

Philipps



Universität
Marburg

Vergleich und Bewertung von Dichtevorhersagen
makroökonomischer Kennzahlen anhand von
„proper scoring rules“

Diplomarbeit

vorgelegt von

Martin Reuter
Hauptstr. 27
36132 Eiterfeld

am

Fachbereich Mathematik und Informatik
Philipps-Universität Marburg

betreut von

Prof. Dr. Hajo Holzmann

Marburg, den 27. März 2013

Inhaltsverzeichnis

1. Einleitung	1
2. Instrumente zur Bewertung von Vorhersagen	3
2.1. Mathematische Vorhersagen	3
2.2. Scoring rules	4
3. Darstellung von proper scoring rules über konvexe Funktionen	6
3.1. Proper scoring rules auf diskreten Ereignisräumen	6
3.1.1. Savage Darstellung für proper scoring rules	6
3.1.2. Expected score function und Bregman Differenzfunktion	8
3.1.3. Beispiele für proper scoring rules auf diskreten Ereignisräumen	9
3.1.4. Sonderfall Ja-Nein-Vorhersage	11
3.2. Proper scoring rules auf stetigen Ereignisräumen	13
3.2.1. Darstellung von proper scoring rules über konvexe Funktionale	13
3.2.2. Beispiele für proper scoring rules auf stetigen Ereignisräumen	16
3.2.3. Continuous ranked probability score (CRPS)	19
3.2.4. Energy score: Verallgemeinerung des CRPS	24
4. Gewichtete scoring rules	25
4.1. Konstruktion von gewichteten scoring rules	25
4.2. Schwellenwert und Quantil gewichtete Versionen des continuous ranked probability scores	27
5. Tests und grafische Vergleichsmethoden	31
5.1. Test auf gleiche Vorhersage Performance	31
5.2. Grafischer Vergleich durch Schwellenwert und Quantil Zerlegung des CRPS	33
6. Praktische Anwendung	35
6.1. Datengrundlage für die Berechnungen	35
6.1.1. Inflationsrate	35
6.1.2. Bruttoinlandsprodukt	38
6.1.3. Arbeitslosenquote	39
6.2. Vorhersagemethoden	41
6.2.1. Grundlagen der Zeitreihenanalyse	41
6.2.2. Einfache Vergleichsvorhersagen	44
6.2.3. Autoregressives Zeitreihenmodell	45
6.2.4. Bank-of-England-Vorhersage	46

6.2.5. Gaussian-Mixture-Model	48
6.2.6. t-Verteilung mit Maximum-Likelihood-Schätzung	49
7. Berechnungen und Analysen	50
7.1. Vorgehensweise bei der Analyse	50
7.2. Analyse der Inflationsrate	53
7.3. Analyse des Bruttoinlandsproduktes	62
7.4. Analyse der Arbeitslosenquote	70
8. Ansätze zur Verfeinerung der Vorhersage	78
8.1. Autoregression mit exogenen Variablen	78
8.1.1. Vorhersage des Bruttoinlandsproduktes unter Verwendung des ifo-Geschäftsklimaindexes	79
8.1.2. Vorhersage der Arbeitslosenquote unter Verwendung des Bruttoinlandsproduktes	81
8.2. Weitere Ansätze im Überblick	83
9. Zusammenfassung und Fazit	86
A. Anhang	88
A.1. Herleitungen zu den Beispielen aus Abschnitt 3.1.3	88
A.1.1. Quadratic score	88
A.1.2. Pseudospherical score	89
A.1.3. Logarithmic score	90
A.2. Herleitung zu den Beispielen aus Abschnitt 3.2.2	91
A.2.1. Quadratic score	91
A.2.2. Pseudospherical score	91
A.2.3. Logarithmic score	92
Literaturverzeichnis	94
Eigenständigkeitserklärung	98
Danksagung	98

Abbildungsverzeichnis

6.1.1.Harmonisierter Verbraucherpreisindex im Zeitverlauf	36
6.1.2.Inflationsrate im Zeitverlauf	37
6.1.3.Saisonbereinigte Inflationsrate im Zeitverlauf	37
6.1.4.Wachstumsraten des Bruttoinlandsproduktes im Zeitverlauf	38
6.1.5.Arbeitslosenquote im Zeitverlauf	40
6.1.6.Wachstumsraten der Arbeitslosenquote in Zeitverlauf	40
6.2.1.Dichtefunktionen der Vorhersagemethoden	48
7.2.1.Dichte der saisonbereinigten Inflationsraten und Dichtevorhersagen	56
7.2.2.Autokorrelationen der saisonbereinigten Inflationsraten	56
7.2.3.Mittlere scores in Abhängigkeit von der Länge des training windows (Inflationsraten)	57
7.2.4.Schwellenwert Zerlegung des CRPS (Inflationsraten)	58
7.2.5.Quantil Zerlegung des CRPS (Inflationsraten)	58
7.3.1.Dichte der Wachstumsraten des Bruttoinlandsproduktes und Dichtevorhersagen	65
7.3.2.Autokorrelationen der Wachstumsraten des Bruttoinlandsproduktes . . .	65
7.3.3.Mittlere scores in Abhängigkeit von der Länge des training windows (Bruttoinlandsprodukt)	66
7.3.4.Schwellenwert Zerlegung des CRPS (Bruttoinlandsprodukt)	67
7.3.5.Quantil Zerlegung des CRPS (Bruttoinlandsprodukt)	67
7.4.1.Dichte der Wachstumsraten der Arbeitslosenquoten und Dichtevorhersagen	73
7.4.2.Autokorrelationen der Wachstumsraten der Arbeitslosenquote	73
7.4.3.Mittlere scores in Abhängigkeit von der Länge des training windows (Arbeitslosenquote)	74
7.4.4.Schwellenwert Zerlegung des CRPS (Arbeitslosenquote)	75
7.4.5.Quantil Zerlegung des CRPS (Arbeitslosenquote)	75
8.1.1.Autokorrelationen zwischen Bruttoinlandsprodukt und ifo-Geschäftsklimaindex	80
8.1.2.Autokorrelationen zwischen Arbeitslosenquote und Bruttoinlandsprodukt	82

Tabellenverzeichnis

4.2.1.Gewichtungsfunktionen	30
7.2.1.Scores Inflationsrate (Einschrittvorhersage)	59
7.2.2.Tests Inflationsrate (Einschrittvorhersage)	59
7.2.3.Gewichtete Scores Inflationsrate	60
7.2.4.Tests Inflationsrate (gewichtete scores)	60
7.2.5.Scores Inflationsrate (Zweischrittvorhersage)	61
7.2.6.Tests Inflationsrate (Zweischrittvorhersage)	61
7.3.1.Scores Bruttoinlandsprodukt (Einschrittvorhersage)	68
7.3.2.Tests Bruttoinlandsprodukt (Einschrittvorhersage)	68
7.3.3.Gewichtete Scores Bruttoinlandsprodukt	68
7.3.4.Tests Bruttoinlandsprodukt (gewichtete scores)	69
7.3.5.Scores Bruttoinlandsprodukt (Zweischrittvorhersage)	69
7.3.6.Tests Bruttoinlandsprodukt (Zweischrittvorhersage)	69
7.4.1.Scores Arbeitslosenquote (Einschrittvorhersage)	76
7.4.2.Tests Arbeitslosenquote (Einschrittvorhersage)	76
7.4.3.Gewichtete Scores Arbeitslosenquote	76
7.4.4.Tests Arbeitslosenquote (gewichtete scores)	77
7.4.5.Scores Arbeitslosenquote (Zweischrittvorhersage)	77
7.4.6.Tests Arbeitslosenquote (Zweischrittvorhersage)	77
8.1.1.Scores des erweiterten AR Modells (Bruttoinlandsprodukt)	80
8.1.2.Scores des erweiterten AR Modells (Arbeitslosenquote)	82

1. Einleitung

Prognosen über die wirtschaftliche Entwicklung eines Landes spielen in der heutigen Zeit eine wichtige Rolle. Das Bundesfinanzministerium benötigt sie für die Festlegung des Bundeshaushaltes, Unternehmen sind auf sie angewiesen, um ihre Finanz- und Produktionsplanungen durchzuführen, und Privatpersonen machen Investitions- und Anlageentscheidungen von ihnen abhängig. Aus diesem Grunde erstellen und veröffentlichen beispielsweise das Finanzministerium oder auch Wirtschaftsforschungsinstitute in regelmäßigen Abständen Vorhersagen zur Entwicklung von makroökonomischen Kennzahlen innerhalb der nächsten Periode. Die Einschätzungen der Vorhersagenden spiegeln sich am besten in einer Dichte wider, welche die erwartete Verteilung der Kennzahl beschreibt. Doch wie erzeugt man entsprechende Wahrscheinlichkeitsvorhersagen und woher weiß man, ob eine Vorhersagemethode gute Vorhersagen liefert? Mit diesen Fragen werden wir uns in dieser Diplomarbeit beschäftigen.

Dazu werden wir im nachfolgenden Kapitel zunächst darauf eingehen, was man sich unter einer mathematischen Vorhersage vorzustellen hat. Anschließend werden so genannte *proper scoring rules* eingeführt, mit denen man Wahrscheinlichkeitsvorhersagen bewerten kann.

Im zweiten Kapitel befassen wir uns mit der Theorie zu *proper scoring rules* und zeigen, wie man *proper scoring rules* konstruieren kann. Dabei wird zunächst der diskrete Fall betrachtet, in dem nur von einer endlichen Anzahl an Ereignissen ausgegangen wird, die in der Zukunft eintreten können. Danach setzen wir uns auch mit dem stetigen Fall auseinander. Es werden jeweils entsprechende Beispiele für *proper scoring rules* vorgestellt, wobei wir uns in Abschnitt 3.2.3 speziell mit dem *continuous ranked probability score* (CRPS) beschäftigen wollen, der in der Literatur in jüngster Zeit an Bedeutung gewonnen hat.

Je nachdem, ob man im Risikomanagement tätig ist und sich deshalb besonders mit möglichen Extremereignissen befasst, oder ob man an der Produktionsplanung arbeitet, bei der die mit hoher Wahrscheinlichkeit erwartete wirtschaftliche Entwicklung entscheidend ist, legt man vornehmlich Wert auf einen bestimmten Bereich der Dichtevorhersage. In Kapitel 4 werden deshalb gewichtete *scoring rules* vorgestellt, die spezielle Abschnitte der Dichtevorhersage besonders kritisch betrachten, in dem sie eine Gewichtungsfunktion in die Bewertung einbeziehen.

