$	rate $\alpha + \beta < 1$
start	0.00012	0.05	0.8	0.85	
<i>MLE</i>	3.373e-05 (8.84e-06)	0.495 (0.056)	0.604 (0.033)	1.099 (0.037)	0.06%
<i>CRPSE</i>	3.098e-05 (5.92e-06)	0.285 (0.025)	0.573 (0.032)	0.859 (0.018)	100%
<i>QSE</i> _{0.01}	7.809e-05 (3.614e-05)	0.525 (0.146)	0.622 (0.081)	1.147 (0.099)	4.04%
<i>QSE</i> _{0.05}	4.442e-05 (1.311e-05)	0.356 (0.063)	0.596 (0.054)	0.952 (0.039)	90.12 %
<i>QSE</i> _{0.075}	3.835e-05 (1.105e-05)	0.322 (0.054)	0.587 (0.052)	0.909 (0.035)	99.7%
<i>QSE</i> _{0.1}	3.451e-05 (9.83e-06)	0.3 (0.05)	0.58 (0.052)	0.88 (0.033)	100%

Table 7: Mean estimates of a misspecified GARCH model in the HafHerwSV sample of size $T = 5,000$ based on $R = 5,000$ replications. Standard errors are given in parentheses below the mean estimates.

	ω	α	β	$\alpha + \beta$	rate $\alpha + \beta < 1$
start	0.00012	0.05	0.8	0.85	
MLE	6.85e-05 (0.00016009)	0.103 (0.106)	0.852 (0.137)	0.955 (0.104)	87.8%
CRPSE	2.444e-05 (6.491e-05)	0.107 (0.068)	0.711 (0.265)	0.817 (0.232)	100%
<i>QSE</i> _{0.01}	9.121e-05 (0.00025211)	0.12 (0.095)	0.856 (0.123)	0.976 (0.086)	59.2%
<i>QSE</i> _{0.05}	4.164e-05 (0.00010526)	0.087 (0.062)	0.847 (0.136)	0.935 (0.103)	99.1%
<i>QSE</i> _{0.075}	3.575e-05 (9.515e-05)	0.087 (0.071)	0.83 (0.162)	0.917 (0.126)	99.7%
<i>QSE</i> _{0.1}	3.108e-05 (8.413e-05)	0.087 (0.074)	0.811 (0.189)	0.898 (0.153)	100%

Table 8: Mean estimates of GARCH(1,1) model estimation based on 3,164 stock data samples (see Table 1. Standard errors are given in parentheses below the mean estimates.

Value at Risk Backtesting Evaluation Results

The backtesting results of Table 9 present rates the backtests fail to reject the H0 in the correct specified GARCH analysis. The rates are rather high which means the tests identify the correct specification. An exception occurs for backtesting the $VaR^{0.01}$ with the ChrT. The tests fail to provide a decision in many replications (MLE 3027 times, CRPSE 3021 times, and $QSE_{0.01}$ 3043 times) and therefore these results are not representative at all. Whereas the rates in this application between the models are quite similar, this changes under misspecification. Table 10 presents the results for the HafHerwSV, which resembles the results of the HafPremSV included in the main text above. Only the QSE achieves high rates, while the MLE and the CRPSE are frequently rejected by both tests. Again, the ChrT itself fails to produce test decisions for the $VaR^{0.01}$ (MLE 2067 times, CRPSE 75 times, and $QSE_{0.01}$ 3539) and, thus, distorts the significance of the corresponding cluster.

Forecast Objective	Estimator	ChrT fails to reject H0	KupT fails to reject H0
$VaR^{0.01}$	<i>MLE</i>	38.20	97.56
	<i>CRPSE</i>	37.90	95.72
	$QSE_{0.01}$	38.38	100.00
$VaR^{0.05}$	<i>MLE</i>	96.96	98.30
	<i>CRPS</i>	96.36	97.36
	$QSE_{0.05}$	98.38	100.00
$VaR^{0.075}$	<i>MLE</i>	96.68	98.30
	<i>CRPSE</i>	96.52	97.80
	$QSE_{0.075}$	98.92	100.00
$VaR^{0.1}$	<i>MLE</i>	96.90	98.08
	<i>CRPSE</i>	96.88	98.08
	$QSE_{0.1}$	98.74	100.00

Table 9: Percentages of “fails to reject the H0” (H0: correct VaR model) of $R = 5,000$ backtests with 95% confidence. GARCH(1,1) samples are used based on parameters $(\omega, \alpha, \beta)' = (0.0001, 0.075, 0.9)'$ and size $T = 5,000$. Forecast model is the GARCH(1,1), in which three competing estimators (the MLE, CRPSE, and QSE_p) are compared with four different VaR levels p .

Forecast Objective	Estimator	ChrT fails to reject H0	KupT fails to reject H0
$VaR^{0.01}$	<i>MLE</i>	0.38	0.14
	<i>CRPSE</i>	0.00	0.00
	<i>QSE</i> _{0.01}	29.22	100.00
$VaR^{0.05}$	<i>MLE</i>	40.54	32.02
	<i>CRPS</i>	0.00	0.00
	<i>QSE</i> _{0.05}	98.96	100.00
$VaR^{0.075}$	<i>MLE</i>	0.22	0.04
	<i>CRPSE</i>	2.10	0.76
	<i>QSE</i> _{0.075}	98.68	100.00
$VaR^{0.1}$	<i>MLE</i>	0.00	0.00
	<i>CRPSE</i>	50.30	39.68
	<i>QSE</i> _{0.1}	97.92	100.00

Table 10: Percentages of “fails to reject the H0” (H0: correct VaR model) of $R = 5,000$ backtests with 95% confidence. HafHerwSV samples are used with size $T = 5,000$. Forecast model is the GARCH(1,1), in which three competing estimators (the MLE, CRPSE, and QSE_p) are compared with four different VaR levels p .

„(Lifetime) Expected Credit Losses“ als Mechanismus der Verlustantizipation nach IFRS 9 – Operationalisierung der Schätzparameter in kritischer Betrachtung –

Erschienen in: *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung IFRS 17(6)*, S. 249-255, 2017.

Zusammenfassung

Mit dem „(Lifetime) Expected Credit Loss“ (LECL) verlangt der Rechnungslegungsstandard IFRS 9 ab dem Jahr 2018 einen komplexen Wertminderungsmaßstab, der sich stärker an dem finanzwirtschaftlichen Risikomanagement von Banken orientiert, dabei allerdings insbesondere zeitpunktsabhängig und zukunftsorientiert ausgerichtet werden soll. Die Diskussion dieser neuartigen Sichtweise bei der Verlustantizipation wird in diesem Beitrag sowohl durch die konzeptionelle Operationalisierung der möglichen Modellparameter, als auch durch eine intensive Auseinandersetzung mit den Anwendungsvoraussetzungen und -möglichkeiten geführt. Im Resultat können erhebliche Schwierigkeiten zwischen den Ansprüchen des Standards und den verfügbaren Methoden zur Umsetzung des LECL herausgestellt werden.

Inhaltsverzeichnis

1	Einleitung	45
2	(Lifetime) Expected Credit Losses nach IFRS 9 – Ein Überblick	46
2.1	Grundstruktur des Wertminderungsmodells	46
2.2	Modellierung des (Lifetime) Expected Credit Loss	48
3	Operationalisierung der Parameter des (Lifetime) Expected Credit Loss	49
3.1	Grundsätzliche Überlegungen	49
3.2	Probability of Default	51
3.3	Loss Given Default	54
3.4	Exposure at Default	54
3.5	Vergleichende Würdigung der Ansätze	55
4	Fazit	57

*Dipl.-Math. oec. Tobias Filusch, Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, ifG Marburg, Am Plan 2, D-35032 Marburg.

†Prof. Dr. Sascha H. Mölls, Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, Professur für Allgemeine Betriebswirtschaftslehre und Rechnungslegung, Barfüßertor 2, D-35032 Marburg.

1 Einleitung

Im Zuge der Subprime- und der sich anschließenden Finanzmarktkrise (Bartmann et al., 2009, S. 127 f.) ist die stichtagbezogene Bilanzierung von Finanzaktiva nach dem ausschließlich retrospektiven Incurred Loss-Modell gemäß der kapitalmarktorientierten IFRS-Rechnungslegung heftig kritisiert worden (Schildbach, 2009, S. 592 f.; FCAG, 2009, S. 4 f.; FSF, 2009, S. 1). Neben der Komplexität und der sich im Rahmen der praktischen Umsetzung zwangsläufig ergebenden Subjektivität der Regelungen des einschlägigen IAS 39 (FCAG, 2009, S. 4 f.; FSF, 2009, S. 4 f.) geht es in der Kritik vor allem um die prozyklische Mechanik des Incurred Loss-Modells, das in konjunkturellen Hochphasen zu hohe Gewinne ausweist und in Abschwüngen anfänglich geringe, jedoch in der Folge massive Abschreibungen verursacht. Entsprechend finden sich Bezeichnungen wie „too little too late“ und „cliff effect“ (Henselmann et al., 2014, S. 355; Große, 2011, S. 419 f.; Grünberger, 2013, S. 195 f.), die die Forderung nach einer Bewertungskonzeption untermauern, die über die Risikoposition von Finanzaktiva frühzeitig informiert und dadurch den zukunftsorientierten Informationsinteressen der Adressaten eher nachkommen soll. Der Standardsetzer IASB hat in der Folge die skizzierten Überlegungen aufgegriffen und in dem im Juli 2014 verabschiedeten Nachfolgestandard IFRS 9 mit dem „(Lifetime) Expected Credit Loss“ (LECL) ab dem Jahr 2018 einen prospektiven Wertminderungsmaßstab für alle diejenigen Finanzinstrumente geschaffen, die zu fortgeführten Anschaffungskosten oder erfolgsneutral zum beizulegenden Wert zu bewerten sind (IASB, 2014). Entsprechend sind in diesen Fällen die zukünftig erwarteten Zahlungsausfälle unter Einbeziehung aller relevanten unternehmensinternen und -externen Informationen zu antizipieren und zu jedem Bewertungszeitpunkt neu zu schätzen, was den neuen Mechanismus zur Verlustantizipation nach IFRS 9 zumindest prinzipiell in die Logik des aus dem HGB a.F. bekannten „nahen Zukunftswert“ stellt (§253 Abs. 3 und Abs. 4 Satz 2 HGB a.F.).¹

Obwohl im Zuge des Entstehungsprozesses von IFRS 9 für das prospektive Wertminderungskonzept verschiedene Modelle vorgeschlagen und gewürdigt worden sind, hat die Diskussion sowohl die konzeptionelle Operationalisierung der Modellparameter als auch eine intensive Auseinandersetzung mit den Anwendungsvoraussetzungen und etwaigen -konsequenzen weitestgehend vernachlässigt.² Der Bewertungsmaßstab des LECL gleicht daher nach wie vor einer wohlklingenden Phrase, die die rechnungslegungsbezogene Bewertung zwar in die Nähe des methodisch anspruchsvollen Risikomanagements von Banken rückt (Schröder, 2015, S. 245), andererseits aber auch Gefahr läuft, dass der Ansatz eine Genauigkeit suggeriert, die sich in der praktischen Anwendung kaum erreichen lassen dürfte. An dieser Stelle setzt der vorliegende Beitrag an: Es sollen – ausgehend von einer kurzen Vorstellung des „neuen“ Wertminderungskonzepts – verschiedene Ansätze für die Bestimmung der erforderlichen Parameter des LECL vorgestellt und hinsichtlich ihrer Umsetzbarkeit untersucht werden, um einerseits eine realistische Umsetzung des Bewertungsmaßstabs zu skizzieren und andererseits die Diskussion um den IFRS 9 auf eine abgesicherte Grundlage zu stellen.

Im folgenden Abschnitt 2 wird zunächst das Wertminderungskonzept gemäß IFRS 9 kompakt in seiner Grundstruktur sowie den sich daraus ergebenden Bewertungskonsequenzen vorge-

¹Vgl. zu diesen Überlegungen auch Krag und Mölls (2017)

²Vgl. zu Ansätzen, diese Lücke zu schließen, u.a. Grünberger (2013, S. 9 f.); Börstler und Mölls (2015a); Börstler und Mölls (2015b); Krag und Mölls (2017).

stellt und in Anlehnung an das Vorgehen in der (bank-)betrieblichen Praxis grundlegend in die Bestimmung des „erwarteten Verlustes“ eingeführt. Die damit geschaffenen Grundlagen ermöglichen es, in Abschnitt 3 die für die Verlustantizipation erforderlichen Modellparameter im Kontext unterschiedlicher Verfahrensweisen kritisch mit Blick auf Anwendungsmöglichkeiten und -probleme zu diskutieren und zu würdigen. Ein Fazit sowie ein Ausblick beschließen den Beitrag.

2 (Lifetime) Expected Credit Losses nach IFRS 9 – Ein Überblick

2.1 Grundstruktur des Wertminderungsmodells

Mit dem neuen Wertminderungsmodell des IFRS 9 verfolgt das IASB sowohl das – bereits erwähnte – Ziel einer im Vergleich zum Incurred Loss-Modell frühen Erfassung erwarteter Kreditausfälle als auch eine regelmäßige Überprüfung der Entwicklung der Kreditqualität insofern, als der Umfang der Verlustantizipation i.S. einer Wertberichtigung bzw. Rückstellungsbildung von signifikanten Veränderungen (Verschlechterungen sowie ggf. Verbesserungen) des Ausfallrisikos abhängt. Eine derartig ausdifferenzierte Verlustantizipation ist nach den Vorgaben des IFRS 9 auf Schuldinstrumente, deren Folgebewertung bei Erfüllung der vorgegebenen Kriterien³ zu fortgeführten Anschaffungskosten oder erfolgsneutral zum beizulegenden Wert erfolgt (z.B. Bankeinlagen, Kredite, Schuldverschreibungen und Forderungen aus Lieferungen und Leistungen), sowie auf Leasingforderungen, aktive Vertragsposten, Kreditzusagen und nicht erfolgswirksam zum beizulegenden Zeitwert bewertete Finanzgarantien anzuwenden. Ausgangspunkt des Bilanzierungskonzepts ist dann zunächst die Erkenntnis, dass der Anschaffungspreis P_0 eines Finanzinstruments die Summe der zum Effektivzinssatz auf den betrachteten Zeitpunkt diskontierten zukünftigen Zahlungen widerspiegelt. Der Effektivzinssatz zum Zugangszeitpunkt i_{eff} ergibt sich dabei, sofern nicht ein „vereinfachtes Verfahren“⁴ zur Anwendung kommt oder finanzielle Vermögenswerte, die bereits bei Erwerb oder Ausreichung wertgemindert sind⁵, zu bewerten sind, unmittelbar aus den vertraglich zugesicherten Zahlungen (Rückflüssen) $Z_{v,t}$ abzüglich direkt zuordenbarer Transaktionskosten TK (z.B. Gebühren). Zur Ermittlung des Effektivzinssatzes muss die Preisgleichung

$$P_0 = \sum_{t=1}^T \frac{E(Z_{v,t})}{(1 + i_{eff})^t} - TK \quad (1)$$

³Voraussetzung ist die Erfüllung der Kriterien (1.) „Geschäftsmodell“ und (2.) „Zahlungsstrom“. Vgl. IFRS 9.4.1.2 und 4.1.2A (IASB, 2014) sowie Märkl und Schaber (2010, S. 65).

⁴Im „vereinfachten Verfahren“ müssen Unternehmen die Änderungen des Kreditrisikos nicht dezidiert erfassen und können stattdessen sowohl bei der erstmaligen Erfassung als auch im Rahmen der Folgebewertung eine Risikovorsorge i.H. des LECL vornehmen. Der Anwendungsbereich dieser Vereinfachung erstreckt sich auf Forderungen aus Lieferungen und Leistungen oder aktive Vertragsposten, die aus Transaktionen resultieren, die in den Anwendungsbereich von IFRS 15 fallen, und die keine wesentliche Finanzierungskomponente enthalten oder die Ausnahmeregelung des IFRS 15 für Verträge mit einer maximal einjährigen Laufzeit in Anspruch genommen wird. Vgl. IFRS 9.5.5.15 (IASB, 2014) sowie für eine aktuelle Kommentierung der IFRS 9 EY (2015, S. 23 ff.).

⁵Im Falle finanzieller Vermögenswerte, die bereits bei Erwerb oder Ausreichung wertgemindert sind, wird ein adjustierter Effektivzinssatz verwendet, der die im Zeitpunkt der Erstbewertung über die Gesamtlaufzeit erwarteten Kreditausfälle widerspiegelt (EY, 2015, S.24 ff.).

– ggf. mithilfe numerischer Verfahren – nach i_{eff} aufgelöst werden. Ein so berechneter Effektivzinssatz i_{eff} wird wegen der Existenz von Transaktionskosten regelmäßig etwas kleiner als der vertragliche Zins i_v sein und stellt im allgemeinen Wertminderungsmodell des IFRS 9 sowohl die Grundlage zur Berechnung der Zinserträge als auch die Basis für die Bewertung von Wertberichtigungen i.S. diskontierter erwarteter Kreditausfälle dar (vgl. IFRS 9.B5.4.1-7).

Mit Blick auf die Veränderung des Kreditrisikos müssen Unternehmen zu jedem Abschlussstichtag auf der Basis einer Zuordnung der Instrumente zu einer von drei Stufen („Three Bucket Approach“) eine Risikovorsorge für zukünftige Kreditausfälle entweder auf der Basis eines „12-Monats-ECL“ oder eines LECL erfassen. In Stufe 1 liegt der Zeithorizont für ein mögliches Verlustszenario bei 12 Monaten. Entsprechend ist eine Verlusterwartung in Höhe der Kreditausfälle zu erfassen, deren Eintritt in den nächsten 12 Monaten erwartet wird. Eine solche „12-Monats-ECL“ stellt einen Teil des LECL dar und umfasst sämtliche künftigen Zahlungen, die aus in den nächsten 12 Monaten eingetretenen Verlustereignissen resultieren. Stufe 1 werden Kredite zugeordnet, bei deren Kreditqualität keine signifikante Verschlechterung zu beobachten war oder die als Investment Grade gelten, also ein geringes Ausfallrisiko aufweisen. Für Kredite abseits des Investment Grade wird der Übergang zur Stufe 2 durch eine signifikante⁶ Erhöhung des Kreditrisikos ausgelöst. Dieser Wechsel in Stufe 2 führt zur Anwendung des LECL, was den Zeithorizont für Ausfallszenarien auf die gesamte Restlaufzeit verlängert und infolgedessen die zu erfassende Risikovorsorge erhöht. Zinserträge und Wertberichtigungen werden auf den Stufen 1 und 2 unter Verwendung des i_{eff} getrennt voneinander erfasst. Basis für die Zinserträge ist jeweils der Bruttobuchwert ohne Abzug der Risikovorsorge. Wertberichtigungen ergeben sich aus den mit i_{eff} diskontierten erwarteten Kreditausfällen.⁷ Für Stufe 3, für die ebenfalls der LECL anzuwenden ist, müssen – in Analogie zu den Wertminderungskriterien des IAS 39 – objektive Hinweise auf eine Wertminderung vorliegen (vgl. IFRS 9.IN6). In diesem Fall sind in den Folgeperioden die Zinserträge unter Verwendung des i_{eff} anstatt auf der Basis des Bruttobuchwerts auf der Grundlage der fortgeführten Anschaffungskosten (Nettobuchwert = Bruttobuchwert – Risikovorsorge) zu berechnen (vgl. IFRS 9.IN6). Sollte sich schließlich der finanzielle Vermögenswert infolge der Wertminderung nicht mehr realisieren lassen, ergibt sich ein Ausbuchungsereignis durch eine unmittelbare Abschreibung des Bruttobuchwerts auf null.

Mit der Anwendung des LECL wird der mögliche Verlustzeitpunkt nicht mehr nur auf die nahe Zukunft beschränkt, wodurch sich eine sehr große Spannweite möglicher Ausfallszenarien oder -zeitpunkte ergibt. Ein früher Totalausfall würde zu sehr großen Verlusten führen, wohingegen ein später Ausfall – zumindest im exemplarischen Fall nicht-endfälliger Schuldinstrumente – eher geringe Verluste impliziert, da die Tilgung zu diesem Zeitpunkt schon weit fortgeschritten sein dürfte. Innerhalb dieser Extremfälle bewegen sich die Möglichkeiten, die jeweils mit bestimmten Wahrscheinlichkeiten eintreten. Darüber hinaus erhöhen sich die Wertminderungsbeträge beim Übergang in Stufe 2 durch zwei simultane Effekte: Zum einen steigen die Wahrscheinlichkeitsgewichte zukünftiger Ausfallszenarien für Kredite nach der signifikant-

⁶Vgl. zu einer Übersicht über Indikatoren für eine signifikante Änderung des Kreditrisikos IFRS 9.B5.5.17 (IASB, 2014) sowie EY (2015, S. 50 ff.). Vgl. zu einer inhaltlichen Bestimmung von „Signifikanz“ IFRS 9.B5.5.9-10 (IASB, 2014) sowie EY (2015, S. 56 ff.).

⁷Erwartete Ausfälle (=Wertminderungen) sind grundsätzlich mit dem Effektivzinssatz zu diskontieren. Sofern aber die Bestimmung des Effektivzinssatzes mit Schwierigkeiten verbunden ist (z.B. bei offenen Portfolios), kann ein Näherungswert verwendet werden (IASB, 2014, IFRS 9.B5.5.47).

ten Erhöhung des Kreditrisikos. Zum anderen erhöht die Ausdehnung der Betrachtungsperiode auf die verbleibende Restlaufzeit die Anzahl der möglichen Ausfallszenarien.

2.2 Modellierung des (Lifetime) Expected Credit Loss

Das Wertminderungskonzept des IFRS 9 greift konzeptionell zwar auf den LECL zurück, verzichtet aber darauf, eine konkrete Berechnungslogik vorzugeben, sodass prinzipiell zahlreiche Umsetzungsfragen offenbleiben und infolgedessen bilanzpolitische Spielräume entstehen. Durch den Rückgriff auf existierende strukturähnliche Konzepte des finanzwirtschaftlichen Risikomanagements sowie deren Anpassung an die spezifischen Vorgaben des IFRS 9⁸ lassen sich diese Spielräume allerdings insofern eingrenzen, als die vorgenommenen Bewertungen auf ein methodisch abgesichertes Fundament gestellt werden können (Grünberger, 2011, S. 29 (36)). In der betrieblichen (Banken-)Praxis wird unter einem Expected Credit Loss (ECL) der Erwartungswert zukünftiger Verluste verstanden, der aus der Ausfallwahrscheinlichkeit (Probability of Default – PD), dem prozentualen Verlust bei Ausfall (Loss Given Default – LGD)⁹ und der Höhe des begebenen Kreditvolumens (Exposure at Default – EAD) besteht und regelmäßig für ein Jahr wie folgt berechnet wird (Henking et al., 2006, S. 23; Rolfes und Kirmße, 2000, S. 647 f.; Hartmann-Wendels et al., 2015, S. 499 f.):¹⁰

$$ECL = PD \cdot LGD \cdot EAD. \quad (2)$$

Für die Übertragung der bisher unterstellten einjährigen Betrachtungsweise auf den nach IFRS 9 geforderten Mehrperiodenfall bietet sich – in Anlehnung an Grünberger (2011, S. 29) – die Schätzung periodenspezifischer ECL für jeden Zeitpunkt des Instruments/Portfolios an. Der LECL entspricht dann dem Barwert der zukünftig erwarteten Verluste. Werden die vorstehenden Überlegungen auf das konkrete Bewertungsmodell des IFRS 9 übertragen, ergibt sich unter Berücksichtigung der unterschiedlichen Erfassung von Abwertungsbeträgen in den Stufen S_1 , S_2 und S_3 folgende Summenzerlegung, die sich ausgehend von stochastischen Ausfallereignissen innerhalb gewählter Perioden und der vereinfachenden Annahme der Unabhängigkeit der Parameter herleiten lässt als:

$$\begin{aligned} LECL_T &= \sum_{t=1}^T \frac{PD_t \cdot LGD_t \cdot EAD_t}{(1 + i_{eff})^t} \\ &= \underbrace{\frac{PD_1 \cdot LGD_1 \cdot EAD_1}{(1 + i_{eff})}}_{S_1} + \sum_{t=2}^T \underbrace{\frac{PD_t \cdot LGD_t \cdot EAD_t}{(1 + i_{eff})^t}}_{S_2/S_3} \end{aligned} \quad (3)$$

mit i_{eff} = Effektivzinssatz (IASB, 2014, IFRS 9.5.4.1-2). Für Instrumente in Stufe 1 (S_1) wird nach dem zweiten Teil von Formel (3) ein Abschreibungsbetrag in Höhe eines einperiodigen

⁸Die Übernahme der bestehenden Logik des finanzwirtschaftlichen Risikomanagements für Zwecke der Rechnungslegung nach IFRS 9 erfordert Anpassungen und Erweiterungen insbesondere deshalb, da die Modellierung gemäß IFRS 9 einer prospektiven und konjunkturabhängigen Point-in-Time (PIT)-Sicht entsprechen soll, wohingegen klassische Kreditrisikomodelle durchschnittliche- oder konservativ-gewählte Downturn-Parameter, d.h. Größen, die in konjunkturell schwachen Phasen anfallen bzw. geschätzt werden, verwenden. Das zuletzt genannte Vorgehen wird häufig als Through-the-Cycle (TTC)-Ansatz bezeichnet.

⁹Der LGD ist abhängig von der Verwertbarkeit der Aktiva (Insolvenzmasse) sowie der Vorrangigkeit der Gläubigerposition, sodass diese Größe auch über eine Rückgewinnungsquote (RR) ausgedrückt werden kann (Schuermann, 2005, S. 5 (11 f.); Pawlowski und Burmester, 2001, S. 349 f.).

¹⁰Vgl. zu einer ähnlichen Herleitung Krag und Mölls (2017).

ECL gebildet. Sobald die in IFRS 9 definierten, aber bisher nur wenig präzisierten Auslösetatbestände erfüllt sind, erfolgt der Übergang in Stufe 2 (S_2) oder Stufe 3 (S_3), wodurch sich die Abwertung um den hinteren Summenterm zum LECL vergrößert (EY, 2015, S. 47 ff.).

Die in diesem Abschnitt skizzierten Komponenten des LECL bedürfen im Rahmen der vorgegebenen Bewertungsstruktur sowie der Forderung des Standardsetzers, auch zukünftige Informationen in die Schätzung einfließen zu lassen, einer möglichst objektivierten und damit nachvollziehbaren Umsetzung, um über verlässliche marktorientierte Bewertungsansätze die Entscheidungsnützlichkeit einer kapitalmarktorientierten Rechnungslegung steigern zu können. Die nachfolgenden Überlegungen greifen diese Anforderung(en) auf und analysieren – im Anschluss an einige grundsätzliche Überlegungen zur Operationalisierung der erforderlichen Parameter – im Sinne einer kapitalmarkt- und zukunftsorientierten Umsetzung des LECL-Modells mögliche Verfahrensweisen¹¹, mit denen die beschriebenen Schätzspielräume eingegrenzt und prinzipiell nachvollziehbare Ergebnisse generiert werden können.

3 Operationalisierung der Parameter des (Lifetime) Expected Credit Loss

3.1 Grundsätzliche Überlegungen

Die zukunftsbezogene Berechnung des LECL wird seitens des Standardsetzers nicht mittels einer konkreten Berechnungsformel vorgegeben. Vielmehr wird lediglich generell auf die Modellierung über einen wahrscheinlichkeitsgewichteten Betrag möglicher Zahlungsausfälle verwiesen und zugleich betont, dass auch Zahlungsverzögerungen als Ausfälle aufzufassen und je nach Dauer der Verzögerung entsprechend gewichtet einzubeziehen seien (IASB, 2014, IFRS 9.B5.5.28-35). Die Wahl und die Anzahl möglicher Szenarien, die in die Betrachtung Einzug halten sollen, wird nicht vorgegeben, sondern lediglich einschränkend angemerkt, dass in vielen Bewertungssituationen eine detaillierte Abgrenzung von Szenarien nicht erforderlich sein sollte. So kann z.B. bei einer großen Anzahl homogener Kredite der durchschnittliche Ausfall als Schätzer für die Erwartung verwendet werden, wohingegen in weniger einheitlichen Konstellationen die Schätzung von Ausfallhöhe und -zeitpunkt sowie der zugehörigen Wahrscheinlichkeiten erforderlich sind (IASB, 2014, IFRS 9.B5.5.42). Die hinter dieser Vorgabe stehende konzeptionelle Überlegung ist dabei einfacher nachzuvollziehen als sich die praktische Umsetzung gestalten dürfte: Zunächst wird angenommen, dass ein Ausfall in einem zukünftigen Zeitraum mit einer gewissen Wahrscheinlichkeit auftritt. Sofern dies der Fall ist, ergibt sich der ausfallende Betrag durch die zu diesem Zeitpunkt noch ausstehende vertraglich festgelegte Forderungshöhe als Barwert abzüglich eines gewissen Betrags, der aus der Verlustquote resultiert. Die Verlustquote gibt also den Anteil an der verbleibenden Forderungshöhe an, der tatsächlich verloren ist, also auch nicht durch die Verwertung unter Umständen vorhandener Ansprüche wie etwa vorliegenden Sicherheiten geltend gemacht werden kann. Ein Ausfallszenario bezieht sich dann stets auf eine konkrete zukünftige Zeitperiode (z.B. ein bestimmtes Kalenderjahr). Für solche Zeiträume müssen die im vorherigen Abschnitt vorgestellten Parameter PD (= Ausfallwahrscheinlichkeit), LGD (= prozentualer Verlust bei Ausfall) und EAD (= Höhe des begebenen Kreditvolumens) jeweils bestimmt werden. Mit Blick auf den Para-

¹¹Vgl. zu einer ähnlichen Darstellung und Würdigung der Verfahrensweisen auch Krag und Mölls (2017).

meter PD ist insbesondere die genaue Bedeutung dieser Wahrscheinlichkeit zu beachten.

Während im Hinblick auf die unmittelbar bevorstehende Periode lediglich zwischen „Ausfall“ und „Nicht-Ausfall“ unterschieden werden muss, ist für die sich anschließenden Jahre jeweils das Überleben des Kredits bis in die jeweilige Periode Voraussetzung, um einen Ausfall überhaupt in Betracht ziehen zu können. Hintergrund dieses Umstands ist die häufig in der praktischen Modellierung verwendete, aber nicht zwingend erforderliche Eigenschaft des Ausfalls als absorbierender Zustand, d.h. eines Zustands, der nicht mehr verlassen werden kann (Bluhm et al., 2003, S. 71). In dieser Logik werden wiedergesundete Kredite analog zu einem neuen Kredit aufgefasst und weisen keine Vorgeschichte auf. Für eine abgegrenzte, zukünftige Betrachtungsperiode t besteht die Ausfallwahrscheinlichkeit PD_t dementsprechend aus den Überlebenswahrscheinlichkeiten bis und der Ausfallwahrscheinlichkeit in t .¹² Die Bestimmung einer solchen Ausfallstruktur ist insbesondere im Hinblick auf die von IFRS 9 geforderte Einbeziehung konjunktureller Einflüsse nicht trivial, sondern dürfte Anwender regelmäßig vor umsetzungsorientierte Probleme stellen. Das Grundniveau und der Verlauf der PD-Struktur werden dabei maßgeblich durch die initiale Bonitätsbewertung determiniert, wobei die vielfältigen Möglichkeiten der Bonitätsverbesserung oder -verschlechterung allerdings nicht unberücksichtigt bleiben dürfen. Ein besonders guter Kredit hat in naher Zukunft eine geringe Ausfallwahrscheinlichkeit, die im weiteren Verlauf ansteigt, da diese Schuldner tendenziell nur schlechter, aber nicht besser werden können. Im Umkehrschluss dürften sich die im Rahmen der initialen Bewertung als schlecht eingestuften Kredite entweder verbessern oder ausfallen, wobei der zweite Fall gleichbedeutend mit einem Ausschluss aus der Betrachtung ist. Aus den vorstehenden Überlegungen folgt ein abnehmender PD-Verlauf mit einem Grundniveau, das oberhalb des Niveaus der guten Bonitätsklasse liegen sollte. Schlussendlich werden die Verläufe, unabhängig von der ursprünglichen Bonität, aber in einer gemeinsamen langfristigen PD^* münden. Ein solcher Verlauf lässt sich damit begründen, dass es mit den aktuellen Informationen kaum möglich sein dürfte, die genaue Qualität eines nicht ausgefallenen Kredits in weiter Zukunft einzuschätzen. Folgerichtig sollte die Einschätzung daher auf eine sehr durchschnittliche Kreditqualität hinauslaufen. Für die Einbeziehung konjunktureller Einflüsse heißt das entsprechend, dass eine Berücksichtigung nur in aktuelleren PD sinnvoll sein kann. Zur Illustration der vorstehenden Überlegungen zeigt Abbildung 1 mögliche PD-Strukturverläufe sowie die PD^* als langfristigen Trend. Konjunkturelle Einflüsse verschieben die Kurven bzw. führen zu Schwankungen des Verlaufes, wobei die langfristige PD^* von diesen Einflüssen allerdings (weitestgehend) unabhängig ist und für ein bestimmtes Segment an Krediten konstant bleiben dürfte.

Zur generellen Veranschaulichung der Verlustquote LGD und der Forderungshöhe bei Verlust EAD sollen schließlich zwei verschiedene Ausfallszenarien während derselben, zuvor gewählten, Periode t angenommen werden. Im ersten Szenario erfolgt der Ausfall (A1) gleich zu Beginn und im zweiten Szenario tritt der Ausfall (A2) kurz vor Ende der Periode ein. Grundsätzlich kann davon ausgegangen werden, dass sich die realisierte LGD und der EAD dieser beiden Szenarien unterscheiden. Zur Ermittlung von Werten für LGD_t und EAD_t , die für die gesamte Periode t repräsentativ sind, muss jeweils derjenige Erwartungswert gesucht werden, der sämtliche Szenarien inklusive A1 und A2 einbezieht. Die Wahl der Periodenlänge steht dem-

¹² An dieser Stelle wird häufig von einer „unbedingten“ Ausfallwahrscheinlichkeit gesprochen, um diese von dem auf das Überleben des Kredits bedingten Pendant abzugrenzen (Grünberger, 2011, S. 29 (30)).

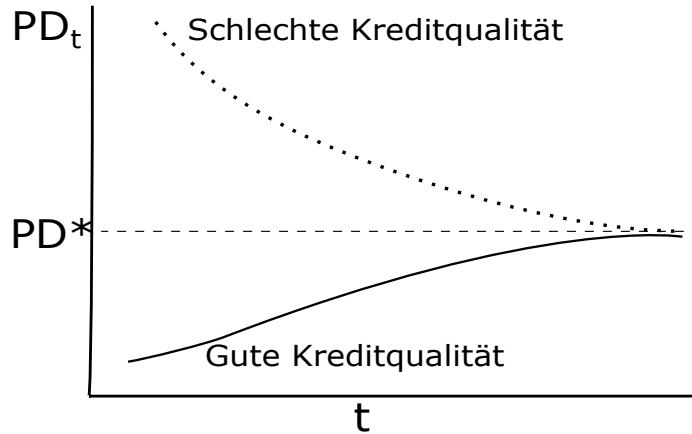


Abbildung 1: Hypothetische Ausfallstruktur eines initial guten (geringe PD, durchgezogene Linie) und eines schlechten (hohe PD, gestrichelte Linie) Kredites (Quelle: Eigene Darstellung)

nach folgendem trade-off gegenüber: Eine zu große Periodenlänge verzerrt die Präzision der drei Kernparameter aufgrund der Mittelung der vorhandenen Schwankungen, wohingegen es sehr kurzen Periodenlängen an aussagekräftigen Mechanismen zur Einschätzung der Kernparameter mangelt und diese dann vermutlich zwangsläufig sehr ähnlich gewählt werden müssen. Vor diesem Hintergrund stellt die Modellierung in Jahresschritten einen guten Kompromiss dar. Eine Orientierung an den Stichtagen des erwarteten Zahlungsstroms und infolgedessen an einer nicht zwingend konstanten Periodenlänge ist allerdings denkbar, sofern es möglich ist, unterjährige Ausfallwahrscheinlichkeiten abzuleiten und diese an die gewählten Zeiträume anzupassen.