Im anschließenden Kapitel werden wir uns mit Tests und grafischen Darstellungsmethoden auseinandersetzen, die zum Vergleich von Wahrscheinlichkeitsvorhersagen herange-

zogen werden können. Dabei wird zunächst auf einen von Gneiting & Ranjan (2011) vorgeschlagenen Test zum Vergleich zweier Vorhersagemethoden auf gleiche Vorhersage Performance eingegangen. Anschließend werden zwei Methoden von Gneiting & Ranjan (2011) vorgestellt, welche die Bewertung von Vorhersagen grafisch darstellen.

Die nachfolgenden Kapitel enthalten den praktischen Teil der Arbeit, in welchem wir uns mit der Erzeugung von Vorhersagen für makroökonomische Kennzahlen befassen und diese auf Grundlage historischer Daten mit den vorgestellten proper scoring rules bewerten.

Zunächst werden in Kapitel 6 die makroökonomischen Daten vorgestellt, für die Wahrscheinlichkeitsvorhersagen erzeugt werden sollen. Wir beschränken uns dabei auf die Inflationsrate, das Bruttoinlandsprodukt und die Arbeitslosenquote der Bundesrepublik Deutschland. Danach werden wir uns mit den Vorhersagemethoden auseinander setzen, die bei der späteren Analyse verglichen werden sollen. Es handelt sich dabei um drei einfache, intuitive Ansätze, ein Zeitreihenmodell und drei komplexere Verfahren, bei denen die Dichtevorhersage mithilfe einer Maximum-Likelihood-Schätzung ermittelt wird. Bevor die einzelnen Verfahren beschrieben werden, erfolgt eine kurze Einführung in die Zeitreihenanalyse, bei der wir auf Modellierungsannahmen eingehen und auf Eigenschaften unserer historischen Daten hinweisen, die bei der Analyse berücksichtigt werden müssen.

In Kapitel 7 geht es schließlich um die Berechnungen und die Analyse der makroökonomischen Kennzahlen. Zunächst werden die Analyseschritte erläutert, die im Anschluss an allen drei Datenreihen durchgeführt werden. Die Rechenergebnisse und die gewonnenen Erkenntnisse werden für jede Kennzahl in einem separaten Unterkapitel beschrieben. Die zugrunde liegenden Berechnungen erfolgen mit der Statistik-Software R.

Im vorletzten Kapitel wollen wir auf weitere Ansätze eingehen, die zu einer Verbesserung der Vorhersagequalität führen können. Dabei wird zu Beginn ein Verfahren vorgestellt, das durch Aufnahme zusätzlicher Inputvariablen das bereits bei der Analyse angewandte Zeitreihenmodell erweitert. Mit diesem neuen Modell erzeugen wir ebenfalls Vorhersagen und vergleichen diese mit den bisherigen Ergebnissen. Hierbei beschränken wir uns jedoch auf zwei Fälle. Zum einen wird die Entwicklung des Bruttoinlandsproduktes unter Verwendung des ifo-Geschäftsklimaindexes prognostiziert und zum anderen werden Vorhersagen für die Arbeitslosenquote unter Berücksichtigung der historischen Werte des Bruttoinlandsproduktes ermittelt. Im weiteren Verlauf dieses Kapitels werden wir auf weitere Ansätze und Verfahren eingehen, die in der Praxis zur Vorhersage unserer drei Kennzahlen herangezogen werden. Es erfolgt jeweils eine kurze Beschreibung der Verfahren und es werden entsprechende Referenzen angegeben.

Abgeschlossen wird die Diplomarbeit durch eine kurze Zusammenfassung und ein Fazit.

2. Instrumente zur Bewertung von Vorhersagen

Heutzutage werden oftmals mathematische Methoden verwendet, um Vorhersagen über die zukünftige Entwicklung von makroökonomischen Kennzahlen zu erzeugen. Diese Methoden können prinzipiell als Inputvariablen alle zum Zeitpunkt der Vorhersage vorhandenen Informationen enthalten, die man in irgendeiner Art und Weise quantifizieren kann. Dazu zählen im Allgemeinen die Vergangenheitswerte der Zeitreihe, aber auch andere Indikatoren, die aus Sicht des Vorhersagenden einen Einfluss auf die zukünftige Entwicklung der Zielgröße haben können. Einige solcher Vorhersagemethoden werden im späteren Verlauf noch vorgestellt. Wir wollen uns aber zunächst damit auseinandersetzen, wie das Resultat einer Vorhersage aussehen kann. Dazu gehen wir im nächsten Abschnitt auf den Unterschied zwischen Punkt- und Wahrscheinlichkeitsvorhersagen ein. Anschließend schauen wir uns an, wie man die Qualität von Wahrscheinlichkeitsvorhersagen mithilfe von scoring rules bewerten kann.

2.1. Mathematische Vorhersagen

Im mathematischen Sinne geht man bei der Ermittlung von Vorhersagen davon aus, dass es eine Menge Ω von Ereignissen gibt, die in der Zukunft eintreten können. Da man im Allgemeinen nicht weiß, welches $\omega \in \Omega$ sich in der Zukunft ereignen wird, versucht man unter Berücksichtigung bekannter Informationen eine Aussage über das Eintreten der möglichen Ereignisse zu treffen. Eine einfache Möglichkeit um seine Erwartungen über die Zukunft zu äußern besteht darin, ein bestimmtes Ereignis ω' aus der Ereignismenge Ω als zu erwartendes Ereignis anzugeben. Eine solche Vorhersage nennt man Punktvorhersage. Punktvorhersagen sind in der Praxis weit verbreitet, da sie eine einfache und leicht verständliche Form der Vorhersage darstellen. Sie haben jedoch den Nachteil, dass man von dem Vorhersagenden weder eine Angabe darüber erhält, wie hoch er die Wahrscheinlichkeit für das Eintreten des Ereignisses einschätzt, noch, wie seine Einschätzungen bezüglich anderer Ereignisse sind. Außerdem muss man bei Punktvorhersagen beachten, dass oftmals nicht ausreichend spezifiziert wird, was der Vorhersagende genau vorhersagt. Vielmehr wird nur seine Prognose erfragt bzw. angegeben. Es wird deshalb beispielsweise nicht deutlich, ob es sich bei dem vorhergesagten Ereignis um das Ereignis handelt, welches aus Sicht des Vorhersagenden unter allen Ereignissen die höchste Wahrscheinlichkeit besitzt, oder ob das Ereignis vielmehr ein Stellvertreter dafür ist, was man im Durchschnitt erwarten kann. Hier können sich Unterschiede ergeben, wenn man für die Vorhersage eine Verteilung annimmt, bei der der Erwartungswert der Verteilung und die Stelle, an der die Dichte der Verteilung maximal ist, nicht über-

einstimmen.

Um solche Probleme zu umgehen, kann man als Vorhersage auch die Verteilung angeben, die man den zukünftigen Ereignissen zugrunde legt. In einem solchen Fall spricht man von einer Wahrscheinlichkeitsvorhersage. Mathematisch gesehen gibt man ein Wahrscheinlichkeitsmaß beziehungsweise eine Wahrscheinlichkeitsdichte auf dem messbaren Raum (Ω, \mathcal{A}) an, wobei Ω die Ereignismenge und \mathcal{A} eine σ -Algebra auf Ω darstellt. Ein solches Wahrscheinlichkeitsmaß beziehungsweise eine solche Wahrscheinlichkeitsdichte spiegelt alle Erwartungen des Vorhersagenden an die Entwicklung der Zielgröße wider. Der Nutzer einer solchen Vorhersage kann sich beispielsweise durch die Ermittlung des Erwartungswertes oder der Varianz oder durch die Berechnung von Quantilen der Verteilung ein genaues Bild über die Einschätzung des Vorhersagenden machen. Die beschriebenen Erläuterungen, sowie weiterführende Informationen zu mathematischen Vorhersagen und zu den Vorteilen von Wahrscheinlichkeitsvorhersagen gegenüber Punktvorhersagen sind zum Beispiel Elliott & Timmermann (2008) und Gneiting (2011) zu entnehmen.

Sowohl bei der Erzeugung, als auch bei der Analyse von Vorhersagen gilt es zu beachten, dass diese zeitpunktbezogen sind. Zum einen ist der Zeitpunkt der Erstellung der Vorhersage entscheidend, da nur die bis dato bekannten Informationen in die Vorhersage einfließen. Zum anderen bezieht sich eine Vorhersage immer auf einen bestimmten Zeitpunkt oder auf ein bestimmtes Zeitfenster in der Zukunft. So geben beispielsweise Wirtschaftsforschungsinstitute an, wie hoch die Zahl der Arbeitslosen zu einem bestimmten Stichtag voraussichtlich sein wird oder um wie viel Prozent das Bruttoinlandsprodukt ihrer Meinung nach innerhalb des nächsten Jahres zulegen wird. Derartige Zeitbezüge werden im weiteren Verlauf der Diplomarbeit, soweit sie relevant sind, entsprechend dargelegt und berücksichtigt.

2.2. Scoring rules

In den vergangenen Jahrzehnten entwickelte sich eine umfangreiche Theorie zur Wahrscheinlichkeitsvorhersage und es wurden verschiedenste Vorhersagemethoden erarbeitet. Um die Qualität der durch die einzelnen Methoden erzeugten Vorhersagen bewerten zu können, führt man einen Abgleich mit den real eingetretenen Ereignissen durch und verwendet dabei so genannte *scoring rules*.

Sei \mathcal{P} eine konvexe Klasse von Wahrscheinlichkeitsmaßen auf dem Raum (Ω, \mathcal{A}) und sei $P \in \mathcal{P}$ eine Wahrscheinlichkeitsvorhersage. Dann versteht man unter einer scoring rule eine reellwertige Abbildung $S : \mathcal{P} \times \Omega \rightarrow \overline{\mathbb{R}}$, wobei $S(P, \cdot)$ \mathcal{P} -quasi-integrierbar sei. Es ergibt sich entsprechend ein score von $S(P, \omega)$, falls $P \in \mathcal{P}$ als Wahrscheinlichkeitsvorhersage angenommen wurde und $\omega \in \Omega$ als reales Ereignis eintritt.