3.2 Probability of Default

Verfahren zur Bestimmung der Ausfallstrukturkurve lassen sich bezogen auf die Art der in die Berechnung eingehenden Datengrundlagen in retro- und prospektive sowie unternehmensinterne und -externe Verfahren einteilen (Bluhm et al., 2003, S. 183; Hayden und Porath, 2011, S. 1 f.), die jeweils auf historische Daten (Scorings/Ratings) (Hayden und Porath, 2011, S. 1 f.; Everling, 1991, S. 24 ff.), modelltheoretische Überlegungen (Strukturmodelle¹³) oder marktorientierte (sogenannte implizite) Verfahren zurückgreifen (Bluhm et al., 2003, S. 205 f.). Eine

¹³Vgl. für eine Einteilung modellgestützter Verfahren Duffie und Singleton (2003, S. 53 ff.).

breite Anwendung in der Praxis finden Scorings¹⁴ und Ratings (Everling, 1991, S. 29 ff.). Beide Verfahren ermitteln die Bonität eines Schuldners, wobei erste individuell (durch das bewertende Institut) ausgestaltbar (Löffler und Posch, 2011, S. 1 f.; Hayden und Porath, 2011, S. 4 ff.), zweite hingegen öffentlich verfügbar sind, jedoch auf einer im Detail unbekanntem Kriterienevaluation beruhen (Everling, 1991, S. 35 f (114, 150 f.); Serfling et al., 2007, S. 717). Im Vertrauen auf ihre Zuverlässigkeit sowie mit dem Ziel einer höheren Präzision existieren in der Bankenpraxis Ansätze zur Integration von Ratings in bestehende bankinterne Scoringssysteme.¹⁵ Im Zeitablauf ermöglichen beide Verfahren die Messung einer Bonitätsveränderung mittels Migrationsmatrizen, die in einer gängigen praktischen Interpretation als empirische Übergangswahrscheinlichkeiten zwischen Bonitätsklassen, die den Ausfall mit einschließen, verstanden werden. Die Einträge der Migrationsmatrix beschreiben für alle möglichen Klassenänderungen während eines Zeitraums die zugehörige Wahrscheinlichkeit dieser Ereignisse, wobei sich der direkte Ausfall in der betrachteten Periode ebenfalls darunter befindet und ablesen lässt (Engelmann und Ermakov, 2011, S. 105 f; Duffie und Singleton, 2003, S. 86).¹⁶ Für den konkreten Aufbau einer gesamten Ausfallstruktur wird schließlich eine Folge von Migrationsmatrizen M_t benötigt, die jeweils die betrachteten und sich aneinander anschließenden Übergangszeiträume t beschreiben. Für die konkrete Ermittlung der PD kann unter der Annahme der Stationarität ausgehend vom Bewertungszeitpunkt und den dort vorliegenden Informationen eine Matrix M_1 durchgehend für alle weiteren Zeiträume beibehalten werden. Über einen mehrere Einzelperioden umfassenden Zeitraum T ergibt sich dann die kumulative Migrationsmatrix basierend auf der Matrizenmultiplikation durch (4):

$$\bar{M}_T = (M_1)^T. \quad (4)$$

Eine auf diesen Überlegungen basierende Ausfallstruktur enthält mit Blick auf den Einfluss von Konjunkturschwankungen allerdings lediglich die durch M_1 vorgegebene Entwicklung, sodass Schwankungen in einzelnen Perioden mit diesem Vorgehen nicht dezidiert abgebildet werden. Im Sinne einer Erweiterung lassen sich derartige Veränderungen unter Aufgabe der Annahme der Stationarität aber über die Wahl mehrerer sich unterscheidender Matrizen M_t berücksichtigen. Entsprechend kann bei diesem Vorgehen für jeden Übergangshorizont eine dort jeweils repräsentative und durch die dann angenommene Konjunktur (mit-)bestimmte Migrationsmatrix erstellt werden. Die kumulative Migrationsmatrix mit wechselndem Periodeneinfluss ergibt sich dann ebenfalls durch Matrizenmultiplikation als (5):

$$\bar{M}_T = \prod_{t=1}^T M_t. \quad (5)$$

Mit den beiden vorgestellten Vorgehensweisen lassen sich ausgehend vom Bewertungszeitpunkt $t = 0$ die kumulativen Ausfallwahrscheinlichkeiten jeder Bonitätsklasse bis zum Ende jeder

¹⁴Bankinterne Scorings (sogenannter IRB-Ansatz) nutzen verschiedene Informationsquellen, um Kreditnehmer bei Abschluss eines Vertrags durch quantitative und qualitative Faktoren bezüglich ihrer Bonität einzuschätzen. Mithilfe ökonomischer, familiärer und weiterer persönlicher Kriterien, die je nach gewählter Relevanz mit verschiedenen Gewichten einbezogen werden, lassen sich Scores zunächst zu einem Gesamturteil verdichten. Auf dieser Basis wird die Bonität des Schuldners eingeordnet und kann schließlich in eine prognostizierte PD überführt werden (Hartmann-Wendels et al., 2015, S. 517 f.).

¹⁵Vgl. für ein ausführliches Beispiel Hartmann und Schwarzhaupt (2007, S. 32 ff.).

¹⁶Konzeptionell ist dies möglich, da Standard & Poor's sowie Fitch ihre Ratings auf Grundlage der PD erstellen. Moody's bewertet Schuldner anhand des Expected Loss (S&P, 2017).

Einzelperiode t direkt aus den kumulativen Migrationsmatrizen \bar{M}_t ablesen.¹⁷ Die Differenzen der kumulativen Ausfallwahrscheinlichkeiten von aufeinanderfolgenden Intervallen sind genau die PD, die für die LECL-Formel (3) benötigt werden.¹⁸

Eine zweite modelltheoretische Ermittlungsmöglichkeit für die PD stellen Strukturmodelle dar, die häufig auch als Insolvenzprognoseverfahren in der Unternehmensanalyse verwendet werden (Merton, 1974, S. 449). Ein Ausfallereignis liegt in diesen Modellen – in Analogie zur buchmäßigen Überschuldung – immer dann vor, wenn die zu Marktpreisen bewerteten Aktiva des Unternehmens unter die bestehenden (marktbewerteten) Zahlungsverpflichtungen fallen (Duffie und Singleton, 2003, S. 43 (53 ff.)). Die Entwicklung des Werts sowohl der Vermögensgegenstände als auch der Zahlungsverpflichtungen (Schulden) werden in diesen Ansätzen typischerweise über stochastische Prozesse abgebildet, die den „wahren“ (d.h. den vermuteten) Wertverlauf bestmöglich approximieren sollen und regelmäßig als eine „Geometrische Brown’sche Bewegung“ erfasst werden. Für die konkrete Bewertung sind dann weitere Parameter wie u.a. der aktuelle (Markt-)Preis, die Volatilität sowie etwaige Trends (Drift) zu schätzen.¹⁹ Mit einer derartigen Bewertung lässt sich – unter den getroffenen (neoklassischen) Annahmen – im Ergebnis das Ausfallereignis mathematisch beschreiben als der erste Schnittpunkt des Vermögensprozesses mit dem Prozess der Zahlungsverpflichtungen innerhalb eines betrachteten Intervalls $(0, t]$. Daraus lässt sich die zugehörige kumulative Ausfallwahrscheinlichkeit ($kumPD_t$) direkt berechnen (Löffler und Posch, 2011, S. 27 ff.) sowie in Analogie zum beschriebenen Vorgehen die PD der LECL-Formel ableiten. Die gesuchte PD für die Periode t ergibt sich wiederum als: $PD_t = kumPD_t - kumPD_{t-1}$.

Die dritte Ermittlungsmöglichkeit zur PD-Bestimmung umfasst die Nutzung von geeigneten Credit Spreads und verwendet hierfür beobachtbare Kurse von Anleihen oder Credit Default Swaps (CDS). Die Transformation beobachtbarer Marktpreise in sogenannte implizite (risikoneutralisierte) Ausfallwahrscheinlichkeiten entspricht der Darstellung eingepreister Kreditausfallwahrscheinlichkeiten (Bluhm et al., 2003, S. 205 f.). In der praktischen Ermittlung werden jeweils für Unternehmensanleihen und CDS-Notierungen folgende Approximationsformeln verwendet (Hull et al., 2005, S. 2; Löffler und Posch, 2011, S. 221 ff.):

$$PD_{UA} = \frac{i_{UA} - i_{rf}}{1 - RR} \text{ und } PD_{CDS} = \frac{PCDS}{1 - RR}. \quad (6)$$

Für die Ausfallstrukturkurvenermittlung gemäß (6) bedarf es mehrerer Anleihen mit unterschiedlichen Laufzeiten sowie fristenkongruenter Staatsanleihen bzw. CDS-Kontrakte mit unterschiedlichen Laufzeitenden. Alle Zwischenperioden müssen interpoliert werden. Implizite Ermittlungsverfahren stellen einen Marktbewertungszugang dar, dessen Funktionsfähigkeit vom aktiven Handel der herangezogenen Unternehmens- und Staatsanleihen oder CDS abhängig ist. Neben der nicht zwangsläufig gegebenen Existenz von verschiedenen (fristenkongruenten) Finanzinstrumenten stellt die Marktliquidität einen zusätzlichen beeinträchtigenden und verzerrenden Faktor dar²⁰, den obige Bewertungsformeln allesamt über die Nichtberücksichtigung einer Liquiditätsprämie ignorieren.

¹⁷Diese stehen in der Spalte, die den Übergang in den Ausfallzustand beschreibt.

¹⁸ $PD_t = kumPD_t - kumPD_{t-1}$ ist dann die gesuchte PD für die Periode t (Grünberger, 2011, S. 30).

¹⁹Vgl. Black-Scholes Modell in Hull (2015, S. 403 ff.).

²⁰Die Liquiditätsprämie ist schwierig zu bestimmen (Hull et al., 2005, S. 6 f.).

3.3 Loss Given Default

Eine intertemporale Modellierung des LGD, wie sie bei der PD vorgenommen wird, ist nicht möglich, da selbst unter großen Anstrengungen das Schuldnervermögen nur schwierig einsehbar und prognostizierbar ist (Hahn und Reitz, 2011, S. 188). Für die Berechnung des LECL bietet sich deshalb grundsätzlich eine im Zeitablauf konstante Schätzung des LGD an. Gängige LGD-Ermittlungsansätze unterteilen sich in Workout LGD²¹, Market LGD²² und Implied Market LGD²³. Neben diesen Verfahren sind weiterhin Schätzungen der externen Ratingagenturen verfügbar, welche als Indikatoren bzw. Vergleichsmaßstäbe herangezogen werden können.²⁴

Ein praktisch weit verbreitetes Verfahren stellt das parametrische Beta-Transformationsverfahren dar, welches für die empirische LGD-Verteilung die Gestalt einer Beta-Verteilung $B(\alpha, \beta)$ annimmt (Henking et al., 2006, S. 100 f.; Schuermann, 2005, S. 13). Die Parameter α und β lassen sich am einfachsten über den empirischen Mittelwert und die Varianz gemäß

$$\begin{aligned}\alpha &= \frac{E(LGD_t)}{var(LGD_t)} \cdot (E(LGD_t) \cdot (1 - E(LGD_t)) - var(LGD_t)) \text{ und} \\ \beta &= \frac{1 - E(LGD_t)}{var(LGD_t)} \cdot (E(LGD_t) \cdot (1 - E(LGD_t)) - var(LGD_t))\end{aligned}\tag{7}$$

bestimmen, wodurch die Verteilung am gegebenen Datensatz kalibriert wird (Löffler und Posch, 2011, S. 122 f.). Dieser generellen Intuition folgend, lässt sich der anzuwendende LGD über den Modus der kalibrierten Dichtefunktion bestimmen.

Renault und Scaillet (2004) zeigen, dass empirische LGD-Verteilungen in verschiedenen Branchen eine bimodale Form annehmen, wonach entweder relativ hohe oder relativ niedrige LGD realisiert werden, die von der Annahme einer unimodalen Beta-Verteilung nicht adäquat abgebildet werden. Alternativ wird daher die Anwendung einer nicht-parametrischen LGD-Schätzung vorgeschlagen, die eine Glättung der empirischen LGD-Stichprobe vornimmt (Renault und Scaillet (2004, S. 2915 (2917-2918, 2930)). Die Eigenschaften dieser Schätzmethodik umfassen eine einfache, auch auf eine kleine Datenbasis anwendbare Umsetzung. Zudem minimiert sie die Verzerrung an den Intervallgrenzen und ist aufgrund der Glättung der empirischen Stichprobe weniger anfällig für Ausreißer (Renault und Scaillet, 2004, S. 2915 (2917 f)).

3.4 Exposure at Default

Der EAD beschreibt die Höhe des begebenen Kreditvolumens und damit den im möglichen Ausfallzeitpunkt verbleibenden erwarteten Zahlungsstrom ($Z_{v,t}$). Basierend auf den im Kreditvertrag vereinbarten Details erhält dieser Parameter einen stärker deterministischen Charakter und lässt damit nur einen schwachen konjunkturellen Einfluss vermuten. Zusammen mit dem Diskontierungsfaktor auf der Basis des Effektivzinses bildet der EAD in der LECL-Formel (3)

²¹Dieser stellt eine individuelle Schätzung zukünftig einbringbarer Zahlungsströme auf Grundlage subjektiver Beurteilung oder historischer Daten dar (Peter, 2011, S. 165 ff (176 f.)).

²²Hierbei werden beobachtbare Preise börsennotierter Fremdkapitaltitel (insbesonder Anleihen) nach ihrem Ausfallereignis zu ihrem Nominalbetrag ins Verhältnis gesetzt, was einer Marktbewertung gleichkommt (Hamerle et al., 2011, S. 137 f; Gupton und Stein, 2005, S. 19).

²³Vgl. für eine ausführliche Darstellung der Modellierungsansätze Schuermann (2005, S. 6 ff).

²⁴Moody's bietet ein kostenpflichtiges Datenpaket online unter URL: <http://www.moody.com> an.

deshalb auch den risikofernen Teil. Entsprechend kann der (aus Vereinfachungsgründen als konstant angenommene) LGD zur deterministischen Zahlungsstromkomponente hinzugefügt werden, was allein die PD als primäre Risikokomponente des LECL isoliert:

$$LECL_T = \sum_{t=1}^T \underbrace{PD_t}_{\text{Risikokomponente}} \cdot \underbrace{\frac{\overline{LGD} \cdot EAD_t}{(1 + i_{eff})^t}}_{\text{Zahlungsstromkomponente}} \quad (8)$$

3.5 Vergleichende Würdigung der Ansätze

Die Betrachtung der LECL-Modellierungsansätze zeigt, dass lediglich die PD intertemporal modellierbar ist, wohingegen EAD über $Z_{v,t}$ ²⁵ und LGD als Konstante in die Schätzung eingehen sollten. Weiterhin ist deutlich geworden, dass insbesondere die PD als zentrale Entscheidungsvariable für die Wertminderungsmodelle eine ausführliche Kriterienevaluation rechtfertigt. Im Hinblick auf eine frühzeitige Verlusterfassung und prospektive Ausrichtung beanspruchen Ratings konzeptionell eine Zukunftsorientierung, indem sie die zukünftige Bonitätsentwicklung von Schuldnern einschätzen. Entsprechend müssen zur Gewährleistung einer prospektiven Sichtweise die für die Ausfallstrukturkurve erforderlichen Migrationsmatrizen für zukünftige Perioden zunächst ermittelt werden. Demgegenüber erfolgt die Bonitätsbewertung im Rahmen der Strukturmodelle bzw. impliziter Ausfallwahrscheinlichkeiten unmittelbar im Kapitalmarktcontext. Aufgrund der allgemeinen theoretischen Auffassung, dass Marktpreise der Bewertung zukünftiger Zahlungsrückflüsse entsprechen, sowie unter Akzeptanz einer schwachen bzw. mittelstarken Kapitalmarkeffizienz (Copeland et al., 2005, S. 354 f, S. 359 f.; Malkiel, 2003, S. 59) ist die generelle Informationsorientierung prospektiv. Marktpreise spiegeln neue gegenwärtige sowie zukunftsbezogene Informationen unmittelbar wider. Empirische Studien bestätigen einen von Bonitätsveränderungen ausgelösten, beobachtbaren Informationsnutzen an Aktien, Anleihe- und CDS-Märkten (Blanco et al., 2005, S. 2255). Hinsichtlich der Anpassung an neue Informationen ist ein mittels Ratings generierter LECL hingegen tendenziell langsam und wird von einer direkten Marktbewertung dominiert. Empirische Studien weisen auf statistisch signifikante „Antizipationseffekte“ vor Herabstufungen hin, woraus geschlossen werden kann, dass die Informationsverarbeitung am Kapitalmarkt kontinuierlich erfolgt, wohingegen Ratings sequenziell bei Überschreitung kritischer Werte angepasst werden (Norden und Weber, 2004, S. 2813; Steiner und Heinke, 2000, S. 540 (553)).

Ratings und Migrationsmatrizen sind z.T. frei verfügbar. Sofern ein Bankinstitut über ein konsistentes Überleitungssystem zwischen internen Scorings und externen Ratingskalen verfügt, ist eine nachvollziehbare Anwendung der Ratingmodelle auf eine Vielzahl nicht durch Ratings bewerteter Unternehmen möglich, sofern dort eine fundierte, für Migrationsverhalten repräsentative Datenbasis vorliegt. Strukturmodelle sowie die impliziten Verfahren benötigten hingegen Marktpreise, die zwar meist öffentlich verfügbar sind, aber bei Weitem nicht für alle erforderlichen Bewertungen vorliegen. In einer Einschätzung der Komplexitätsdimension kann die Ratingmethodik anhand von Matrizen verschiedener Agenturen durchgeführt werden. Die jeweiligen Ergebnisse können allerdings stark voneinander abweichen und sind nur schwer ineinander überführbar. Aus Einheitlichkeitsgründen empfiehlt sich eine konsequente Anwendung auf Basis der Daten einer Ratingagentur, da dies den Ratingprozess und die

²⁵Hiermit wird die Identität des LECL als Barwert zukünftiger Verluste gewahrt. Außerdem werden nur offene Forderungen zum Bilanzstichtag berücksichtigt (Grünberger, 2011, S. 30).

gewählte Ausfalldefinition eben dieser Agentur akzeptiert und letztendlich Vergleichbarkeit sowie Verständlichkeit seitens der Adressaten ermöglicht. Bei Berücksichtigung solcher Anwendungsfragen sind Ratings leicht verständlich und allgemein akzeptiert (Everling, 1991, S. 309). Strukturmodelle sind in ihrem theoretischen Konstrukt ebenfalls klar verständlich, lassen jedoch auch einige praktische Spielräume in der Operationalisierung aufkommen (Duffie und Singleton, 2003, S. 54; Löffler und Posch, 2011, S. 33 (35)). Die Strukturkurvenbestimmung mittels impliziter Verfahren bedarf einer Reihe von gehandelten Instrumenten mit unterschiedlichen Endfälligkeiten. Die Verwendung von Unternehmensanleihen erfordert überdies fristenkongruente Staatsanleihen. Im Methodenvergleich weisen implizite Verfahren somit den höchsten, Strukturmodelle einen mittleren und ratingorientierte Verfahren den niedrigsten Komplexitätsgrad auf.

Aus einer empirischen Perspektive lässt sich der Wirtschaftszyklus²⁶ als wesentlicher Einflussfaktor auf den erwarteten Verlust identifizieren. Dieser Zusammenhang behält bei der Erweiterung auf den LECL Gültigkeit und kann im Rahmen der drei vorgestellten Ansätze über die PD abgebildet werden (Pearce, 1983, S. 7 (8)). Konjunkturelle Einflüsse in der PD-Modellierung auf der Grundlage von Ratings müssen allerdings nachträglich über einen adäquaten Mechanismus in die verwendeten Migrationsmatrizen der nahen Zukunft eingepreist werden. Langfristig dürften Durchschnittsmatrizen die beste Informationsgrundlage bieten, da diese auf hinreichend großen Datensätzen basieren und die Prognosefähigkeit für die ökonomische Entwicklung ohnehin zeitlich stark begrenzt ist. Problematisch an Strukturmodellen ist im vorliegenden Kontext vor allem, dass eine Anwendung auf allgemeine Kredite aufgrund der Marktwertbetrachtung insofern kritisch sein dürfte, als Vermögenswerte nicht unbedingt liquide gehandelt werden, wodurch die Einschätzung der genauen Dynamiken – gemessen an den zu schätzenden Parametern – erheblich erschwert wird. Weiterhin unterstellen Strukturmodelle einen idealtypischen Ausfallmechanismus, der realiter so nicht zwingend auftritt. Implizite Verfahren sind schließlich sehr stark auf liquide und umfassende Märkte bei Kreditderivaten angewiesen. Mit abnehmender Liquidität verlieren die abgeleiteten Ergebnisse an Bedeutung und erschweren die Anwendung auf den ohnehin eingeschränkten Bereich.

In einer abschließenden Würdigung verbleibt die Schlussfolgerung, dass implizite Verfahren sowie Strukturmodelle zukunftsorientierte Indikatoren mit nachgewiesener Entscheidungsnützlichkeit darstellen, jedoch an ihrer Komplexität, einem kleinen Anwendungsbereich sowie in der betrieblichen Praxis kaum vertretbaren Annahmen scheitern. Ratings bieten einen anwendbaren kapitalmarktorientierten Ansatzpunkt, das informative Zielsystem muss allerdings auch durch eine Reihe von Annahmen bezüglich der zukunftsorientierten Entwicklung gestützt und selbst im Fall grundsätzlicher Konformität mit gängigen Risikomanagementsystemen vermutlich durch Modifikationen an die konkreten Vorgaben des IFRS 9 angepasst werden. Die direkte Anwendung einer ratingorientierten Modellierung bezieht sich zudem nur auf kapitalmarktorientierte Unternehmen mit vorhandenem Rating. Diese stellen jedoch innerhalb einer obligatorischen und flächendeckenden Anwendung von IFRS 9 eine geringe Teilmenge dar, sodass die Nutzung von bankinternen Scorings und heuristischen Durchschnittsverfahren die unmittelbare Folge ist. Der im Rahmen von IFRS 9 ermittelte LECL kann damit überwiegend aus unternehmensinternen Datensätzen und über heuristische Vereinfachungen ermittelt werden. Die Komplexität eines LECL führt bei konsequenter Operationalisierung zu weiter-

²⁶Für einen Überblick vgl. Altman et al. (2005, S. 50 ff).

führenden, methodisch anspruchsvollen Modellen, deren Integration in die Rechnungslegung nicht problemlos möglich sein dürfte.

4 Fazit

Die vorliegende Untersuchung beschäftigt sich mit dem „(Lifetime) Expected Credit Loss“ als Bewertungsmaßstab für Wertminderungen gemäß IFRS 9. Die Standardsetzer nutzen diesen Begriff zur idealistischen Diskussion eines komplexen Abschreibungsmechanismus, welcher an das finanzwirtschaftliche Risikomanagement von Banken angelehnt ist und zukunftsorientierte Informationen berücksichtigen soll. Die gesamte Diskussion um den LECL kann deshalb einerseits als ein weiteres Beispiel für eine zunehmende Orientierung der Rechnungslegung nach IFRS an den Methoden des internen Controllings identifiziert werden, bedeutet im Falle einer großflächigen Umsetzung zugleich aber eine Ausweitung der vorzunehmenden Abschreibungen auf den „nahen Zukunftswert“. Die sich daraus potenziell ergebenden Konsequenzen für die Ertragslage bei den betroffenen Unternehmen rechtfertigt ohne Zweifel eine frühzeitige Beschäftigung mit der Umsetzung dieses auf internationaler Ebene vorgegebenen Wertminderungsmodells.

Die vorliegende Untersuchung greift deshalb die zuvor skizzierten Aspekte dergestalt auf, dass die Vorgaben des IFRS 9 mit den einschlägigen Modellen des finanzwirtschaftlichen Risikomanagements operationalisiert werden sollen. Im Ergebnis zeigt sich, dass der LECL in der praktischen Anwendung (bisher) faktisch eher einer einfachen, heuristischen Umsetzung gleicht, da die verfügbaren Modelle aus verschiedenen Gründen nur bedingt einsatzfähig sind und mitunter einer rechnungslegungsbezogenen Evaluation nicht standhalten. Ungeachtet dessen, ist die wesentliche Neuerung des IFRS 9 allerdings in der imparitätischen Verlustantizipation zu sehen, die das kaufmännische Vorsichtsprinzip stärkt und die mit dem Incurred Loss-Modell einhergehende Kritik einer prozyklischen Verlusterfassung zu reduzieren vermag. Aus konzeptioneller Sicht kann der Vorstoß des IASB hingegen insbesondere im Hinblick auf eine mögliche Reduzierung existierender Spielräume bei der Bilanzierung nicht überzeugen, da der IFRS 9 im Zusammenhang mit dem Wertminderungsmodell keine detaillierten Vorgaben hinsichtlich der konkreten Berechnungslogik des LECL vorsieht und deshalb – abseits der hier vorgestellten Ansätze – grundsätzlich noch eine Reihe alternativer Umsetzungsmöglichkeiten zulässig sind.

Literatur

- Altman, E., Resti, A. und Sironi, A. (2005), Recovery Risk: The Next Challenge in Credit Risk Management, in: Altman, E., Resti, A. und Sironi, A. (Hrsg.), Recovery Risk, Risk Books, London.
- Bartmann, P., Buhl, H. U. und Hertel, M. (2009), Ursachen und Auswirkungen der Subprime-Krise, Informatik Spektrum **32**(2), S. 127–145.
- Blanco, R., Brennan, S. und Marsh, I. (2005), An Empirical Analysis of the Dynamic Relation between Investment-Grade Bonds and Credit Default Swaps, The Journal of Finance: The Journal of the American Finance Association **60**(5), S. 2255–2282.
- Bluhm, C., Overbeck, L. und Wagner, C. (2003), An Introduction to Credit Risk Modeling, Chapman & Hall/ CRC, CRC Press, Boca Raton.
- Börstler, D. und Mölls, S. H. (2015*a*), Modeling Lifetime Expected Credit Losses Under IFRS 9 - Fundamental Questions Answered, Working Paper Series: Accounting & Corporate Governance (11).
- Börstler, D. und Mölls, S. H. (2015*b*), Modeling PD Term Structures - A New Approach Using Scoring Recalibration, Working Paper Series: Accounting & Corporate Governance (12).
- Copeland, T. E., Weston, J. F. und Shastri, K. (2005), Financial Theory and Corporate Policy, 4. Aufl., Pearson, Addison-Wesley.
- Duffie, D. und Singleton, K. J. (2003), Credit Risk: Pricing, Measurement, and Management, Princeton Univ. Press, Princeton, NJ [u.a.].
- Engelmann, B. und Ermakov, K. (2011), Transition Matrices: Properties and Estimation Methods, in: Engelmann, B. und Rauhmeier, R. (Hrsg.), The Basel II Risk Parameters, Springer, Heidelberg [u.a.], S. 103–116.
- Everling, O. (1991), Credit Rating durch internationale Agenturen: Eine Untersuchung zu den Komponenten und instrumentalen Funktionen des Rating, Gabler, Wiesbaden.
- EY (2015), Wertminderung finanzieller Vermögenswerte nach IFRS 9, Ernst & Young, Stuttgart.
- FCAG (2009), Report of the Financial Crisis Advisory Group, Financial Crisis Advisory Group.
URL: <https://www.iasplus.com/en/binary/crunch/0907fcagreport.pdf>
- FSF (2009), Report of the Financial Stability Forum on Addressing Procyclicality in the Financial System, Financial Stability Forum.
URL: http://www.fsb.org/wp-content/uploads/r_0904a.pdf
- Große, J.-V. (2011), Auf der Suche nach dem richtigen Impairmentmodell, IRZ: Zeitschrift für internationale Rechnungslegung **6**(10), S. 419–426.
- Grünberger, D. (2011), Der Lifetime-Expected Loss, IRZ: Zeitschrift für internationale Rechnungslegung **6**(1), S. 29–37.

- Grünberger, D. (2013), Kreditrisiko im IFRS-Abschluss: Handbuch für Bilanzersteller, Prüfer und Analysten, Schäffer-Poeschel, Stuttgart.
- Gupton, G. M. und Stein, R. M. (2005), LossCalc V2: Dynamic Prediction of LGD-Modeling Methodology, Moody's KMV Company.
URL: http://www.defaultrisk.com/_pdf6j4/LCv2_DynamicPredictionOfLGD_fixed.pdf
- Hahn, R. und Reitz, S. (2011), Possibilities of Estimating Exposures, in: Engelmann, B. und Rauhmeier, R. (Hrsg.), The Basel II Risk Parameters, Springer, Heidelberg [u.a.], S. 185–200.
- Hamerle, A., Knapp, M. und Wildenauer, N. (2011), Modelling Loss Given Default, in: Engelmann, B. und Rauhmeier, R. (Hrsg.), The Basel II Risk Parameters, Springer, Heidelberg [u.a.], S. 127–142.
- Hartmann, W. und Schwarzhaupt, O. (2007), Die Bedeutung des Ratings im Bankkreditgeschäft, in: Büschgen, H. E. und Everling, O. (Hrsg.), Handbuch Rating, Gabler, Wiesbaden, S. 25–46.
- Hartmann-Wendels, T., Pfingsten, A. und Weber, M. (2015), Bankbetriebslehre, 6. Aufl., Springer, Berlin [u.a.].
- Hayden, E. und Porath, D. (2011), Statistical Methods to Develop Rating Models, in: Engelmann, B. und Rauhmeier, R. (Hrsg.), The Basel II Risk Parameters, Springer, Heidelberg [u.a.], S. 1–12.
- Henking, A., Bluhm, C. und Fahrmeir, L. (2006), Kreditrisikomessung: Statistische Grundlagen, Methoden und Modellierung, Springer, Berlin [u.a.].
- Henselmann, K., Ditter, D. und Holstein, T. (2014), Too little, too late? Ein empirisch gestützter Beitrag zur Bilanzierung der Kreditrisikoversorge, KoR: Internationale und Kapitalmarktorientierte Rechnungslegung IFRS **14**(7/8), S. 355–362.
- Hull, J. (2015), Optionen, Futures und andere Derivate, 9. Aufl., Pearson Studium, München.
- Hull, J., Predescu, M. und White, A. (2005), Bond Prices, Default Probabilities and Risk Premiums, The Journal of Credit Risk (JCR) **1**(2), S. 53–60.
- IASB (2014), IFRS 9: Financial Instruments, International Financial Accounting Standard Board.
URL: <http://www.ifrs.org>
- Krag, J. und Mölls, S. H. (2017), „Lifetime Expected Credit Losses“ im Rahmen des Impairments nach IFRS 9, in: Kaluza, B., Braun, K. D., Beschorner, H. und Rolfes, B. (Hrsg.), Betriebswirtschaftliche Fragen zu Steuern, Finanzierung, Banken und Management, Springer Fachmedien, Wiesbaden, S. 403–417.
- Löffler, G. und Posch, P. (2011), Credit Risk Modeling Using Excel and VBA, 2. Aufl., Wiley, Hoboken, NJ.
- Malkiel, B. G. (2003), The Efficient Market Hypothesis and Its Critics, Journal of Economic Perspectives **17**(1), S. 59–82.

- Märkl, H. und Schaber, M. (2010), IFRS 9 Financial Instruments: neue Vorschriften zur Kategorisierung und Bewertung von finanziellen Vermögenswerten, KoR: Internationale und Kapitalmarktorientierte Rechnungslegung IFRS **10**(2), S. 65–74.
- Merton, R. C. (1974), On the Pricing of Corporate Debt. The Risk Structure of Interest Rates, *Journal of Finance* **29**, S. 449–470.
- Norden, L. und Weber, M. (2004), Informational Efficiency of Credit Default Swap and Stock Markets: The Impact of Credit Rating Announcements, *Journal of Banking & Finance* **28**(11), S. 2813–2844.
- Pawlowski, N. und Burmester, C. (2001), Ableitung von Standardrisikokosten auf der Basis von Expected-Loss-Kalkulationen, in: Schierenbeck, H., Rolfes, B. und Schüller, S. (Hrsg.), *Handbuch Bankcontrolling*, Gabler, Wiesbaden, S. 345–372.
- Pearce, D. K. (1983), Stock Prices and the Economy, *Economic Review* (9), S. 7–22.
- Peter, C. (2011), Estimating Loss Given Default: Experience from Banking Practice, in: Engelmann, B. und Rauhmeier, R. (Hrsg.), *The Basel II Risk Parameters*, Springer, Heidelberg [u.a.], S. 151–183.
- Renault, O. und Scaillet, O. (2004), On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities, *Journal of Banking & Finance* **28**(12), S. 2915–2931.
- Rolfes, B. und Kirmße, S. (2000), Risikomanagement in Banken, in: Dörner, D., Horváth, P. und Kagermann, H. (Hrsg.), *Praxis des Risikomanagements*, Schäffer-Poeschel, Stuttgart, S. 623–668.
- Schildbach, T. (2009), Fair Value-Statik und Information des Kapitalmarktes, *Betriebswirtschaftliche Forschung und Praxis: BFuP* **61**(6), S. 581–598.
- Schröder, T. (2015), Expected Loss im Kreditgeschäft nach IFRS 9 und Basel III: Kompatibilität von Rechnungslegung und Regulatorik?, *KoR: Internationale und Kapitalmarktorientierte Rechnungslegung IFRS* **15**(5), S. 245–251.
- Schuermann, T. (2005), What do we know about Loss Given Default?, in: Altman, E., Resti, A. und Sironi, A. (Hrsg.), *Recovery Risk*, Risk Books, London, S. 3–24.
- Serfling, K., Badak, E. und Jeiter, V. (2007), Möglichkeiten und Grenzen des Credit Ratings, in: Büschgen, H. E. und Everling, O. (Hrsg.), *Handbuch Rating*, Gabler, Wiesbaden.
- S&P (2017), *Default, Transition, and Recovery: 2016 Annual Global Corporate Default Study And Rating Transitions*, Standard & Poor's Financial Services LLC.
URL: <https://www.spratings.com>
- Steiner, M. und Heinke, V. G. (2000), Der Informationswert von Ratings - eine empirische Analyse am Markt für internationale DM-Anleihen, *Journal of Business Economics: JBE* **70**(5), S. 540–565.

Konjunkturabhängige PD-Bestimmung nach IFRS 9

– Zum Informationsgehalt öffentlich verfügbarer Ratings –

Erschienen in: *Handbuch Kapitalmarktorientierte Unternehmensbewertung – Grundlagen, Methoden, Regulierung und Branchentrends*, S. 429-447, Stuttgart, 2018.

Zusammenfassung

Mit dem „(Lifetime) Expected Credit Loss“ (LECL) verlagert der IFRS 9 seit dem Jahr 2018 einen Wertminderungsmaßstab, der anders als im finanzwirtschaftlichen Risikomanagement von Banken zukunftsorientierte Informationen explizit für die zeitpunktabhängige Schätzung einbeziehen soll. Die geforderte Berücksichtigung konjunktureller Einflüsse rückt dabei insbesondere die im Kern des LECL fundamentale Ausfallstruktur stark in den Fokus, regelmäßig basierend auf Ratingdaten großer Agenturen (z.B. S&P) bestimmt wird. Vor diesem Hintergrund werden die Ratingdaten im vorliegenden Beitrag auf ihren Zusammenhang mit Konjunkturindikatoren untersucht und darauf aufbauend ein Modellierungskonzept für konjunkturbedingte Ausfallstrukturkurven entwickelt.

Inhaltsverzeichnis

1	Einleitung	62
2	Lifetime Expected Credit Loss	63
2.1	Anforderungen an den Bewertungsmaßstab nach IFRS 9	63
2.2	Ausfallwahrscheinlichkeit und Ausfallstrukturkurven	66
3	Standard & Poor’s Ratingdaten	68
3.1	Ratingmigration und Ausfall	68
3.2	Konjunkturabhängigkeit der Ratingänderungen	69
3.3	Zeitliche Dynamik der Ausfallraten	71
4	Konjunkturabhängige Modellierung der Probability of Default	72
4.1	Regressionsbasierte Schätzung der Ausfallraten	72
4.2	Erweiterung zur migrationsbasierten Ausfallstruktur	74
5	Fazit	78

*Dipl.-Math. oec. Tobias Filusch, Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, ifG Marburg, Am Plan 2, D-35032 Marburg.

†Prof. Dr. Sascha H. Mölls, Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, Professur für Allgemeine Betriebswirtschaftslehre und Rechnungslegung, Barfußertor 2, D-35032 Marburg.