Da der score $S(P, \omega)$ sehr stark von dem Ereignis ω abhängt und im Vorhinein nicht bekannt ist, welches Ereignis eintreten wird, betrachtet man auch den erwarteten score

unter Annahme der wahren Verteilung Q

$$S(P, Q) = \int S(P, \omega) dQ(\omega). \quad (2.2.1)$$

In dem speziellen Fall, dass die vorhergesagte Verteilung P die Dichte f und die tatsächliche Verteilung Q die Dichte g bezüglich des Lebesgue-Maßes besitzen, entspricht die obige Darstellung (2.2.1) der folgenden:

$$\mathbb{E}_g S(f, Y) = \int_{\Omega} S(f, y) g(y) dy \quad (2.2.2)$$

Es existieren im Allgemeinen viele Möglichkeiten eine scoring rule zu konstruieren. Um allerdings eine vernünftige praktische Anwendung der scoring rules gewährleisten zu können, ist eine Einschränkung bzw. eine Charakterisierung der scoring rules von Nöten. Erste Ansätze gehen auf Good (1952) zurück und besagen, dass scoring rules den Vorhersager aus rationaler Sicht dazu verleiten sollten, seine wahren Erwartungen an die Zukunft als Vorhersage anzugeben. Als eine Folge dieser Überlegungen entstand die Definition von 'geeigneten', *proper*, scoring rules (siehe z.B. Gneiting & Raftery (2007)).

Definition 2.2.1 Eine scoring rule $S(P, \omega)$ ist *proper* bezüglich \mathcal{P} , falls

$$S(Q, Q) \geq S(P, Q), \quad \forall P, Q \in \mathcal{P}. \quad (2.2.3)$$

Sie ist *strictly proper* bezüglich \mathcal{P} , falls genau dann Gleichheit herrscht, wenn $P=Q$ gilt.

In der Praxis werden meist in regelmäßigen Abständen Prognosen über den Verlauf einer Zielgröße erstellt. Ein Unternehmen möchte beispielsweise für die innerbetriebliche Finanzplanung jeden Monat eine Einjahres-Vorhersage der Inflation oder des Zinssatzes heranziehen, um die kurzfristige Finanzentwicklung des Unternehmens einschätzen zu können. Gehen wir nun einmal davon aus, dass wir einen Zeitraum von n Perioden betrachten und uns für die Qualitätsbewertung unserer Vorhersagemethode n Wahrscheinlichkeitsvorhersagen P_1, \dots, P_n und n tatsächlich eingetretene Inflationsraten bzw. Zinssätze y_1, \dots, y_n vorliegen. Aus diesen Daten können wir für jede Periode i einen score $S(P_i, y_i)$ berechnen und anschließend durch Mittelwertbildung den *mittleren score*

$$\bar{S}_n = \frac{1}{n} \sum_{i=1}^n S(P_i, y_i) \quad (2.2.4)$$

bestimmen (siehe z.B. Gneiting & Raftery (2007)). Ein einzelner mittlerer score besitzt oftmals wenig Aussagekraft. Interessant wird es erst, wenn man unterschiedliche Vorhersagemethoden zur Verfügung hat. Dann kann man mit jedem Verfahren die n Vorhersagen berechnen und den mittleren score bestimmen. Durch einen Vergleich der ermittelten mittleren scores lässt sich nun die Vorhersage Performance der Verfahren bewerten. Für eine sinnvolle Bewertung sollte man allerdings darauf achten, dass man für die Vorhersagen die jeweils gleiche Datenbasis verwendet.

3. Darstellung von proper scoring rules über konvexe Funktionen

In Kapitel 2 wurde erläutert, warum wir uns bei der Wahl der scoring rule auf die Klasse der proper scoring rules beschränken sollten. Die ersten scoring rules, welche die Eigenschaft proper besitzen, wurden bereits durch Brier (1950) und Good (1952) konstruiert. McCarthy (1956) begann damit, eine allgemeine Darstellung für proper scoring rules herzuleiten. Bei den anfänglichen Überlegungen beschränkte man sich auf diskrete Ereignisräume und Savage (1971) gelang es eine allgemeine Formel für proper scoring rules über konvexe Funktionen zu entwickeln. Hendrickson & Buehler (1971) verallgemeinerten die Theorie auf stetige Ereignisräume und Gneiting & Raftery (2007) konstruierten eine zu Savage (1971) äquivalente Darstellung von proper scoring rules auf stetigen Ereignisräumen. Im Folgenden wird zuerst auf den diskreten Fall eingegangen, während in Abschnitt 3.2 der stetige Fall erläutert wird.

3.1. Proper scoring rules auf diskreten Ereignisräumen

Der Ereignisraum, auf dem eine Wahrscheinlichkeitsvorhersage durchgeführt werden soll, kann sehr unterschiedlich sein. Manchmal können nur endlich viele Ereignisse eintreten, wie beispielsweise bei dem Ausgang eines Fußballspiels (Sieg, Unentschieden, Niederlage), während in anderen Situationen unendlich viele Ereignisse möglich sind. Hierzu wäre als Beispiel die Schätzung der Inflationsrate oder des Bruttoinlandsproduktes zu nennen. Im diesem Abschnitt wird zunächst die Savage Darstellung für proper scoring rules vorgestellt, welche die Eigenschaften von proper scoring rules auf den diskreten Ereignisräumen charakterisiert. Anschließend werden entsprechende Beispiele für scoring rules dargestellt.

3.1.1. Savage Darstellung für proper scoring rules

Im Folgenden wird von einem Ereignisraum $\Omega = \{1, \dots, m\}$ mit einer endlichen Anzahl m sich gegenseitig ausschließender Ereignisse ausgegangen. Eine entsprechende Wahrscheinlichkeitsvorhersage auf diesem Raum besteht aus einem Vektor $p = (p_1, \dots, p_m)^T$, wobei dem Ereignis i die Eintrittswahrscheinlichkeit p_i zugeordnet wird. Die Menge der möglichen Wahrscheinlichkeitsmaße kann zusammengefasst werden in der konvexen Klasse $\mathcal{P} = \mathcal{P}_m$, mit

$$\mathcal{P}_m = \left\{ p = (p_1, \dots, p_m)^T : p_1, \dots, p_m \geq 0, \quad p_1 + \dots + p_m = 1 \right\}. \quad (3.1.1)$$

Eine scoring rule für eine solche Wahrscheinlichkeitsvorhersage ergibt sich aus der Kollektion von m Funktionen der Form

$$S(\cdot, i) : \mathcal{P} \rightarrow \overline{\mathbb{R}}, \quad i = 1, \dots, m. \quad (3.1.2)$$

Nimmt man nun als Vorhersage den Wahrscheinlichkeitsvektor p an, dann ergibt sich bei dem Eintreten von Ereignis i ein score in Höhe von $S(p, i)$. Die allgemeine Definition 2.2.1 für proper scoring rules überträgt sich entsprechend auf den diskreten Fall, dass heißt eine scoring rule S ist proper, falls

$$S(q, q) = \sum_{i=1}^m q_i S(q, i) \geq \sum_{i=1}^m q_i S(p, i) = S(p, q), \quad \forall p, q \in \mathcal{P}_m \quad (3.1.3)$$

gilt (Gneiting & Raftery, 2007).

Die Savage Darstellung für proper scoring rules beruht auf der Verwendung von konvexen Funktionen und deren Eigenschaften. Deshalb erfolgt nun eine kurze Einführung in die Theorie der konvexen Funktionen (siehe z.B. Rockafellar (1970)).

Definition 3.1.1 Eine Funktion $G : \mathcal{P}_m \rightarrow \mathbb{R}$ ist konvex, falls gilt

$$G((1 - \lambda)p + \lambda q) \leq (1 - \lambda)G(p) + \lambda G(q), \quad \forall \lambda \in (0, 1), p, q \in \mathcal{P}_m. \quad (3.1.4)$$

G ist streng konvex, falls in Ungleichung (3.1.4) genau dann Gleichheit eintritt, falls $p = q$ vorliegt.

Sei im Folgenden $G : \mathcal{P}_m \rightarrow \mathbb{R}$ eine konvexe Funktion.

Definition 3.1.2 Ein Vektor $G'(p) = (G'_1(p), \dots, G'_m(p))^T$ ist ein Subgradient von G in Punkt $p \in \mathcal{P}_m$, wenn

$$G(q) \geq G(p) + \langle G'(p), q - p \rangle, \quad \forall q \in \mathcal{P}_m \quad (3.1.5)$$

erfüllt ist. Dabei bezeichnet $\langle \cdot, \cdot \rangle$ das Standard Skalarprodukt und es wird davon ausgegangen, dass die Komponenten von $G'(p)$ reellwertig sind, wobei als Ausnahme $G'_i(p) = -\infty$ im Falle $p_i = 0$ erlaubt ist.

Eine konvexe Funktion kann in einem Punkt mehrere Subgradienten haben (Rockafellar, 1970). Nach Theorem 25.1 aus Rockafellar (1970) wird jedoch durch den Gradienten der Funktion G im Punkt p der eindeutige Subgradient von G in p beschrieben, falls G in p differenzierbar und $G(p)$ endlich ist.

Definition 3.1.3 Eine scoring rule S für diskrete Wahrscheinlichkeitsvorhersagen ist regular, falls $S(\cdot, i)$ für $i = 1, \dots, m$ reellwertig ist, mit Ausnahme, dass $S(p, i) = -\infty$ für $p_i = 0$ erlaubt ist.

Die vorangegangenen Definitionen und Erläuterungen dienen als Grundlage für das nachfolgende Theorem, welches die Savage Darstellung für proper scoring rules enthält (Gneiting & Raftery (2007), Theorem 3.2).

Theorem 3.1.1 *Eine regular scoring rule S für diskrete Wahrscheinlichkeitsvorhersagen ist genau dann (strictly) proper, falls gilt*

$$S(p, i) = G(p) - \langle G'(p), p \rangle + G'_i(p), \quad i = 1, \dots, m, \quad (3.1.6)$$

wobei $G : \mathcal{P}_m \rightarrow \mathbb{R}$, $p \mapsto S(p, p)$, eine (streng) konvexe Funktion ist und $G'(p)$ ein Subgradient von G in Punkt p für alle $p \in \mathcal{P}_m$.