‡Christopher Ratz, M.Sc., Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, Professur für Allgemeine Betriebswirtschaftslehre und Rechnungslegung, Barfußertor 2, D-35032 Marburg.

1 Einleitung

Die Ereignisse der letzten Subprime- und der sich anschließenden globalen Finanzmarktkrise haben das Ausfallrisiko von Investitionsprojekten (wieder) verstärkt in das Bewusstsein vieler Investoren gerückt und in der Folge zu einer höheren Risikosensibilität geführt. Während die Krise zunächst vorwiegend dem Finanzsektor zuzuordnen war, weitete sie sich in ihrem weiteren Verlauf auf die Realwirtschaft aus, so dass auch Unternehmen außerhalb des Finanzsektors zunehmend mit ausfallgefährdeten Forderungen konfrontiert waren und nach wie vor sind. Die sich daraus ergebenden aktuellen und potenziellen Vermögensverluste sowie die ggf. resultierende Erhöhung der Insolvenzwahrscheinlichkeit sind wichtige Informationen für Investoren, die in die Entscheidungskalküle über das Kaufen, das Halten oder den Verkauf von Wertpapieren eingehen sollten. Das zentrale Informationsinstrument, das Investoren zur Verfügung steht, um sich über das Ausfallrisiko kapitalmarktorientierter Unternehmen zu informieren, ist der Konzernabschluss, der für eine Vielzahl von Marktteilnehmern nach den internationalen Rechnungslegungsvorschriften, den International Financial Reporting Standards (IFRS), aufgestellt werden muss.

Innerhalb der IFRS sind die viele Jahre praktizierten Vorschriften zur Abbildung des Ausfallrisikos von finanziellen Vermögenswerten nach IAS 39 im Rahmen der Finanzmarktkrise als vermeintliche Schwachstelle der internationalen Rechnungslegungsstandards identifiziert worden, wobei insbesondere die Wertminderungsvorschriften auf der Basis der fortgeführten Anschaffungskosten im Kreuzfeuer der Kritik stehen. Nach diesen Regelungen müssen die aufgrund des Ausfallrisikos bei finanziellen Vermögenswerten zu erfassenden Wertminderungen anhand des sog. „Incurred-Loss“-Modells bestimmt werden, nach dem Wertminderungen nur dann zu erfassen sind, wenn sie sich auf bereits eingetretene Verlustereignisse beziehen. Damit lassen sich jedoch erhöhte Ausfallerwartungen bis zum Eintreten eines konkreten Ausfallereignisses nicht berücksichtigen. Die Kritiker des „Incurred-Loss“-Modells führen deshalb auch an (Henselmann et al., 2014, S. 355; Große, 2011, S. 419; Grünberger, 2013, S. 195 f.), dass die nach diesem Modell zu berücksichtigenden Wertminderungen einen zu geringen Umfang einnehmen und regelmäßig erst verspätet durchgeführt werden. Dies würde zu einer prozyklischen Abbildung des Ausfallrisikos führen.

Nicht zuletzt aufgrund des politischen Drucks seitens der Europäischen Union (EU) ist eine Neuausrichtung der Bilanzierung von Finanzinstrumenten eine fast zwangsläufige Konsequenz der jüngsten Krise. Das International Accounting Standards Board (IASB) hat sich deshalb bereits vor einigen Jahren entschlossen, die Regelungen des IAS 39 vollständig zu überarbeiten und für das neue Regelwerk den Rechnungslegungsstandard IFRS 9 „Financial Instruments“ zu schaffen. Inhaltlich bezieht sich das „Replacement“-Projekt in seiner zweiten Phase auf Wertminderungsvorschriften von zu fortgeführten Anschaffungskosten oder erfolgsneutral zum beizulegenden Zeitwert bewerteten finanziellen Vermögenswerten und sieht für diese Fälle die Einführung eines „Lifetime Expected Credit Loss“-Modells (LECL) vor. Dieses zukunftsgerichtete Bewertungsmodell soll das bisher vorherrschende „Incurred-Loss“-Modell ersetzen und ab dem 1. Januar 2018 verbindlich von kapitalmarktorientierten Unternehmen angewendet werden. Im Gegensatz zu den früheren Wertminderungsvorschriften nach IAS 39 wird im IFRS 9 ein erwarteter Kreditverlust bereits frühzeitig und nicht erst nach dem Eintritt eines konkreten Ausfallereignisses berücksichtigt. Hierzu soll bereits zum Ansatzzeitpunkt

eines finanziellen Vermögenswertes der erwartete Kreditverlust ermittelt werden, für den während der Laufzeit des finanziellen Vermögenswertes sukzessiv eine Risikovorsorge zu bilden ist.

Einen wichtigen Bestandteil der LECL-Ermittlung stellt die „Probability of Default“ (PD) dar. Diese beschreibt die Wahrscheinlichkeit für ein Ausfallereignis in einem bestimmten zukünftigen Zeitraum. PDs variieren dabei nicht nur wegen der Bonität des Schuldners, sondern insbesondere auch aufgrund konjunktureller Einflüsse oder der Länge der betrachteten Zeitperiode. Entsprechend wird die Berücksichtigung des Konjunktureinflusses in den Parameteranforderungen des IFRS 9 zwar explizit vorgegeben, wirft aber viele methodische Umsetzungsfragen seitens der Anwender auf, die sich z.B. auf die zu verwendenden ökonomischen Faktoren oder auch die Spezifikationen eines Modells zur Schätzung der PDs beziehen. An dieser Stelle setzt der vorliegende Beitrag an, in dem das Augenmerk auf die Bestimmung der PD als zentralem Kernparameter des neuen LECL-Modells gelegt und der Frage nach einer geeigneten Ermittlungsmethode der PD auf der Basis öffentlich verfügbarer und damit leicht zugänglicher (Rating-)Daten nachgegangen wird. Im Ergebnis zeigt sich ein Einfluss der Konjunktur auf durch Standard & Poor's (S&P) bereitgestellte Ausfallraten, deren zeitliche Dynamik sich durch ein autoregressives Zeitreihenmodell mit externen Einflussfaktoren darstellen lässt. In diesem Modell ermöglichen insbesondere konjunkturelle Frühindikatoren eine Prognose der zukünftigen Ausfallraten. Basierend auf diesen Analysen wird ein Markov-Modell für den Aufbau einer „Point-In-Time“-PD-Struktur vorgestellt, an die vorliegenden S&P-Daten angepasst und anschließend diskutiert. Eine IFRS-konforme Modellierung ist damit grundsätzlich möglich, im Hinblick auf die Präzision der PD-Prognose können aber auch Schwächen festgestellt werden, die in künftigen Arbeiten weitergehend untersucht werden müssen.

2 Lifetime Expected Credit Loss

2.1 Anforderungen an den Bewertungsmaßstab nach IFRS 9

Unternehmen sind im Rahmen ihrer Tätigkeit zahlreichen finanzwirtschaftlichen Risiken ausgesetzt, wobei die Steuerung dieser Risiken in der jüngeren Vergangenheit zunehmend an Bedeutung gewonnen hat. Charakteristisch für eine Risikosituation ist die Möglichkeit unterschiedlicher zukünftiger Umweltzustände, denen sich im besten Fall gewisse Eintrittswahrscheinlichkeiten zuordnen lassen (Läger, 2002, S. 14). Ein bedeutsames Risiko, mit dem sowohl Industrie- und Handelsunternehmen bei Forderungen gegenüber Kunden und Lieferanten sowie bei Investitionen in Schuldtitel als auch Banken im Rahmen ihrer originären Geschäftstätigkeit konfrontiert werden, stellt das Kreditrisiko dar (Büschgen, 1998, S. 928). In einem weiter differenzierenden Begriffsverständnis kann dieses Kreditrisiko in (1.) das Ausfallrisiko und (2.) das Bonitätsrisiko aufgespalten werden (Bachmann, 2004, S. 46). Das Ausfallrisiko stellt auf das Risiko ab, dass ein Schuldner seinen Zahlungsverpflichtungen gar nicht, nicht zeitgerecht oder nicht vollständig nachkommt (Hartmann-Wendels et al., 2015, S. 499 ff.). Es besteht somit die Gefahr, dass die tatsächlich eintretenden Zahlungen des Schuldners von den erwarteten Zahlungen (negativ) abweichen. Eine mögliche Verschlechterung der Bonität wird durch das Bonitätsrisiko beschrieben und kann zu einem entsprechenden Risikoabschlag (auf den Kurs der betreffenden Wertpapiere) führen.

Das neue Wertminderungskonzept des IFRS 9 für finanzielle Vermögenswerte, die zu fort-

geführten Anschaffungskosten oder erfolgsneutral zum beizulegenden Wert („Fair Value“) bewertet werden, bezieht sich auf die zuvor skizzierte Präzisierung des Kreditrisikos und fordert eine umfassende Berücksichtigung erwarteter Kreditausfälle im Bewertungsmaßstab des „Lifetime Expected Credit Loss“ (LECL). Der Standardsetter verzichtet an dieser Stelle allerdings auf die Vorgabe einer konkreten Methode für die Schätzung des LECL, sondern steckt den Rahmen für die Quantifizierung nur grob ab (IASB, 2014, S. 29). Danach soll der LECL folgende Bestandteile widerspiegeln (IASB, 2014, S. 111):

1. einen unverzerrten, wahrscheinlichkeitsgewichteten Betrag, der durch eine Vielzahl an potenziellen Ergebnisausprägungen zukunftsorientiert geschätzt wird;
2. den Zeitwert des Geldes;
3. sinnvolle und begründete Informationen vergangener Ereignisse, aktuelle Verhältnisse und Prognosen zukünftiger makro- und mikroökonomischer Gegebenheiten, die ohne übermäßige Kosten zum Abschlussstichtag verfügbar sind.

Es wird demnach also ein LECL-Modell eingeführt, ohne dieses jedoch tatsächlich zu konkretisieren, wobei sich zugleich wiederholte Verweise auf bekannte und verständliche Kennzahlen aus der finanzwirtschaftlichen Praxis ausmachen lassen (Grünberger, 2011, S. 29). Entsprechend kann für die Ermittlung des vom IASB geforderten „Expected Credit Loss“ (ECL) auf die bereits in Kreditinstituten – i.d.R. für den Zeitraum von einem Jahr – angewandte Methode zur Bestimmung der Mindesteigenkapitalanforderungen für Kreditrisiken zurückgegriffen werden. Das diesbezügliche Vorgehen ist im „Internal Ratings-Based Approach“ (IRB-Approach) der Basel II-Regularien zur Berechnung der Eigenmittelunterlegung von Kreditrisiken vorgeschrieben (BCBS, 2004, Tz. 211 ff.) und berechnet sich gemäß

$$ECL = PD \cdot EaD \cdot LGD, \quad (1)$$

wobei EaD (= „Exposure at Default“) den gesamten ausfallbedrohten Betrag zum Zeitpunkt des Ausfalls und LGD (= „Loss Given Default“) die Verlustquote bei Eintritt eines Ausfalls umschreibt. Die PD gibt die Wahrscheinlichkeit eines Ausfalls während der direkt anschließenden Zeitperiode (i.d.R. innerhalb des nächsten Jahres) an.

Mit Blick auf die Veränderung des Kreditrisikos müssen Unternehmen nach dem neuen Wertminderungsmodell des IFRS 9 zu jedem Abschlussstichtag auf der Basis einer Zuordnung der Instrumente zu einer von drei Stufen („Three Bucket Approach“) eine Risikovorsorge für zukünftige Kreditausfälle entweder auf der Basis eines „12-Monats-ECL“ oder eines LECL erfassen (IASB, 2014, S. 111 ff.). In Stufe 1 liegt der Zeithorizont für ein mögliches Verlustszenario bei 12 Monaten. Eine solche „12-Monats-ECL“ stellt einen Teil des LECL dar und umfasst sämtliche künftigen Zahlungen, die aus in den nächsten 12 Monaten eingetretenen Ereignissen resultieren. Stufe 1 werden Kredite zugeordnet, bei deren Kreditqualität keine signifikante Verschlechterung zu beobachten war oder die als „Investment Grade“ gelten, also ein geringes Ausfallrisiko aufweisen. Für Kredite abseits des Investment Grade wird der Übergang zur Stufe 2 durch eine signifikante¹ Erhöhung des Kreditrisikos ausgelöst. Dieser Wechsel in Stufe 2 führt zur Anwendung des LECL, was den Zeithorizont für Ausfallszenarien

¹Vgl. zu einer Übersicht über Indikatoren für eine signifikante Änderung des Kreditrisikos EY (2015, S. 50 ff.). Anhang B des IFRS 9 enthält allgemeine Leitlinien zur Anwendung des Standards. Vgl. zu einer inhaltlichen Bestimmung von „Signifikanz“ EY (2015, S. S. 56 ff.).

auf die gesamte Restlaufzeit verlängert und infolgedessen die zu erfassende Risikovorsorge erhöht. Zinserträge und Wertberichtigungen werden auf den Stufen 1 und 2 unter Verwendung des Effektivzinssatzes getrennt voneinander erfasst. Basis für die Zinserträge ist jeweils der Bruttobuchwert ohne Abzug der Risikovorsorge. Wertberichtigungen ergeben sich aus den mit dem Effektivzinssatz diskontierten erwarteten Kreditausfällen. Für Stufe 3, für die ebenfalls der LECL anzuwenden ist, müssen – in Analogie zu den Wertminderungskriterien des IAS 39 – objektive Hinweise auf eine Wertminderung² vorliegen. In diesem Fall sind in den Folgeperioden die Zinserträge unter Verwendung des Effektivzinssatzes anstatt auf der Basis des Bruttobuchwertes auf der Grundlage der fortgeführten Anschaffungskosten (Nettobuchwert = Bruttobuchwert \cdot / Risikovorsorge) zu berechnen (IASB, 2014, S. 25 ff.). Sollte sich schließlich der finanzielle Vermögenswert infolge der Wertminderung nicht mehr realisieren lassen, ergibt sich ein Ausbuchungsereignis durch eine unmittelbare Abschreibung des Bruttobuchwertes auf null.

Beginnend mit der Zuordnung jedes Instruments in die Stufe 1 ist an jedem folgenden Abschlussstichtag zu prüfen, ob die Transferbedingungen, d.h. eine „signifikante“ Erhöhung des Ausfallrisikos mit Blick auf die Restlaufzeit des Instruments, erfüllt sind und somit ein Wechsel in Stufe 2 bzw. Stufe 3 erforderlich wird (IASB, 2014, S. 26). Im Sinne eines symmetrischen Wertminderungsmodells sind schließlich auch Rücktransfers eines Finanzinstruments in eine niedrigere Stufe vorzunehmen, sobald die Voraussetzungen für einen Wechsel in eine höhere Stufe entfallen (IASB, 2014, S. 27 ff.).

Der LECL kann auf Basis von (1) und den für die entsprechenden Jahre in der Zukunft prognostizierten Parametern PD, EaD und LGD abgeleitet werden. Die mehrjährige Erweiterung auf T Perioden ist – mit r als risikolosem Zins – durch (2) dargestellt (Grünberger, 2011, S. 29):

$$LECL_T = \sum_{t=1}^T \frac{PD_t \cdot EaD_t \cdot LGD_t}{(1+r)^t}. \quad (2)$$

Bei Anwendung des LECL kann sich durch die Ausweitung des Betrachtungszeitraums eine sehr große Spannweite von möglichen Ausfallszenarien ergeben. Ein früher Ausfall könnte zu sehr großen Verlusten führen, wohingegen ein sehr später Ausfall bei entsprechendem Zahlungsstrom sehr geringe Ausfälle induziert. Innerhalb dieser Extremfälle bewegen sich die Möglichkeiten, die – mit bestimmten Wahrscheinlichkeiten gewichtet – eintreten können. In Bezug auf die oben formulierten Anforderungen an den LECL kann (2) bereits die Berücksichtigung der Diskontierung entnommen werden. Die Verwendung angemessener Informationen und die Einbeziehung der konjunkturellen Lage lässt sich auf die drei Kernparameter übertragen: Erfüllen diese die Kriterien, so gilt dies auch für den LECL. Der EaD beschreibt die Höhe des begebenen Kreditvolumens und damit den im möglichen Ausfallzeitpunkt verbleibenden erwarteten Zahlungsstrom. Basierend auf den im Kreditvertrag vereinbarten Details erhält dieser Parameter einen stärker deterministischen Charakter und lässt damit nur einen schwachen konjunkturellen Einfluss vermuten. Zusammen mit dem Diskontierungsfaktor auf der Basis des Effektivzinssatzes bildet der EaD in der LECL-Formel (2) deshalb auch den risikofernen

²Wertminderungen liegen in einem zukunftsorientierten Verständnis vor, wenn eines oder mehrere Ereignisse eintreten, die die geschätzten künftigen „Cashflows“ des betreffenden finanziellen Vermögenswertes negativ beeinflussen. Vgl. zu einer exemplarischen Aufzählung – in Anlehnung an IAS 39 – IASB (2014, Appendix A).

Teil. Analog kann mit vereinfachenden Annahmen auch der LGD zur deterministischen Zahlungsstromkomponente hinzugefügt werden, was allein die PD als primäre Risikokomponente des LECL isoliert. Derartige Vereinfachungen sollten nach den Vorgaben des IFRS 9 zulässig sein, da die Bestimmung der Kernparameter im Einzelnen nicht spezifiziert wird und das hier vorgestellte Umsetzungskonzept des LECL im Einklang mit den allgemeinen Anforderungen des Standards steht.

2.2 Ausfallwahrscheinlichkeit und Ausfallstrukturkurven

Angesichts der lediglich allgemein gehaltenen Forderung nach der Einbeziehung makro- und mikroökonomischer Gegebenheiten verzichtet der Standardsetter auf die Vorgabe einer konjunkturbezogenen Berechnungslogik der PD. Mit den zuvor beschriebenen Vereinfachungen des EaD und der LGD hin zu einer deterministischen Modellierung muss dieser Aspekt aber über die PD – als dann einzigem risikobehafteten Kernparameter – Einzug in das LECL-Modell erhalten. Da Ausfälle einen grundsätzlich nicht zu vernachlässigenden Bezug zur Konjunktur vermuten lassen, stellt dieser Parameter eine naheliegende Einbeziehungsmöglichkeit dar. Im Hinblick auf den gewünschten Charakter des LECL muss bei der PD-Struktur für Formel (2) insbesondere – und dabei sehr präzise – auf die tatsächliche Aussagekraft der gebildeten Wahrscheinlichkeit geachtet werden. Während bei der Betrachtung einer unmittelbar bevorstehenden Periode lediglich die Ereignisse „Ausfall“ oder „Nicht-Ausfall“ eintreten können, ist dies bei der Ausweitung auf mehrere anschließende Jahre deutlich komplexer. In diesem Fall muss für jeden betrachteten Ausfall zunächst das Überleben des Kredits bis zur betreffenden Periode vorausgesetzt werden. Für eine abgegrenzte, zukünftige Betrachtungsperiode t besteht die Ausfallwahrscheinlichkeit in t – auf die sich in der Folge der Terminus „PD“ bezieht – dementsprechend aus den Überlebenswahrscheinlichkeiten bis und der Ausfallwahrscheinlichkeit in t . An dieser Stelle wird von der „unbedingten“ Ausfallwahrscheinlichkeit gesprochen (Grünberger, 2011, S. 29). Das Grundniveau und der Verlauf dieser PD-Struktur sollten dabei maßgeblich durch die initiale Bonitätsbewertung determiniert werden und nicht nur auf einer uneingeschränkten Einzelbewertung beruhen, die durch den Standard aber auch nicht gefordert wird. Die vielfältigen Möglichkeiten der Bonitätsverbesserung oder -verschlechterung basierend auf der Einteilung in mehrere Bonitätsklassen dürfen in diesem Zusammenhang ebenfalls nicht vernachlässigt werden. Ein Kredit besonders guter Klassen hat in naher Zukunft naturgemäß eine sehr geringe Ausfallwahrscheinlichkeit, die im weiteren Verlauf aber ansteigen wird, da diese Schuldner tendenziell nur schlechter, aber nicht besser werden können. Im Umkehrschluss dürften sich die Kredite einer schlechten Bonität entweder verbessern oder zeitnah ausfallen, wobei der Ausfall zunächst gleichbedeutend mit einem Ausschluss aus der Betrachtung sein sollte. Aus diesen Überlegungen folgt ein abnehmender PD-Verlauf mit einem Grundniveau, das oberhalb des Niveaus der guten Bonitätsklasse liegen sollte. Schlussendlich werden die Verläufe aller Bonitätsklassen in einer gemeinsamen langfristigen PD* münden. Ein solcher Verlauf lässt sich damit begründen, dass es mit den aktuellen Informationen kaum möglich sein dürfte, die genaue Qualität eines nicht ausgefallenen Kredits in weiter Zukunft einzuschätzen. Folgerichtig sollte die Einschätzung daher auf eine sehr durchschnittliche Kreditqualität – insbesondere in weiterer Zukunft – hinauslaufen. Für die Einbeziehung konjunktureller Einflüsse bedeutet dies entsprechend, dass eine konkrete Berücksichtigung nur in aktuelleren Perioden sinnvoll sein kann. Zur Illustration der vorstehenden Überlegungen zeigt Abbildung 1 zwei mögliche PD-Strukturverläufe sowie die PD* als langfristigen Trend. Konjunkturelle Einflüsse verschieben die Kurven bzw. führen zu Schwankungen des Verlaufes, wobei die langfristige

PD^* von diesen Einflüssen weitestgehend unabhängig ist und für ein bestimmtes Segment an Krediten konstant bleiben dürfte.

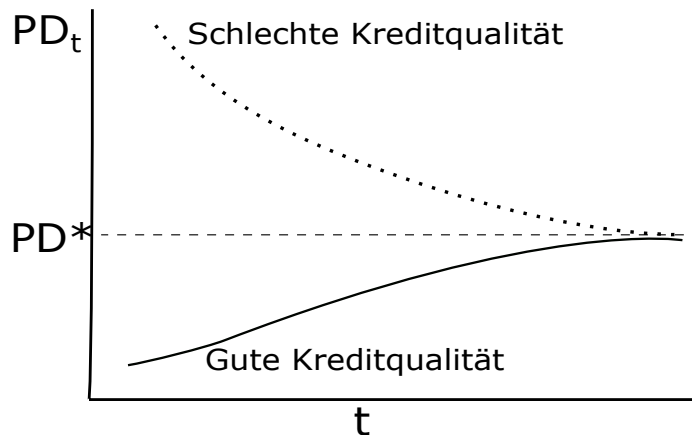


Abbildung 1: Hypothetische Ausfallstruktur eines initial guten (geringe PD, durchgezogene Linie) und eines schlechten (hohe PD, gestrichelte Linie) Kredites (Quelle: Eigene Darstellung).

Die unbedingte Ausfallwahrscheinlichkeit (PD) steht in enger Beziehung zur „kumulierten“ Ausfallwahrscheinlichkeit (kumPD), welche sich für einen Zeitpunkt t aus den aufsummierten unbedingten Ausfallwahrscheinlichkeiten ergibt. Damit sind diese Kurven stets und für alle Bonitätsklassen monoton steigend. Beide Wahrscheinlichkeiten lassen sich einfach ineinander überführen und bieten somit äquivalente Möglichkeiten, die für den LECL benötigten Parameterwerte zu generieren. Ein dritter relevanter Wahrscheinlichkeitsbegriff ergibt sich schließlich aus der auf das Überleben des Kredites bis zur Vorperiode „bedingten“ Ausfallwahrscheinlichkeit (bedPD). Diese in einer Periode t betrachtete Ausfallwahrscheinlichkeit beschreibt ausschließlich das Ausfallverhalten in diesem Zeitfenster. Ob der Kredit bis dato schon länger überlebte oder erst gerade frisch abgeschlossen wurde, findet keine Berücksichtigung.

Die PD-Modellierung im Markov-Modell mit einer endlichen Anzahl von Zuständen (Bonitätsklassen) und obligatorischen, die Übergänge beschreibenden Migrationsmatrizen stellt eine intuitiv geeignete Bestimmungsmöglichkeit dar, die die zuvor formulierten Basiseigenschaften des Ausfallverhaltens beinhaltet (Filusch und Mölls, 2017, S. 254 f.). Solche Methoden finden hinsichtlich der Ausgestaltung für Formel (2) bereits seit Jahren im Risikomanagement der betrieblichen Praxis Anwendung und fügen sich daher in ein stimmiges Gesamtkonzept ein (Engelmann und Ermakov, 2011, S. 105 ff.). Wie die in Abschnitt 2.1 aufgeführten Anforderungen des IFRS 9 zeigen, ergibt sich für die dort vorgeschriebene Ausfallstruktur allerdings

ein entscheidender Unterschied zum traditionellen Vorgehen: Während das Risikomanagement mehr an konservativen und konjunkturunabhängigen Aussagen („Through-The-Cycle“ (TTC)-Perspektive) interessiert ist, schreiben die IFRS-Standardsetter eindeutig vor, auch die zum Zeitpunkt der Bewertung vorherrschenden Bedingungen explizit mit einzubeziehen („Point-In-Time“ (PIT)-Perspektive). Dieser markante Unterschied muss demnach bei der Modellierung der IFRS-konformen Ausfallstruktur berücksichtigt werden, um die Effekte in den LECL einfließen lassen zu können.

3 Standard & Poor’s Ratingdaten

3.1 Ratingmigration und Ausfall

Die für die Zuteilung von Kreditnehmern zu geeigneten Bonitätsklassen verwendete Berechnungslogik eröffnet Kreditinstituten grundsätzlich Spielräume, wobei insbesondere Scorings (Hartmann-Wendels et al., 2015, S. 517 f.) und Ratings (Everling, 1991, S. 29 ff.) in der betrieblichen Anwendung eine breite Verwendung erfahren. Beide Verfahren sind auf die Ermittlung der Bonität des Schuldners ausgerichtet, unterscheiden sich aber hinsichtlich ihrer Transparenz und Beeinflussbarkeit: Während Scorings individuell durch das bewertende Institut ausgestaltbar und dadurch transparent sind (Löffler und Posch, 2011, S. 1 f.; Hayden und Porath, 2011, S. 4 ff.), stellen Ratings teilweise öffentlich verfügbare und intransparente Informationen dar, da sie auf einer im Detail unbekanntem Kriterienevaluation externer Ratingagenturen beruhen (Everling, 1991, S. 35 f., 114, 150 f.; Serfling et al., 2007, S. 717). Im Vertrauen auf ihre Zuverlässigkeit sowie mit dem Ziel einer höheren Präzision existieren in der Bankenpraxis Ansätze zur Integration von Ratings in bestehenden bankinternen Scoring-Systeme (Hartmann und Schwarzhaupt, 2007, S. 32 ff.). Im Zeitablauf ermöglichen beide Verfahren die Messung einer Bonitätsveränderung mittels relativer Klassifikationsänderungen, die in der gängigen praktischen Interpretation als empirische Übergangswahrscheinlichkeiten zwischen den vordefinierten Bonitätsklassen verstanden werden können. Die Einträge der Migrationsmatrix beschreiben für alle möglichen Klassenänderungen die zugehörige Wahrscheinlichkeit dieser Ereignisse während eines bestimmten Zeitraums, wobei sich der direkte Ausfall in der betrachteten Periode ebenfalls darunter befindet (Duffie und Singleton, 2003, S. 86). Fällt ein Kredit in einer Vorperiode aus, so wird davon ausgegangen, dass zunächst kein weiterer Ausfall desselben Kredites stattfinden kann. Dieser Umstand geht in der praktischen Markov-Modellierung mit der verwendeten Annahme des absorbierenden Zustands, d.h. eines Zustandes, der nicht mehr verlassen werden kann, einher (Bluhm et al., 2003, S. 71). In dieser Logik werden wiedergesundete Kredite analog zu einem neuen Kredit aufgefasst und weisen keine Vorgeschichte auf. Eine solche Wiedergesundung ist dabei mehr als nur theoretisch relevant, da das Ausfallereignis zwangsläufig an eine gewisse Definition gebunden sein muss. Die gängige Umsetzung sieht vor, dass unerfüllte Zahlungsverpflichtungen nach einer vordefinierten Wartezeit dem Ausfallzustand zugeordnet werden (S&P, 2017, S. 5 f.), wobei aber natürlich grundsätzlich die uneingeschränkte Möglichkeit besteht, zukünftig wieder ein Rating jenseits von D zu erhalten. Für das Markov-Modell wird dieser Übergang allerdings stets ausgeblendet, wodurch die Bestimmung der Ausfallstruktur sehr viel einfacher möglich ist. Die zunächst über den Zustand D ausgeschiedenen Unternehmen können stets als neue Unternehmen ohne Vorgeschichte wieder in die Modellwelt aufgenommen werden.

Ratings und ihre Veränderungen im jährlichen Verlauf sind über einen repräsentativen Zeithorizont insbesondere seitens der großen Ratingagenturen Moody's, Standard & Poor's (S&P) sowie Fitch verfügbar (S&P, 2017). Da die Agenturen aber (teilweise) eine Gebühr von den zu ratenden Gesellschaften verlangen, wird die Aussagekraft von Ratingkennzahlen regelmäßig kritisch gesehen. Es ist deshalb auch zu vermuten, dass infolge der Auftragsbeziehung zwischen Ratersteller und der das Rating nachfragenden Gesellschaft nicht von unabhängigen und damit unverzerrten Beurteilungen ausgegangen werden kann. Zudem fehlen Ratings für eine Vielzahl von insbesondere kleinen und kapitalmarktfernen Unternehmen. Probleme ergeben sich aber auch im Zusammenhang mit internen Bewertungen und Klassenzuteilungen basierend auf Scorings. Derartige Modelle implizieren einerseits eine aufwendige Erstellung und fortlaufende Pflege und liegen andererseits aktuell wohl in kaum ausreichend langen Historien vor. Entsprechend scheint dieser Weg erst in der (fernen) Zukunft nutzbar zu sein.

Eine Alternative zu den zuvor diskutierten Verfahrensweisen scheint schließlich ein Rückgriff auf die Daten der jährlich von der Ratingagentur S&P herausgegebenen „Default Study“ zu sein, obwohl bei den dort angegebenen Ratings auch von einer konjunkturunabhängigen Bonitätsbewertung die Rede ist und die empirischen Bonitätsänderungen ausschließlich in Kalenderjahresschritten zurückverfolgt werden können. Für das angestrebte Ziel, eine LECL-geeignete Ausfallstruktur zu bestimmen, liegen aber die jährlichen Ausfall und Übergangsraten sowie die über alle einbezogenen Jahre kumulierte Migrationsmatrix gesamtlich für alle durch S&P bewerteten Unternehmen weltweit vor („Global Corporates“). Die darüber beobachtete Migrationsdynamik lässt sich daher leicht als Ausgangspunkt für Untersuchungen des konjunkturbedingten Migrationsverhaltens verwenden und kann für die angestrebte Modellierung als Testfeld dienen. Die vorliegenden Daten decken einen Zeitraum von 1981 bis 2016 ab, für den je Kalenderjahr eine konkrete Beobachtung sämtlicher Übergangsraten vorliegt. Die Dimension dieser Matrizen ist an die Anzahl der Bonitätsklassen gebunden, die als die Zustände der Markov-Kette gelten. Seitens S&P wird dabei üblicherweise die folgende Skala – ausgehend von guter Bonität bis hin zum Ausfall – zugrunde gelegt: AAA, AA, A, BBB, BB, B, CCC.C. Eine darauf aufbauende, noch feinere Unterteilung ist bei den Agenturen zwar regelmäßig verfügbar, hat aber eine unerwünscht dünne Beobachtungsgrundlage des Migrationsverhaltens zwischen diesen verfeinerten Klassen zur Folge. Aus diesem Grund sollen für die nachfolgenden Untersuchungen lediglich die sieben genannten Klassen und der Ausfall selbst verwendet werden.

3.2 Konjunkturabhängigkeit der Ratingänderungen

Die beobachteten Migrations- und Ausfallraten (Default Rates – DR) der S&P-Datenbank sind über das Sample hinweg nicht konstant, sondern folgen einer bestimmten Dynamik, die nun auf Zusammenhänge mit Konjunkturzyklen untersucht werden soll. Von besonderer Bedeutung sind dabei die jährlichen DRs, die ausgehend von einer bestimmten Ratingklasse den direkten Ausfall in der Folgeperiode charakterisieren und damit eine konkrete Beobachtungsgrundlage des zeitlichen Verhaltens der bedingten Ausfallwahrscheinlichkeit liefern.

In der Literatur wird eine Vielzahl makroökonomischer Faktoren bzw. Kennzahlen, wie etwa das Bruttoinlandsprodukt, Zinsspreads, Aktienindizes, die Industrieproduktion oder Arbeitslosenquoten genannt, die Aussagen über die konjunkturelle Entwicklung bzw. die wirtschaftliche Situation im Allgemeinen erlauben (Engelkamp und Sell, 2017, S. 302 ff.). Basierend auf die-

ser Vermutung soll zunächst anhand des US-amerikanischen Bruttoinlandsprodukts (GDP)³ überprüft werden, ob und inwieweit Ausfallraten durch den Konjunkturzyklus beeinflusst wer-

Variable	AA	A	BBB	BB	B	CCC.C
GDP_t	-0.21	-0.35*	-0.07	-0.02	-0.20°	-0.42**
GDP_{t-1}	-0.06	0.11	0.17	0.50**	0.14	-0.13
STR_t	-0.01	-0.05	-0.05	-0.08	0.06	0.04
STR_{t-1}	-0.38*	-0.34*	-0.35*	-0.36*	-0.36*	-0.22°
$SP500_t$	-0.44**	-0.45**	-0.24	-0.23	-0.07	-0.12
$SP500_{t-1}$	0.02	-0.21	-0.27°	-0.24°	-0.29°	-0.39*

Tabelle 1: Korrelationsstrukturen der betrachteten makroökonomischen Faktoren im Zeitraum 1981-2016 (eigene Berechnung). *** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$, ° $p < 0.15$.

den. Die Korrelationsstrukturen zwischen S&P-Ausfallraten verschiedener Ratingkategorien und Wachstumsraten im Zeitraum von 1981 bis 2016 zeigen einen signifikanten Zusammenhang (vgl. Tab. 1). Die erwartungsgemäß negative Korrelation lässt sich deutlich erkennen und bestätigt die Intuition, dass im Aufschwung niedrige Ausfallraten und umgekehrt im ökonomischen Abschwung höhere Ausfallraten zu erwarten sind. Der Bezug der beobachteten Ausfallraten zum Konjunkturzyklus wird zusätzlich an der grafischen Gegenüberstellung der Ausfallraten mit den Wachstumsraten deutlich (vgl. Abb. 2). Es zeigt sich eine negative Korrelation der Ausfallraten (hier am Beispiel der B-Ratingkategorie) mit dem GDP (gestrichelte Kurve). Die beiden Kurven verhalten sich erwartungsgemäß stark gegenläufig, was insbesondere sehr gut in Krisenzeiten beobachtet werden kann. So steigen z.B. im Jahr 2009 die Unternehmensausfälle sehr stark an, während die Konjunktur einen Einbruch erleidet und kurzzeitig sogar negative Werte aufweist.

Neben dem GDP existieren noch weitere makroökonomische Faktoren, deren Stärke vor allem in der Eigenschaft eines stärker vorauslaufenden Indikators begründet liegt. Derartige Frühindikatoren sind ein wichtiges Instrumentarium zur Einschätzung konjunktureller Entwicklungen sowie von Trendänderungen und sollten in Bezug auf ein Prognosemodell zu besseren Schätzungen führen, sofern sie Erwartungen frühzeitig bündeln. So zählt der Zinsspread zwischen 10-jährigen und 3-monatigen US-Staatschatzbriefen (STR) zu den sog. Frühindikatoren, die Hinweise auf zukünftige Entwicklungen der ökonomischen Lage geben, bevor diese sich tatsächlich ändert (O’Sullivan, 2003, S. 314). Kurz vor den letzten vier Rezessionen fiel der Spread z.B. jeweils auf ein niedriges Niveau nahe 0% und bei der letzten schweren Krise im Jahr 2008 sogar knapp unter 0%. Aus diesem Grund ist der Zinsspread prinzipiell bestens geeignet, um die Entwicklung der DRs zu erklären.