Im späteren Verlauf werden wir das allgemeinere Theorem 3.2.1 kennen lernen. Da die Beweise der Theoreme äquivalent erfolgen, wird auf den Beweis für den diskreten Fall verzichtet und auf den Beweis des allgemeineren Theorems 3.2.1 verwiesen.

Aus dem Theorem lässt sich die etwas anschaulichere Aussage folgern, dass eine regular scoring rule S genau dann (strictly) proper ist, wenn die Funktion $G(p) := S(p, p)$ (streng) konvex auf \mathcal{P}_m ist und der Vektor mit den Komponenten $S(p, i)$, $i = 1, \dots, m$, ein Subgradient von G in Punkt p ist, und zwar für alle $p \in \mathcal{P}_m$. Hieraus lässt sich wiederum schließen, dass jede beschränkte, (streng) konvexe Funktion G auf \mathcal{P}_m eine regular (strictly) proper scoring rule generiert (Gneiting & Raftery, 2007). Einige solcher konvexen Funktionen und die entsprechenden scoring rules werden in Abschnitt 3.1.3 beschrieben.

3.1.2. Expected score function und Bregman Differenzfunktion

Durch die Savage Darstellung wird eine proper scoring rule eindeutig durch die konvexe, reelwertige Funktion

$$G(p) = S(p, p) = \sup_{q \in \mathcal{P}_m} S(q, p), \quad p \in \mathcal{P}_m, \quad (3.1.7)$$

charakterisiert. Diese Funktion wird in Anlehnung an Grünwald & Dawid (2004) und Buja et al. (2005) *expected score function* genannt und kennzeichnet den höchsten zu erreichenden score in Abhängigkeit von p . Weitere gebräuchliche Bezeichnungen sind *uncertainty measure* oder *generalized entropy function* (Gneiting & Raftery, 2007).

Anhand dieses maximal zu erreichenden scores lässt sich eine Differenzfunktion herleiten, welche bei Vorhersage p und tatsächlicher Verteilung q die entsprechende score Abweichung ermittelt. Für eine regular proper scoring rule S erhält man die nichtnegative Differenzfunktion

$$d(p, q) = S(q, q) - S(p, q), \quad p, q \in \mathcal{P}_m. \quad (3.1.8)$$

In Falle einer strictly proper scoring rule ist diese Differenzfunktion sogar streng positiv für alle $p \neq q$ (Gneiting & Raftery, 2007). Die Verwendung von Differenzfunktionen wurde bereits von Bregman (1967) untersucht. Er studierte diese unter der Annahme eines endlichen Ereignisraums und bezeichnete sie als *Bregman divergence*, falls die entsprechende entropy function ausreichend glatt ist. Bregman divergences spielen besonders in der Optimierung eine wichtige Rolle (Gneiting & Raftery, 2007).

3.1.3. Beispiele für proper scoring rules auf diskreten Ereignisräumen

Im Folgenden werden einige proper scoring rules vorgestellt. Diese lassen sich jeweils mit Hilfe von Theorem 3.1.1 aus der entsprechenden expected score function herleiten und es lässt sich zu jedem score die entsprechende Bregman divergence ermitteln. Da man aus den Herleitungen für die drei Beispiele keine besonderen Erkenntnisse ziehen kann und sie nur den Lesefluss beeinträchtigen würden, sind diese dem Anhang A.1 zu entnehmen. Weiterhin erfolgt der Nachweis, dass es sich bei den drei expected score functions um konvexe Funktionen handelt, im Anhang. Die Beispiele stammen aus Gneiting & Raftery (2007).

Eine der bekanntesten scoring rules ist der *quadratic score*

$$S(p, i) = 2p_i - \sum_{j=1}^m p_j^2, \quad i = 1, \dots, m, \quad (3.1.9)$$

welcher sich aus der konvexen Funktion $G(p) = \sum_{j=1}^m p_j^2$ herleitet. Für den quadratic score ergibt sich entsprechend Formel (3.1.8) eine Bregman divergence von $d(p, q) = \sum_{j=1}^m (p_j - q_j)^2$, welche der quadrierten Euklidischen Norm entspricht. Der quadratic score ist eng verwandt mit dem von Brier (1950) hergeleiteten score

$$S(p, i) = \sum_{j=1}^m (\delta_{ij} - p_j)^2, \quad i = 1, \dots, m, \quad (3.1.10)$$

wobei $\delta_{ij} = 1$ ist, falls $i = j$ vorliegt, und $\delta_{ij} = 0$ sonst. Der *Brier score* ergibt sich durch Negieren des quadratic scores und Addition von 1. Die Negation hat zur Folge, dass der Vorhersager beim Brier score einen minimalen score anstreben sollte, während bei der in dieser Arbeit verwendeten Definition von proper scoring rules ein maximaler score wünschenswert ist.

In Anlehnung an die zur Herleitung des quadratic scores genutzte expected score function $G(p) = \sum_{j=1}^m p_j^2$ lässt sich für $\alpha > 1$ die Funktion

$$G(p) = \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}} \quad (3.1.11)$$

ableiten. Bei dieser handelt es sich wiederum um eine konvexe Funktion, welche den erstmals von Good (1971) vorgestellten *pseudospherical score*

$$S(p, i) = \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} p_i^{\alpha-1}, \quad i = 1, \dots, m \quad (3.1.12)$$

induziert. Die im Allgemeinen in der Literatur als *spherical score* bezeichnete scoring rule ergibt sich für $\alpha=2$. Die entsprechende Bregman divergence lautet

$$d(p, q) = \left(\sum_{j=1}^m q_j^\alpha \right)^{\frac{1}{\alpha}} - \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} \sum_{i=1}^m q_i p_i^{\alpha-1}. \quad (3.1.13)$$

Der bereits von Good (1952) verwendete *logarithmic score*

$$S(p, i) = \log p_i, \quad i = 1, \dots, m \quad (3.1.14)$$

stellt eine weitere proper scoring rule dar, welche sich von der erstmals durch Shannon (1948) als Bewertungsfunktion genutzten negativen *Shannon entropy*

$$G(p) = \sum_{j=1}^m p_j \log p_j \quad (3.1.15)$$

ableitet. Als Bregman divergence ergibt sich die *Kullback-Leibler divergence*

$$d(p, q) = \sum_{i=1}^m q_i \log \frac{q_i}{p_i}. \quad (3.1.16)$$

Der logarithmic score wird auch *ignorance score* genannt und besitzt weit verbreitete Anwendungsmöglichkeiten, vgl. z.B. Roulston & Smith (2002).

Nun wollen wir noch den *linear score* $S(p, i) = p_i$, $i = 1, \dots, m$, betrachten. Dieser scheint auf den ersten Blick zwar sehr vernünftig zu sein, aber er erweist sich als nicht proper. Um dies zu zeigen betrachten wir ein einfaches Beispiel. Dazu beschränken wir unseren Ereignisraum auf die zwei Elemente ω und $\bar{\omega}$ und nehmen an, dass ω mit der Wahrscheinlichkeit $q(\omega) = 3/4$ und $\bar{\omega}$ mit der Wahrscheinlichkeit $q(\bar{\omega}) = 1/4$ eintritt. Wir betrachten als Vorhersage die Verteilung p , welche ω mit Wahrscheinlichkeit 1 annimmt, und zeigen, dass diese unter Annahme der wahren Verteilung q einen höheren erwarteten score erhält.

$$\begin{aligned} S(q, q) &= q(\omega)^2 + q(\bar{\omega})^2 = \left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2 = \frac{10}{16} \\ &< \frac{12}{16} = \frac{3}{4} + 0 = p(\omega)q(\omega) + p(\bar{\omega})q(\bar{\omega}) = S(p, q) \end{aligned}$$

Der linear score ist folglich nicht proper. Dies liegt daran, dass man bei vorliegen der wahren Verteilung q ganz leicht eine geeignete Verteilung p konstruieren kann, welche den Ereignissen i , denen q eine hohe Wahrscheinlichkeit zuordnet, eine entsprechend noch höhere Wahrscheinlichkeit zuordnet, und, dem entgegengesetzt, den Ereignissen i , denen q eine niedrige Wahrscheinlichkeit zuordnet, eine entsprechend noch geringere Wahrscheinlichkeit zuweist. Durch diese extreme Gewichtung der mit hoher Wahrscheinlichkeit eintretenden Ereignisse erhält die konstruierte Verteilung unter dem linear score eine höhere erwartete Bewertung (Gneiting & Raftery, 2007).

3.1.4. Sonderfall Ja-Nein-Vorhersage

Eine Besonderheit unter den diskreten Ereignisräumen stellt der Fall dar, in dem nur über das Eintreten beziehungsweise das Ausbleiben eines Events unterschieden wird. Im Gegensatz zu den vorangegangenen Betrachtungen liegen nur zwei mögliche, sich gegenseitig ausschließende Ereignisse vor, welche zur Vereinfachung im weiteren Verlauf durch 1, Event tritt ein, und 0, Event tritt nicht ein, dargestellt werden. Eine Wahrscheinlichkeitsvorhersage auf dem Ereignisraum $\Omega = \{0, 1\}$ reduziert sich auf die Angabe einer Wahrscheinlichkeit $p \in [0, 1]$ für das Eintreten des Events, die Restwahrscheinlichkeit $1 - p$ entspricht der Vorhersagewahrscheinlichkeit für das Ausbleiben des Events.

Eine scoring rule $S : \mathcal{P} \times \Omega \rightarrow \overline{\mathbb{R}}$ lässt sich für diesen speziellen Fall auf zwei Funktionen $S(\cdot, 1) : [0, 1] \rightarrow \overline{\mathbb{R}}$ und $S(\cdot, 0) : [0, 1] \rightarrow \overline{\mathbb{R}}$ reduzieren, wobei $S(p, 1)$ den score bei tatsächlichem Eintreten des Ereignisses wiedergibt und entsprechend $S(p, 0)$ beim Ausbleiben des Ereignisses. Es lassen sich wie schon im Abschnitt zur Savage Darstellung die Begriffe regular und Subgradient definieren und es existiert eine abgewandelte Version von Theorem 3.1.1 (Gneiting & Raftery, 2007).