Des Weiteren zählt auch der S&P 500-Index (SP500) als Aktienindex zu den sogenannten Frühindikatoren und sollte ebenfalls in das Modell aufgenommen werden. In die Berechnung des Indexes gehen die prozentualen Veränderungen zwischen den einzelnen Jahren ein, um mit dieser Umwandlung bzw. Normalisierung für eine stationäre⁴ Datenreihe zu sorgen. Grund-

³Sämtliche Zeitreihendaten makroökonomischer Variablen sind der Internetdatenbank der Federal Reserve Bank of St. Louis entnommen (abrufbar unter: www.fred.stlouisfed.org).

⁴Inhaltlich bedeutet Stationarität, dass die stochastischen Eigenschaften eines Prozesses invariant unter einer Zeitverschiebung sind, d.h. sich nicht ändern bei einer Verschiebung des Zeitursprungs. Vgl. Hackl (2013, S. 194 ff.) und Greene (2012, S. 953 ff.)

lage für die Berechnung des Indexes sind Jahresschlusskurse, da diese bereits einen gewissen Prognoseanteil für die Folgeperiode beinhalten sollten. Hintergrund dieser These ist die Vorstellung, dass aktuelle Kurse die Erwartungen der Anleger an die (zumindest kurzfristige) zukünftige ökonomische Entwicklung widerspiegeln und dadurch deutlich besser für eine Stichtagsbetrachtung geeignet sein sollten als z.B. durchschnittliche Jahreswerte.

Tabelle 1 gibt einen Überblick über die jeweiligen Korrelationsstrukturen der Faktoren mit den verschiedenen Ausfallraten der unterschiedlichen Ratingkategorien. Analog zum GDP bestehen (bis auf wenige Ausnahmen) durchgehend negative Beziehungen zwischen Konjunktur und Ausfallraten. Es ist deutlich zu erkennen, dass insbesondere die verzögerten Variablen STR_{t-1} und $SP500_{t-1}$ eine besonders starke Erklärungskraft aufweisen. Diese Variablen gehen deshalb jeweils mit einer Periode (einem Jahr) Verzögerung in die Analyse ein und liefern wichtige Ansatzpunkte für die noch durchzuführende Regression. Darüber hinaus wird deutlich, dass sich vor allem in den Ausfallraten der Ratingkategorien schwacher Bonität deutlich erhöhte Korrelationen mit der Konjunktur beobachten lassen. Spätestens an dieser Stelle zeigen sich die Limitationen des GDP als Bestandteil eines Prognosemodells, da bezüglich dieser Variable sämtliche Korrelationen der Ausfallraten mit der verzögerten Datenreihe (GDP_{t-1}) nicht signifikant sind und teilweise keine plausiblen Vorzeichen aufweisen. Folglich spricht viel für die Verwendung der Größen STR und SP500 im Rahmen des Regressionsmodells.

Abbildung 2 verdeutlicht die bisherigen Ergebnisse durch Gegenüberstellung der Ausfallraten mit den Konjunkturindikatoren. Es ist erkennbar, dass sich die Konjunktur in ökonomisch schlechten Zeiten (z.B. im Zeitraum 1995 bis 2002) genau gegenläufig zu den Ausfallraten entwickelt. Weiterhin wird die Eigenschaft von Zinsspreads als Frühindikator deutlich, da die gepunktete Kurve jeweils bereits vor einem Anstieg der Ausfallraten, die durch die schwarze Kurve beschrieben werden, fällt. Die empirisch abgesicherte Auswahl für die geplante „Einschritt“-Vorhersage der Ausfallraten, die später als bedingte PDs zur Entwicklung der PIT-Ausfallstruktur verwendet werden, fällt also auf den STR und den SP500 als Frühindikatoren mit einem starken und zugleich zukunftsorientierten Erklärungsgehalt der konjunkturellen Lage.

3.3 Zeitliche Dynamik der Ausfallraten

Neben der festgestellten zeitlichen Dynamik beobachteter Ausfallraten und der damit verbundenen Abhängigkeit von (verzögerten) makroökonomischen Variablen lässt sich zudem eine gewisse Eigendynamik, d.h. eine Abhängigkeit zur Ausfallrate aus vergangenen Jahren, und demnach ein signifikanter Erklärungsgehalt der eigenen Vergangenheit für die nachfolgende Periode vermuten. Tabelle 2 gibt diesbezüglich einen Überblick über die Korrelationen der Ausfallraten verschiedener Ratingkategorien mit ihren jeweiligen eigenen verzögerten Datenreihen. Es wird deutlich, dass insbesondere die kurzfristige Vergangenheit, d.h. Verzögerungen um eine Periode (DR_{t-1}), signifikante Ergebnisse aufweist. Für sämtliche Ausfallraten der Ratingklassen A–CCC.C zeigen sich in der kurzfristigen Betrachtung signifikante Befunde und damit eine empirisch untermauerte hohe Relevanz der eigenen (kurzfristigen) Vergangenheit. Es liegt daher nahe, diese Dimension in die PD-Modellierung als weitere Komponente einzubeziehen. Weiter in der Vergangenheit zurückliegende Datenreihen (DR_{t-2} und DR_{t-3}) weisen hingegen keine signifikanten Einflüsse auf aktuelle Ausfallraten auf und bleiben deshalb im aufzustellenden Prognosemodell unberücksichtigt.

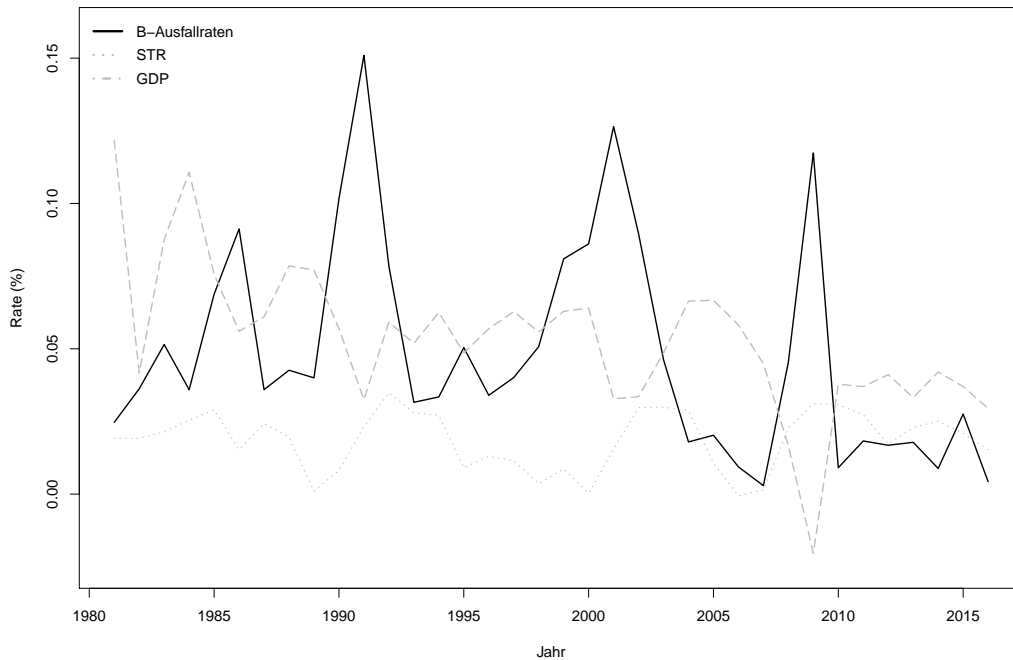


Abbildung 2: Konjunkturelle Abhängigkeit der Ausfallraten (B-Ratingkategorie) im Zeitraum 1981-2016 (Quelle: Eigene Darstellung).

4 Konjunkturabhängige Modellierung der Probability of Default

4.1 Regressionsbasierte Schätzung der Ausfallraten

Die zuvor bestätigte konjunkturelle Abhängigkeit der Ausfallraten soll im Folgenden dazu genutzt werden, um unter Hinzunahme weiterer prognoserelevanter ökonomischer Indikatoren Ausfallwahrscheinlichkeiten für die künftige Periode mithilfe eines Regressionsmodells zu schätzen. Aus den Vorüberlegungen in Kapitel 3 folgt, dass – neben der eigenen kurzfristigen Vergangenheit der Ausfallraten (DR_{t-1}) – insbesondere zwei ökonomische Faktoren (STR_{t-1} und $SP500_{t-1}$) eine hohe Prognoserelevanz besitzen. Auf dieser Basis sollte sich die Forderung des IASB hinsichtlich der Ermittlung des Hauptbestandteils des neuen LECL-Ansatzes, der Probability of Default, fundiert erfüllen lassen.

In methodischer Sicht bietet sich für die Schätzung ein Rückgriff auf das lineare Regressionsmodell auf Basis der Methode der kleinsten Quadrate an (Greene, 2012, S. 66 ff.). In diesem Rahmen stellen dynamische Modelle, die nicht nur durch kontemporäre Variablen, sondern ebenfalls durch verzögerte Variablen gekennzeichnet sind, eine geeignete Ausgestaltung zur PD-Schätzung dar. Ein Spezialfall derartiger Modelle sind autoregressive Ansätze, in denen die erklärte Variable (Ausfallraten) zusätzlich zu verzögerten Variablen (Makrofakto-

Variable	AA	A	BBB	BB	B	CCC.C
DR_{t-1}	-0.0528	0.2731 [°]	0.2535	0.2306 [°]	0.4956**	0.3166**
DR_{t-2}	-0.0545	-0.1369	-0.0093	0.0736	0.0738	0.0849
DR_{t-3}	-0.0562	-0.2229	-0.2292	0.0338	-0.0824	-0.0566

Tabelle 2: Korrelationen verschiedener Ausfallraten mit eigener Historie (1981-2016) (eigene Darstellung). *** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$, [°] $p < 0.15$.

ren) durch ihre eigene Vergangenheit⁵ erklärt wird (Hackl, 2013, S. 221 ff.). Die in Abschnitt 3.2 festgestellten linearen Abhängigkeiten der externen Faktoren und die darauf aufbauend analysierte Eigendynamik der Ausfallraten (Abschnitt 3.3) erlauben es in einer Gesamtschau, die Interaktion zwischen Konjunktur und Ausfallraten über Formel (3) darzustellen:

$$DR_t = \alpha_0 + \delta_1 STR_{t-1} + \delta_2 SP500_{t-1} + \delta_{DR} DR_{t-1} + u_t, \quad (3)$$

wobei Parameter α_0 den konstanten, konjunkturunabhängigen Teil darstellt, δ_{DR} den autoregressiven Einfluss gewichtet und δ_1 sowie δ_2 die Gewichtungen der ökonomischen Faktoren quantifizieren. Der Fehlerterm wird durch u_t repräsentiert, der als normalverteilt mit Erwartungswert Null und Varianz σ^2 angenommen wird.

Die Parameterschätzungen für die (sechs) verschiedenen Regressionsmodelle basierend auf den beobachteten Ausfallraten der einzelnen Ratingkategorien sind in Tabelle 3 zusammengefasst, in der zusätzliche Werte zur Verteilung der resultierenden Residuen, die Aufschluss über die Passgenauigkeiten des gewählten Modells ergeben, aufgeführt sind. Die Modelle AA bis CCC.C stellen die Regressionen der Ausfallraten gemäß Formel (3) dar. Die (fast) durchgehend negativen Vorzeichen der Parameter δ_1 und δ_2 sind erwartungsgemäß auf die bereits analysierte negative Korrelation der Ausfallraten mit den ökonomischen Faktoren zurückzuführen (vgl. Abschnitt 3.2). War der Zinsspread der Vorperiode hoch, werden c.p. durchschnittlich

Variable	Datenreihen					
	AA	A	BBB	BB	B	CCC.C
α_0	0.0007**	0.0013**	0.0043***	0.0185***	0.0597***	0.2908***
STR_{t-1}	-0.0281*	-0.0378*	-0.1016*	-0.4237*	-1.5584**	-3.6445 [°]
$SP500_{t-1}$	-0.0003	-0.0012	-0.0047	-0.01823	-0.0717*	-0.3361**
DR_{t-1}	-0.0828	0.1808	0.1642	0.1478	0.5089***	0.3003*
R^2	0.1530	0.2033	0.2454	0.2309	0.4933	0.3206
Adj. R^2	0.0711	0.1263	0.1724	0.1565	0.4443	0.2548
Standardfehler	0.0007	0.0010	0.0025	0.0105	0.0276	0.1141

Tabelle 3: Modellvergleich mit historischen S&P Daten (1981-2016 – Einzelne Ratingklassen (eigene Darstellung) *** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$, [°] $p < 0.15$.

die erwarteten Ausfallraten sinken. Analoge Überlegungen gelten für die verzögerte Variable $SP500_{t-1}$. Sofern bezüglich dieser Variable die prozentuale Änderung der Vorperiode erhöht ist, werden aktuelle Ausfallraten im Mittel erwartungsgemäß geringer ausfallen. An dieser

⁵Hier wird ein autoregressiver Term erster Ordnung ($AR(1)$), d.h. eine um eine Periode verzögerte Zeitreihe, verwendet.

Stelle wird – wie bereits in der Analyse des autoregressiven Terms in Abschnitt 3.3 – erneut deutlich, dass eine Schätzung basierend auf Ausfallraten höchster Bonität⁶ (AA–A) keine validen Ergebnisse liefert. Die Faktoren weisen in diesen Fällen zwar den erwarteten negativen Einfluss auf, dieser ist jedoch durchgehend auf keinem gewöhnlichen Niveau (alle $> 15\%$) signifikant (vgl. Tab. 3). Dieser Befund kann jedoch – wie bereits gezeigt – insofern nicht überraschen, als es nur sehr wenige Beobachtungen von ausgefallenen Unternehmen gibt, die vorher mit höchster Bonität bewertet worden sind und sich demnach für diese Art der Analyse nicht gut eignen.

Die vier weiteren Modelle, basierend auf den Ausfallraten der Ratingkategorien BBB–CCC.C, stellen sich als sehr erklärungskräftig heraus. Sämtliche im Modell verwendeten Variablen weisen sehr niedrige p -Werte auf (durchgehend $< 15\%$) und lassen daher auf eine hohe Signifikanz der verwendeten Makrofaktoren schließen. Zusätzlich wird deutlich, wie sehr die Ausfallraten von ihrer eigenen Vergangenheit beeinflusst werden. Die Gewichtungsfaktoren (δ) der entsprechenden $AR(1)$ -Terme weisen insbesondere bei den Modellen B und CCC.C durchgängig stark positive Werte (≤ 0.30) auf. Dies bedeutet, dass sie innerhalb des Modells einen hohen Erklärungsgehalt liefern und zusätzlich auch eine hohe Signifikanz (niedrige p -Werte) ausweisen, was einen deutlichen Einfluss der (kurzfristigen) eigenen Vergangenheit auf aktuelle Ausfallraten bestätigt. Die Residuen u_t zeichnen sich bei der durchgeführten Modellanpassung durch einen Erwartungswert nahe Null und eine relativ geringe Standardabweichung aus. Beide Befunde sind Anzeichen für eine gute Passgenauigkeit dieses Modells bezüglich der hier verwendeten Daten. Ferner zeigen die durchgeführten Analysen zur Überprüfung der Residuen, dass es insbesondere bei der Verwendung der erklärten Variablen AA–BBB zu Verletzungen der Annahme der Normalverteilung kommen kann, was sich wiederum auf die im Untersuchungszeitraum geringe Anzahl beobachteter Unternehmensausfälle innerhalb der entsprechenden Ratingkategorien zurückführen lässt. Insgesamt ergibt sich ein aussagekräftiges Modell zur Erklärung zukunftsorientierter Ausfallraten in den einzelnen Ratingkategorien, dessen Struktur sich auf wenige zentrale Parameter beschränkt. Dieses Modell soll im nächsten Abschnitt zu einem Migrationsmodell erweitert werden, das in der Lage ist, eine mehrjährige PD-Struktur im Einklang mit IFRS 9 aufzubauen.

4.2 Erweiterung zur migrationsbasierten Ausfallstruktur

Die vorstehende Analyse der Ausfallraten zeigt auf, wie sich deren zeitliche Dynamik über eine Autoregression und den Einfluss zusätzlicher, externer Konjunkturvariablen erklären lässt. Damit ist ein wichtiger Grundstein für eine PIT-Modellierung der Ausfallstrukturkurve gelegt, die im Folgenden durch ein Migrationsmodell ausgebaut wird. Vor allem im Hinblick auf weiter in der Zukunft liegende Ausfallereignisse werden in diesem Zusammenhang weitergehende Informationen über das Migrationsverhalten zwischen sämtlichen Bonitätsklassen des Nicht-Ausfalls benötigt.

Zunächst sei für jeden beliebigen Kredit angenommen, er folge im Migrationsverhalten einer Markov-Kette $(X_t)_{t>0}$ (Norris, 2009, S. 1 ff.; Henking et al., 2006, S. 150 ff.). Das Migrationsverhalten wird dabei für alle Kreditnehmer als homogen vermutet. Es lässt sich demnach im

⁶Die höchste Bonitätsstufe AAA wird hier nicht betrachtet, da im betrachteten Zeitraum keine Ausfälle beobachtet werden konnten.

Intervall $(t - 1, t]$ für alle Kredite durch die Migrationsmatrix

$$M_t := \begin{pmatrix} m_{11}^{(t)} & \dots & m_{1,7}^{(t)} & m_{1,8}^{(t)} \\ \vdots & \ddots & \vdots & \vdots \\ m_{7,1}^{(t)} & \dots & m_{7,7}^{(t)} & m_{7,8}^{(t)} \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

beschreiben. Die letzte Zeile der Matrix bleibt aufgrund der Modellannahme des absorbierenden Default-Zustandes mit Nullen und einer Eins für alle t fest. Die acht bezifferten Bonitätsklassen inklusive des Ausfalls entsprechen den Rating-Klassen AAA, AA, ..., CCC.C. Die Einträge

$$m_{i,j}^{(t)} := P(X_t = j | X_{t-1} = i)$$

der Migrationsmatrix geben die Wahrscheinlichkeit an, dass die Kette am Ende der Periode t den Zustand j erreicht, bedingt auf den Ausgangszustand i am Ende der Vorperiode $t - 1$. Jede Migrationsmatrix hat ausschließlich positive Einträge und alle Zeileneinträge summieren sich jeweils zu Eins. Jede Matrixzeile ist damit eine diskrete Wahrscheinlichkeitsverteilung, die das Migrationsverhalten ausgehend von dem betreffenden Ausgangszustand vollständig beschreibt.

Für den konkreten Aufbau einer gesamten Ausfallstruktur wird eine Folge von Migrationsmatrizen M_t benötigt, die jeweils die betrachteten und sich aneinander anschließenden Übergangszeiträume $t_0 < t = t_1, t_2, \dots, T$ beschreiben. Diese Matrizen können sich grundsätzlich in jedem dieser Zeitpunkte individuell unterscheiden, was in Anbetracht eines gewissen konjunkturellen Einflusses auf alle Matrixelemente auch die unmittelbare Konsequenz wäre. Bei diesem allgemeinen Vorgehen kann demnach prinzipiell für jeden Übergangshorizont eine dort jeweils repräsentative und durch die dann angenommene Konjunktur (mit-)bestimmte Migrationsmatrix herangezogen werden. Die kumulative Migrationsmatrix im Zeitintervall $(t_0, T]$ mit wechselndem Periodeneinfluss ergibt sich durch Matrixmultiplikation als:

$$\overline{M}_T = \prod_{t=t_1, t_2, \dots}^T M_t.$$

Mit dieser Vorgehensweise lassen sich die kumulierten Ausfallwahrscheinlichkeiten leicht ablesen und die unbedingte Ausfallwahrscheinlichkeit für die Periode t ergibt sich einfach durch: $PD_t = \text{kum}PD_t - \text{kum}PD_{t-1}$, wobei sich $\text{kum}PD_t$ bonitätsklassenspezifisch direkt aus der Default-Spalte der zugehörigen Matrix \overline{M}_t ablesen lässt. Diese Differenzen der kumulierten Ausfallwahrscheinlichkeiten von aufeinanderfolgenden Intervallen sind genau die unbedingten Ausfallwahrscheinlichkeiten, die dann auch für die LECL-Formel (2) benötigt werden.

Würde als naive Vereinfachung für zukünftige Perioden stets auf die aus den initialen Informationen in t_0 prognostizierte Migrationsmatrix M_{t_1} zurückgegriffen, so ergäbe sich die Ausfallstruktur mit $T = 1, 2, 3, \dots$ über:

$$\overline{M}_T = (M_{t_1})^T.$$

Eine auf diesen Überlegungen basierende Ausfallstruktur erfüllt durchaus den gewünschten PIT-Charakter, sobald alle relevanten, zum Bewertungszeitpunkt t_0 vorliegenden Informationen zur generellen Bestimmung von M_{t_1} einbezogen wurden. Andererseits scheint dieser Ansatz

dennoch nicht sonderlich angemessen zu sein, da die für die initiale Matrix M_{t_1} veranschlagten Konjunktoreinflüsse in jedem weiteren Zeitraum erneut Einzug in das Übergangverhalten der Markov-Kette erhalten würden. Dadurch kann es zu einer über- oder unterbewerteten Ausfallstruktur immer dann kommen, wenn in t_0 eine eher schlechte oder eher gute Konjunkturlage zu konstatieren war. Folglich muss unbedingt berücksichtigt werden, dass zukünftig auch andere – heute ggf. nur schwer prognostizierbare Bedingungen – vorherrschen können.

Mithilfe von Modell (3) können unbedingte Ausfallwahrscheinlichkeiten $m_{i,8}^{(t_1)} := PD_{t_1}^{(i)}$ für sämtliche Bonitätsklassen $i = 1, \dots, 7$ zum Informationsstand am Ende von t_0 statistisch vorhergesagt werden. In einem zweiten Schritt besteht dann noch die Notwendigkeit, die Resteinträge der Matrix M_{t_1} abseits der Default-Spalte zu füllen. Einerseits ist davon auszugehen, dass auch diese Migrationswahrscheinlichkeiten unter einem gewissen konjunkturellen Einfluss stehen, andererseits erhöht eine explizite Modellierung auch dieser Abhängigkeiten die Anzahl der Parameter und damit die Komplexität des Modells nicht unwesentlich. Eine Korrelationsanalyse der einzelnen Einträge der historischen Migrationsraten im Zeitverlauf – analog zu den (Regressions-)Analysen der Ausfallraten – zeigt nur schwache lineare Abhängigkeiten auf. Dieses Ergebnis ist insofern wenig überraschend, da die Ratingmigrationen auf durch die Agenturen intern beschlossenen Ratingänderungen basieren. Ausfälle dagegen werden maßgeblich durch die deutlich objektivere Definition charakterisiert und mit der direkten Beobachtung einer Zahlungsverzögerung ausgelöst. Der vermeintlich stärkste Konjunktoreinfluss an dieser Stelle hält folglich bereits über die Ausfallspalte Einzug in die Matrix. Die verbleibenden Einträge werden – als erster Vereinfachungsschritt – mit durchschnittlichen und konjunkturunabhängigen Migrationswahrscheinlichkeiten aufgefüllt. Dazu wird die empirische Durchschnittsmatrix $M = (m_{ij})_{i,j=1,\dots,8}$ herangezogen, um damit die Matrix

$$M'_t = \begin{pmatrix} m_{11} & \dots & m_{1,7} & m_{1,8}^{(t_1)} \\ \vdots & \ddots & \vdots & \vdots \\ m_{7,1} & \dots & m_{7,7} & m_{7,8}^{(t_1)} \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

zu bilden. Im Allgemeinen erfüllt diese zusammengesetzte Matrix die obligatorischen Eigenschaften der Migrationsmatrix nicht, da etwa die Zeilensummen wohl kaum gerade Eins getragen werden. Zur Gewährleistung dieser Eigenschaft wird die Zeilennormierung

$$\hat{m}_{i,j} := \frac{m_{j,j}}{m_{i,8}^{(t_1)} + \sum_{j=1}^7 m_{i,j}}, \quad \text{mit } i = 1, \dots, 8$$

angewendet, um Migrationsmatrix $M_{t_1} := (\hat{m}_{i,j})_{i,j=1,\dots,8}$ zu finalisieren. Dieser einfache Ansatz erlaubt es, durch deutliche Vereinfachungen die erste, aber entscheidende Hürde in Richtung PIT-Ausfallstrukturkurve zu nehmen. Es ist insbesondere davon auszugehen, dass diese den aktuellen PIT in einem wesentlich stärkeren Maße charakterisiert als es die anschließenden Migrationsmatrizen $(M_t)_{t>t_1}$ für die gesuchte Struktur tun, die zudem auch nur mit sehr hoher Unsicherheit prognostizierbar sind. Dadurch wird die Idee zu einem zweiten Vereinfachungsschritt motiviert, für alle weiter in der Zukunft liegenden Perioden eine zeithomogene und damit konjunkturunabhängige Dynamik zu unterstellen. Dieser Ansatz ist äquivalent zur Wahl:

$$M =: M_{t_2} =: M_{t_3} =: M_{t_4} =: \dots =: M_T.$$

Die generelle Prognostizierbarkeit der Migrationsmatrizen ab t_2 scheint prinzipiell sehr problematisch zu sein, da für diese Prognose zwei mit Unsicherheit behaftete Prognoseschritte notwendig sind: Erstens wird eine Prognose der erwarteten Konjunkturlage verlangt, zweitens müsste eine geeignete – mit weiterer Unsicherheit versehene – Dynamik der bedingten Migrationsmatrix davon abgeleitet werden. Der vorgeschlagene Ansatz ließe eine solche flexiblere konjunkturbedingte Modellierung von M_{t_2} und auch M_t mit $t_2 > t$ grundsätzlich zu, würde dabei aber verlässliche Konjunkturprognosen voraussetzen.

Das vorgeschlagene Vorgehen zur PIT-Ausfallstrukturkurve soll nun anhand der S&P-Ratingdaten getestet werden und damit die Anwendung aus Abschnitt 4.1 erweitern. Die historischen Jahremigrationsdaten der Ratingagentur S&P werden zum einen als Referenzpunkt für Ausfallstrukturkurven und zum anderen als Input für das eigene Vorhersagemodell verwendet. Die Wahl des Jahres 2002 als Beispiel-PIT bietet ein historisches Sample von 20 Jahren zur Parameterschätzung und erlaubt als Vergleich zugleich eine 14-jährige Benchmark-Struktur. Die Vorhersage verwendet dabei lediglich die Informationen, die in 2002 zur Verfügung standen, um die PD-Kurven zu prognostizieren. Die für diesen Informationsstand resultierenden Parameter sind in Tabelle 4 angegeben und beschreiben den gewählten PIT 2002 für die Anwendung von Modell (3). Das skizzierte Vorgehen erlaubt den Aufbau der PIT-Ausfallkurven,

Variable	Datenreihen					
	AA	A	BBB	BB	B	CCC.C
α_0	0.0003	0.0011	0.0058	0.0337	0.0784	0.3290
STR_{t-1}	-0.0127	-0.0267	-0.1217	-0.6624	-18.800	-45.051
$SP500_{t-1}$	0.0005	-0.0007	-0.0055	-0.0498	-0.0597	-0.2815
DR_{t-1}	-0.1359	0.1418	-0.0061	-0.1209	0.4088	0.2507

Tabelle 4: Parameterwerte von Modell (3) angewendet auf die S&P Ausfallraten im Zeitraum von 1981–2002 (eigene Darstellung).

die als kumulative PDs in Abbildung 3 für die vier schlechtesten Ratingklassen B–CCC.C im Vergleich zur historischen Benchmark, d.h. der ab 2002 tatsächlich beobachteten Ausfallkurven, dargestellt sind.

Die Abbildung zeigt, dass die historischen Kurven einen schwankenden Verlauf aufweisen, d.h. mal stärker und mal schwächer ansteigen, und darüber hinaus zum Teil deutlich unterhalb der getroffenen Prognosen liegen. In Richtung einer langfristigen Prognose wird ersichtlich, dass der homogene Migrationstrend ab 2004, erfasst über die Durchschnittsmatrix des Zeitraums von 1981 bis 2001, das tatsächliche Ausfallrisiko über den nächsten Zeitpunkt hinaus offenbar überschätzt. Schließlich kann in Bezug auf die Modellierungsanforderungen des LECL nach IFRS 9 festgehalten werden, dass der PIT-Charakter über die vorgestellte PD-Struktur sinnvoll einbezogen werden kann. Der Einsatz dieser Methode ist demnach zu einem aktuellen Bewertungszeitpunkt möglich und greift dafür lediglich auf zu diesem Zeitpunkt beobachtbare Informationen zurück.

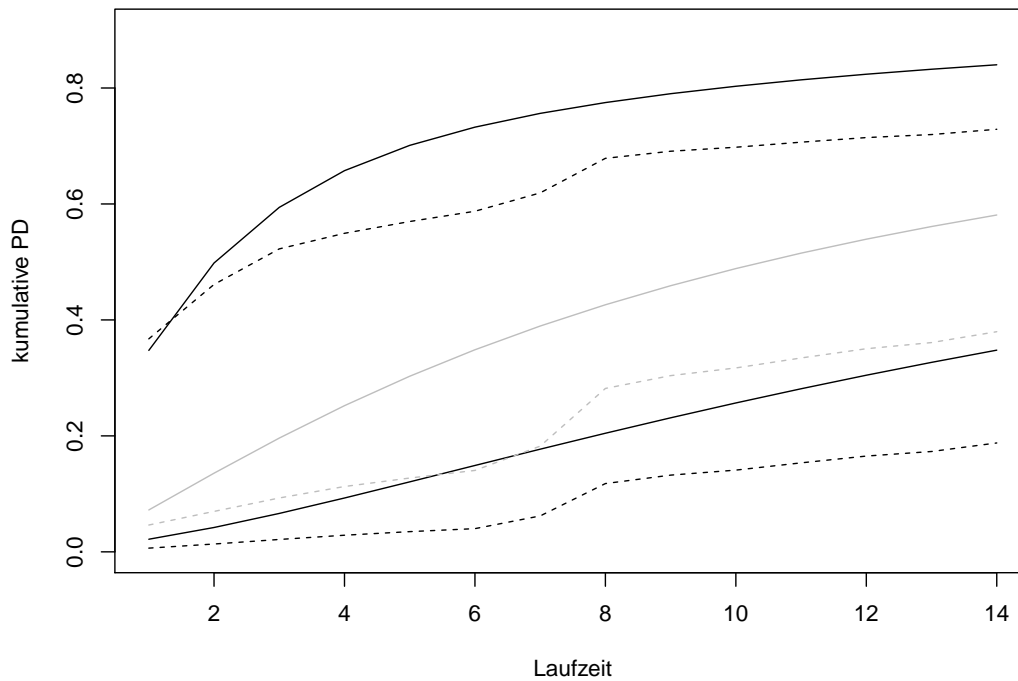


Abbildung 3: Die kumulativen S&P Benchmark Kurven (gestrichelt) zu den Ratingklassen CCC.C (schwarz oben), B (grau) und BB (schwarz unten) im Vergleich zu den Prognosen (durchgezogen) (Quelle: Eigene Darstellung).

5 Fazit

Die Berücksichtigung konjunktureller Effekte in der Bestimmung des LECL als Bewertungsmaßstab für die Erfassung der Wertminderung gemäß IFRS 9 stellt sowohl theoretisch als auch in der praktischen Umsetzung eine große Herausforderung dar. Der vorliegende Beitrag präsentiert bezüglich dieser Fragestellung eine standardkonforme Umsetzungsmöglichkeit mit Fokussierung auf die Probability of Default als zentralem Kernparameter für stichtagsbezogene „Point-in-Time“-Prognosen basierend auf der erwarteten Konjunkturentwicklung. Der entwickelte Ansatz bettet sich dabei in bestehende Konzepte des Risikomanagements ein bzw. erweitert diese, um den in Banken typischerweise zur Anwendung kommenden „Through-The-Cycle“-Ansatz zu überwinden.

Nach einer kompakten Vorstellung des Wertminderungskonzepts des IFRS 9 sowie einer Präzisierung des Begriffs der Ausfallwahrscheinlichkeit zeigt die durchgeführte empirische Untersuchung von S&P-Ausfallraten einen (linearen) Zusammenhang mit Konjunkturindikatoren, von denen insbesondere Frühindikatoren wie Zinsspreads und Aktienindizes (S&P500) eine starke Vorhersagekraft für bevorstehende Ausfälle aufweisen. In einem nächsten Schritt wird auf

Basis der Vorüberlegungen ein kompaktes regressionsbasiertes Prognosemodell abgeleitet, in das auch die festgestellte autoregressive Eigendynamik der Ausfallraten eingeht. Dieses Modell wiederum ist die Grundlage für ein Migrationsmodell, mit dem die Ausfallstruktur nach IFRS 9 – inklusive zusätzlicher Intuitionen bezüglich der PD-Verläufe – erfasst und mithilfe von S&P-Daten demonstriert werden kann. Die zentralen Schritte des vorgestellten zweistufigen PD-Modellierungskonzeptes sind demnach (1.) die „Ein-Schritt“-Prognose der Ausfallraten, die prognostizierte Ausfallwahrscheinlichkeiten der direkt an den Bewertungszeitpunkt anschließenden Periode ermöglicht und (2.) die anschließende Erweiterung zu periodenspezifischen Migrationsmatrizen unter Vernachlässigung weiter in der Zukunft liegender Konjunkturprognosen. Im Ergebnis ist das vorgestellte PD-Modell in der Lage, selbst unter vereinfachenden Annahmen für die verbleibenden Parameter EaD und LGD eine den Anforderungen des IFRS 9 entsprechende konjunkturabhängige LECL-Schätzung zu ermöglichen.

Zwangsläufig gehen mit jeder Komplexitätsreduktion aber auch Schwächen in Bezug auf die Aussagekraft des Modells einher: So führt im vorgeschlagenen Lösungsansatz die Verwendung von Migrationsmatrizen durchschnittlicher Konjunkturausprägung ab der übernächsten Periode häufig zu einer Unter- oder Überschätzung der zukünftigen Ausfallkurven. Hintergrund dieser Unschärfe ist offenbar ein initialer „Schock“ in Schritt (1.), der im weiteren Verlauf recht stark aufrechterhalten wird, wohingegen er in den empirischen Kurven vermutlich durch weitere konjunkturelle Schwankungen in die Gegenrichtung ausgeglichen wird. Eine weitere anfängliche Ungenauigkeit, verursacht durch die Prognose aus (1.), betrifft den teilweise großen Standardfehler (vgl. Tab. 3), der eine ungünstige Unschärfe bereits in der initialen PD-Prognose induziert und sich angesichts der zuvor angesprochenen Problematik im weiteren Verlauf noch ausdehnen wird. Trotz dieser Schwächen beschreibt das präsentierte Modell einen Rahmen, in dem unpräzisen und damit offenen Vorgaben des IFRS 9 zumindest etwas eingegrenzt und empirisch getestet werden können.