Definition 3.1.4 *Eine scoring rule S für eine binäre Variable ist regular, falls $S(\cdot, 1)$ und $S(\cdot, 0)$ reellwertig sind, mit Ausnahme, dass $S(0, 1) = -\infty$ oder $S(1, 0) = -\infty$ erlaubt sind.*

Sei $G : [0, 1] \rightarrow \mathbb{R}$ eine konvexe Funktion.

Definition 3.1.5 $G'(p)$ heißt Subgradient von G in Punkt $p \in [0, 1]$, falls

$$G(q) \geq G(p) + G'(p)(q - p), \quad \forall q \in [0, 1]. \quad (3.1.17)$$

Theorem 3.1.2 *Jede regular (stictly) proper scoring rule ist von der Form*

$$S(p, 1) = G(p) + (1 - p)G'(p), \quad S(p, 0) = G(p) - pG'(p), \quad (3.1.18)$$

wobei $G : [0, 1] \rightarrow \mathbb{R}$ eine (streng) konvexe Funktion und $G'(p)$ ein Subgradient von G in $p \in [0, 1]$ darstellt.

Beweis: Die in diesem Theorem beschriebene Darstellung für eine proper scoring rule ist eine spezielle Version der Savage Darstellung. Zur Vereinfachung der folgenden Erklärung bezeichnen alle mit einer Tilde '˜' versehenen Funktionen und Parameter die entsprechenden Elemente aus dem vorangegangenen Theorem 3.1.2. In diesem wurde ausgenutzt, dass im Falle einer Ja-Nein-Vorhersage nur die Wahrscheinlichkeit \tilde{p} für das Eintreten des Ereignisses angegeben werden muss, da die Gegenwahrscheinlichkeit $(1 - \tilde{p})$ durch die Wahl von \tilde{p} schon festgelegt ist. Die auf die zweidimensionale Vorhersage $p = (1 - \tilde{p}, \tilde{p})^T$ angewendete Funktion aus Theorem 3.1.1 $G : \mathcal{P}_2 \rightarrow \mathbb{R}$, $(1 - \tilde{p}, \tilde{p}) \mapsto G(1 - \tilde{p}, \tilde{p})$, ist eine ausschließlich von \tilde{p} abhängige Funktion. Wir können uns deshalb eine neue Funktion $\tilde{G} : \mathcal{P}_1 \rightarrow \mathbb{R}$, $\tilde{p} \mapsto G(1 - \tilde{p}, \tilde{p})$, definieren. Die Ableitung von \tilde{G} ergibt sich entsprechend aus den partiellen Ableitungen G_1 bzw. G_2 der Funktion G :

$$\tilde{G}'(\tilde{p}) = -G_1(1 - \tilde{p}, \tilde{p}) + G_2(1 - \tilde{p}, \tilde{p})$$

Der Nachweis, dass es sich bei \tilde{G}' um einen Subgradienten der Funktion G handelt, lässt sich durch Zurückführen auf die Subgradienten-Eigenschaft von $G' = (G_1, G_2)^T$ für G erbringen. Seien $\tilde{p}, \tilde{q} \in [0, 1]$ beliebig, dann folgt mit $p = (1 - \tilde{p}, \tilde{p})^T$ und $q = (1 - \tilde{q}, \tilde{q})^T$

$$\begin{aligned}
 & \tilde{G}(\tilde{p}) + \tilde{G}'(\tilde{p})(\tilde{q} - \tilde{p}) \\
 &= G(1 - \tilde{p}, \tilde{p}) + (-G_1(1 - \tilde{p}, \tilde{p}) + G_2(1 - \tilde{p}, \tilde{p}))(\tilde{q} - \tilde{p}) \\
 &= G(1 - \tilde{p}, \tilde{p}) + \left\langle \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}, \begin{pmatrix} 1 - \tilde{q} \\ \tilde{q} \end{pmatrix} - \begin{pmatrix} 1 - \tilde{p} \\ \tilde{p} \end{pmatrix} \right\rangle \\
 &= G(p) + \langle G'(p), (q - p) \rangle \\
 &\leq G(q) \\
 &= \tilde{G}(1 - \tilde{q}, \tilde{q}).
 \end{aligned}$$

Wendet man nun die Savage Darstellung aus Theorem 3.1.1 auf die Vorhersage $p = (1 - \tilde{p}, \tilde{p})^T$ an, so erhält man durch leichte Umformung die Darstellung aus Theorem 3.1.2. Man beachte die unterschiedliche Indizierung der beiden Darstellungen.

$$\begin{aligned}
 S(p, 1) &= G(p) - \langle G'(p), p \rangle + G'_1(p) \\
 &= G(1 - \tilde{p}, \tilde{p}) - (G_1(1 - \tilde{p}, \tilde{p})(1 - \tilde{p}) + G_2(1 - \tilde{p}, \tilde{p})\tilde{p}) + G_1(1 - \tilde{p}, \tilde{p}) \\
 &= \tilde{G}(\tilde{p}) - \tilde{p}(-G_1(1 - \tilde{p}, \tilde{p}) + G_2(1 - \tilde{p}, \tilde{p})) \\
 &= \tilde{G}(\tilde{p}) - \tilde{p}\tilde{G}'(\tilde{p}) \\
 &= \tilde{S}(\tilde{p}, 0) \\
 S(p, 2) &= G(p) - \langle G'(p), p \rangle + G'_2(p) \\
 &= G(1 - \tilde{p}, \tilde{p}) - (G_1(1 - \tilde{p}, \tilde{p})(1 - \tilde{p}) + G_2(1 - \tilde{p}, \tilde{p})\tilde{p}) + G_2(1 - \tilde{p}, \tilde{p}) \\
 &= \tilde{G}(\tilde{p}) + (1 - \tilde{p})(-G_1(1 - \tilde{p}, \tilde{p}) + G_2(1 - \tilde{p}, \tilde{p})) \\
 &= \tilde{G}(\tilde{p}) + (1 - \tilde{p})\tilde{G}'(\tilde{p}) \\
 &= \tilde{S}(\tilde{p}, 1)
 \end{aligned}$$

□

Als ein Beispiel für eine proper scoring rule für eine Ja-Nein-Vorhersage sei der *negative Brier score* genannt. Er besitzt die konvexe entropy function $G(p) = -p(1 - p)$ und lässt sich mit Hilfe von Theorem 3.1.2 wie folgt herleiten:

$$\begin{aligned}
 S(p, 1) &= G(p) + (1 - p)G'(p) & S(p, 0) &= G(p) - pG'(p) \\
 &= -p(1 - p) + (1 - p)(2p - 1) & &= -p(1 - p) - p(2p - 1) \\
 &= -p + p^2 + 2p - 1 - 2p^2 + p & &= -p + p^2 - 2p^2 + p \\
 &= -(1 - p)^2 & &= -p^2
 \end{aligned}$$

3.2. Proper scoring rules auf stetigen Ereignisräumen

Während wir uns in den vorangegangenen Überlegungen aus Abschnitt 3.1 auf Ereignisräume mit einer endlichen Anzahl an Ereignissen beschränkt haben, wollen wir nun dazu übergehen, stetige Ereignisräume zu betrachten. Man nehme einmal an, man möchte eine Wahrscheinlichkeitsvorhersage für die allgemeine Preisentwicklung in den nächsten 12 Monaten erstellen. Die jährliche Preisentwicklung wird durch die Inflationsrate angegeben und diese kann aus theoretischer Sicht jeden Wert innerhalb der reellen Zahlen annehmen. Unser Ereignisraum $\Omega = \mathbb{R} = (-\infty, \infty)$ besitzt folglich unendlich viele Elemente, sodass wir nicht mehr jedem Ereignis eine entsprechende Wahrscheinlichkeit zuordnen können, sondern eine Wahrscheinlichkeitsverteilung auf den reellen Zahlen angeben müssen. Zur Bewertung dieser Wahrscheinlichkeitsvorhersagen benötigen wir eine Verallgemeinerung der in Abschnitt 3.1 erläuterten Savage Darstellung (3.1.6) für proper scoring rules. Hendrickson & Buehler (1971) entwickelten einen solchen Ansatz und Gneiting & Raftery (2007) konstruierten eine zu Savage (1971) äquivalente Darstellung für proper scoring rules über konvexe Funktionale.

3.2.1. Darstellung von proper scoring rules über konvexe Funktionale

Im weiteren Verlauf gehen wir bei dem Ereignisraum Ω von einer beliebigen Menge mit möglicherweise unbeschränkter Anzahl an Elementen aus und nehmen außerdem an, dass \mathcal{A} eine σ -Algebra auf Ω sei. Weiterhin sei \mathcal{P} eine konvexe Klasse von Wahrscheinlichkeitsmaßen auf dem Raum (Ω, \mathcal{A}) . Ein Funktional ist eine Funktion $G : V \rightarrow \mathbb{R}$ von einem Vektorraum V in die reellen Zahlen. Dieser Vektorraum entspricht in unserem Fall der konvexen Klasse \mathcal{P} der Wahrscheinlichkeitsmaße. Die Definition eines konvexen Funktionals ist äquivalent zur Definition 3.1.1 für konvexe Funktionen auf der Klasse \mathcal{P}_m (Gneiting & Raftery, 2007).

Definition 3.2.1 Ein Funktional $G : \mathcal{P} \rightarrow \mathbb{R}$ ist konvex, falls gilt

$$G((1 - \lambda)P + \lambda Q) \leq (1 - \lambda)G(P) + \lambda G(Q), \quad \forall \lambda \in (0, 1), \quad P, Q \in \mathcal{P}. \quad (3.2.1)$$

G ist streng konvex, falls in Ungleichung (3.2.1) genau dann Gleichheit eintritt, falls $P = Q$ gilt.

Definition 3.2.2 Eine scoring rule $S : \mathcal{P} \times \Omega \rightarrow \overline{\mathbb{R}}$ ist regular bezüglich der Klasse \mathcal{P} , falls $S(P, Q)$ reellwertig ist für alle $P, Q \in \mathcal{P}$, mit der Ausnahme, dass $S(P, Q) = -\infty$ erlaubt ist.