Literatur

- Bachmann, U. (2004), Die Komponenten des Kreditspreads: Zinsstrukturunterschiede zwischen ausfallbehafteten und risikolosen Anleihen, Gabler Edition Wissenschaft, Deutscher Universitätsverlag, Wiesbaden.
- BCBS (2004), International Convergence of Capital Measurement and Capital Standards, Bank for International Settlements.
URL: <https://www.bis.org/publ/bcbs107.pdf?noframes=1>
- Bluhm, C., Overbeck, L. und Wagner, C. (2003), An Introduction to Credit Risk Modeling, Chapman & Hall/ CRC, CRC Press, Boca Raton.
- Büschgen, H. E. (1998), Bankbetriebslehre: Bankgeschäfte und Bankmanagement, 5. Aufl., Gabler, Wiesbaden.
- Duffie, D. und Singleton, K. J. (2003), Credit Risk: Pricing, Measurement, and Management, Princeton Univ. Press, Princeton, NJ [u.a.].
- Engelkamp, P. und Sell, F. L. (2017), Einführung in die Volkswirtschaftslehre, 7. Aufl., Springer, Berlin.
URL: <http://dx.doi.org/10.1007/978-3-662-53961-3>
- Engelmann, B. und Ermakov, K. (2011), Transition Matrices: Properties and Estimation Methods, in: Engelmann, B. und Rauhmeier, R. (Hrsg.), The Basel II Risk Parameters, Springer, Heidelberg [u.a.], S. 103–116.
- Everling, O. (1991), Credit Rating durch internationale Agenturen: Eine Untersuchung zu den Komponenten und instrumentalen Funktionen des Rating, Gabler, Wiesbaden.
- EY (2015), Wertminderung finanzieller Vermögenswerte nach IFRS 9, Ernst & Young, Stuttgart.
- Filusch, T. und Mölls, S. H. (2017), „(Lifetime) Expected Credit Losses“ als Mechanismus der Verlustantizipation nach IFRS 9: Operationalisierung der Schätzparameter in kritischer Betrachtung, KoR IFRS Kapitalmarktorientierte Rechnungslegung **17**(6), S. 249–255.
- Greene, W. H. (2012), Econometric Analysis, Pearson Series in Economics, 7. Aufl., Pearson, Boston [u.a.].
- Große, J.-V. (2011), Auf der Suche nach dem richtigen Impairmentmodell, IRZ: Zeitschrift für internationale Rechnungslegung **6**(10), S. 419–426.
- Grünberger, D. (2011), Der Lifetime-Expected Loss, IRZ: Zeitschrift für internationale Rechnungslegung **6**(1), S. 29–37.
- Grünberger, D. (2013), Kreditrisiko im IFRS-Abschluss: Handbuch für Bilanzersteller, Prüfer und Analysten, Schäffer-Poeschel, Stuttgart.
- Hackl, P. (2013), Einführung in die Ökonometrie, 2. Aufl., Pearson, München.
URL: <http://lib.myilibrary.com/detail.asp?id=650988>

- Hartmann, W. und Schwarzhaupt, O. (2007), Die Bedeutung des Ratings im Bankkreditgeschäft, in: Büschgen, H. E. und Everling, O. (Hrsg.), Handbuch Rating, Gabler, Wiesbaden, S. 25–46.
- Hartmann-Wendels, T., Pfingsten, A. und Weber, M. (2015), Bankbetriebslehre, 6. Aufl., Springer, Berlin [u.a.].
- Hayden, E. und Porath, D. (2011), Statistical Methods to Develop Rating Models, in: Engelmann, B. und Rauhmeier, R. (Hrsg.), The Basel II Risk Parameters, Springer, Heidelberg [u.a.], S. 1–12.
- Henking, A., Bluhm, C. und Fahrmeir, L. (2006), Kreditrisikomessung: Statistische Grundlagen, Methoden und Modellierung, Springer, Berlin [u.a.].
- Henselmann, K., Ditter, D. und Holstein, T. (2014), Too little, too late? Ein empirisch gestützter Beitrag zur Bilanzierung der Kreditrisikovorsorge, KoR: Internationale und Kapitalmarktorientierte Rechnungslegung IFRS **14**(7/8), S. 355–362.
- IASB (2014), IFRS 9: Financial Instruments, International Financial Accounting Standard Board.
URL: <http://www.ifrs.org>
- Läger, V. (2002), Bewertung von Kreditrisiken und Kreditderivaten, Risikomanagement und Finanzcontrolling, Band 6, Uhlenbruch, Bad Soden.
- Löffler, G. und Posch, P. (2011), Credit Risk Modeling Using Excel and VBA, 2. Aufl., Wiley, Hoboken, NJ.
- Norris, J. R. (2009), Markov Chains, Cambridge Series on Statistical and Probabilistic Mathematics, 15. Aufl., Cambridge Univ. Press, Cambridge.
- O’Sullivan, A. (2003), Economics: Principles in Action, Pearson/Prentice Hall, Needham, Massachusetts.
- Serfling, K., Badak, E. und Jeiter, V. (2007), Möglichkeiten und Grenzen des Credit Ratings, in: Büschgen, H. E. und Everling, O. (Hrsg.), Handbuch Rating, Gabler, Wiesbaden.
- S&P (2017), Default, Transition, and Recovery: 2016 Annual Global Corporate Default Study And Rating Transitions, Standard & Poor’s Financial Services LLC.
URL: <https://www.spratings.com>

Tobias Filusch*

Calibrating and Forecasting Point in Time Probability of Default Term Structures

Abstract

The lifetime expected credit loss has recently come to the fore because of a new impairment model included in the International Financial Reporting Standard (IFRS) 9. A corresponding model can be implemented around the focus on the probability of default (PD) term structure, which in this case must be predicted from the perspective of a particular point in time (PIT). Such a forecast requires the inclusion of the current as well as an outlook for the future economic condition. This paper introduces and applies suitable forecast models that can reflect the course of historical PD term structures and predict them based on economic indicators. The two presented models perform similarly well and provide features necessary for IFRS 9 impairment but lacked in the academic literature so far.

Contents

1	Introduction	83
2	Lifetime Probability of Default Structure for the IFRS 9 Standard	85
2.1	Lifetime Expected Credit Losses	85
2.2	Transition Modeling and Model Validation	87
3	Modeling Point In Time Probability of Default Term Structures	89
3.1	Transition Matrix Model	89
3.2	The Time-Dependent Generator Model	90
3.3	Model Parameter Forecast	92
4	Model Application and Validation Study	93
4.1	Model Calibration to Historical Data	93
4.2	Economic Indication	97
4.3	Probability of Default Term Structure Forecasting	100
5	Conclusions	104

*Dipl.-Math. oec. Tobias Filusch, Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, ifG Marburg, Am Plan 2, D-35032 Marburg.

1 Introduction

In the aftermath of the latest global financial crisis, accounting standard setters initiated major reforms to address the dangers of “too little, too late” accounting (FSF, 2009; FCAG, 2009). In this regard, the International Accounting Standard Board (IASB) published the new International Financial Reporting Standard (IFRS) 9 as the successor for the International Accounting Standard (IAS) 39. IFRS 9 introduces many modifications, including a more early consideration of financial risk for the impairment of financial instruments. The proposed impairment model is based on expected credit losses (ECLs) for a wide range of financial instruments measured at amortized costs, but especially credits (IASB, 2014). The Basel Committee on Banking Supervision (BCBS) supports the proposed use of ECLs (BCBS, 2017) and requests to extend the concepts of existing credit risk models like their internal ratings-based (IRB) approach (BCBS, 2004). Therefore, most approaches so far propose to split the ECL into three core parameters: the exposure at default (EAD), loss given default (LGD), and probability of default (PD) (Grosskord et al., 2013; Grünberger, 2013). Furthermore, two IFRS 9 requirements stand out as crucial, yet possibly cumbersome differences compared to the IRB methodology. These include considering a lifetime expected credit loss (LECL) for instruments whose default risk has significantly increased and the general scope of a point in time (PIT) outlook that is explicitly required to include the economic conditions (IASB, 2014, 5.5.3-5.5.4, 5.5.17). Since the targeted LECL covers multiple periods, a term structure meeting those requirements must be in place for the three parameters.

The PD term structure seems to be the most important element in this scenario for two reasons. First, it describes the default risk which is also used to measure a significant deterioration of the credit quality and, second, concerning the LECL methodology itself, it may be the most closely related parameter to the given economic phase. Without advanced models for EAD and LGD in this regard, this requirement of the IFRS 9 concerning the LECL impairment can already be established by using a sufficient PD term structure. Existing models from risk management reflect a through-the-cycle (TTC) approach (Gerhold et al., 2017). Hence, an appropriate PD term structure forecast can be considered the focal point for LECLs. Three important steps must be accomplished by any model approach, yet all three remain insufficiently treated in the academic literature: precisely replicating observed PD term structures corresponding to different PITs, extending this to allow for reasonable PD term structure forecasts based on available, forward-looking economic information at any particular PIT, and, ultimately, establishing an appropriate validation procedure composed to evaluate and compare forecasts and corresponding forecast models.

Previous research has concentrated on the estimation of PDs. Notable model classes encompass survival models (Shumway, 2001; Stepanova and Thomas, 2002; Jafry and Schuermann, 2004; Banasik et al., 2017), Bayesian models (Kadam and Lenk, 2008; Stefanescu et al., 2009), regression models (Crook and Bellotti, 2010; Leow et al., 2014), and intensity models (Jarrow et al., 1997; Lando and Skødeberg, 2002; Duffie et al., 2007; Leow and Crook, 2014). In addition to PDs, some studies focus on transition probabilities or intensities in accordance with a certain number of risk classes (e.g. rating classes). This leads to transition matrices describing the dynamic between these classes; they are also used for the pricing of credit derivatives (Jarrow et al., 1997) or in risk management (Gavalas and Syriopoulos, 2014).

Rating transition probabilities have been found unstable over time (Nickell et al., 2001), even though agencies usually decide upon rating changes with a TTC point of view (S&P, 2018).

Belkin et al. (1998), Kim (1999), Wei (2003), and Trück (2008) propose and extend the factor model class that takes time-varying transition matrices driven by business cycle effects into account. Another approach for such a dynamic is suggested by Bangia et al. (2002) and Fei et al. (2012), who use a Markov regime switching model to include two or three different business cycle phases. However, an accurate forecast of the transition matrix – howsoever such a forecast is actually evaluated (Jafry and Schuermann, 2003) – may be unnecessary with regard to the goal of forecasting PD term structures. Bluhm (2007) introduces a non-homogeneous continuous time Markov chain model (NHCTMCM) approach.¹ They in particular present an efficient way to calibrate it to a historical PD term structure resulting from Standard and Poor’s default data. In contrast to the approaches above, the NHCTMCM uses a TTC transition matrix, or at least its time-continuous counterpart, as a starting point.² This starting point seems to be a reasonable extension of already existing concepts in risk management. Furthermore, the calibration only requires few parameters in addition to the TTC matrix. Even though the NHCTMCM provides such strong properties, it has not been revived to establish IFRS 9-compliant PIT PD term structures. This paper focuses on providing this step.

We demonstrate how the NHCTMC model can be reduced in complexity to the rNHCTMCM while maintaining its flexibility to reflect observed PD term structures. We introduce an extension of the rNHCTMCM that allows to forecast PIT PD term structures based on economic indicators. Our reduced version only requires one parameter for each non-default risk class for calibration; the forecast mechanism adds just another two parameters, respectively. As a benchmark for calibrating and forecasting PD term structures, we introduce a straightforward PIT generalization of the TTC approach: the transition matrix model (TMM) starts with a TTC transition matrix, whereby we employ the first subsequent period’s default probabilities as the model parameters. For all future periods, average and, in particular, constant transition behavior is assumed by applying the TTC matrix. For both models, we propose the same forecast add-on. It establishes a connection to economic indicators, e.g. interest rate spreads and stock indices (Ivanova et al., 2000; Pearce and Roley, 1985; Pearce, 1983). We validate PD term structure models with the (equally-weighted) mean squared error (MSE) principle. It provides a distance measure between the model-induced PD term structure and the historical counterparts over the same horizon.

Our contribution is to introduce IFRS 9-compliant model approaches that forecast PD term structures based on leading indicators for the economy. Our application study shows how the emerging PD term structure is predicted far better than the TTC forecast can. While the latter approach is not PIT-oriented and, consequently, not IFRS 9-compliant, we add models that provide this aspect. We apply the models to S&P rating transition data, which enables the analysis of PD term structures starting from different PITs within the history. Comparing the two models, the rNHCTMCM appears to be more sophisticated. In fact, it provides continuous PD term structures, while the discrete TMM provides PDs only at certain grid

¹This approach is also included in Bluhm and Overbeck (2007).

²A time-independent generator matrix is required, which can be calculated from the TTC transition matrix or directly estimated from transition data (Israel et al., 2001).

points. In addition, the forecast error of the TMM is slightly smaller. All in all, the analysis suggests that both models are almost equivalent.

The paper is structured as follows. Section 2 provides the general framework of LECL in accordance with IFRS 9 accounting. Section 3 then moves on to introduce our model approaches, which we eventually apply and compare in Section 4. In the concluding Section 5, we outline our findings.

2 Lifetime Probability of Default Structure for the IFRS 9 Standard

2.1 Lifetime Expected Credit Losses

The IASB prescribes a LECL methodology without providing details on a concrete calculation concept. The use of the core parameters LGD, EAD and PD familiar from risk management applications is a reasonable – if not necessary – attempt to link the accounting standard to already established methods. Drawing on this natural approach, we first introduce how to deduce the corresponding LECL representation. Loss calculations start by considering a potential loss event that may occur over the lifetime T of a certain instrument; it is represented by the stochastic variable L . Considering the current point in time $t_0 = 0$, a default may occur during the subsequent lifetime $(t_0, T]$. The indicator variable $\mathbb{1}_{(t_0, T]} = 1$ remains zero as long as no default occurs. When a loss occurs, its amount is quantified by the random variables *realized exposure at default* $rEAD_{(t_0, T]}$ and *realized loss given default* $rLGD_{(t_0, T]}$. The rEAD takes the outstanding future cash flow as well as the time value of money into account. Corresponding to the interval at hand, rLGD determines the portion of rEAD that is actually lost.³ Thus, the loss (random) variable is given by

$$L = \mathbb{1}_{(t_0, T]} \cdot rLGD_{(t_0, T]} \cdot rEAD_{(t_0, T]}. \quad (1)$$

If lifetime T covers a rather short period of time, e.g. 12 months, it seems reasonable to only include such a single period. Unfortunately, the longer the lifetime is, the worse this approach will turn out. Consequently, this necessitates a more detailed evolution of the risk over the lifetime. For example, one scenario may be an early default within this interval, where nothing has been re-paid yet. Otherwise, the default may happen directly preceding the end date T , in which case several payments would have already been made. Thus, $rEAD$ differs significantly between these two scenarios. Taking care of this issue, dates $t_0 < t_1, \dots, t_{n-1} < t_n = T$ can be established to divide the lifetime into n intervals. During any (potentially irregularly spaced) interval $(t_{i-1}, t_i], i = 1, \dots, n$, a loss event may occur, which is characterized by the indicator variable $\mathbb{1}_{(t_{i-1}, t_i]}$ as above. Once the indicator turns one, it remains zero for the remaining intervals, which means this approach excludes the possibility of multiple default events. Finally, the sum over the intervals' marginal loss variables forms the lifetime loss variable through

$$L = \sum_{i=1}^n \mathbb{1}_{(t_{i-1}, t_i]} \cdot rLGD_{(t_{i-1}, t_i]} \cdot rEAD_{(t_{i-1}, t_i]}. \quad (2)$$

³Financial securities may work against a loss of the entire exposure at default.

Note that for a given interval, $rEAD$ takes all outstanding payments from the remaining lifetime into account. Here, the exact number of remaining intervals and their length is crucial for precision. As stated above: fewer, longer intervals require averages (over time) to be considered and lead to rather inaccurate estimates of actual losses.

The three core parameters are not necessarily stochastically independent. However, if we make this assumption, the LECL is given by

$$\begin{aligned}
E(L) &= E\left(\sum_{i=1}^n \mathbb{1}_{(t_{i-1}, t_i]} \cdot rLGD_{(t_{i-1}, t_i]} \cdot rEAD_{(t_{i-1}, t_i]}\right) \\
&= \sum_{i=1}^n E\left(\mathbb{1}_{(t_{i-1}, t_i]}\right) \cdot E\left(rLGD_{(t_{i-1}, t_i]}\right) \cdot E\left(rEAD_{(t_{i-1}, t_i]}\right) \\
&= \sum_{i=1}^n PD_{(t_{i-1}, t_i]} \cdot LGD_{(t_{i-1}, t_i]} \cdot EAD_{(t_{i-1}, t_i]}, \tag{3}
\end{aligned}$$

where the last step defines the expected per period values of the LGD and EAD. Moreover, the probability of default (PD) results from the expected value of the indicator variable.⁴ In this formula the variable PD characterizes the probability of a loss during a certain period, prior to which no default event has happened. The term “unconditional PD” is often used to distinguish this PD from the PD conditional on the survival of the credit until the end of the previous period. The PD term structure describes the evolution of the unconditional PD or equivalently its sum, the cumulative PD, over all relevant future periods. Even if the exposures following a given future loss are known – for example based on an observable credit contract – major uncertainty still lurks because of the PD term structure, which weighs every potential loss scenario of future periods.

The IASB does not prescribe calculation requirements towards the PD term structure. They generally focus on the LECL determination, for which they prescribe several requirements. Every financial instrument should be treated on an individual basis to begin with, although collective treatment may be applicable in some cases (IASB, 2014, 5.5.11). The lifetime of the instrument must be exactly reflected: a period that is longer or shorter cannot be considered instead (IASB, 2014, 5.5.19). The time value of money must be included with a “probability-weighted amount” that considers a certain “range of possible outcomes” (IASB, 2014, 5.5.17 (a,b)). LECL implementation should also include “reasonable and supportable information that is available without undue cost or effort at the reporting date about past events, current conditions and forecasts of future economic conditions” (IASB, 2014, 5.5.17 (c)). If any core parameter in equation (3) – e.g. the PD term structure – meets these requirements, it implicitly holds for the LECL as well. However, further details on the exact implementation of a reasonable PD term structure are not provided. Before we can actually suggest concrete models, we first discuss the desired properties like the expected shape of PD term structures. We also discuss how different risk classes evolve and how to treat them. Finally, we introduce a validation principle for evaluating PIT PD term structure forecasts, which ultimately helps to identify a reasonable model approach.

⁴We define $LGD_{(t_{i-1}, t_i]} := E(rLGD_{(t_{i-1}, t_i]})$ and $EAD_{(t_{i-1}, t_i]} := E(rEAD_{(t_{i-1}, t_i]})$

2.2 Transition Modeling and Model Validation

Since any credit portfolio covers a wide spectrum of different credit qualities, we must define classes of different risk profiles. The mapping to a certain risk class requires a definition, where every class represents a different default probability. Credit risk may also change over time, which the transitions between risk classes characterize. We consider a certain number of risk classes k for establishing the frequencies of relative defaults and transitions from one risk class to another. Then in general, the cohort method enables to determine empirical transition matrices, which is equivalent to the maximum likelihood estimator of a transition matrix (Hanson and Schuermann, 2006). A time-homogeneous transition matrix characterizes the average transition dynamic over a certain time span. It treats different periods of the same length homogeneously and accumulates the observations. Matrices for individual periods allow to distinguish the dynamic of the different periods regarded, but they often lack a sufficient data basis for estimation. Consequently, the estimates may contain a rather high standard deviation, which in particular leads several entries of the resulting matrix to have a transition or default probability of zero. This seems inappropriate because any transition should be at least possible. Nevertheless, capturing this varying transition dynamic period by period appears to be necessary and, basically, helps to create empirical PD term structures for different time windows. Such a PIT PD term structure is completely defined by a specific series of $k \times k$ -transition matrices $(M_t)_{t=1, \dots, T}$, with each $M_t = (m_{ij,t})_{i,j=1, \dots, k}$ representing the transition probabilities between the classes $\{1, \dots, k\}$ in an interval $(t-1, t]$. We arrange the classes corresponding to their risk profile in the following way: we implement an ascending order and set k as the default state. Thus, $m_{ik,t}$ represents the default probability for class i during period $(t-1, t]$.⁵ Note that we will only take equidistant periods of length equal to one into account, whereby intervals $(0, 1], \dots, (T-1, T]$ cover the lifetime T of the regarded instrument.

Since the matrix multiplication $(*)$ enables to link the transition over subsequent periods, it leads to cumulative transition matrices

$$\overline{M}_t := M_1 * \dots * M_t. \quad (4)$$

The PD term structure based on matrices $(\overline{M}_t)_{t=1, \dots, T}$ is then extracted by the respective k -th columns of those matrices, namely

$$cPD_t := (\overline{M}_t)_{\cdot k}.^6 \quad (5)$$

This PD term structure describes the (cumulative) default probability in $(0, t]$, which may consist of different sub-periods already accumulated in matrix \overline{M}_t . The differences in those cumulative probabilities lead to unconditional PDs by

$$PD_t := cPD_t - cPD_{t-1}. \quad (6)$$

The unconditional PD describes the default probability in a certain period after surviving, but not conditional on the survival, in all previous periods. Moreover, it is required for the LECL (3). Both the unconditional and the cumulative PD are basically suitable as the forecast

⁵For all t , we set $m_{k,t} = (0, \dots, 1)'$.

⁶ $(M)_{\cdot k}$ denotes the vector equal to the k -th column of a matrix M .

objective, as in particular (6) makes it easy to convert. The cumulative PD appears to be the better choice, though, because it has two favorable properties: it is monotonically increasing in t and converges to one (for $t \rightarrow \infty$).⁷ The latter is easy to explain because over an infinite number of periods, one should expect all credits to default at one point. Statistically, this holds true for any credit with default probability bigger than zero that does not converge to zero. These assumptions seem reasonable. First, defaults should at least be possible in each period, and second, PDs tend to fluctuate within their range $(0, 1)$, e.g. related to the business cycle (Nickell et al., 2001), rather than converging towards one of the two limits.

The attention of this section now shifts to the question how to evaluate PD term structure forecasts based on empirical defaults during predefined periods. At PIT $t = 0$, we assume the term structure $(c\overline{PD}_{t,T}^{(i)})_{t>0}$ of length T is predicted for risk classes $i = 1, \dots, k - 1$, which leads to $k - 1$ curves. The validation at T measures the quality of this forecast after observing the portfolio movement. In other words, it compares the forecast contingent upon information set \mathcal{F}_0 (e.g. the information given at $t = 0$) to the at time T observed PD term structure $(cPD_{t,T}^{(i)})_{t>0}$. We propose evaluating the quality of forecasts by measuring the difference between the forecast and observation with the mean squared error (MSE). By accumulating means over the term structures' length and the included risk classes, our overall measure for validation is given by

$$\frac{1}{T(k-1)} \sum_{t=1}^T \sum_{i=1}^{k-1} (c\overline{PD}_{t,T}^{(i)} - cPD_{t,T}^{(i)})^2. \quad (7)$$

MSE (7) provides a simple evaluation principle that determines the PD term structure forecasts' distance from a benchmark and can be used to compare the accuracy of different models. As discrete benchmarks – e.g. extracted from the rating transitions – are more accessible than continuous, only grid points $1, \dots, T$ are considered and accumulated. However, it is easy to establish a continuous generalization if necessary, too. The models' goal is to achieve a small MSE and, in particular, IFRS 9 practitioners expect any PIT model to inevitably outperform the TTC PD term structure resulting from unconditional and time-independent transition probabilities (Gerhold et al., 2017). We denote the TTC PD term structure by $cPD_{t,T}^{(i)}(M)$, with M being the unconditional transition matrix.

The MSE allows to compare forecasts and the associated models. The next step is to introduce models that are capable of reflecting empirical PD term structures. This means we can calibrate them to given data in such a way that it reflects the data reasonably well i.e. with a small MSE. We must consider different PITs and, subsequently, different PD term structures to verify that the models are able to react to different conditions. Finally, we are interested in extending these models to predict PD term structures at different PITs and the information given then.

⁷Note that according to (6), PD_t then converges to zero.

3 Modeling Point In Time Probability of Default Term Structures

The starting point in modeling PD term structures is the TTC approach, which is well-known in practice (IRB-approach) and allows to forecast PD term structures following a single transition matrix M . However, we must develop this approach and include PIT-specific information. We now introduce a direct extension to the TTC approach, which serves the outlined purpose. In addition, we take on the NHCTMCM introduced by Bluhm (2007) and introduce a reduced version – the rNHCTMCM. This section adds the calibration routines of the models and, finally, proposes a forecast feature that includes economic indicators for a PIT prediction.

3.1 Transition Matrix Model

Equation (4) establishes a PD term structure with a sequence of (empirical) transition matrices. For a corresponding forecast based on a given information set \mathcal{F}_{t-1} , reasonable transition matrices $M_t, M_{t+1}, \dots, M_{t+T}$ must be predicted for future periods $(t-1, t], \dots, (T-1, T]$. This allows a PD term structure to be set-up. Unfortunately, we would have to estimate a total of $k^2 - k$ parameters for each period, which equals the dimension of a single, unknown transition matrix (without the default state's row).⁸ For the complete PD term structure T periods, such an element-by-element approach is unwieldy and, moreover, even the prediction of the first subsequent period's transition matrix appears unfeasible. However, as the original objective is to predict PD term structures, we may focus on just few of those parameters, while keeping the remaining entries of the matrix constant; we mainly stay close to the TTC probabilities.

From the perspective of $t-1$, we focus on the first matrix M_t of the targeted matrix sequence. In most prediction attempts, it has the priority as we cannot expect better prediction results for periods further ahead. Moreover, out of this matrix, the default row (the conditional PDs) should have the most impact on the PD term structure. In fact, these parameters even present the first points of the PD term structure. Consequently, we use them as parameters for the *Transition Matrix Model* (TMM) and add the TTC transition probabilities from matrix M to complete matrix M'_t in the following way. We denoted the parameters by $p_t \in [0, 1]^{k-1}$ and, thus, we deduce

$$M'_t = \begin{pmatrix} m'_{11,t} & \dots & m'_{1\ k-1,t} & m'_{1\ k,t} \\ \vdots & \ddots & \vdots & \vdots \\ m'_{k-1\ 1,t} & \dots & m'_{k-1\ k-1,t} & m'_{k-1\ k,t} \\ 0 & \dots & 0 & 1 \end{pmatrix} := \begin{pmatrix} m_{11} & \dots & m_{1\ k-1} & p_t^{(1)} \\ \vdots & \ddots & \vdots & \vdots \\ m_{k-1\ 1} & \dots & m_{k-1\ k-1} & p_t^{(k-1)} \\ 0 & \dots & 0 & 1 \end{pmatrix}. \quad (8)$$

In general, this matrix will not meet the requirements of a stochastic matrix, i.e. a matrix with row sums equal to one. We normalize M'_t to get the forecast transition matrix $M_t :=$

⁸Note, the k -th row of the matrix is assumed to represent the absorbing default state. This means all transition probabilities to states other than k are fixed at zero by default.

$(m_{ij,t})_{i,j=1,\dots,k}$ by

$$m_{ij,t} := \frac{m'_{ij,t}}{\sum_{j=1}^k m'_{ij,t}}. \quad (9)$$

Thereafter, we set all remaining matrices of the sequence equal to the TTC matrix, i.e. $M_{t+1} = \dots = M_{t+T} = M$. In other words, we are adjusting the TTC PD term structure approach by keeping the first years' PDs variable for calibration purposes. Conditional on TMM parameter $p \in [0, 1]^{k-1}$, we denote the resulting cumulative PD term structure of length T by $cPD_{t,T}^{(i)}(p)$. The TMM calibration to observed PD term structures based on the OLS method is then given by

$$\hat{p} = \arg \min_{p \in [0,1]^{k-1}} \left\{ \frac{1}{T(k-1)} \sum_{t=1}^T \sum_{i=1}^{k-1} (c\widehat{PD}_{t,T}^{(i)} - cPD_{t,T}^{(i)}(p))^2 \right\}, \quad (10)$$

where $c\widehat{PD}_{t,T}^{(i)}$ is the observed and $cPD_{t,T}^{(i)}(p)$ the model induced PD term structure.

First and foremost, the TMM approach focuses on the immediate future and on this basis develops a PIT PD term structure. Due to the implementation that uses matrix multiplication, however, the model parameter essentially affects the entire PD term structure. A TMM calibration for a multi-periodic range correspondingly leads to a parameter that reflects the course of a complete PD term structure. The TMM only requires $(k-1)$ parameters for calibration, which seems adequate to begin with. This approach furthermore leads to a general recognition towards implementing a parametrization within the matrix sequence of period-specify transition matrices in order to calculate PD term structures. Apparently, we are able to alter parts of the series or just parts of M to be able to establish a more flexible model that extends the TTC approach.

The TMM makes covering exact lifetimes of financial instruments difficult, because it requires fixed and equidistant period lengths. Thus, this model does not directly provide fractional period's transition or default probabilities. More precisely, users must overcome the "embedding problem", which stems from the problem of finding the underlying time-continuous process described by a generator matrix (Israel et al., 2001). As a matter of fact, many estimated transition matrices actually do not have an exact solution to this problem available. But since the continuous process is ultimately helpful or even necessary to cover exact lifetimes, we may be better off by extending the TTC approach with a continuous alternative to the just introduced TMM.

3.2 The Time-Dependent Generator Model

In contrast to the time-discrete approach above, a time-continuous transition model allows for a more precise coverage of arbitrary PD term structures. A generator matrix $Q = (q_{ij})_{i,j=1,\dots,k}$ describes a time-independent transition process through transition intensities q_{ij} between risk classes i and j . In general, the task to find a generator from a transition matrix may be troublesome. Even if such a generator exists it may not be unique (Israel et al., 2001). However, finding a generator Q following TTC matrix M is at least approximately possible by applying the matrix logarithm in combination with numerical smoothing. Alternatively, continuous transition observations can serve as the foundation for a direct estimation of the

generator (Albert, 1962). Given a generator matrix Q of a continuous process, determining the embedded time-discrete process of any time grid is easy: the transition matrix of a period $T \in \mathbb{R}$ is given by

$$M_T = \exp(T Q) := \sum_{l=0}^{\infty} \frac{(T Q)^l}{l!}. \quad (11)$$

The NHCTMC model of Bluhm (2007) then adds a time-dependent component to matrix Q through the diagonal matrix

$$\Phi_t(\alpha, \beta) = (\varphi_{ij,t}(\alpha, \beta))_{1 \leq i, j \leq k} := \begin{cases} 0 & \text{falls } i \neq j \\ \frac{(1-e^{-\alpha_i t})t^{\beta_i-1}}{1-e^{-\alpha_i}} & \text{falls } i = j < k \\ 1 & \text{falls } i = j = k \text{ (the default state),} \end{cases} \quad (12)$$

where $\alpha, \beta \in [0, \infty)^{k-1}$ are two $(k-1)$ -dimensional parameters. The product of this matrix and the generator Q that describes the time-independent transition of the portfolio eventually leads to the time-inhomogeneous generator matrix

$$Q_t(\alpha, \beta) := \Phi_t(\alpha, \beta) * Q. \quad (13)$$

Basically, the time-dependent component Φ_t scales every entry of generator Q by a row-specific factor. In particular, this transformation retains all generator matrix properties from Q , including rows that add up to zero with negative entries only on the diagonal.

With given parameters, equations (11) and (6) allow to determine the PD term structure of the NHCTMCM for any time grid t_0, \dots, t_n . Step one based on (11) leads to cumulative PDs, which are represented by the k -th columns of matrices $(\exp(Q_t(\alpha, \beta)))_{t=t_1, \dots, t_n}$. Then with (6), the differences of these cumPDs, i.e. $(\exp(Q_{t_i}(\alpha, \beta)))_{.k} - (\exp(Q_{t_{i-1}}(\alpha, \beta)))_{.k}$ for $i = 1, \dots, n$, provide the unconditional PDs.⁹ A clear advantage of this approach is that it can target arbitrary time grids and even reflect varying period lengths. Without losing generality, an equidistant time grid is again considered, namely the grid corresponding to the targeted historical PD term structure. Depending on parameter $(\alpha, \beta)'$, the model leads to the cumulative PD term structure

$$cPD_{t,T}^{(i)}(\alpha, \beta) := (\exp(Q_t(\alpha, \beta)))_{ik}, \text{ with } i = 1, \dots, k-1. \quad (14)$$

With this model-induced PD term structure, the NHCTMCM calibration to an observed PD term structure $\widehat{cPD}_{t,T}^{(i)}$ based on the OLS method is then given by

$$(\widehat{\alpha}, \widehat{\beta})' = \arg \min_{\alpha, \beta \in [0, \infty)^{k-1}} \left\{ \frac{1}{T(k-1)} \sum_{t=1}^T \sum_{i=1}^{k-1} (\widehat{cPD}_{t,T}^{(i)} - cPD_{t,T}^{(i)}(\alpha, \beta))^2 \right\}. \quad (15)$$

In total, the NHCTMCM requires 2 $(k-1)$ parameters for calibration purposes. Compared to the earlier introduced TMM, the NHCTMCM seems to be more flexible due to twice as many parameters. However, the two parameters per risk class makes it difficult to evaluate and later reflect the shifts of the PIT PD term structures through the model parameter.

⁹Note, $(\exp(Q_{t_0}(\alpha, \beta)))_{.k} = (0, \dots, 0)'$.

In this regard we hence introduce a reduced version, the *reduced NHCTMCM* (rNHCTMCM).

Taking a closer look at (12), one identifies how parameters α_i and β_i both affect row i – the transition intensities of instruments in risk class i . For the rNHCTMCM, we focus on parameter β , while keeping α constant at a certain choice $\alpha^* \in [0, \infty)^{k-1}$; this way we keep the row-specific scaling. The calibration of the model, thus, is given by

$$\widehat{\beta} = \arg \min_{\beta \in [0, \infty)^{k-1}} \left\{ \frac{1}{T(k-1)} \sum_{t=1}^T \sum_{i=1}^{k-1} (\widehat{cPD}_{t,T}^{(i)} - cPD_{t,T}^{(i)}(\alpha^*, \beta))^2 \right\}. \quad (16)$$

The rNHCTMCM has just $k-1$ model parameters – the same number as the TMM. Additionally, we have just one parameter for each risk class, which allows to establish a straight-forward connection to economic variables. Our next task is to develop such a forecast mechanism that general applicable for both the TMM and the rNHCTMCM.

3.3 Model Parameter Forecast

The last section introduced the calibration mechanism of the rNHCTMCM and the TMM, which allows to fit the models to observed PD term structures with respect to different PITs. Unfortunately, this result cannot provide PDs necessary for LECL measurement. This requires a mechanism that forecasts the future, unknown PD term structure based on given information. In fact, these information should include the economic conditions, which we proxy by using observable leading indicators.

The calibration results help to establish a certain connection to the state of the economy. Any historical PIT $t \in \{1, \dots, n\}$ can serve as a starting point for calibration as long as the available sample still includes the PD observation for $n+T$. For a fixed T , this allows to perform n calibrations and end up with the same number of model parameters representing the corresponding PIT t . More precisely, these parameters reflect the PD term structure that must be predicted in t and on the basis of economic information available by then. Let the calibrated model parameter – either for the TMM or rNHCTMCM – at PIT t be $\theta_t \in D^{k-1}$, with D being the domain of the respective model's parameter. Subsequently, periods $t+1, \dots, t+T$ belong to PIT t , which means only the information until t – information set \mathcal{F}_t – is available.

Let $z_t = (z_{1,t}, \dots, z_{l,t})'$ be a vector of available explanatory variables in \mathcal{F}_t . To establish a connection between the model parameter and a number of l indicators, we propose to use a linear model. This explains the timely dynamic of each $\theta^{(i)}$ with coefficients $\psi_{i,0}, \dots, \psi_{i,l}$ through equation

$$\theta_t^{(i)} = \psi_{i,0} + \psi_{i,1} z_{1,t} + \dots + \psi_{i,l} z_{l,t} + \epsilon_{i,t}, \quad (17)$$

where $\epsilon_{i,t} \sim i.i.d. \mathcal{N}(0, \sigma_i^2)$ is assumed for all $i = 1, \dots, k-1$. Moreover, the OLS estimation for this model is given by

$$\widehat{\psi}^{(i)} = \arg \min_{\psi \in \mathbb{R}^{l+1}} \left\{ \frac{1}{n} \sum_{t=1}^n (\theta_t^{(i)} - \psi_0 + \psi_1 z_{1,t-1} + \dots + \psi_l z_{l,t-1})^2 \right\}. \quad (18)$$

Next, we explain how this approach also allows to forecast the model parameter θ , which in turn provides a model-induced PD term structure forecast. Let $\widehat{\psi} \in \mathbb{R}^{k-1}$ be the estimate following (18), then the one-step ahead forecast in PIT t^* is deduced by

$$\bar{\theta}_{t^*}^{(i)} = \widehat{\psi}_{i,0} + \widehat{\psi}_{i,1} z_{1,t^*} + \dots + \widehat{\psi}_{i,l} z_{l,t^*}, \quad i = 1, \dots, k-1. \quad (19)$$

The thereby determined PD term structure forecast corresponds to periods $t^* + 1, \dots, t^* + T$. Note that, in general, another lifetime T would require another set-up. However, we expect shorter lifetimes – e.g. $T' < T$ – to be usable reasonably well. This model parameter forecast mechanism leads to a PD term structure forecast $cPD_{t^*+1,T}^{(i)}(\bar{\theta}_{t^*})$. Applied to the two particular models, this denotes the TMM-forecast by $cPD_{t^*+1,T}^{(i)}(\bar{p}_{t^*})$ and the rNHCTMCM-forecast by $cPD_{t^*+1,T}^{(i)}(\bar{\beta}_{t^*})$.

For both models, this forecast mechanism allows for PIT PD term structures that reflects the LECL requirements IASB (2014, 5.5.17 (c)). Indeed, we include information about ”past events” for the parameter estimation (18), then account for ”current conditions” for the forecast (19) and, finally reflect ”forecasts of future economic conditions”. The latter is achieved through the use of leading economic indicators that particularly represent ”reasonable and supportable information that is available without undue cost or effort at the reporting date”. The upcoming application section suggests a suitable selection with respect to these requirements.