Die folgende Definition für einen Subgradienten stammt aus Hendrickson & Buehler (1971) und ist eine Verallgemeinerung des von Rockafellar (1970) geprägten Begriffs Subgradient aus Definition 3.1.2.

Definition 3.2.3 Eine Funktion $G^*(P, \cdot) : \Omega \rightarrow \overline{\mathbb{R}}$ ist ein Subgradient von G in $P \in \mathcal{P}$, falls sie bezüglich P integrierbar ist, quasiintegrierbar für alle $Q \in \mathcal{P}$ und

$$G(Q) \geq G(P) + \int G^*(P, \omega) d(Q - P)(\omega), \quad \forall Q \in \mathcal{P} \quad (3.2.2)$$

gilt.

Mit Hilfe dieser Definitionen lässt sich das Theorem 2.2 aus Gneiting & Raftery (2007) zur Darstellung von proper scoring rules über konvexe Funktionale herleiten.

Theorem 3.2.1 Eine regular scoring rule $S : \mathcal{P} \times \Omega \rightarrow \overline{\mathbb{R}}$ ist genau dann (strictly) proper bezüglich der Klasse \mathcal{P} , falls für alle $P \in \mathcal{P}$ und $\omega \in \Omega$ gilt

$$S(P, \omega) = G(P) - \int G^*(P, \omega) dP(\omega) + G^*(P, \omega), \quad (3.2.3)$$

wobei $G : \mathcal{P} \rightarrow \mathbb{R}$, $P \mapsto S(P, P)$, ein (streng) konvexes, reellwertiges Funktional auf \mathcal{P} ist und $G^*(P, \cdot) : \Omega \rightarrow \overline{\mathbb{R}}$ ein Subgradient von G in $P \in \mathcal{P}$.

Für den Beweis des Theorems benötigen wir das Theorem 2.1 aus Hendrickson & Buehler (1971).

Theorem 3.2.2 Sei $G : \mathcal{P} \rightarrow \mathbb{R}$ ein Funktional auf der konvexen Menge \mathcal{P} . Falls G in jedem Punkt P aus der konvexen Menge \mathcal{P} einen Subgradienten G^* besitzt, dann ist G konvex in \mathcal{P} .

Beweis: Seien $P, Q \in \mathcal{P}$ beliebig und sei G_1^* der Subgradient von G in $P_1 = (1 - \lambda)P + \lambda Q$. Dann gilt $G(P) \geq G(P_1) + \int G^*(P_1, \omega) d(P - P_1)(\omega)$ und $G(Q) \geq G(P_1) + \int G^*(P_1, \omega) d(Q - P_1)(\omega)$. Hieraus folgt die Konvexität:

$$\begin{aligned} & (1 - \lambda)G(P) + \lambda G(Q) \\ & \geq (1 - \lambda) \left(G(P_1) + \int G^*(P_1, \omega) d(P - P_1)(\omega) \right) \\ & \quad + \lambda \left(G(P_1) + \int G^*(P_1, \omega) d(Q - P_1)(\omega) \right) \\ & = (1 - \lambda) \left(G(P_1) + \int G^*(P_1, \omega) dP(\omega) - G(P_1) \right) \\ & \quad + \lambda \left(G(P_1) + \int G^*(P_1, \omega) dQ(\omega) - G(P_1) \right) \\ & = \int G^*(P_1, \omega) d((1 - \lambda)P + \lambda Q)(\omega) \\ & = G((1 - \lambda)P + \lambda Q) \end{aligned}$$

□

Beweis von Theorem 3.2.1:

„ \Leftarrow “:

Seien $P, Q \in \mathcal{P}$ und sei S eine scoring rule entsprechend der Darstellung aus Formel (3.2.3):

$$S(P, \omega) = G(P) - \int G^*(P, \omega) dP(\omega) + G^*(P, \omega)$$

Aufgrund der Ungleichung (3.2.2) für die konvexe Funktion G folgt:

$$\begin{aligned}
 & G(Q) \geq G(P) + \int G^*(P, \omega) \, d(Q - P)(\omega) \\
 \Leftrightarrow & \quad G(Q) - \int G^*(P, \omega) \, dQ \geq G(P) - \int G^*(P, \omega) \, dP \\
 \Rightarrow & \quad S(P, \omega) \leq G(Q) - \int G^*(P, \omega) \, dQ(\omega) + G^*(P, \omega) \\
 \Rightarrow & \quad S(P, Q) = \int S(P, \omega) \, dQ(\omega) \\
 & \leq \int \left(G(Q) - \int G^*(P, \omega) \, dQ(\omega) + G^*(P, \omega) \right) dQ(\omega) \\
 & = G(Q) - \int G^*(P, \omega) \, dQ(\omega) + \int G^*(P, \omega) \, dQ(\omega) \\
 & = S(Q, Q)
 \end{aligned}$$

Somit wäre bewiesen, dass S proper ist.

„ \Rightarrow “:

Sei S eine regular proper scoring rule, seien $P, Q \in \mathcal{P}$ und sei $G^*(P, \omega) := S(P, \omega)$. Dann lässt sich mit Hilfe der Eigenschaft proper von S die Subgradienten-Ungleichung (3.2.2) nachweisen:

$$\begin{aligned}
 & G(P) + \int G^*(P, \omega) \, d(Q - P)(\omega) \\
 = & G(P) + \int S(P, \omega) \, dQ(\omega) - \int S(P, \omega) \, dP(\omega) \\
 = & G(P) + \int S(P, \omega) \, dQ(\omega) - G(P) \\
 = & \int S(P, \omega) \, dQ(\omega) \\
 \leq & \int S(Q, \omega) \, dQ(\omega) \\
 = & G(Q)
 \end{aligned}$$

Nun lässt sich mit Theorem 3.2.2 folgern, dass G auf \mathcal{P} konvex ist. Da sich weiterhin durch $G(P) := S(P, P)$ und $G^*(P) := S(P, \omega)$ die Darstellung (3.2.3) erzeugen lässt, erfüllen G und G^* die geforderten Eigenschaften.

$$\begin{aligned}
 & G(P) - \int G^*(P, \omega) \, dP(\omega) + G^*(P, \omega) \\
 &= S(P, P) - \int S(P, \omega) \, dP(\omega) + S(P, \omega) \\
 &= S(P, \omega)
 \end{aligned}$$

Die Tatsache, dass eine strictly proper scoring rule entsprechend über ein streng konvexes Funktional dargestellt werden kann, geht zurück auf ein Argument von Hendrickson & Buehler (1971), wonach strenge Ungleichheit in der Ungleichung (2.2.3) der Definition von proper äquivalent dazu ist, dass ein Subgradient von G in P niemals gleichzeitig auch ein Subgradient von G in Q sein kann, für $P, Q \in \mathcal{P}$ und $P \neq Q$. Dies ist jedoch wiederum äquivalent dazu, dass G streng konvex auf \mathcal{P} ist. \square

Aus diesem Beweis wird deutlich, dass, etwas vereinfacht ausgedrückt, eine regular scoring rule S genau dann (strictly) proper bezüglich der Klasse \mathcal{P} ist, falls für alle $P \in \mathcal{P}$ die expected score function $G(P) = S(P, P)$ (streng) konvex und $S(P, \omega)$ ein Subgradient von G in P ist. Der Begriff der *expected score function* $G(P)$ und der Begriff der *Bregman divergence* $d(P, Q)$ überträgt sich entsprechend Formel (3.1.7) beziehungsweise Formel (3.1.8) auf die Klasse \mathcal{P} mit $P, Q \in \mathcal{P}$.

3.2.2. Beispiele für proper scoring rules auf stetigen Ereignisräumen

In diesem Abschnitt wollen wir einige Beispiele für proper scoring rules betrachten. Bevor wir dies tun können, müssen wir jedoch erst auf die Klasse der Wahrscheinlichkeitsmaße, die wir für die Vorhersage als zulässig erachten, eingehen. Sei deshalb μ ein σ -endliches Maß auf dem messbaren Raum (Ω, \mathcal{A}) . Für $\alpha \geq 1$ bezeichne \mathcal{L}_α die Klasse der Wahrscheinlichkeitsmaße auf (Ω, \mathcal{A}) , die absolut stetig bezüglich μ sind und die μ -Dichte p besitzen, sodass

$$\|p\|_\alpha = \left(\int p(\omega)^\alpha \, d\mu(\omega) \right)^{\frac{1}{\alpha}} \quad (3.2.4)$$

endlich ist. Die in Abschnitt 3.1.3 beschriebenen scoring rules *quadratic score*, *pseudospherical score* und *logarithmic score* für proper scoring rules auf diskreten Ereignisräumen existieren in entsprechend abgewandelten Formen auch für den stetigen Fall (siehe z.B. Gneiting & Raftery (2007)).

Der *quadratic score* besitzt die Darstellung

$$\text{QuadS}(p, \omega) = 2p(\omega) - \|p\|_2^2 \quad (3.2.5)$$

und ist strictly proper bezüglich der Klasse \mathcal{L}_2 . Die expected score function ergibt sich als $G(p) = \|p\|_2^2$ und die Bregman divergence lautet

$$d(p, q) = \|p - q\|_2^2. \quad (3.2.6)$$

Die entsprechenden Herleitungen der expected score function, der Nachweis für dessen Konvexität, sowie die Ermittlung der Bregman divergence für dieses und für die nachfolgenden Beispiele sind dem Anhang A.2 zu entnehmen.

Der *pseudospherical score*

$$\text{PseudoS}(p, \omega) = \frac{p(\omega)^{\alpha-1}}{\|p\|_{\alpha}^{\alpha-1}} \quad (3.2.7)$$

ist für $\alpha > 1$ strictly proper bezüglich der Klasse \mathcal{L}_{α} und entspricht dem spherical score im Falle $\alpha = 2$. Für den pseudospherical score lässt sich die expected score function $G(p) = \|p\|_{\alpha}$ ermitteln und die Bregman divergence entspricht

$$d(p, q) = \|q\|_{\alpha} - \int_{\Omega} \frac{p(\omega)^{\alpha-1}}{\|p\|_{\alpha}^{\alpha-1}} q(\omega) \, d\mu(\omega). \quad (3.2.8)$$

Auch der *logarithmic score*

$$\text{LogS}(p, \omega) = \log(p(\omega)) \quad (3.2.9)$$

ist auf dem stetigen Ereignisraum strictly proper, aber er besitzt diese Eigenschaft nur bezüglich der Wahrscheinlichkeitsmaße aus \mathcal{L}_1 (Gneiting & Raftery, 2007). Die expected score function lautet $G(p) = \int_{\Omega} \log(p(\omega)) p(\omega) \, d\mu(\omega)$. Die Bregman divergence, welche bei dem logarithmic score auch Kullback-Leibler divergence genannt wird (Gneiting & Raftery, 2007), lautet

$$d(p, q) = \int_{\Omega} \log\left(\frac{q(\omega)}{p(\omega)}\right) q(\omega) \, d\mu(\omega). \quad (3.2.10)$$

Neben den drei genannten scoring rules existieren natürlich noch viele weitere proper scoring rules. Um einige weitere Möglichkeiten zu zeigen, wie man proper scoring rules konstruieren kann, beschränken wir uns im Folgenden auf den stetigen Ereignisraum der reellen Zahlen, $\Omega = \mathbb{R}$. Auf diesem Raum lässt sich durch jede proper scoring rule für Ja-Nein-Vorhersagen eine proper scoring rule auf dem stetigen Ereignisraum erzeugen. Um dies zu demonstrieren, nehmen wir einmal an, wir hätten eine proper scoring rule $S : [0, 1] \times \{0, 1\} \rightarrow \overline{\mathbb{R}}$, welche bei Vorhersagewahrscheinlichkeit $p \in [0, 1]$ für das Eintreten eines Ereignisses den score $S(p, 1)$ zurückgibt, falls das Ereignis wirklich eingetreten ist, und $S(p, 0)$ sonst. Nun definieren wir das Ereignis als „der wahre Parameter y ist kleiner oder gleich einem Wert $x \in \mathbb{R}$ “. Weiterhin sei $\mathbb{I}\{y \leq x\}$ die Indikatorfunktion, die den Wert 1 zurück gibt, falls $y \leq x$ gilt, und 0 sonst. Außerdem sei F die Verteilungsfunktion der Wahrscheinlichkeitsvorhersage auf dem stetigen Ereignisraum. Dann lässt sich für $p = F(x)$ durch Integration über den Parameter x der score

$$S^*(F, y) := \int_{-\infty}^{\infty} S(F(x), \mathbb{I}\{y \leq x\}) \, dx \quad (3.2.11)$$

herleiten, welcher eine proper scoring rule zur Bewertung von Verteilungsfunktionen auf dem stetigen Ereignisraum darstellt (Matheson & Winkler, 1976; Gerds, 2002).

Äquivalent dazu lassen sich auch durch die Verwendung von *proper quantile scores* weitere proper scoring rules auf stetigen Ereignisräumen konstruieren. Dabei stellt ein quantile score $\text{QS}_\alpha : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ eine Bewertungsfunktion dar, welche bei Vorhersage $q \in \mathbb{R}$ für das α -Quantil einer Verteilung und bei Eintritt des Ereignisses $y \in \mathbb{R}$ den score $\text{QS}_\alpha(q, y)$ zurückgibt. Bezeichne nun wieder F die Verteilungsfunktion der Wahrscheinlichkeitsvorhersage, dann ergibt sich bei Schätzung des Quantils durch $q = F^{-1}(\alpha)$ eine neue proper scoring rule auf dem stetigen Ereignisraum, wenn man über den Parameter α integriert (Matheson & Winkler, 1976; Cervera & Muñoz, 1996):

$$S^\alpha(F, y) := \int_0^1 \text{QS}_\alpha(F^{-1}(\alpha), y) \, d\alpha \quad (3.2.12)$$

Dass sich sowohl durch Darstellung (3.2.11), als auch durch Darstellung (3.2.12) tatsächlich scoring rules entwickeln lassen, welche die Eigenschaft proper erfüllen, wird in einem späteren Teil der Arbeit in den Theoremen 4.2.1 und 4.2.2 gezeigt. Als ein Beispiel für einen score, der mithilfe dieser beiden Ansätze hergeleitet werden kann, sei der *continuous ranked probability score* (CRPS) genannt. Dieser lässt sich sowohl durch Verwendung eines speziellen quantile scores als auch unter Verwendung des Brier scores erzeugen und wird in Abschnitt 3.2.3 vorgestellt (Gneiting & Raftery, 2007).

Schließlich sei als Beispiel für einen score, der nicht proper ist, der linear score

$$\text{LinS}(p, \omega) = p(\omega) \quad (3.2.13)$$

genannt, welchen wir in Abschnitt 3.1.3 bereits für den diskreten Falls untersucht hatten. Wir beschränken uns dazu auf den Ereignisraum der reellen Zahlen und führen den Nachweis, dass der linear score nicht proper ist, entsprechend den Überlegungen aus Abschnitt 3.1.3 durch. Als wahre Verteilung wählen wir die Standardnormalverteilung mit Dichte ϕ und konstruieren uns für die Vorhersage eine Verteilung, die genau den Ereignissen eine überproportional hohe Wahrscheinlichkeit zuordnet, die auch bezüglich der wahren Verteilung sehr wahrscheinlich sind. In Anlehnung an Gneiting & Raftery (2007) wählen wir die Gleichverteilung auf $(-\epsilon, \epsilon)$ mit entsprechender Dichte $u(x) = \frac{1}{2\epsilon} \mathbb{I}\{x \in (-\epsilon, \epsilon)\}$, wobei $\mathbb{I}\{x \in (-\epsilon, \epsilon)\}$ die Indikatorfunktion auf $(-\epsilon, \epsilon)$ darstellt. Für $\epsilon \leq \sqrt{\log 2}$ lässt sich zeigen, dass

$$\begin{aligned} \text{LinS}(u, \phi) &= \int_{\mathbb{R}} \frac{1}{2\epsilon} \mathbb{I}\{x \in (-\epsilon, \epsilon)\} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2}\right) \, dx = \frac{1}{(2\pi)^{\frac{1}{2}}} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} \exp\left(-\frac{x^2}{2}\right) \, dx \\ &> \frac{1}{2\pi^{\frac{1}{2}}} = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-x^2) \, dx = \int_{\mathbb{R}} \phi(x)^2 \, dx = \text{LinS}(\phi, \phi) \end{aligned}$$

gilt, womit bewiesen wäre, dass der linear score nicht proper ist (Gneiting & Raftery, 2007).

3.2.3. Continuous ranked probability score (CRPS)

In diesem Abschnitt wollen wir uns dem *continuous ranked probability score* (CRPS) widmen, welcher erstmals durch Epstein (1969) eingeführt wurde und in der Literatur zur Bewertung von Wahrscheinlichkeitsvorhersagen auf den reellen Zahlen zunehmend an Bedeutung gewonnen hat. Während beispielsweise der logarithmic score die Dichte p der vorhergesagten Verteilung $P \in \mathcal{P}$ nur an dem tatsächlich realisierten Wert $y \in \mathbb{R}$ ausgewertet ($\text{LogS}(p, y) = \log(p(y))$) und deshalb den *local scores* zugeordnet wird (vgl. Bröcker & Smith (2007)), berücksichtigt der CRPS die gesamte Verteilung der Vorhersage. Dies ist insbesondere dann von Bedeutung, wenn man die Wahrscheinlichkeitsvorhersage nicht nur an Hand eines einzelnen realisierten Wertes y bewertet, sondern als Datenbasis viele Realisierungen y_1, \dots, y_n nutzt und den mittleren score als Bewertungsmaßstab ansetzt. Wie wir bei den späteren Analysen sehen werden, wirken sich Ausreißer bei Verwendung des logarithmic scores stärker auf den mittleren score aus als beim CRPS.

Für die weitere Betrachtung sei \mathcal{P} die Klasse der Borel Wahrscheinlichkeitsmaße auf den reellen Zahlen \mathbb{R} und F die Verteilungsfunktion von $P \in \mathcal{P}$. Dann ist der *continuous ranked probability score* (CRPS) definiert als

$$\text{CRPS}(F, y) = - \int_{-\infty}^{\infty} (F(x) - \mathbb{I}\{x \geq y\})^2 dx. \quad (3.2.14)$$

Vereinfacht ausgedrückt könnte man sagen, dass der CRPS die Differenz misst zwischen der vorhergesagten Verteilungsfunktion F und einer Verteilungsfunktion, welche dem Ereignis y die Wahrscheinlichkeit 1 zuordnet (Gneiting & Raftery, 2007; Gneiting & Ranjan, 2011).

Wie schon in Abschnitt 3.2.2 angedeutet wurde, lässt sich der CRPS sowohl mithilfe eines scores für binäre Variablen als auch durch die Verwendung eines quantile scores konstruieren (Gneiting & Ranjan, 2011). Bei ersterem nutzt man den Brier score für binäre Ereignisse

$$\text{BS}(p, \mathbb{I}\{x \geq y\}) = (p - \mathbb{I}\{x \geq y\})^2, \quad (3.2.15)$$

wobei $p = F(x)$ die Wahrscheinlichkeitsvorhersage für das binäre Event $\{Y \leq x\}$ darstellt. Durch Integration über den Parameter x erhält man den CRPS

$$\text{CRPS}(F, y) = - \int_{-\infty}^{\infty} \text{BS}(F(x), \mathbb{I}\{x \geq y\}) dx. \quad (3.2.16)$$

Bei letzterem verwendet man den in Gneiting & Raftery (2007) angegebenen quantile score

$$\text{QS}_{\alpha}(q, y) = 2(\mathbb{I}\{y < q\} - \alpha)(y - q). \quad (3.2.17)$$

Es handelt sich hierbei um eine proper scoring rule und durch Integration über α und Ersetzen von q durch $F^{-1}(\alpha)$ ergibt sich der score

$$\text{CRPS}(F, y) = \int_0^1 \text{QS}_{\alpha}(F^{-1}(\alpha), y) d\alpha. \quad (3.2.18)$$

Bei diesem score handelt es sich um eine äquivalente Darstellung des CRPS. Der entsprechende Nachweis dafür geht auf Laio & Tamea (2007) zurück.