4 Model Application and Validation Study

During this section, we examine the models’ performance in three steps. First, the model calibration to default data demonstrates how well the models reflect observed PD term structures. This study in particular considers 21 PD term structures of lifetime $T = 15$ corresponding to different PITs 1983 until 2002. Second, we use the estimates resulting from calibration and estimate the linear model for the forecast mechanism. This establishes the connection to the economic condition. Third, we use this model to forecast PIT PD term structures, which enables us to evaluate the forecast capabilities of the models. Ultimately, we clarify which indicator is best-suited for this attempt within the used data.

4.1 Model Calibration to Historical Data

Rating agencies like Standard & Poor’s (S&P) evaluate the creditworthiness of corporates. Based on their rating system, they aggregate issuers of similar risk profiles into a finite number of rating classes. Issuers are regularly reevaluated, which may happen on any given business day. Hence, a corresponding rating process is basically a time-continuous process with a finite number of states. S&P publishes annual default and transition information as well as several statistics free of charge, e.g. in S&P (2018). Since the early 80s, S&P evaluated about 19,725 companies, whereby the total number of rated companies has almost constantly increased from the initial 1,383. From low to high default risk, we consider their system of rating classes AAA, AA, A, BBB, BB, B, CCC.C, and D. Class D is the default class, which indicates past-due payments of the issuer according to a certain definition. Table 1 presents the average, one-year transition frequencies of the ”Global Company” sector for the period 1983

to 2017.¹⁰ Average transition matrices include several irregularity problems. For instance, there are frequencies of zero percent, which signifies an impossible transition between these classes over a one-year span (e.g. a direct default from class AAA is impossible according to Table 1). To avoid such irregularities, practitioners use adjustment algorithms to create normative correct transition matrices (Ong, 2000).¹¹ Further insights into the S&P rating methodologies as well as the exact definition of default are given in the appendix of the Default Study (S&P, 2018, App. I).

%	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	88.33	10.96	0.48	0.06	0.07	0.03	0.07	0.00
AA	0.54	90.88	7.87	0.53	0.07	0.08	0.02	0.02
A	0.04	1.94	91.80	5.55	0.42	0.17	0.01	0.06
BBB	0.01	0.17	4.42	89.72	4.54	0.79	0.14	0.22
BB	0.01	0.06	0.26	6.13	83.66	8.08	0.80	1.00
B	0.00	0.05	0.23	0.32	6.09	84.03	4.43	4.85
CCC/C	0.00	0.00	0.20	0.38	0.92	14.11	56.82	27.56
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Table 1: Average transition matrix according to transition data from S&P between 1983 and 2017 presented in percent. Source: own calculation based on S&P data.

The attempt of forecasting PD term structures with such an average transition matrix was pointed out earlier (TTC approach with M equal to Table 1). However, S&P also provides historical year-by-year transition rates within their study (e.g. S&P (2018, Table 20)). These frequencies indicate how the cohorts of each rating class evolved over a particular year. Correspondingly, they also allow to deduce estimates of conditional default probabilities for every calendar year between 1983 and 2017. We use these 35 matrices to define the 15-year PD term structure benchmarks $cPD_{t^*+1,15}, \dots, cPD_{t^*+15,15}$ associated to PITs $t^* = 1982, \dots, 2002$.¹²

On this basis, we estimate the model parameters, i.e. we calibrate the models to the data by applying (10), (15), and (16). Figures 6 and 7 illustrate the estimated values resulting from the calibration and their temporal developments over the various PITs. These estimates vary over time according to the different historical PD term structures they reflect within the model universe. Figure 1 illustrates the PD term structure benchmarks and the TMM calibration of rating classes A, BBB, BB, and B for PIT 2000. At this particular PIT, the derived MSE (over all seven rating classes and $T = 15$) for the TMM is 0.034%. If we complement this measurement by including all 21 PITs, the mean MSE (and standard deviation (SD)) of the TMM calibration is 0.065% (0.033%). For comparison, the left side of Figure 2 shows the results of the same application for the NHCTMCM. In this case, the MSE for PIT 2000 is 0.113%, whereas the mean MSE over all 21 lifetimes of the sample is 0.056% (0.03%). The third calibration applies the rNHCTMCM, which is initialized by using the NHCTMCM

¹⁰Due to only few issuers and, correspondingly, only sparse transition information within the first two years, we only analyze S&P data beginning from 1983.

¹¹In the first place, a corresponding procedure is typically necessary to exclude the "NR" class. Within this class, the agencies accumulate issuers that withdrawn their rating. The "NR"-class is not easy to integrate into the general hierarchy of classes, because it hardly allows to assess credit risk.

¹²Note that the last PIT we can consider depends on T . For $T = 15$, 2002 is the final PIT.

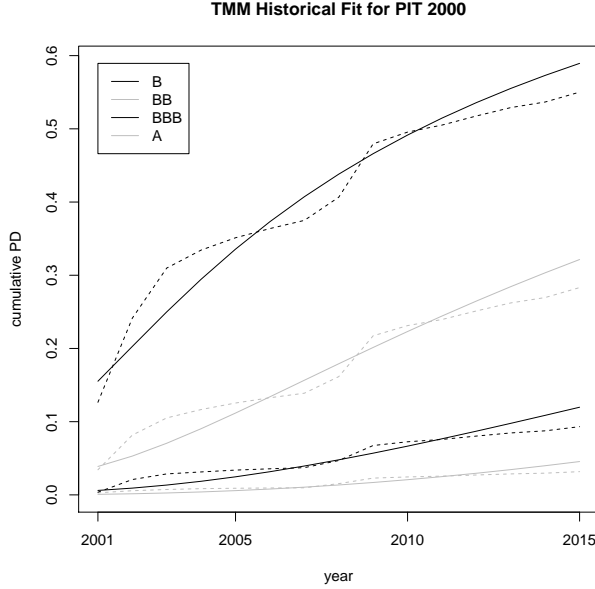


Figure 1: PIT 2000: historical PD term structures from the S&P data are dashed, while the fitted curves based on the TMM are solid.

average of parameters α_t . Thus, we set

$$\begin{aligned} \alpha^* &:= \frac{1}{21} \sum_{t=1982}^{2002} \hat{\alpha}_t \\ &= (3.634, 3.263, 2.074, 1.689, 1.936, 2.811, 3.606)'. \end{aligned} \quad (20)$$

The right-hand side of Figure 2 illustrates the results of calibrating the rNHCTMCM. The MSE for this PIT is 0.131%. Over the complete PIT history, the rNHCTMCM has a mean MSE (SD) of 0.065% (0.031%). The model calibration apparently leads to smoother PD term structures compared to the benchmarks; they do not reflect the benchmarks' regime switches. Hence, we in principle expect a certain MSE-level the models cannot outperform.

Table 2 summarizes the results of all individual PITs as well as provides additional statistics below. It can be seen from this table that the TMM and the rNHCTMCM perform similarly well. At the same time, the NHCTMCM appears to be superior; it has the smallest mean MSE and the smallest SD. To put the results into more context, Table 2 also indicates the MSE of the TTC approach. Note that using just matrix M leads to a single PD term structure, which in this application is compared to the different benchmarks of the associated PITs. This results in a mean MSE (SD) of 0.235% (0.162%). The left side of Figure 3 shows annual MSEs of the TTC approach and the NHCTMCM calibration. The mean MSEs (dashed) indicate the average level. The SD of the TTC model is rather high and relatively low MSEs only occur at few PITs. On the contrary, the calibration of the NHCTMCM is much less volatile. This observation is not surprising and can be interpreted as follows. Only average lifetimes correspond to the TTC PD term structure and consequently provide adequate MSEs

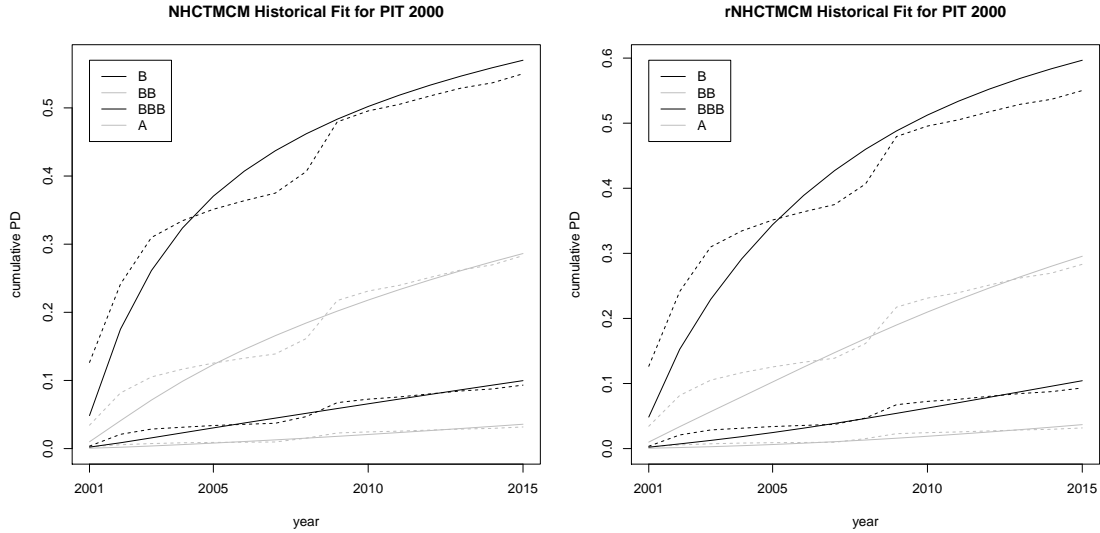


Figure 2: PIT 2000: historical PD term structures from the S&P data are dashed, while the fitted curves based on the NHCTMCM (left) and rNHCTMCM (right) are solid.

(e.g. at PIT 1993). The calibration, on the other hand, reacts to individual courses and finds appropriate model specifications that reflects them. Thus, it's MSE remains below the TTC outcome in every year. However, we must distinguish the TTC approach from the calibration and carry out the comparison with caution. The implementation of the calibration minimizes the distance between observation and model, resulting in the parameter that meets this criterion best. In contrast, the TTC approach is a time-homogeneous Markov chain forecast based on a single transition matrix; it only provides transition information and due to that makes the calculation of a PD term structure possible.

The right side of Figure 3 contrasts the calibration MSEs for the TMM and rNHCTMCM.¹³ Both provide similar good fits, but differ over the course of the included PITs. Interestingly, a conclusion that can be drawn from this comparison is that overall the TMM and the rNHCTMCM are on par in calibration. However, there are obvious differences in achieving this comparable good precision. The TMM has less flexibility, as in the long run, the course of the term structures is primarily determined by the constant TTC transition probabilities. The comparison of PD term structures corresponding to rating class B, for example, leads to the conclusion that the TMM is not suitable for creating a sufficient curvature to replicate the history, while the rNHCTMCM seems to have more flexibility for this (See Figure 2). To compensate for this weakness, the TMM massively adjusts the PD of the first year within the calibration, which at times produces weak results. However, Figure 2 shows that the rNHCTMCM also has difficulty capturing the PD of the first period. In contrast to the TMM, the parameter in the rNHCTMCM adjusts the transition intensities of generator Q , resulting in a more uniform influence on the entire PD term structure. This seems to offer better flexibility in the long run but overall does not lead to better results in the calibration

¹³Note that for the rNHCTMCM, α^* is chosen as in (20).

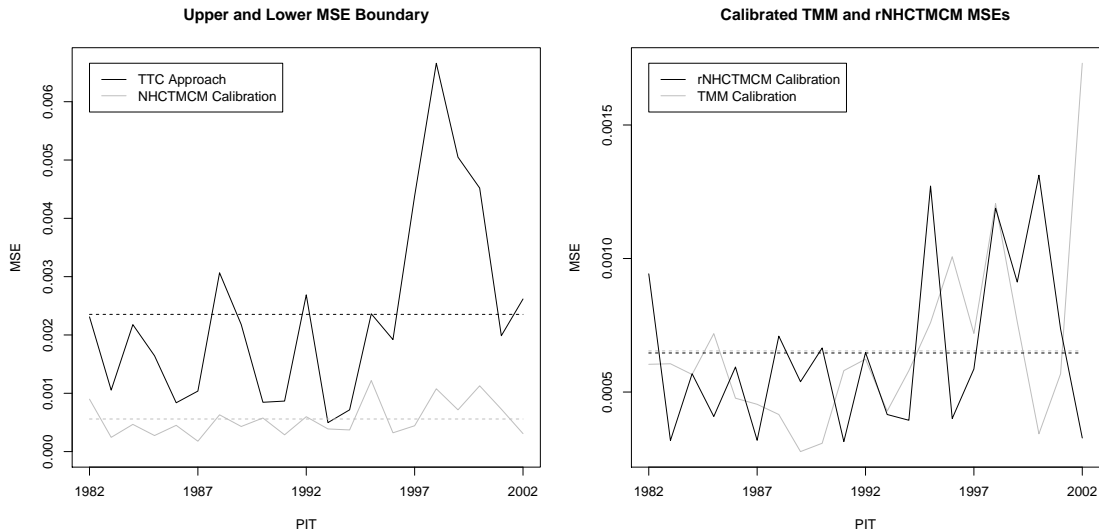


Figure 3: MSE for each PIT from 1982 until 2002. Left, the upper and lower limits for the planned forecasts are outlined by the TTC and NHCTMCM calibrations. Right, the rNHCTMCM and TMM calibration outcomes are presented. Means for every PIT are dashed.

study compared to the TMM. The next section analyses economic indicators to prepare the application of the forecasting model.

4.2 Economic Indication

The model calibration of the last section verified the capabilities of the TMM and the rNHCTMCM to reflect historical PD term structures that belong to different PITs. The NHCTMCM outperforms both models, but requires twice as many parameters. In addition, it is based on two parameters for each risk class, which makes it difficult to link the PIT effects contained in the PD term structures with these parameters; interdependency may be present in this regard. Therefore, we focus on the TMM and the rNHCTMCM within the forecast section, which, in particular, have the exact same number of parameters. Given the estimates from calibration, we establish a connection to economic conditions through economic indicators. We restrict the search for suitable predictors with respect to the following two requirements. First, in order to predict PD term structures, any indicator should provide forward-looking information. Second, they should be available at the corresponding PIT. The first requirement suggests the use of leading indicators, which have already shown that they indicate future developments. The second excludes variables like GDPs, which are summary in nature and, therefore, not immediately available.

Predictors for economic phases are frequently derived from market data like interest rate spreads or index returns (Fama, 1990; Ivanova et al., 2000). In particular, both indicator classes are a common component of composite leading economic indicators (Lahiri and Wang, 1994). Worth mentioning here are the Composite Leading Indicator (CLI) and the Leading Economic Index (LEI) (EU and UN, 2017). Prices from financial markets are available in real

in %	PDTS Calibration			
	TTC	NHCTMCM	rNHCTMCM	TMM
1982	0.231	0.090	0.094	0.060
1983	0.105	0.024	0.032	0.061
1984	0.218	0.047	0.057	0.057
1985	0.164	0.028	0.041	0.072
1986	0.084	0.045	0.059	0.048
1987	0.104	0.018	0.032	0.045
1988	0.307	0.063	0.071	0.042
1989	0.218	0.043	0.054	0.028
1990	0.085	0.058	0.067	0.031
1991	0.087	0.029	0.031	0.058
1992	0.269	0.060	0.065	0.062
1993	0.050	0.039	0.042	0.043
1994	0.072	0.037	0.039	0.058
1995	0.236	0.122	0.127	0.076
1996	0.192	0.032	0.040	0.101
1997	0.438	0.044	0.059	0.072
1998	0.666	0.108	0.119	0.121
1999	0.505	0.072	0.091	0.076
2000	0.452	0.113	0.131	0.034
2001	0.199	0.073	0.074	0.057
2002	0.262	0.031	0.033	0.173
MEAN	0.235	0.056	0.065	0.065
SD	0.162	0.030	0.031	0.033
MAX	0.666	0.122	0.131	0.173
MIN	0.050	0.018	0.031	0.028

Table 2: Calibration MSEs in % over 15-year horizons corresponding to single PITs within the sample of S&P transition data between 1983 and 2017.

time and, thus, suit our purposes.¹⁴ In total, we consider three variables. First, the December averages of spreads between ten-year and two-year US treasuries (T10Y2Y).¹⁵ Second, we derive the annual return of the S&P 500 index. The S&P 500 indicates the movement of the American stock market and clearly is one of the most recognized indices for the US economy. On that note, it appears to be crucial for the global economy, too, especially as US companies make up the majority of S&P’s rated companies which make up the default data. In order to capture the change in the market corresponding to the S&P 500 in a certain year, we consider the relative difference in monthly averages of daily index returns (JanS&P500).¹⁶ Finally, we include the amplitude-adjusted CLI for OECD countries – a publicly available

¹⁴E.g. from information providers like Bloomberg, Yahoo! Finance or Thomson Reuters as well as from other financial institutions like the FRED.

¹⁵We calculate these averages from daily data provided by the Federal Reserve Bank of St. Louis (FRED) and retrieved from “<https://fred.stlouisfed.org/series/>” on May 10, 2018.

¹⁶We calculate these returns from daily closing prices (in USD, adjusted for dividends and splits) from Yahoo! Finance, Historical Data, Instrument S&P 500 (GSPC). Retrieved from <https://de.finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC> on May 10, 2018.

leading economic indicator provided by the OECD.¹⁷

These indicators form the foundation of the linear model (17). Since the CLI contains both index and spread data, we only combine the other two and, otherwise, keep $l = 1$. The estimates of the linear model explain the temporal development of the observed PD term structures.¹⁸ Observations of the indicators then lead to forecasts of PD term structure following the linear one step ahead forecast. If we just include a single indicator, it adds 14 parameters to the model: the intercept and one coefficient for each non-default rating class. For each additional indicator, it adds another 7. Table 3 summarizes the results of the linear model estimation

Parameter	Intercept	T10Y2Y Coefficient	Adj. R^2
$\beta_t^{(1)}$	1.742***	11.773	-0.005
$\beta_t^{(2)}$	1.88***	8.164**	0.309
$\beta_t^{(3)}$	1.938***	-0.903	-0.037
$\beta_t^{(4)}$	1.926***	1.395	-0.045
$\beta_t^{(5)}$	1.872***	2.1	-0.007
$\beta_t^{(6)}$	2.223***	-12.048**	0.337
$\beta_t^{(7)}$	3.32***	-74.017*	0.15
$p_t^{(1)}$	0	0.059	0.065
$p_t^{(2)}$	0.001	0.026	-0.047
$p_t^{(3)}$	0.005*	-0.044	-0.048
$p_t^{(4)}$	0.016**	-0.283	-0.027
$p_t^{(5)}$	0.054***	-2.328***	0.481
$p_t^{(6)}$	0.186***	-7.891***	0.649
$p_t^{(7)}$	0.606***	-23.193**	0.283

Table 3: Estimation results of the calibrated TMM and rNHCTMCM model parameters and the indicator T10Y2Y. Significance codes indicate p-values $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*).

with respect to indicator T10Y2Y and Table 4 complements corresponding results for indicators T10Y2Y and JanS&P500 combined.¹⁹ In fact, it is hard to confirm a close connection between the indicators and the estimates on basis of these results because only few parameter series reveal a significant linear dependency (e.g. $p_t^{(6)}$). We can give three possible reasons for this. First, the historical PD term structure corresponding to at least some risk classes may simply not relate to economic phases. For example, since direct defaults from class AAA never occurred over a one-year span the term structures always start at PD zero and after that remain rather stable over the following years as well; in the long run they only shift due to the other classes' dynamics. Transferred to the parameters of the calibration this means there is also no significant temporal dynamic present. Therefore, we may debate the general approach of even including a time-varying PD term structure for credits within such risk

¹⁷Retrieved from <https://data.oecd.org/leadind/composite-leading-indicator-cli.htm> on April 10, 2019.

¹⁸We use the function "lm" from the "stats" package in R.

¹⁹See the Appendix for the estimates of the other indicators.

Parameter	Intercept	T10Y2Y Coefficient	JanS&P500	Adj. R^2
$\beta_t^{(1)}$	1.97***	0.311	-1.377*	0.151
$\beta_t^{(2)}$	1.896***	7.341*	-0.099	0.288
$\beta_t^{(3)}$	1.896***	1.194	0.252**	0.29
$\beta_t^{(4)}$	1.875***	3.968	0.309	0.02
$\beta_t^{(5)}$	1.848***	3.349	0.15	0.013
$\beta_t^{(6)}$	2.147***	-8.209*	0.461*	0.484
$\beta_t^{(7)}$	3.869***	-101.696*	-3.325	0.234
$p_t^{(1)}$	0.001	0.021	-0.005*	0.242
$p_t^{(2)}$	0.001	0.048	0.003	-0.088
$p_t^{(3)}$	0.003	0.067	0.013	0.023
$p_t^{(4)}$	0.012	-0.053	0.028	-0.006
$p_t^{(5)}$	0.048***	-2**	0.039	0.501
$p_t^{(6)}$	0.178***	-7.498***	0.047	0.638
$p_t^{(7)}$	0.713***	-28.544**	-0.643	0.326

Table 4: Estimation results of the calibrated TMM and rNHCTMCM model parameters and indicators T10Y2Y and JanS&P500 combined. Significance codes indicate p-values $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*).

classes. A second reason for a missing relationship may originate from the linear model itself. It only reflects linear dependencies that may not exist. Instead, there may only be non-linear dependencies the applied model cannot detect. More sophisticated, non-linear models could solve this potential issue. Thirdly and finally, the indicators may only provide short-term outlooks and, hence, fail to predict multi-periodic term structures reflected by each model parameter in the presented application. This explanation seems quite obvious and, indeed, is backed by the comparison of the two models. Taking a closer look at Table 3 it seems that only the TMM parameters have – at least for some risk classes’ parameters – a significant linear dependency to the indicator. The TMM’s parameters are particularly well linked to the short term PD and, therefore, reveal a stronger dependency to the spread. In contrast, the rNHCTMCM parameters appear to be more strongly embedded in the functional terms of the inhomogeneous generator construct, which is why linear dependency measurement does not come up with a lot.

At this point we would like to recall our main goal of making PD term structure forecasts. Basically, the linear model allows the identification of the PITs and also leads to a PD term structure prediction, which results from an integrated combination of several of these linear predictions. Therefore, the model valuation should be based on the PD term structure predictions carried out in the next section. Only after that a fair judgment is possible.

4.3 Probability of Default Term Structure Forecasting

The objective of this section is to let the two models compete against each other in forecasting. The presented indicators establish the model parameter forecast (19) at time t^* .

Consequently, observations of the indicators fuel the two superordinate models, which forms the group of competing forecast models. Implicitly we use the TTC transition matrix M (Table 1) for both models. This is easy to see for the TMM, which relies on the TTC transition matrix for all future periods. In contrast, the rNHCTMCM needs M as the starting point to determine the time-independent generator matrix Q . Since we cannot built on any data before 1982, the forecast study relies on M to produce a meaningful forecast sample.

Let $\bar{p}_{t^*,z}$ (TMM) and $\bar{\beta}_{t^*,z}$ (rNHCTMCM) be the model parameter forecasts based on indicator $z \in \{\text{T10Y2Y, JanS\&P500, CLI}\}$ and PITs $t^* = 1982, \dots, 2002$. Then the resulting 15-year PD term structure forecasts are denoted by

$$cPD_{t^*+1,15}(\bar{p}_{t^*,z}), \dots, cPD_{t^*+15,15}(\bar{p}_{t^*,z}) \quad (21)$$

and

$$cPD_{t^*+1,15}(\bar{\beta}_{t^*,z}), \dots, (cPD_{t^*+15,15}(\alpha^*, \bar{\beta}_{t^*,z})). \quad (22)$$

in %	T10Y2Y		JanS&P500		CLI		T10Y2Y & JanS&P500	
	rNHCTMCM	TMM	rNHCTMCM	TMM	rNHCTMCM	TMM	rNHCTMCM	TMM
1982	0.322	0.249	0.341	0.254	0.182	0.134	0.280	0.199
1983	0.090	0.064	0.105	0.072	0.170	0.134	0.077	0.057
1984	0.169	0.131	0.223	0.186	0.236	0.195	0.202	0.159
1985	0.059	0.072	0.071	0.064	0.068	0.070	0.036	0.069
1986	0.137	0.098	0.159	0.099	0.139	0.099	0.125	0.073
1987	0.027	0.038	0.035	0.050	0.069	0.054	0.037	0.050
1988	0.078	0.044	0.065	0.094	0.069	0.053	0.067	0.039
1989	0.065	0.028	0.061	0.037	0.050	0.031	0.076	0.025
1990	0.114	0.037	0.069	0.035	0.052	0.053	0.088	0.047
1991	0.027	0.055	0.384	0.190	0.229	0.108	0.111	0.163
1992	0.099	0.073	0.513	0.357	0.331	0.203	0.050	0.060
1993	0.062	0.038	0.162	0.095	0.164	0.094	0.050	0.038
1994	0.264	0.189	0.145	0.087	0.228	0.146	0.229	0.237
1995	0.381	0.354	0.350	0.268	0.282	0.229	0.345	0.263
1996	0.055	0.098	0.045	0.081	0.047	0.089	0.050	0.085
1997	0.044	0.067	0.057	0.123	0.054	0.104	0.045	0.075
1998	0.108	0.152	0.143	0.247	0.205	0.345	0.111	0.169
1999	0.085	0.094	0.113	0.177	0.087	0.094	0.084	0.092
2000	0.145	0.061	0.136	0.180	0.141	0.152	0.099	0.034
2001	0.167	0.276	0.090	0.084	0.158	0.144	0.087	0.168
2002	0.083	0.169	0.266	0.231	0.324	0.211	0.148	0.129
MEAN	0.123	0.114	0.168	0.143	0.156	0.131	0.114	0.106
SD	0.094	0.089	0.131	0.089	0.091	0.074	0.084	0.071
MAX	0.381	0.354	0.513	0.357	0.331	0.345	0.345	0.263
MIN	0.027	0.028	0.035	0.035	0.047	0.031	0.035	0.025

Table 5: PD term structure forecast MSEs (%) over 15-year horizons corresponding to single PITs within the sample of S&P transition data between 1982 and 2002.

The historical PD term structure $(cPD_t)_{t=t^*+1, \dots, t^*+15}$ allows to validate the made predictions; it reveals how far away from the observations each forecast is – measured by the MSE. Thereby, this study compares the two models and its' implicitly used predictor variable. This

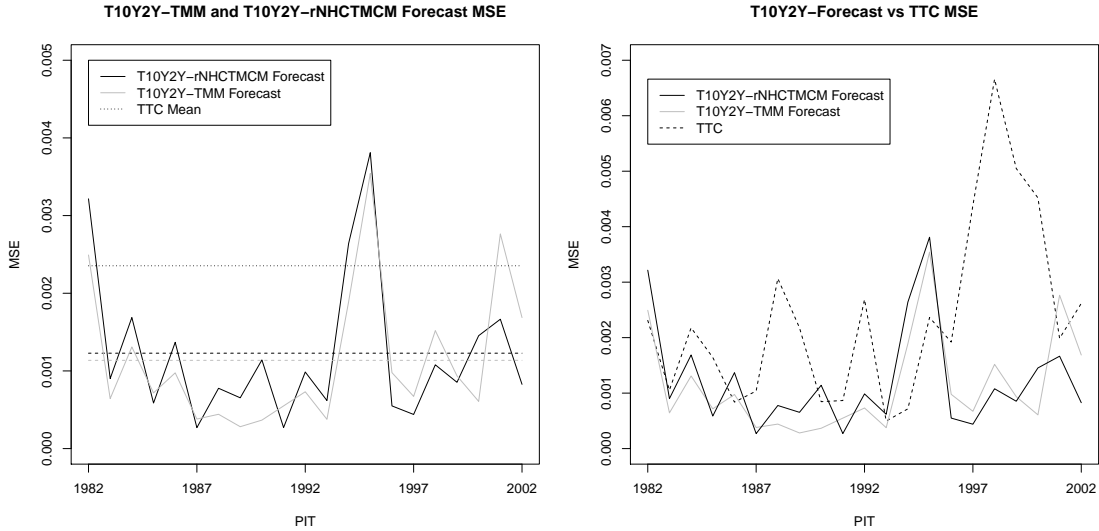


Figure 4: Forecast MSE for each PIT from 1982 to 2002. The T10Y2Y-rNHCTMCM (black) and the T10Y2Y-TMM (gray) forecast models are displayed corresponding to the values of Table 5. Mean MSEs are shown dashed. TTC MSE values can be found in Table 2 above.

measurement is presented in Table 5. In addition to the annual MSEs corresponding to the individual PITs, the table also includes statistics such as the mean MSE over all 21 PITs. The overall lowest mean MSE of 0.106% has the TMM based on the combined indicator T10Y2Y-JanS&P500. It also has the lowest SD (0.071%). Closely followed by the rNHCTMCM based on the same indicator, this model is superior to all others. Of course, it also refers to more parameters, which is why this result is not surprising.

Comparing the TMM and the rNHCTMCM, it can be seen from Figure 4 (left) or Figure 5 that the forecast MSEs of the two models differ over time, but also struggle at the same PITs (e.g. in 1982 or 1995). On average they almost coincide. The TMM appears to provide somewhat better results, which may be due to the stronger linear dependence of the model parameters and a correspondingly better forecast. The mean MSE of the TTC approach can also be seen on the figures. It is important to point out that the forecast models only exceed this level a few times. Two of these exceptions belong to PITs 1994 and 1995. We observe a peak during this time for both models and can give a good explanation for it. As a matter of fact, at this time the financial crisis becomes part of the benchmark. It includes an increase in default risk within the corresponding PD term structure, particularly in 2009 and 2010. Interestingly, the right plots, respectively, illustrate how the TTC approach is even less capable to reflect the looming crisis; the TTC's MSE continues to rise, while the other models seem to be responding to the new situation from 1996 onwards.

Solely based on the S&P 500, the forecast capabilities of the two models turn out to be slightly worse, respectively. Thus, the interest rate spread appears to have a better predictive power concerning the PD term structure; it also outperforms the CLI, which is designed to predict economic conditions. We conclude that we are dealing with a different issue here,

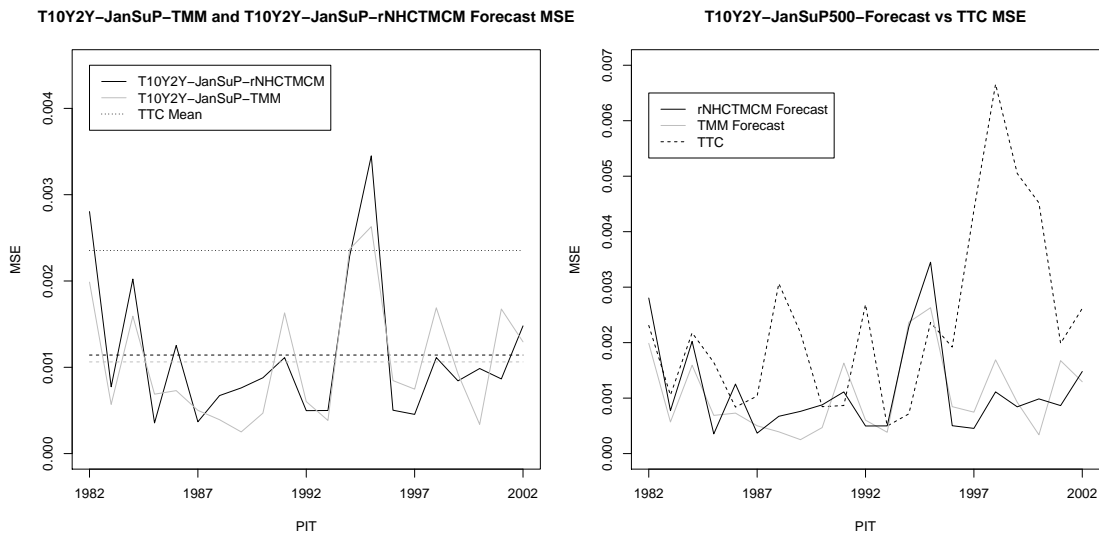


Figure 5: Forecast MSE for each PIT from 1982 until 2002. The T10Y2Y-JanSuP500-rNHCTMCM (black) and the T10Y2Y-JanSuP500-TMM (gray) forecast models are displayed corresponding to the values of Table 5. Mean MSEs are shown dashed. TTC MSE values can be found in Table 2 above.

which also requires a different indication compared, for example, to business cycles. As this is a new and rarely explored topic, further research will be needed to figure out the best way to forecast PIT PD term structures, which is what we start with the examination at hand. In further interpreting the outcomes of this forecast study, a comparison of Tables 2 and 5 reveals how close the forecasts get to the optimal level represented by the calibration results, i.e. a mean MSE of 0.065% and an SD of 0.031%. Evaluating the mean, we see that the best forecast outcome of 0.106% exceeds the benchmark value. The two models produce values that lie roughly between the benchmark and the TTC approach.

We conclude that the two proposed models provide a PIT forecast that takes the economic condition into account and, therefore, serves well to implement LECL requirements. However, the approaches need further testing and can certainly be improved. Practitioners can address the following areas going forward. First, other economic indicators may be found and included that provide better indication of the PD term structure outlook. The question arises as to which concrete horizon should be directly related to the economic situation? Closely related to this issue is the used concept of parameter estimation within our framework. We use a parameter estimation based on equally weighted mean squared deviation compared to the benchmark. An adjustment could increase the focus on the short term PDs in calibration. This would result in two effects. On the one hand, the short-term PDs would be adapted more precisely and, thus, the estimates would primarily reflect this part of the term structure, which could lead to a closer relationship with indicators, as they may not provide an indication of the more distant future. On the other hand, this can lead to a too high long-term deviation and, thus, to an undesired bias of the LECL estimate. Second, altering the forecast add-on in general may help to improve the linkage between the model parameter and the indicators

and, consequently, establish a better connection to the economic conditions. Especially with regard to the rNHCTMCM, the linear model does not seem to fit well. Calibration with the time-varying generator matrix may require the use of a non-linear methodology. Third, keeping parameters of risk classes that are not particularly time-varying fixed could improve the forecast models and reduce the complexity at the same time. PD term structures of risk classes like the S&P rating class AAA appears to be fairly constant over time. In addition, the TMM shows how the worse risk classes in particular are strongly related to the indicators considered. This can be used to further reduce complexity by focusing the models more on these classes as the main drivers of default risk. Fourth, it can be stated that the forecast evaluation should be supplemented by more practice-relevant applications. For example, the in-sample approach should be complemented by an out-of-sample analysis. This desire for such applications is made more difficult by the lack of a sufficient data universe of rating transitions and the relatively short histories available.

5 Conclusions

This paper introduces models that can be calibrated to and predict PD term structures. This contributes to the field of credit risk measurement in accounting, where effective forecast models for predicting PIT-specific PD term structures are needed for international reporting, but hardly to be found. We show how the TMM and the rNHCTMCM can reflect the course of observed PD term structures and introduce a forecast mechanism based on these results; this mechanism uses indicators of economic conditions. Thus, the models enable to predict from the perspective of a certain PIT and on the basis of the (forward looking) information given then. Both models help to meet the IFRS 9 requirements towards the LECL impairment. In addition, the models extend the IRB approach of the Basel framework as they build on a time-independent transition matrix. This leads to a meaningful integration of the two areas of risk assessment in financial institutions.

The TMM parameterizes PDs of the successive period and combines them with average transition and default probabilities. Primarily, the model vertically aligns the PD term structure depending on the economic condition characterized by the indicators. This affects the location of the term structure, yet does not alter the curvature in a significant way. As a time-discrete model, the TMM has difficulty reflecting arbitrary lifetimes in a satisfying way. As a time-continuous alternative, we introduce the rNHCTMCM, which is a less complex version of the NHCTMCM – a calibration approach based on a time-varying generator matrix and already used in the literature. Our simplification halves the number of parameters, but still retains the flexibility to calibrate the model precisely to data. The time-continuous model can easily reflect arbitrary lifetimes of given instruments.

The forecast study of this paper shows that the TMM and the rNHCTMCM can be calibrated equally good to data. For the same PITs, we make forecasts based on three different indicator variables. The first is derived from US treasury spreads, the second from S&P index returns and as the third we include the known leading economic indicator CLI. For a complete PD term structure forecast, the models require only a few parameters. The forecast accuracy lies between the calibration outcome as the best possible adaptation following the models and the TTC approach, which predicts on the basis of a single average transition matrix. Both

the TMM and the rNHCTMCM identify individual PITs on the basis of economic indicators and, thus, provide better forecasts. In identifying economic conditions with regard to the PD term structure outlook, we find that interest rate spreads achieve the best result. Combining the spreads with the index returns of the S&P 500 increases effectiveness, but also requires more parameters. Comparing the performance of the two models, the TMM turns out to predict slightly better. However, the rNHCTMCM is more flexible and, therefore, seems to be superior, in particular if we consider that we only apply the models to discrete data and this excludes the rNHCTMCM's greatest strength.

In detail, there are several opportunities for improvement. High-risk classes could be focused as the main drivers of default risk, an early return to average default trends over the course of each PD term structure could be integrated, or more effective indicators could be found. In addition, altering the forecasting mechanism fundamentally could help. The linear model is limited and in our study shows a significant dependence on the indicators for only few risk classes, especially in the rNHCTMCM. Nevertheless, the PIT indication is useful and stimulates further development of the model approaches presented in this paper. The evaluation of the models is mainly based on a relative comparison with the results from calibration and the TTC PD term structure. It is not surprising that the mean MSE improves from the level of the TTC forecast since we add flexibility for the different PITs. How reliable the PD term structure forecasts really are in practice will have to be verified by further studies that also include pure out-of-sample analyses. This is of crucial importance and yet problematic, as most available data histories only cover transitions between risk classes in recent years. This leads to the recommendation that users of the IFRS standard should collect data in such a way to be prepared for more comprehensive applications in the future.

References

- Albert, A. (1962), Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process, *The Annals of Mathematical Statistics* **33**(2), pp. 727–753.
- Banasik, J., Crook, J. N. and Thomas, L. C. (2017), Not If But When Will Borrowers Default, *Journal of the Operational Research Society* **50**(12), pp. 1185–1190.
- Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C. and Schuermann, T. (2002), Ratings Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing, *Journal of Banking & Finance* **26**(2-3), pp. 445–474.
- BCBS (2004), International Convergence of Capital Measurement and Capital Standards, Basel Committee on Banking Supervision.
URL: <https://www.bis.org/publ/bcbs107.pdf?noframes=1>
- BCBS (2017), Regulatory Treatment of Accounting Provisions – Interim Approach and Transitional Arrangements, Basel Committee on Banking Supervision.
URL: <https://www.bis.org/bcbs/publ/d401.pdf>
- Belkin, B., Suchower, S. and Forest, L. (1998), A One-Parameter Representation of Credit Risk and Transition Matrices, JP Morgan.
URL: <https://www.z-riskengine.com/media/1032/a-one-parameter-representation-of-credit-risk-and-transition-matrices.pdf>
- Bluhm, C. (2007), Calibration of PD Term Structures: To Be Markov Or Not To Be.
URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.114.3942>
- Bluhm, C. and Overbeck, L. (2007), Structured Credit Portfolio Analysis, Baskets & CDOs, Chapman & Hall / CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL.
- Crook, J. and Bellotti, T. (2010), Time Varying and Dynamic Models for Default Risk in Consumer Loans, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **173**(2), pp. 283–305.
- Duffie, D., Saita, L. and Wang, K. (2007), Multi-Period Corporate Default Prediction with Stochastic Covariates, *Journal of Financial Economics* **83**(3), pp. 635–665.
- EU and UN (2017), Handbook on Cyclical Composite Indicators: For Business Cycle Analysis, United Nations, European Commission, Conference Board Inc.
URL: <https://ec.europa.eu/eurostat/documents/3859598/8232150/KS-GQ-17-003-EN-N.pdf/868e9a5b-9bae-450b-a2d9-c668524c738d>
- Fama, E. F. (1990), Term-Structure Forecasts of Interest Rates, Inflation, and Real Returns, *Journal of Monetary Economics* **25**(1), pp. 59–76.
- FCAG (2009), Report of the Financial Crisis Advisory Group, Financial Crisis Advisory Group.
URL: <https://www.iasplus.com/en/binary/crunch/0907fcagreport.pdf>

- Fei, F., Fuertes, A.-M. and Kalotychou, E. (2012), Credit Rating Migration Risk and Business Cycles, *Journal of Business Finance & Accounting* **39**(1-2), pp. 229–263.
- FSF (2009), Report of the Financial Stability Forum on Addressing Procyclicality in the Financial System, Financial Stability Forum.
URL: http://www.fsb.org/wp-content/uploads/r_0904a.pdf
- Gavalas, D. and Syriopoulos, T. (2014), Bank Credit Risk Management and Rating Migration Analysis on the Business Cycle, *International Journal of Financial Studies* **2**(1), pp. 122–143.
- Gerhold, P., Kleppe, A., Seifert, M. and Thakkar, D. (2017), Constructing the PD Term Structure.
URL: <https://ssrn.com/abstract=2998824>
- Grosskord, M., Mach, P. and Reher, D. G. (2013), Lifetime Expected Loss: Anwendungsfelder und Berechnungsmethoden, Deloitte.
URL: <https://www2.deloitte.com/de/de/pages/financial-services/articles/White-Paper-No-58.html>
- Grünberger, D. (2013), Kreditrisiko im IFRS-Abschluss: Handbuch für Bilanzersteller, Prüfer und Analysten, Schäffer-Poeschel, Stuttgart.
- Hanson, S. and Schuermann, T. (2006), Confidence Intervals for Probabilities of Default, *Journal of Banking & Finance* **30**(8), pp. 2281–2301.
- IASB (2014), IFRS 9: Financial Instruments, International Accounting Standard Board.
URL: <http://www.ifrs.org>
- Israel, R. B., Rosenthal, J. S. and Wei, J. Z. (2001), Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings, *Mathematical Finance* **11**(2), pp. 245–265.
- Ivanova, D., Lahiri, K. and Seitz, F. (2000), Interest Rate Spreads as Predictors of German Inflation and Business Cycles, *International Journal of Forecasting* **16**(1), pp. 39–58.
- Jafry, Y. and Schuermann, T. (2003), Measurement and Estimation of Credit Migration Matrices, Wharton Financial Institutions Center Working Paper No. 03-08.
URL: <https://ssrn.com/abstract=394021>
- Jafry, Y. and Schuermann, T. (2004), Measurement, Estimation and Comparison of Credit Migration Matrices, *Journal of Banking & Finance* **28**(11), pp. 2603–2639.
- Jarrow, R. A., Lando, D. and Turnbull, S. M. (1997), A Markov Model for the Term Structure of Credit Risk Spreads, *The Review of Financial Studies* **10**(2), pp. 481–523.
- Kadam, A. and Lenk, P. (2008), Bayesian Inference for Issuer Heterogeneity in Credit Ratings Migration, *Journal of Banking & Finance* **32**(10), pp. 2267–2274.
- Kim, J. (1999), Conditioning the Transition Matrix, Credit Risk Special Report pp. 37–40.
- Lahiri, K. and Wang, J. G. (1994), Predicting Cyclical Turning Points with Leading Index in a Markov Switching Model, *Journal of Forecasting* **13**(3), pp. 245–263.

- Lando, D. and Skødeberg, T. M. (2002), Analyzing Rating Transitions and Rating Drift with Continuous Observations, *Journal of Banking & Finance* **26**(2-3), pp. 423–444.
- Leow, M. and Crook, J. (2014), Intensity Models and Transition Probabilities for Credit Card Loan Delinquencies, *European Journal of Operational Research* **236**(2), pp. 685–694.
- Leow, M., Mues, C. and Thomas, L. (2014), The Economy and Loss Given Default: Evidence from two UK Retail Lending Data Sets, *Journal of the Operational Research Society* **65**(3), pp. 363–375.
- Nickell, P., Perraudin, W. and Varotto, S. (2001), Stability of Ratings Transitions, *Bank of England Quarterly Bulletin* **41**(2), pp. 216.
- Ong, M. K. (2000), *Internal Credit Risk Models: Capital Allocation and Performance Measurement*, Risk Books, London.
- Pearce, D. K. (1983), Stock Prices and the Economy, *Economic Review*.
URL: <https://www.kansascityfed.org/publications/research/er>
- Pearce, D. K. and Roley, V. V. (1985), Stock Prices and Economic News, *The Journal of Business* **58**(1), pp. 49–67.
- Shumway, T. (2001), Forecasting Bankruptcy More Accurately: A Simple Hazard Model, *The Journal of Business* **74**(1), pp. 101–124.
- S&P (2018), *Default, Transition, and Recovery: 2017 Annual Global Corporate Default Study And Rating Transitions*, Standard & Poor’s Financial Services LLC.
URL: <https://www.spratings.com>
- Stefanescu, C., Tunaru, R. and Turnbull, S. (2009), The Credit Rating Process and Estimation of Transition Probabilities: A Bayesian Approach, *Journal of Empirical Finance* **16**(2), pp. 216–234.
- Stepanova, M. and Thomas, L. (2002), Survival Analysis Methods for Personal Loan Data, *Operations Research* **50**(2), pp. 277–289.
- Trück, S. (2008), Forecasting Credit Migration Matrices with Business Cycle Effects: A Model Comparison, *The European Journal of Finance* **14**(5), pp. 359–379.
- Wei, J. Z. (2003), A Multi-Factor, Credit Migration Model for Sovereign and Corporate Debts, *Journal of International Money and Finance* **22**(5), pp. 709–735.

Appendix

Calibration Estimates

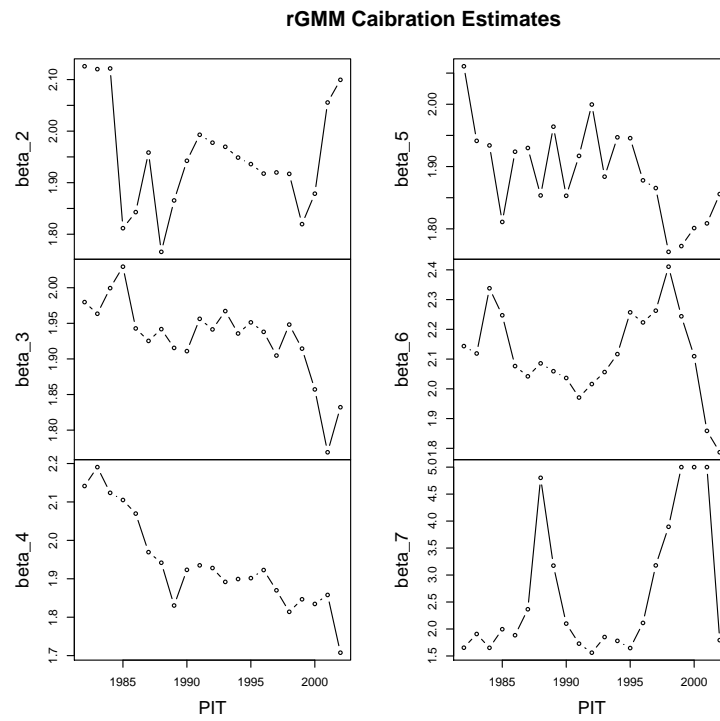


Figure 6: Temporal evolution of the estimates in calibration.

TMM Caibration Estimates

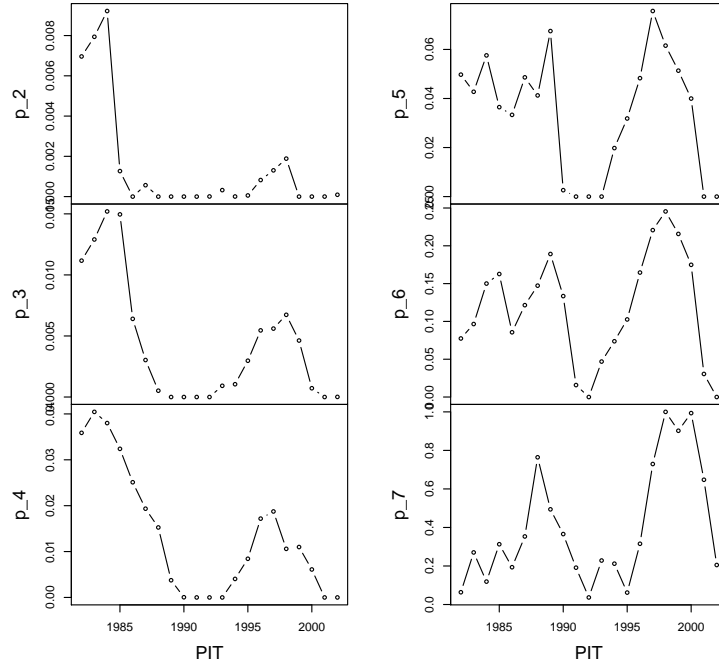


Figure 7: Temporal evolution of the estimates in calibration.

Linear Model Estimates

Parameter	Intercept	JanS&P500 Coefficient	Adj. R^2
$\beta_t^{(1)}$	1.973***	-1.384*	0.196
$\beta_t^{(2)}$	1.976***	-0.263	0.087
$\beta_t^{(3)}$	1.909***	0.225**	0.306
$\beta_t^{(4)}$	1.918***	0.221	0.02
$\beta_t^{(5)}$	1.884***	0.075	-0.031
$\beta_t^{(6)}$	2.058***	0.644***	0.364
$\beta_t^{(7)}$	2.768***	-1.059	-0.037
$p_t^{(1)}$	0.001***	-0.005***	0.27
$p_t^{(2)}$	0.001	0.002	-0.046
$p_t^{(3)}$	0.003*	0.012	0.066
$p_t^{(4)}$	0.011**	0.029	0.047
$p_t^{(5)}$	0.026***	0.084***	0.207
$p_t^{(6)}$	0.097***	0.214	0.141
$p_t^{(7)}$	0.403***	-0.007	-0.053

Table 6: Estimation results of the calibrated TMM and rNHCTMCM model parameters and the indicator JanS&P500. Significance codes indicate p-values $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*).

Parameter	Intercept	CLI Coefficient	Adj. R^2
$\beta_t^{(1)}$	16.122	-0.143	0.033
$\beta_t^{(2)}$	7.148*	-0.052	0.124
$\beta_t^{(3)}$	0.468	0.015	-0.004
$\beta_t^{(4)}$	-0.378	0.023	-0.027
$\beta_t^{(5)}$	3.141	-0.013	-0.033
$\beta_t^{(6)}$	-4.482	0.066	0.088
$\beta_t^{(7)}$	-40.983	0.437	0.032
$p_t^{(1)}$	0.057	-0.001	0.076
$p_t^{(2)}$	-0.027	0	-0.046
$p_t^{(3)}$	-0.117	0.001	-0.012
$p_t^{(4)}$	-0.537	0.006	0.064
$p_t^{(5)}$	-1.65**	0.017*	0.284
$p_t^{(6)}$	-4.387*	0.045*	0.223
$p_t^{(7)}$	-11.666	0.121	0.057

Table 7: Estimation results of the calibrated TMM and rNHCTMCM model parameters and the indicator CLI. Significance codes indicate p-values $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*).

Tobias Filusch*/Sascha H. Mölls†

The Bargaining Game of the Audit

– Regulatory Provisions for High-Quality Financial Statements and an
Independent Audit –

Abstract

Reliable audits that ensure the truthfulness of financial statements are an important driver for functioning capital markets. This paper focuses on a model for the auditor-client relationship, in which a coalition on the exact content of the audit opinion could have attractive advantages for both parties. Since it contradicts the independence of the auditor, such a coalition must be prevented in order to avoid manipulation and to achieve a high quality of published financial statements. The model is designed as a (Nash) bargaining game concerning the sharing rule of the potential coalition benefits. The analysis of the paper evaluates the effects of regulatory provisions on the decisions of the auditor and client management. It leads to the development of an effective regime of provisions that encourages client management to report truthfully.

Contents

1	Introduction	113
2	Modeling the Auditor Client Management Negotiation	115
2.1	The Bargaining Model	115
2.2	Treatment of Uncertain Model Parameters	117
2.3	Model Mechanisms	119
3	Achieving Correct Financial Statements with Effective Regulation	120
3.1	Threats to the Independence and its Regulation	120
3.2	An Effective Regulatory Framework	123
3.3	Comparative Statics	124
4	Conclusions	126

*Dipl.-Math. oec. Tobias Filusch, Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, ifG Marburg, Am Plan 2, D-35032 Marburg.

†Prof. Dr. Sascha H. Mölls, Philipps-Universität Marburg, Fachbereich Wirtschaftswissenschaften, Professur für Allgemeine Betriebswirtschaftslehre und Rechnungslegung, Barfußertor 2, D-35032 Marburg.

1 Introduction

Due to the agency problem resulting from the separation of a company's ownership and control, doubts may arise as to the quality of the financial information published by the company (Jensen and Meckling, 1976; Berle and Means, 1991). In order to encourage potential investors and to stabilize capital markets in general, the reliability of financial statements is essential (Benston et al., 2006*b*). A valuable tool for ascertaining the quality of financial statements is the audit by a third party (Benston et al., 2006*a*). However, this concept only functions properly if auditors really carry out their mission, which is characterized by two main criteria, namely the "ability to discover errors or breaches in the accounting system" and "the ability to withstand client pressures to disclose selectively in the event a breach is discovered" (DeAngelo, 1981*a*). The second criterion only applies if the auditor is independent in appearance from the influence of the auditee. Otherwise, audit findings are essentially worthless to the public. Multiple recent bankruptcy scandals have aroused a general suspicion of the audit industry shirking its responsibilities.¹ Therefore, the independence of the audit has repeatedly been the subject of public debates and even earned a spot on the agenda of regulatory authorities, who are interested in effective provisions to solve the issue (Fischbacher and Stefani, 2007). Common threats to the independence are a high client importance, the (excessive) possibility to perform simultaneous non-audit services, a long auditor tenure, and any kind of affiliation between auditor and auditee (Tepalagul and Lin, 2015). Correspondingly, some regulatory provisions are under discussion that target these issues; a few even have been implemented in certain areas. With regard to the issue of a too close association or non-audit services carried out at the same time, authorities implemented complete or partial prohibitions (e.g a revenue limit for non-audit services). To make these attempts effective, it is necessary to monitor the whole process and/ or at least create high incentives for whistleblowing combined with severe punishments for violations. On the other hand, a mandatory auditor rotation after a certain number of performed audits or the provision of exclusively permitting joint-audits by different audit companies target the regulation of the business relationships of the audit. How such provisions really affect this relationship, however, is rarely analyzed.

There is a substantial body of recent literature on the relationship of the client management and the auditor. The final stage of the audit period appears to be a closed-door meeting between the client management and audit partner where the content of the audit opinion is discussed (Gibbins et al., 2001, 2005, 2007). During this auditor-client management negotiation (ACMN), the auditor was found to play the role of the "relationship manager" (McCracken et al., 2008), who attempts to be reappointed for the next period. Factual, the management is in charge of that decision (Beattie et al., 2004), which raises concerns about how much pressure the management is able to impose on the auditor. This pressure could also target a confirmation of falsified financial statements (Bennett et al., 2015; Fearnley et al., 2005). Consequently, the ACMN may lead to a coalition of the two parties to the benefit of both sides. This could even be further encouraged. Let us assume that intentionally falsified financial statements of the management exist at the beginning of the audit. This only makes sense if it results in a surplus for management, which management in turn can use in part as an additional incentive for the coalition; management could lure the audit partner with

¹See Sauviat (2006) among others for insights in the Enron scandal.

future mandates or other benefits. Therefore, effective regulation must consider the tendency of forming a coalition on falsified financial statements during the ACMN. Such a coalition impairs the independence and clearly leads to unreliable financial statements, too.

Previous research has already concentrated on audit quality and the independence of the auditor in a theoretical fashion. Since DeAngelo (1981*b,a*) has identified quasi-rents as crucial drivers of an impaired independence, many authors have joined the associated discussion (e.g. Magee and Tseng (1990), Kanodia and Mukherji (1994), Schatzberg and Sevcik (1994), Schatzberg (1994), and Lee and Gu (1998)). Mainly (non-cooperative) strategic games were applied, where the audit market setting (competitive or noncompetitive), the personality of the auditor and client (honest or dishonest), the effort of the auditor (high or low), and/ or the audit technology (perfect or imperfect) were common simplified characteristics (Fellingham and Newman, 1985; Shibano, 1990; Matsumura and Tucker, 1992; Patterson, 1993; Holler and Nguyen, 2007). Demski and Seieringa (1974), Hatherly (1996), and Cook et al. (1997) debated whether the relationship between auditor and auditee is actually non-cooperative as they have persuasively argued for a cooperative audit game. This paper expands upon the sparse literature on a cooperative auditor-client management relationship, because only that can lead to a formation of a coalition in a game theoretic framework. The intention of this paper is to develop a model for this relationship, including in particular a looming coalition of the two parties.

Our model considers a two-stage decision process. First, the management decides whether or not to (intentionally) falsify the financial statements prior to the audit. This decision could be part of the management's active earnings management, which may have several motives (Healy and Wahlen, 1999). If the financial statements are undisputable correct, however, the audit can only confirm it and there is virtually no issue regarding the independence. In principle, the only plausible reason for intentionally falsifying financial statements appears to be the obtaining of a surplus controlled by management. Only in this case there is a relevant second stage. It describes the decision of the auditor, who we assume to find any issue due to a high audit effort. During the ACMN, the auditor is subject to management's efforts to confirm its version of the financial statements, thus, to form a coalition. To achieve that goal, management can hand over a share of the surplus, but how much exactly must be clarified in a two-player bargaining game. If the auditor refuses the offer, unattractive consequences could evolve for both parties. In particular, auditors who refuse to cooperate must fear that they will not be reappointed by the company (Goldman and Barlev, 1974). This puts a lot of pressure on auditors and reduces their bargaining power. Due to the integrated bargaining game, the model of in this paper is related to Zhang (1999), who considers the company value the auditor is willing to report to the public as the objective of a bargaining game. Similar to DeAngelo (1981*a*), the expectations about positive quasi-rents are found to be the main driver for the auditor-accepted nominal value. As a clear extension of this examination, our model explicitly includes the individual considerations of the parties with regard to their decisions as well as relevant parameters of real world situations in more detail. The decisions to be made determine the actions of the two parties and, thus, also the outcome of the audit as well as the quality of the information provided to the public. Furthermore, our focus is on discussing and evaluating a regulatory framework that effectively prevents a coalition, which none of the linked references has done yet.

This paper contributes to the academic literature by a (game) theoretical model on the behavior and the decision process during the ACMN. We analyze the influence of regulation on the actions of both audit parties. The main finding is that regulatory provisions affect the negotiation in different ways. Some provisions reduce the attractiveness of all possible arrangements of the negotiation, which is basically a good result. But even then, the attractiveness of a coalition could be higher than its refusal. Our analysis isolates the opportunity to explicitly target the solution of the negotiation, which we model through the well-known Nash bargaining game (Nash, 1950). Several regulatory provisions are able to shift the bargaining solution, which is largely determined by the distribution of the bargaining power. By regulating this bargaining power, the allocation of the surplus is changing. For example, a more powerful auditor is able to claim a larger share. With this principle, our model allows to determine exactly the threshold above which management cannot take advantage of its surplus, which appears to be a reasonable goal of effective regulation. Because if such an outlook is anticipated by management, it significantly reduces the incentives for management to deliberately prepare incorrect financial statements.

The paper begins by describing the model for the ACMN in Section 2. After the introduction of the bargaining game, it explains our approach of treating the uncertain model parameters, i.e. those that describe the risk of forming a coalition regarding incorrect financial statements (2.2). Section 2.3 then visualizes and analyzes the model mechanisms. Section 3 goes on to discuss threats to the independence, regulatory provisions, and how they influence the ACMN. First, threats to the independence are discussed and analyzed (3.1). Section 3.2 then moves on to propose effective provisions. To further illustrate the model effects, Section 3.3 provides a comparative static analysis. Section 4 concludes the paper.

2 Modeling the Auditor Client Management Negotiation

2.1 The Bargaining Model

The actors in our model are the auditor (A) and the client management (M). Both parties make decisions exclusively on the basis of the benefits they expect as rational agents and act as a single person. Prior to the initial audit by a newly assigned auditor, management has the option to produce a surplus resulting from falsified financial statements. We only consider this first period's situation and do not consider repeating the same process in subsequent periods. Additionally, we exclude the option of reverting decisions in the future. Basically, the model focuses on intentional falsification with the goal of gaining benefits. Alternatively, management can stick to the truth, which is in the best interest of society and, correspondingly, the most favorable outcome of regulatory efforts. If true statements are present during the audit, the “honest” outcome h materializes corresponding to (1), where r is the risk-free rate.

$$h = (h_M, h_A)' := \left(\sum_{t=1}^{T_M} \frac{S}{(1+r)^t}, \sum_{t=1}^{T_A} \frac{F}{(1+r)^t} \right)' \quad (1)$$

The outcome (1) describes a situation in which there are no significant objections to the financial statements provided by management. The auditor maintains the mandate multiple

periods T_A and earns a fixed audit fee (F). Similarly, management earns a fixed salary (S) per period and stays under contract with the company for several periods T_M .

If management falsifies the financial statements prior to the audit and, thus, produces a surplus, we expect all infringements being discovered during the audit (perfect audit quality). Therefore, the second stage is where the auditor decides to either qualify the opinion and to truthfully disclose any findings or form a coalition with the management. The latter option means any findings remain undisclosed. If the auditor decides to qualify the opinion, the parties earn $d = (d_M, d_A)'$ given by (2). We expect this outcome to have unpleasant consequences for the two parties involved, assuming that (a) the company will appoint another auditor for the next audit period (Goldman and Barlev, 1974) and (b) the client's management will be replaced. Replacing of the management – either through resignation or dismissal by other bodies – seems to be only appropriate for more severe issues.² In this case, we furthermore expect both parties to look for new appointments in the market. Those market expectations are included by present values E_M and E_A . Thus, the value of the described scenario for each of the two parties is given by

$$d = (d_M, d_A)' := \frac{1}{(1+r)} (S + E_M, F + E_A)' \quad (2)$$

In general, outcome d results if either the client management or the auditor refuses to form the coalition. This decision is exclusively in the hands of the auditor as the management decides earlier whether to create the surplus and, therefore, will always prefer to form the coalition at this point. The outcome d is called the status quo of the impending bargaining game, which characterizes the ACMN.

In our interpretation, the ACMN describes the phase in which the auditor confronts the management with the audit results. Due to a perfect audit quality, these include the entire extend of the falsifications within the financial statements. Consequently, the surplus can only be claimed if the auditor agrees to form a coalition and leaves the opinion unqualified. This coalition surplus is denoted by c_s and characterizes, for example, more attractive bonuses, additional financial flexibility (e.g. for projects or other mandates), or reputational gains for the client's management. Note that the surplus is expressed as the constant present value that exists in each period in which the infringements are maintained.

The next step is to clarify which value a coalition really generates. Part of the equation is the created surplus and, in particular, how exactly it is shared among the parties; this is characterized by the sharing rule θ . Accordingly, the coalition value is designated as $c(\theta) = (c_M(\theta), c_A(\theta))'$ and deduced in three steps. First, we include the uncertainty of prosecution, due to which we then establish the cash flows as a random variable. Second, we must calculate the expected value taken into account by the decision makers. Third, we derive a reasonable bargaining outcome, i.e. a sharing rule on which the parties agree during the ACMN. Figure 1 shows a decision tree that describes the procedure for making the decisions.

²After a new management is introduced, “opinion shopping” is a well documented reason that can immediately lead to an auditor change as well (Magee and Tseng, 1990); this connects both effects. Additionally, note that Fischkin and Gassen (2011) summarizes the various forms of possible auditor change scenarios.

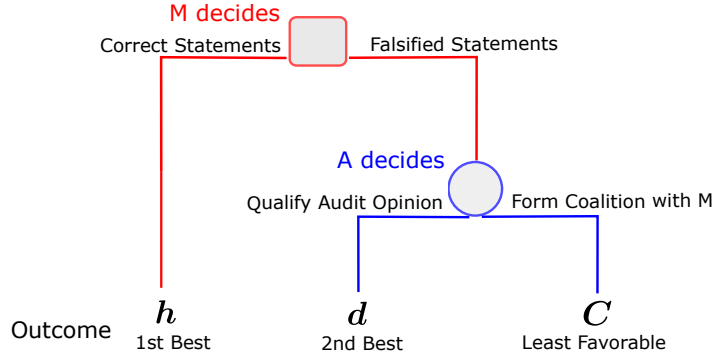


Figure 1: The two-step decision tree leading to three different outcomes.

2.2 Treatment of Uncertain Model Parameters

The audit is supposed to be an official and independent process, which means any collusion may result in prosecution and penalties if it is detected.³ This detection is subject to uncertainty. It is uncertain when or if violations will be detected at all. Therefore, the model includes a stochastic variable. We consider the number of potential detections until a certain point in time t as a Poisson process $B_{t>0}$ with intensity parameter λ . Hence, the time of the first detection is given by a stopping time $\tau := \inf\{t \mid B_t = 1\}$. In this case, the following properties hold (Çinlar and Sollenberger, 2013):

$$\tau \sim \exp(\lambda), \quad (3)$$

$$P(\{\tau \leq t\}) = 1 - e^{-\lambda t} \text{ for all } t \geq 0, \quad (4)$$

and

$$E(\tau) = \frac{1}{\lambda}, \quad Var(\tau) = \frac{1}{\lambda^2}. \quad (5)$$

It is assumed that the two parties will face two major consequences in the event of detection: first, the cash flow out of the assignment stops for both. Second, penalties are imposed that may arise from a loss of reputation or through direct litigation. As long as there is no detection, they maintain the coalition and collect all the benefits.

Following these assumptions, the coalition values C_A and C_M are given by

$$C_M(\theta) = \sum_{t=1}^{T_M} \left[\frac{S + \theta c_s}{(1+r)^t} \mathbb{1}_{\{\tau > t\}} \right] - \frac{\alpha_M}{(1+r)^{\tau+T_d}} \mathbb{1}_{\{\tau \leq T_p\}} \text{ and} \quad (6)$$

$$C_A(\theta) = \sum_{t=1}^{T_A} \left[\frac{F + (1-\theta) c_s}{(1+r)^t} \mathbb{1}_{\{\tau > t\}} \right] - \frac{\alpha_A}{(1+r)^{\tau+T_d}} \mathbb{1}_{\{\tau \leq T_p\}}. \quad (7)$$

Note that management benefits from the coalition with its share θc_s , while the auditor adds the complement $(1-\theta) c_s$. However, the detection of the ongoing coalition may change the

³Whistleblowing is a potential detection mechanism observed in practice. The incentives of whistleblowers concerning the disclosure of any type of fraudulent agreement depend on the ethical values and the litigation environment of the society at hand. The Sarbanes-Oxley Act (SOX) in the US for example protects whistleblowers who provide sensitive information to the authorities.

situation, which our model involves through the stochastic event $\{\tau > t\}$. If detected, penalties α_M and α_A are imposed after a certain delay T_d . For example, lengthy lawsuits that decrease the present value of the penalty regime may lead to a higher T_d . The coalition can only be detected before the time of prosecution T_p runs out. After that punishment is not possible anymore. Consequently, the term of penalty only materializes when $\tau \leq T_p$ holds, which turns the included indicator variable $\mathbb{1}_{\{\tau \leq T_p\}}$ to one. Essentially, this means that each party can only benefit from the coalition as long as it remains undiscovered. The detection event may lead to serious consequences relative to the benefits of the coalition. As it is not clear when exactly this will happen, cash flows (6) and (7) characterize the uncertainty faced by the actors in their decisions reasonably well. The corresponding expected values then serve as the basis for decision-making.

The Expected Coalition Value

The presented coalition values (6) and (7) are random variables that lead to certain expectations. For the management's expectation $E(C_M)$, we deduce a lower bound in the following way:

$$E[C_M(\theta)] \stackrel{\tau \sim \text{exp}(\lambda)}{=} \sum_{t=1}^{T_M} \left[\frac{S + \theta c_s}{(1+r)^t} e^{-\lambda t} \right] - E\left(\frac{\alpha_M}{(1+r)^{\tau+T_d}} \mathbb{1}_{\{\tau \leq T_p\}} \right)$$

$$\stackrel{\tau > 0}{\geq} \sum_{t=1}^{T_M} \left[\frac{S + \theta c_s}{(1+r)^t} e^{-\lambda t} \right] - E\left(\frac{\alpha_M}{(1+r)^{T_d}} \mathbb{1}_{\{\tau \leq T_p\}} \right) \quad (8)$$

$$\stackrel{\tau \sim \text{exp}(\lambda)}{=} \sum_{t=1}^{T_M} \left[\frac{S + \theta c_s}{(1+r)^t} e^{-\lambda t} \right] - \frac{\alpha_M}{(1+r)^{T_d}} (1 - e^{-\lambda T_p}) =: c_M(\theta). \quad (9)$$

Essentially, this reduces the complexity emerging from the expected penalty

$$E\left(\frac{\alpha_M}{(1+r)^{\tau+T_d}} \mathbb{1}_{\{\tau \leq T_p\}} \right).$$

By cutting the time until the penalty is enforced from $\tau + T_d$ to just T_d , the time of the penalty enforcement and the time of the detection τ became uncoupled from one another. Note, this manipulation does not influence the general results as it basically estimates the minimum value that management expects to receive from the coalition. Analogous to (8), the lower bound of the auditor is deduced as

$$E[C_A(\theta)] \geq \sum_{t=1}^{T_A} \left[\frac{F + (1-\theta) c_s}{(1+r)^t} e^{-\lambda t} \right] - \frac{\alpha_A}{(1+r)^{T_d}} (1 - e^{-\lambda T_p}) =: c_A(\theta).. \quad (10)$$

These two results (9) and (10) define the value pair $c(\theta) := (c_M(\theta), c_A(\theta))$ that is a major part of the bargaining game. The following section focuses on the sharing rule that is expected as the outcome of the negotiation.

The Nash Bargaining Solution

Taking into account the definitions given so far, the bargaining game can be denoted by (d, E) , where d is the status quo (2) and E the efficient set that is defined as

$$E := \{c(\theta) \mid \theta \in [0, 1]\}. \quad (11)$$

It describes the set, consisting of all possible allocations of values from the coalition, on which the parties can agree during the negotiations. Ultimately, they are based on the sharing rule $\theta \in [0, 1]$. This sharing rule is the result of the bargaining game and leads to a single allocation within E . The Nash bargaining solution θ^* offers a well-known solution for such a bargaining game. It provides the only solution that combines Pareto optimality, symmetry as well as independence of both linear transformation and irrelevant alternatives (Nash, 1950). Because of these reasonable properties, we assume this to be the solution of the bargaining. It is calculated through the optimization problem

$$\begin{aligned} \theta^* &= \arg \max_{\theta \in [0,1]} (c_M(\theta) - d_M) \cdot (c_A(\theta) - d_A) \\ (d_M, d_A) &\leq (c_M(\theta), c_A(\theta)). \end{aligned} \tag{12}$$

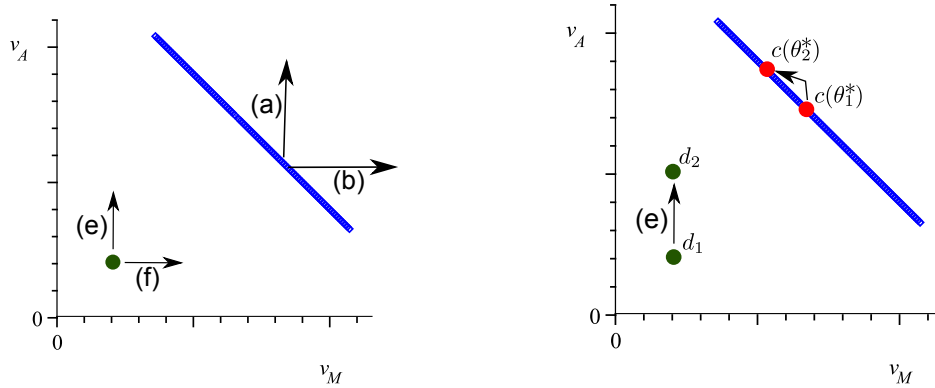
Ideally, this setup returns an analytical solution, which can be written as a function of all included parameters. Equation (14) in the Appendix presents the exact solution in detail. With this explicit solution, the bargaining becomes even more traceable. The effects on the situation caused by changes in the parameter values can now be investigated in detail. These changes can influence three different components of the model, namely the status quo, the efficient set, and/ or the bargaining solution.

2.3 Model Mechanisms

The bargaining game (d, E) is visualized in Figure 2a, where the horizontal and vertical axis represent the game's outcome for management and the auditor, respectively. It visualizes the situation in which each solution within the efficient set appeals more to both parties than the status quo does. From the various model parameters included, the risk-free rate r remains fix. This leaves $S, F, T_A, T_M, T_d, T_p, c_s, \lambda, E_M, E_A, \alpha_M,$ and α_A adjustable and evaluable with regard to the effects on the bargaining game. The arrows in Figure 2a describe possible shift directions indicated by (a) and (b) for E as well as (e) and (f) for d . Note that negative signs, e.g. (-a), also allow opposite directions. Table 1 summarizes the various effects of altering model parameters. Interestingly, some parameters only affect E , while some affect both E and d ; only E_M and E_A determine the location of d without influencing E . For example, increasing T_p results in shift (-a, -b). This means that E narrows the gap to the status quo, which remains stationary. For the two parties, this reduces the attractiveness of any point on E relative to the attractiveness of d .

Shifting the Nash Bargaining Solution

If E becomes less attractive relative to d , the Nash bargaining solution – as one specific element of E – becomes less attractive as well. Despite such a useful provision, the parties could still prefer $c(\theta^*) \in E$. It turns out that shifts of d in favor of the auditor leads to a bigger portion $(1 - \theta^*)$. The opposite is true if d_M is increased relative to d_A : management's share θ^* is increased. Within the current setting, only parameters E_M and E_A produce such an effect as they alter θ^* , which Figure 2b illustrates. Compared to the situation leading to θ_1^* , the sharing rule θ_2^* results after the implementation of a provision increasing E_A . As a consequence, the auditor increases his bargaining outcome ($\theta_2^* < \theta_1^*$) and, thus, his own value out of the coalition ($c_A(\theta_2^*) > c_A(\theta_1^*)$). As a side effect, the opposite is true for the management whose outcome decreases.



(a) Possible shift directions of the efficient set E (b) Effects of a status quo d (green) shift on Nash bargaining solution (red).

Figure 2: The horizontal axis indicates the value of employment from the management’s point of view. The vertical axis indicates the value of the mandate for the auditor. See Table 1 to track the effects of changes in model parameters.

The mechanisms presented provide an overview of how the bargaining game can be specifically influenced. Since this parameter shifts lead to different results, the next step is to identify opportunities for effective regulation. These regulatory provisions should be aimed at influencing the decisions of the parties in order to direct them towards the public interest.

3 Achieving Correct Financial Statements with Effective Regulation

The last section established a model that describes the ACMN by a theoretical model. This section now compares the mechanisms of the model with real-world threats to the independence. Furthermore, the analysis includes regulatory provisions aimed at addressing them. Note that some provisions directly target model parameters, some are not. But the latter category could nevertheless have a significant impact when it comes to creating the framework for the mechanisms taking place. In addition, regulation may not be able to influence some of the model parameters, making it difficult to achieve certain desired objectives. The second part of this section develops a regulatory strategy that takes into account these identified difficulties and yet leads to the desired objectives. Last but not least, a comparative study with static values illustrates our proposal and demonstrates its effectiveness.

3.1 Threats to the Independence and its Regulation

The independence of the auditor has been the objective of many studies. By reviewing the academic literature, Tepalagul and Lin (2015) identified four frequently discussed threats to this independence. They are a high client importance, the possibility of simultaneously performed non-audit services, a long auditor tenure, and a client’s affiliation with the audit firm. Some of the reviewed studies found evidence of the presence of these threats, others did not. We can now discuss the influence of these threats against the background of our model

Parameter	Description	Shift of E	Shift of d	Value
α_M	Penalty for M	(-b)	—	5
α_A	Penalty for A	(-a)	—	7
T_d	Delay of penalty enforcement	(a, b)	—	10
T_p	Time of prosecution	(-a, -b)	—	10
λ	Detection intensity	(-a, -b)	—	0.01
T_A	Auditor tenure	(a)	—	<i>variable</i>
T_M	Management tenure	(b)	—	<i>variable</i>
S	Salary of M	(a)	(f)	6
F	Audit fee	(b)	(e)	8
c_s	Cooperation surplus	(a, b)	—	4
E_M	Short term market expectations M	—	(f)	12
E_A	Short term market expectations A	—	(e)	22

Table 1: Summary of the model-effects induced by an increase of the model parameters. Both the effects on the status quo d and those on the efficient set E are included with the respective direction. An inverse effect of increasing a parameter is marked by the negative sign, e.g. (-a). See Figure 2a for a visualization. The last column shows the fixed values used in Section 3.3 later on.

and also evaluate regulatory provisions that are currently under discussion but whose exact effectiveness is still uncertain.

Impact of Penalty Regimes and Mandatory Auditor Rotations

Any discussion on regulation puts sanctions and their enforcement on the agenda. The model implements a penalty regime based on five parameters. First there is the probability of a detection expressed by the rate λ . It represents the intensity of detection, which allows to easily calculate the expected time until detection through $\frac{1}{\lambda}$. For example, if one sets $\lambda = 0.05$ we expect the detection after 20 years. By increasing the monitoring of the audit or incentivize whistleblowing, the detection rate can be increased through regulation. As long as a violation remains undetected, no punishment can be imposed. After a certain time, the perpetrators could even get away without punishment. This explains the time of prosecution T_p in our model. If an infringement has been detected, parameter T_d affects the actual amount of the penalty, at least from the perspective of those who are to be penalized. If this penalty is only imposed in the distant future, decision makers could tend to ignore it; especially if it may even be outside their expected lifetime. Therefore, any regulation should keep both T_d and T_p small. Only if the prosecution is carried out in a time-efficient manner, the penalties α_M and α_A provide a relevant impact. Of course, higher penalties lead to more severe considerations of the consequences resulting from coalitions. Thus, an efficient litigation process leads to a small T_d and subsequently presents an effective provision. The penalty regime as a whole is able to shift E towards the status quo, but we find that the penalties α_M and α_A do not actually have as high an impact as might be assumed. Uncertainty about the actual discovery rate and a possible delay in prosecution dampen these effects.

A long auditor tenure has been identified as a potential cause that compromises independence. Restricting the auditor tenure has been found to decrease the willingness to form

a coalition during the ACMN (Wang and Tuttle, 2009; Schmidt and Cross, 2014; Bowlin et al., 2015; Booker and Booker, 2016). Indeed, decreasing model parameter T_A lowers the attractiveness to cooperate as E is shifted towards the status quo (-a). This positive effect is identical to improving the penalty regime. A long term in office of the auditor has been identified as a potential cause that compromises independence. Restricting the auditor tenure has already been implemented in certain areas.⁴ These provisions require to change the auditor after a fixed number of periods. We refer to it as the maximum auditor term (MaxAT) provision. A closer look at the appointment of the management leads to parameter T_M . This parameter is rarely fixed in practice, as shareholders, supervisory boards, or other committees could be responsible for an exchange. Age limits could be present as a general upper limit. A regulation of this parameter is rarely discussed, but seems possible.

Non-Audit Services and Affiliation with the Audit Firm

A close business relationship between auditor and auditee can increase the general cooperative behavior during the ACMN. An affiliation even connects the mutual benefits, which in turn incentivizes a coalition that indeed benefits both parties. Moreover, it provides possibilities for the exchange of payments, which is also important to consider with regard to non-audit services. In practice, managers from the client could be affiliates of the audit company, e.g. could be shareholders or even be part of the supervisory board. Non-audit services are arranged mandates for the audit firm with the client additional to their core business. Thus they provide lucrative additional income, too. Frequently, auditing firms also offer consulting services, which in particular focus on topics of the client's business.

Both non-audit services and an affiliation with the audit firm are binding contracts that allow to hand-over certain amounts from management to the auditor. Thus, both are crucial of providing opportunities to reliably divide the surplus among the parties, which in our model is characterized by θ . Otherwise, it will be more difficult for the parties to actually implement the outcome of the negotiations. Therefore, limiting or even prohibiting companies to simultaneously order non-audit services from the appointed audit firm seems reasonable in order to prevent coalitions. But even then, there still could be other ways of transferring payments. Furthermore, prohibiting non-audit services is often debated to increase client importance (Joe and Vandervelde, 2007), which is considered next.

Client Importance and the Auditor's Bargaining Power

Client importance occurs when the audit firm generates the majority of its revenues from the mandate with a particular firm. It is only natural to assume that such a scenario would allow management to exert high pressure on the auditor during the ACMN. From the perspective of the auditor, this is expressed by a small E_A relative to F and $(1 - \theta^*) c_s$ as the audit market is rather unattractive compared to the current mandate with the important client. Changing this ratio is difficult given the current status quo (2). Parameters E_A and E_M are given by the respective markets for auditors and managers. Therefore, efforts should be implemented to keep these markets liquid. This would create opportunities to find (attractive) alternative assignments as compensation for a lost mandate. How this can be achieved is difficult to say

⁴A maximum auditor tenure has been considered by the European Commission as well as the Sarbanes-Oxley Act (See "Restoring confidence in financial statements: the European Commission aims at a higher quality, dynamic, and open audit market" by the European Commission and the Sarbanes-Oxley Act of 2002, Section 203).

and goes beyond the objectives of this paper. With regard to the audit of large companies, regulators face a major problem. They are not able to improve the auditors' negotiation position and, thus, reduce the client's potential pressure. This problem is present in the model because there is no possibility to influence only d except to change the hard-to-regulate parameters E_A and E_M , which characterizes limited options for shifting the bargaining solution in general. The regulatory authorities appear to have their hands tied at this point. Recall that the status quo resulted from the assumption that the auditor is released after refusing to form the coalition. This flexibility seems to be in the best interest of the client's management.

3.2 An Effective Regulatory Framework

After analyzing the threats and corresponding regulatory provisions, we were able to identify several efforts to narrow the gap between the efficient set and the status quo in our model. As a matter of fact, this reduces the attractiveness of any pair of possible coalition values and seemingly strengthens the desired independence. Unfortunately, it might not be enough to completely prevent coalitions. Even in cases of a rather small surplus, the benefits of forming a coalition could still be positive. Thus the identified difficulties to influence the negotiation power appears to be very unfortunate. It could alter the decisions of the parties in a more fundamental way. We will now turn the attention to opportunities of implementing more options. The target is the Nash solution that we expect to be the outcome if the parties negotiate. The current choice d (see (2)) must be changed. We propose the generalization d' given by (13).

$$d' = (d'_M, d'_A)' := \left(\frac{S + E_M}{(1+r)}, \sum_{t=1}^{T'_A} \frac{F}{(1+r)^t} + E'_A \right)'. \quad (13)$$

This status quo includes the additional parameter T'_A – a minimum auditor term (MinAT) – as a counterpart to the MaxAT discussed in the last section. Note that the old status quo (2) is closely related as a special case of 13 with $T'_A = 1$. This representation can guarantee the auditor multiple audits regardless of the audit opinion. The provision to force firms to commit to such a MinAT has already been discussed and found to be effective in cases where really the management is responsible for hiring the auditor, which we assume, too (Stefani, 2002).

With a regulatory provisions that results in a MinAT $T'_A > 1$, the status quo is shifted in favor of the auditor (e). This can lead to $d'_A > c_A(\theta)$ for all $\theta \in [0, 1]$, which means there is no bargaining outcome that is as attractive as qualifying the opinion. The auditor's decision would implement the second best outcome – an independent audit (see Figure 1). One step earlier it is up to the management to decide whether there will be a surplus at all. If management refuses to create it, there can be no bargaining. They will do so if they anticipate outcome d' and if, furthermore, $h_M > d'_M$ holds. This describes a scenario where management stays with the company long terms by providing correct financial statements in advance to the audit. Obviously, this is the first best outcome for the public. Moreover, there can be no threat to the independence of the auditor in this case. Therefore, the MinAT appears to provide a strong tool for regulation.

Unfavorably, it may be that the outlined approach requires a very long tenure T'_A . It seems to be quite inefficient to simply implement a high T'_A , especially in terms of maintaining market

mechanisms of audits. Therefore, we now develop an efficient strategy – the optimal MinAT provision. It should encourage management to opt for h while keeping T'_A at a rather small level. With a fixed choice of remaining model parameters, the desired solution \hat{T}'_A can be determined with criterion $\theta^* = 0$. It characterizes an outcome of the negotiation that excludes the management from the additional benefits. Instead, the entire surplus is transferred to the auditor. If management expects this impending outcome, it is only logical to expect them to stay with h . The model allows to determine an analytical solution for the optimal MinAT, which is presented in (15).

Simply put, the effect of the proposed strategy can be described as follows: if one of two parties in a game realizes that the prize of the game is unattainable, then he is unlikely to play. However, if this price depends on the effort of this particular player, then no price can be expected. The provision improves the auditor’s position just as much as necessary to skim off the entire surplus. Obviously, this threshold cannot be exactly met in practice, but an approximation of it may also be very helpful. In our model, the coalition value takes into account an additional income from the surplus for each upcoming period. This could be unrealistic because, for example, this additional value could only be attainable in a few periods or even just in a single one. Thus, the coalition value used can be seen as the upper limit of what a surplus can yield, namely additional periodic income over several periods. In future periods, however, the situation may generally change and the actors could revert their decisions after re-evaluating the different benefits. It is difficult to prevent management from waiting until its own bargaining power increases compared to that of the auditor before working towards a coalition just then. This will undoubtedly happen as the time of the mandatory rotation of the auditor approaches. Therefore, the implementation of these provisions in practice requires a dynamic adjustment that takes into account the changes over time. However, the proposed strategy massively limits the time frame for manipulative opportunities because as rotation approaches, another auditor is appointed and successfully encouraged to resist the temptations of management. Thus, despite the shortcomings described, a positive effect prevails.

3.3 Comparative Statics

The transition from an initially attractive game setting that promotes coalitions to one that leads to correct financial statements is now demonstrated with an example of static values. The goal is to make h given in (1) the favorable outcome for management as it leads to the desired goal. Choosing $T_M = 10$ in combination with the given static values of Table 1 leads to $h_M = 68.83$, which is the threshold for management that regulation must target.

Starting with MaxAT $T_A = 12$ and the situation without a MinAT provision (i.e. $T'_A = 1$) leads to the green setting in Figure 3. The efficient set (green line) is much more attractive for both players compared to the status quo (green point). Hence, the parties have obvious incentives to negotiate the surplus. Note that this setting is only possible if a cooperation surplus of $c_s = 4$ has already been created by management. The Nash bargaining solution (12) of this setting derived by the analytical solution (14) is $\theta_1^* = 0.614$, which means that management keeps more than 61% of the surplus. The corresponding values the parties gain are $c_M(0.614) = 94.73$ and $c_A(0.614) = 106.83$ (see the red point on the green set). Obviously, this outcome attracts management much more than h_M and, thus, falsified financial

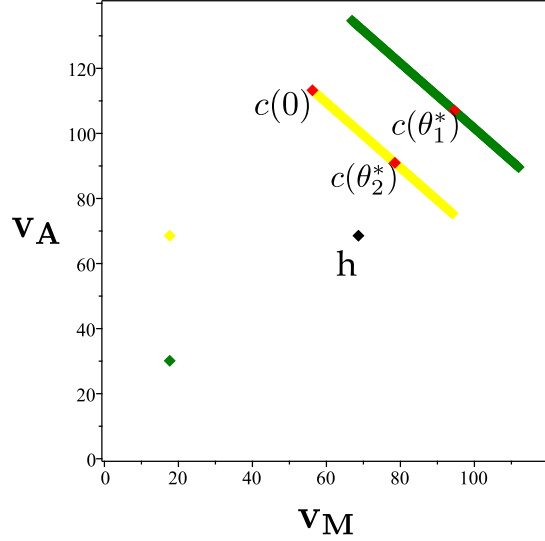


Figure 3: The vertical axis represents the value of A, the horizontal the value of M. The shift of the efficient set if the MaxAT is shortened from $T_A = 12$ (green) to $T_A = 10$ (yellow). The Nash bargaining solutions are marked red, respectively. The corresponding status quo for the green setting ($T'_A = 1$) is shifted by a provision to chose a proper MinAT in order to get $\theta^* = 0$ for the yellow setting. This requires to choose $T'_A \approx 6.82$. Fixed parameter choices are $T_d = T_p = 10$, $r = 0.01$, $\lambda = 0.01$, $S = 6$, $F = 8$, $c_s = 4$, $E_M = 12$, $E_A = 22$, $\alpha_A = 7$, and $\alpha_M = 5$.

statements emerge as the best strategy. In contrast to this static example, management could be able to even increase the surplus in advance. This changes the setting and makes it even more difficult to match the attractiveness of a coalition with outcome h . Nevertheless, we move on with these statics.

The first regulatory step to prevent coalitions is to impose a provision on the MaxAT. It shifts the efficient set towards the origin. The yellow efficient set results from a reduction T_A from 12 to 10. By allowing fewer audit periods, the benefit of every point on E is lower than before. The new Nash bargaining solution is $\theta_2^* = 0.588$, which means that this provision barely reduces the management's share and still, $c_M(0.588) = 78.67 > 68.83 = h_M$ holds. In other words, without reaching the threshold the gap has been narrowed. In some cases, this provision may already be sufficient. To ensure that the desired outcome is obtained, we impose an additional MinAT provision on the status quo.

The MinAT provision is established by using the more flexible form of the status quo, namely d' as given in (13). In order to make a proper choice $\hat{T}'_A > 1$, we can consider Figure 4. It shows the Nash bargaining solution θ^* as a function of T_A and T'_A . The intended outcome $\theta^* = 0$ is given by the plane at the bottom of the graph. There, the black line marks the choice of the first provision step $T_A = 10$, which leads to the unique choice of $\hat{T}'_A = 6.82$ fol-

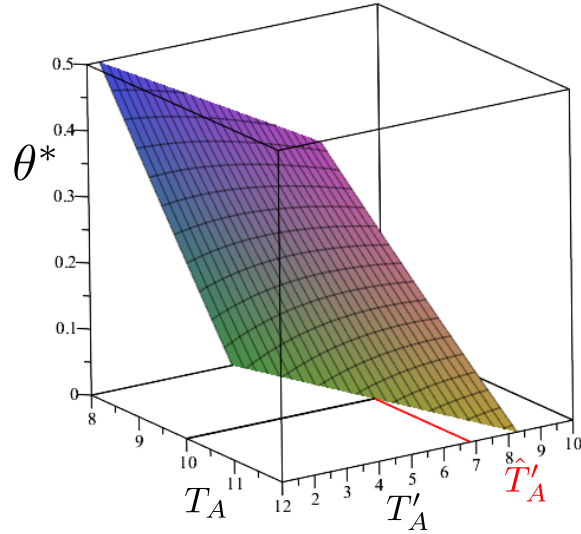


Figure 4: Fixed parameter choices are $T_d = T_p = 10$, $r = 0.01$, $\lambda = 0.01$, $S = 6$, $F = 8$, $c_s = 4$, $E_M = 12$, $E_A = 22$, $\alpha_A = 7$, and $\alpha_M = 5$. The threshold for T'_A to prevent the bargaining corresponding to $T_A = 10$ is given by $\hat{T}'_A \approx 6.82$.

lowing the red line. Analytically, equation (15) provides the exact solution. As a consequence of that provision, the Nash bargaining outcome for the collaboration parties then results in $c_A(0) = 113.05$ and $c_M(0) = 56.39$, respectively.

Referring once again to Figure 3 and the way this catalog of provisions started, the choice of MinAT $T'_A = 6.83$ leads to the yellow mark as the new status quo. Compared to the green, it is shifted vertically in favor of A. The value just derived is also included in Figure 3 labeled with a red mark at the upper left edge of the yellow efficient set. Compared to the original Nash solution (θ_1^*) following the green setting, the combination of the two coordinated provisions dramatically reduces the value of the management. At the same time, the value of the auditor increases. In particular, the targeted threshold is now undercut by $h_M - c_M(0) = 12.43$. So returning to the point in time when management is deciding whether to create a surplus. Now the more favorable strategy is clearly to strive for value $h_M = 68.83$.

4 Conclusions

A close review of the literature exploring the independence of financial statement audits identified two essential gaps: first, there is a general lack of theoretical research that considers a collusive relationship between the auditor and the client's management. Second, there has never been a comprehensive analysis addressing the issue of intentionally falsified financial statements involving collusion between the parties.

Incorrect financial statements are only attractive to a company's management if it expects

an outcome greater than the value of true statements. However, this surplus can only be claimed if the appointed auditor, who is in principle capable of uncovering manipulations, agrees to a coalition. Through a bargaining game, we clarify how the management and the auditor could share this surplus. This could be the main objective of the ACMN to conclude the audit. Our model distinguishes two types of influence from regulation. First, some regulatory provisions alter the setting of the game, for example, through the oft-discussed mandatory auditor rotation. We reveal how this reduces the attractiveness of any formed coalition, yet only prevents it in special cases. Second, other regulatory provisions instead influence the bargaining solution by shifting the negotiating power between the actors. The academic literature to date has not explicitly dealt with this negotiation and, thus, failed to separate these two facets. However, this consideration is very important in the light of the decisions and actions of the parties involved regarding the audit opinion and the resulting financial reporting.

The goal is to incentivize management to produce correct financial statements and provide them to the auditor in advance to the audit. In this case, the independence of the auditor is irrelevant. To achieve this goal, we propose a two-stage strategy of regulation that, first, establishes an appropriate setting for the bargaining game and, second, directly influences its looming solution. To modify the bargaining game so that it reaches our goal, we suggest introducing a minimum auditor tenure provision. This means that each newly hired auditor is obliged to remain with the client for a guaranteed number of periods; he cannot be replaced before this tenure expires. Such a provision increases the auditor's bargaining power and, thus, it increases his share of the surplus achievable in the ACMN. However, the auditor tenure should not be arbitrarily long since it would undermine natural market mechanisms and could over-benefit audit firms. Therefore, we specify the audit tenure exactly at the level that just eliminates the management's share of the surplus. Our model allows us to determine this level precisely. This process gives management an incentive to reconsider its strategy. In fact, the management is responsible for the existence of the surplus, but since it will not benefit from that surplus, we can expect them to issue accurate financial statements from the outset.

This paper also discusses the impact of well-known threats to auditor independence. Unsurprisingly, we find non-audit services provided by the incumbent auditor to be a potential threat to the independence. Such services provide a way to exchange shares of the surplus gained from falsified financial statements between the parties. It therefore seems wise to limit this opportunity. Similarly, any kind of affiliation between the auditor and the auditee may encourage collusion because it directly links the individual interests to the common interest. Thus, it characterizes a cooperative relationship that ultimately triggers the analyzed circumstances of this paper. Furthermore, we find that the more important a client is to the auditor, the greater the management's ability to pressure the auditor becomes. From the auditor's perspective, the degree of client importance can be expressed by the relation between the benefits of the mandate and those of alternative mandates in the audit market – two variables of our model. Unfortunately, it seems impossible to control these variables effectively through regulation. The attractiveness of the audit market would need to be increased, but as large clients generally appear more attractive than several small clients, this objective is unlikely to be achieved. However, client importance primarily reflects a weak bargaining position of the auditor, which we identified as the major reason for forming coalitions. This issue can be

overcome by the minimum auditor tenure provision we introduce in this paper. A high client importance requires increased support of the auditor to stabilize his position. Consequently, an individual level for the minimum auditor tenure is crucial for this provision to succeed. Basically, both intensively discussed regulatory provisions – the minimum and the maximum auditor tenure – as well as the whole regulatory regime must be seen in context and well coordinated in order to achieve desired effects. The analysis of this paper helps to measure effects of regulation and, thus, to accomplish the goal of the preparation of truthful financial statements by companies.

Some findings and criticisms of this paper may stimulate future research projects. The elements of the model like the cash flows used to characterize the values the actors consider to make their decisions could be altered to better represent reality through the model. For example, multi-periodic benefits out of the coalition could be only a special case. The suggested provisions only consider an initial ACMN and neglect interactions of the parties in future periods. Then there could be other conditions and different parameter values would have to be taken into account, leading to changes in the strategies of the actors. In addition, model assumptions such as a perfect audit effort seem to be decisive, but are not necessarily present in reality. If auditors cannot identify all infringements in the financial statements, incorrect financial statements can also pass through the auditing process without any collusion. In such cases, the management can forego negotiations with the auditor, which in general increases its incentive to manipulate. For a practical use, the inaccuracy of parameter estimates may distort the originally intended effects. It seems to be impossible to implement the optimal auditor tenure exactly. In conclusion, further examinations of these issues and further development of the model may be an interesting objective of future research projects.

References

- Beattie, V., Fearnley, S. and Brandt, R. (2004), A Grounded Theory Model of Auditor-Client Negotiations, *International Journal of Auditing* **8**(1), pp. 1–19.
- Bennett, G. B., Hatfield, R. C. and Stefaniak, C. (2015), The Effect of Deadline Pressure on Pre-Negotiation Positions: A Comparison of Auditors and Client Management, *Contemporary Accounting Research* **32**(4), pp. 1507–1528.
- Benston, G. J., Bromwich, M., Litan, R. E. and Wagenhofer, A. (2006*a*), Auditing and Accounting Standards: Their Roles and Usefulness to Independent Public Accountants and Investors, in: Benston, G. J., Bromwich, M., Litan, R. E. and Wagenhofer, A. (eds), *Worldwide Financial Reporting*, Oxford University Press, pp. 39–48.
- Benston, G. J., Bromwich, M., Litan, R. E. and Wagenhofer, A. (2006*b*), Why and How Audited Financial Accounting Statements Are Useful, in: Benston, G. J., Bromwich, M., Litan, R. E. and Wagenhofer, A. (eds), *Worldwide Financial Reporting*, Oxford University Press, pp. 17–38.
- Berle, A. A. and Means, G. C. (1991), *The Modern Corporation and Private Property*, 2nd edn., Taylor & Francis Inc., Milton Keynes.
- Booker, K. and Booker, Q. (2016), The Effects of Auditor Rotation and Client Importance on CPAs' Perceptions of Independence and Financial Statement Reliability, *Journal of Business and Economic Perspectives* **43**(1), pp. 13–37.
- Bowlin, K. O., Hobson, J. L. and Piercey, M. D. (2015), The Effects of Auditor Rotation, Professional Skepticism, and Interactions with Managers on Audit Quality, *The Accounting Review* **90**(4), pp. 1363–1393.
- Çınlar, E. and Sollenberger, N. J. (2013), *Introduction to Stochastic Processes*, Dover Publications Inc., Mineola, New York.
- Cook, J., Hatherly, D., Nadeau, L. and Thomas, L. C. (1997), Does Cooperation in Auditing Matter? A Comparison of a Non-Cooperative and a Cooperative Game Model of Auditing, *European Journal of Operational Research* **103**(3), pp. 470–482.
- DeAngelo, L. E. (1981*a*), Auditor Independence, 'Low Balling', and Disclosure Regulation, *Journal of Accounting & Economics* **3**(2), pp. 113–127.
- DeAngelo, L. E. (1981*b*), Auditor Size and Audit Quality, *Journal of Accounting & Economics* **3**(3), pp. 183–199.
- Demski, J. S. and Seieringa, R. J. (1974), A Cooperative Formulation of the Audit Choice Problem, *The Accounting Review* **49**(3), pp. 506–513.
- Fearnley, S., Beattie, V. A. and Brandt, R. (2005), Auditor Independence and Audit Risk: A Reconceptualization, *Journal of International Accounting Research* **4**(1), pp. 39–71.
- Fellingham, J. C. and Newman, D. P. (1985), Strategic Considerations in Auditing, *The Accounting Review* **60**(4), pp. 634–650.

- Fischbacher, U. and Stefani, U. (2007), Strategic Errors and Audit Quality: An Experimental Investigation, *The Accounting Review* **82**(3), pp. 679–704.
- Fischkin, M. and Gassen, J. (2011), Ökonomie des Abschlussprüferwechsels, *Zeitschrift für Betriebswirtschaft* **81**(7-8), pp. 855–900.
- Gibbins, M., McCracken, S. A. and Salterio, S. E. (2005), Negotiations over Accounting Issues: The Congruency of Audit Partner and Chief Financial Officer Recalls, *Auditing: A Journal of Practice & Theory* **24**(s-1), pp. 171–193.
- Gibbins, M., McCracken, S. and Salterio, S. (2007), The Chief Financial Officer’s Perspective on Auditor-Client Negotiations, *Contemporary Accounting Research* **24**(2), pp. 387–422.
- Gibbins, M., Salterio, S. and Webb, A. (2001), Evidence About Auditor-Client Management Negotiation Concerning Client’s Financial Reporting, *Journal of Accounting Research* **39**(3), pp. 535–563.
- Goldman, A. and Barlev, B. (1974), The Auditor-Firm Conflict of Interests: Its Implications for Independence, *The Accounting Review* **49**(4), pp. 707–718.
- Hatherly, D. (1996), Game Theory and the Auditor’s Penalty Regime, *Journal of Business Finance & Accounting* **23**(1), pp. 29–45.
- Healy, P. M. and Wahlen, J. M. (1999), A Review of the Earnings Management Literature and Its Implications for Standard Setting, *Accounting Horizons* **13**(4), pp. 365–383.
- Holler, M. J. and Nguyen, T. (2007), Regulating Balance Sheet Audit: A Game Theoretical Analysis, *Journal für Rechtspolitik* **15**(1), pp. 42–52.
- Jensen, M. C. and Meckling, W. H. (1976), Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, *Journal of Financial Economics* **3**(4), pp. 305–360.
- Joe, J. and Vandervelde, S. (2007), Do Auditor-Provided Nonaudit Services Improve Audit Effectiveness?, *Contemporary Accounting Research* **24**(2), pp. 467–487.
- Kanodia, C. and Mukherji, A. (1994), Audit Pricing, Lowballing and Auditor Turnover: A Dynamic Analysis, *The Accounting Review* **69**(4), pp. 593–615.
- Lee, C. W. J. and Gu, Z. (1998), Low Balling, Legal Liability and Auditor Independence, *The Accounting Review* **73**(4), pp. 533–555.
- Magee, R. P. and Tseng, M. C. (1990), Audit Pricing and Independence, *The Accounting Review* **65**(2), pp. 315–336.
- Matsumura, E. M. and Tucker, R. R. (1992), Fraud Detection: A Theoretical Foundation, *The Accounting Review* **67**(4), pp. 753–782.
- McCracken, S., Salterio, S. E. and Gibbins, M. (2008), Auditor-Client Management Relationships and Roles in Negotiating Financial Reporting, *Accounting, Organizations and Society* **33**(4-5), pp. 362–383.
- Nash, J. F. (1950), The Bargaining Problem, *Econometrica* **18**(2), pp. 155–162.

- Patterson, E. R. (1993), Strategic Sample Size Choice in Auditing, *Journal of Accounting Research* **31**(2), pp. 272–293.
- Sauviat, C. (2006), The Demise of Andersen: A Consequence of Corporate Governance Failure in the Context of Major Changes in the Accounting Profession and the Audit Market, in: Dembinski, P. H., Lager, C., Cornford, A. and Bonvin, J.-M. (eds), *Enron and World Finance*, Vol. 1, Palgrave Macmillan UK, London, pp. 143–158.
- Schatzberg, J. W. (1994), A New Examination of Auditor Low Ball Pricing Theoretical Model and Experimental Evidence, *Auditing: A Journal of Practice & Theory* **13**(1), pp. 33–55.
- Schatzberg, J. W. and Sevcik, G. R. (1994), A Multiperiod Model and Experimental Evidence of Independence and Lowballing, *Contemporary Accounting Research* **11**(1), pp. 137–174.
- Schmidt, R. N. and Cross, B. E. (2014), The Effects of Auditor Rotation on Client Management’s Negotiation Strategies, *Managerial Auditing Journal* **29**(2), pp. 110–130.
- Shibano, T. (1990), Assessing Audit Risk from Errors and Irregularities, *Journal of Accounting Research* **28**, pp. 110–140.
- Stefani, U. (2002), Abschlussprüfung, Unabhängigkeit und strategische Interdependenzen: Eine ökonomische Analyse institutioneller Reformen zur Steigerung der Prüfungsqualität, *Moderne Betriebswirtschaftslehre*, Schäffer-Poeschel, Stuttgart.
- Tepalagul, N. and Lin, L. (2015), Auditor Independence and Audit Quality: A Literature Review, *Journal of Accounting, Auditing and Finance* **30**(1), pp. 101–121.
- Wang, K. J. and Tuttle, B. M. (2009), The Impact of Auditor Rotation on Auditor-Client Negotiation, *Accounting, Organizations and Society* **34**(2), pp. 222–243.
- Zhang, P. (1999), A Bargaining Model of Auditor Reporting, *Contemporary Accounting Research* **16**(1), pp. 167–184.

Appendix

Nash Solution of the Bargaining Game

For a fixed set of model parameters in the Nash bargaining game (d, E) , where the status quo is defined through (13) and the efficient set through (11), the Nash bargaining solution is given by

$$\theta^* = \frac{N}{2(1+r)^{T_d} c_s \left(\left(\frac{1}{1+r} \right)^{T+1} r^2 + 2 \left(\frac{1}{1+r} \right)^{T+1} r + \left(\frac{1}{1+r} \right)^{T+1} - r - 1 \right)} \quad (14)$$

with

$$\begin{aligned} N := & \alpha_A r - \alpha_M r^2 - \alpha_M r + \alpha_A r^2 + (1+r)^{T_d} S - (1+r)^{T_d} c_s - e^{-\lambda T_p} \alpha_A r^2 + e^{-\lambda T_p} \alpha_M r^2 \\ & - (1+r)^{T_d} c_s r - \alpha_A r e^{-\lambda T_p} + e^{-\lambda T_p} \alpha_M r + (1+r)^{T_d} r^2 E_A + E_A r (1+r)^{T_d} \\ & - r E_M (1+r)^{T_d} - \left(\frac{1}{1+r} \right)^{T_A+1} (1+r)^{T_d} F [r^2 + 2r + 1] \\ & + \left(\frac{1}{1+r} \right)^{T+1} (1+r)^{T_d} [F - S + c_s + F r^2 - S r^2 + c_s r^2 + 2F r - 2S r + 2c_s r]. \end{aligned}$$

Minimum Auditor Tenure

Let a fixed set of model parameters in the Nash bargaining game (d, E) , where the status quo is defined through (13) and the efficient set through (11) be given. The MinAT \hat{T}'_A is the smallest T'_A leading to $\theta^* = 0$, where θ^* is given following (14). The analytical solution is

$$\hat{T}'_A = -\frac{1}{\log\left(\frac{1}{1+r}\right)} \left(\log\left(\frac{1}{1+r}\right) - \log \left[\frac{1}{(1+r)^2 F} \left((1+r)^{-T_d} B + C \right) \right] \right) \quad (15)$$

with

$$\begin{aligned} B := & e^{-\lambda T_p} \cdot (\alpha_M r^2 - \alpha_A r^2 - \alpha_A r + \alpha_M r) + \alpha_A r^2 + \alpha_A r - \alpha_M r^2 - \alpha_M r \\ C := & \left(\frac{1}{1+r} \right)^T \cdot \left[F r - S r + c_s r + F - S + c_s \right] + E_A r^2 + E_A r - E_M r + S - c_s - c_s r. \end{aligned}$$

Auditor Reward Regime

Similar to the MinAT, the bargaining can also be influenced by implementing another mechanism that incentivizes honest action. For example, the original status quo (2) can be enhanced by including a reward regime based on a certain $\beta_A > 0$. This leads to status quo

$$d'' := (d_M, d_A)' = \frac{1}{(1+r)} (S + E_M, F + E_A + \beta_A)' \quad (16)$$

In this situation, reward β_A compensates the costs of the auditor for not being reappointed after qualifying the audit opinion. Again, it shifts the status quo in direction (e) and, thus, increases the bargaining power of the auditor. So the basic principle here is comparable to

the one provided by the MinAT provision, however, it corresponds more to an a posteriori mechanism that rewards for certain audit findings. The nominal level of the reward can also be determined in order to prevent overcompensating of the auditor. The difficulty is that the reward must be integrated regardless of the actual reason for the refused reappointment. Consequently, it is more of a principle compensation than it is a reward.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorgelegte Dissertation selbst und ohne fremde Hilfe verfasst, eventuelle Beiträge von Ko-Autoren dokumentiert habe, nicht andere als die in ihr angegebenen Quellen oder Hilfsmittel benutzt, alle vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet sowie die Dissertation in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht habe.

Marburg, im Juni 2019

Dipl.-Math. oec Tobias Filusch