Theorem 3.2.3 *Sei F eine Verteilungsfunktion auf den reellen Zahlen mit endlichem ersten Moment, dann folgt*

$$\int_0^1 QS_\alpha(F^{-1}(\alpha), y) \, d\alpha = - \int_{-\infty}^{\infty} (F(x) - \mathbb{I}\{x \geq y\})^2 \, dx. \quad (3.2.19)$$

Beweis: Aus der Formel für die partielle Integration (siehe z.B. Elstrodt (2011)) lässt sich direkt folgende Gleichung herleiten

$$\int_{-\infty}^x F(u) \, dF(u) = F(x)F(x) - \int_{-\infty}^x F(u) \, dF(u).$$

Diese ist wiederum äquivalent zu

$$F(x)^2 = 2 \int_{-\infty}^x F(u) \, dF(u).$$

Neben diesem Zusammenhang verwenden wir bei den nachfolgenden Umformungsschritten den Satz von Fubini (siehe z.B. Elstrodt (2011)) und wir führen zu Beginn mithilfe des Transformationssatzes (siehe z.B. Elstrodt (2011)) die Substitution $F(u) = \alpha$ durch. Am Ende wird noch die Eigenschaft der Verteilungsfunktion, $F(x) = 1$ für $x \rightarrow \infty$, genutzt.

$$\begin{aligned} & \int_0^1 QS_\alpha(F^{-1}(\alpha), y) \, d\alpha \\ &= 2 \int_0^1 (\mathbb{I}\{y < F^{-1}(\alpha)\} - \alpha) (y - F^{-1}(\alpha)) \, d\alpha \\ &= 2 \int_{-\infty}^{\infty} (\mathbb{I}\{y \leq u\} - F(u)) (y - u) \, dF(u) \\ &= -2 \int_{-\infty}^{\infty} F(u) \mathbb{I}\{y > u\} (y - u) \, dF(u) + 2 \int_{-\infty}^{\infty} (1 - F(u)) \mathbb{I}\{y \leq u\} (y - u) \, dF(u) \\ &= -2 \int_{-\infty}^{\infty} F(u) \mathbb{I}\{y > u\} \int_{-\infty}^{\infty} \mathbb{I}\{u \leq x < y\} \, dx \, dF(u) \\ &\quad - 2 \int_{-\infty}^{\infty} (1 - F(u)) \mathbb{I}\{y \leq u\} \int_{-\infty}^{\infty} \mathbb{I}\{y \leq x \leq u\} \, dx \, dF(u) \\ &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 2F(u) \mathbb{I}\{u \leq x\} \, dF(u) \mathbb{I}\{x < y\} \, dx \\ &\quad - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 2(1 - F(u)) \mathbb{I}\{x \leq u\} \, dF(u) \mathbb{I}\{y \leq x\} \, dx \\ &= - \int_{-\infty}^{\infty} F(x)^2 \mathbb{I}\{x < y\} \, dx - \int_{-\infty}^{\infty} \left(2 \int_{-\infty}^{\infty} \mathbb{I}\{x \leq u\} \, dF(u) \right. \\ &\quad \left. - \left(2 \int_{-\infty}^{\infty} F(u) \, dF(u) - 2 \int_{-\infty}^{\infty} F(u) \mathbb{I}\{u \leq x\} \, dF(u) \right) \right) \mathbb{I}\{y \leq x\} \, dx \end{aligned}$$

$$\begin{aligned}
&= - \int_{-\infty}^{\infty} F(x)^2 \mathbb{I}\{x < y\} \, dx \\
&\quad - \int_{-\infty}^{\infty} \left(2(1 - F(x)) - F(\infty)^2 + F(x)^2 \right) \mathbb{I}\{y \leq x\} \, dx \\
&= - \int_{-\infty}^{\infty} F(x)^2 \mathbb{I}\{x < y\} \, dx - \int_{-\infty}^{\infty} \left(2 - 2F(x) - 1 + F(x)^2 \right) \mathbb{I}\{y \leq x\} \, dx \\
&= - \int_{-\infty}^{\infty} (F(x) - \mathbb{I}\{x \geq y\})^2 \, dx
\end{aligned}$$

□

Neben den zwei beschriebenen Darstellungen (3.2.14) und (3.2.18) für den CRPS gibt es noch eine weitere Darstellung, die Erwartungswert Darstellung

$$\text{CRPS}(F, y) = \frac{1}{2} \mathbb{E}_F |X - X'| - \mathbb{E}_F |X - y|, \quad (3.2.20)$$

wobei X und X' zwei unabhängige, identisch verteilte Zufallsvariablen mit Verteilungsfunktion F darstellen (Gneiting & Raftery, 2007). Es ist auf den ersten Blick nicht offensichtlich, dass es sich bei der Darstellung (3.2.20) ebenfalls um den CRPS handelt. Um die Äquivalenz zeigen zu können, benötigen wir Lemma 2.1 aus Baringhaus & Franz (2004).

Theorem 3.2.4 *Seien X und X' zwei unabhängige, reellwertige Zufallsvariablen mit endlichem Erwartungswert. Weiterhin sei F die Verteilungsfunktion von X und G die Verteilungsfunktion von X' . Dann gilt*

$$\mathbb{E} |X - X'| = \int_{-\infty}^{\infty} F(x)(1 - G(x)) \, dx + \int_{-\infty}^{\infty} G(x)(1 - F(x)) \, dx. \quad (3.2.21)$$

Beweis: Sei $P_{XX'}$ die gemeinsame Verteilung von X und X' . Dann folgt aufgrund der Unabhängigkeit von X und X' und unter Verwendung des Satzes von Fubini (siehe z.B. Elstrodt (2011)) die Behauptung.

$$\begin{aligned}
&\mathbb{E} |X - X'| \\
&= \int_{-\infty \times -\infty}^{\infty \times \infty} |X - X'| \, dP_{XX'} \\
&= \int_{-\infty \times -\infty}^{\infty \times \infty} \int_{-\infty}^{\infty} \mathbb{I}\{X \leq x < X'\} + \mathbb{I}\{X' \leq x < X\} \, dx \, dP_{XX'} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{I}\{X \leq x\} \mathbb{I}\{x < X'\} + \mathbb{I}\{X' \leq x\} \mathbb{I}\{x < X\} \, dF(X) \, dG(X') \, dx
\end{aligned}$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \mathbb{I}\{X \leq x\} dF(X) \int_{-\infty}^{\infty} \mathbb{I}\{x < X'\} dG(X') \right) dx \\
 &\quad + \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \mathbb{I}\{X' \leq x\} dG(X') \int_{-\infty}^{\infty} \mathbb{I}\{x < X\} dF(X) \right) dx \\
 &= \int_{-\infty}^{\infty} F(x)(1 - G(x)) dx + \int_{-\infty}^{\infty} G(x)(1 - F(x)) dx.
 \end{aligned}$$

□

Mithilfe dieses Theorems lässt sich das folgende Theorem und somit die Äquivalenz der Darstellungsformen (3.2.14) und (3.2.20) beweisen.

Theorem 3.2.5 *Sei F eine Verteilungsfunktion auf den reellen Zahlen mit endlichem ersten Moment. Außerdem seien X, X' unabhängig und identisch verteilt mit Verteilungsfunktion F und $y \in \mathbb{R}$ sei eine Konstante. Dann folgt*

$$\frac{1}{2} \mathbb{E}_F |X - X'| - \mathbb{E}_F |X - y| = - \int_{-\infty}^{\infty} (F(x) - \mathbb{I}\{x \geq y\})^2 dx. \quad (3.2.22)$$

Beweis: Aus Theorem 3.2.4 folgt für X, X' unabhängig und identisch verteilt mit Verteilungsfunktion F und y konstant mit entsprechender Verteilungsfunktion $\mathbb{I}\{y \leq x\}$ die Behauptung:

$$\begin{aligned}
 &\frac{1}{2} \mathbb{E} |X - X'| - \mathbb{E} |X - y| \\
 &= \frac{1}{2} \left(\int_{-\infty}^{\infty} F(x)(1 - F(x)) dx + \int_{-\infty}^{\infty} F(x)(1 - F(x)) dx \right) \\
 &\quad - \left(\int_{-\infty}^{\infty} F(x)(1 - \mathbb{I}\{y \leq x\}) dx + \int_{-\infty}^{\infty} \mathbb{I}\{y \leq x\}(1 - F(x)) dx \right) \\
 &= \int_{-\infty}^{\infty} F(x) - F(x)^2 dx \\
 &\quad - \int_{-\infty}^{\infty} F(x) - 2F(x)\mathbb{I}\{y \leq x\} + \mathbb{I}\{y \leq x\} dx \\
 &= - \int_{-\infty}^{\infty} F(x)^2 - 2F(x)\mathbb{I}\{y \leq x\} + \mathbb{I}\{y \leq x\}^2 dx \\
 &= - \int_{-\infty}^{\infty} (F(x) - \mathbb{I}\{x \geq y\})^2 dx \\
 &= \text{CRPS}(F, y)
 \end{aligned}$$

□

Aus der Erwartungswert Darstellung des CRPS wird deutlich, dass der CRPS für jede Verteilungsfunktion mit nicht endlichem ersten Moment den Wert minus unendlich annimmt. In der Klasse aller Verteilungsfunktionen mit endlichem ersten Moment ist der

CRPS jedoch endlich und besitzt die Eigenschaft, eine strictly proper scoring rule zu sein. Nachfolgende Berechnung beweist dies unter Verwendung des Satzes von Fubini (siehe z.B. Elstrodt (2011)), wobei das Kleinerzeichen in der Umformung nur unter der Voraussetzung $F \neq G$ korrekt ist.

$$\begin{aligned}
 \text{CRPS}(F, G) &= \int_{-\infty}^{\infty} \left(- \int_{-\infty}^{\infty} (F(x) - \mathbb{I}\{x \geq y\})^2 dx \right) dG(y) \\
 &= - \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} F(x)^2 - 2F(x) \mathbb{I}\{x \geq y\} + \mathbb{I}\{x \geq y\} dG(y) \right) dx \\
 &= - \int_{-\infty}^{\infty} F(x)^2 - 2F(x)G(x) + G(x) dx \\
 &= - \int_{-\infty}^{\infty} (F(x) - G(x))^2 + G(x) - G(x)^2 dx \\
 &< - \int_{-\infty}^{\infty} G(x)(1 - G(x)) dx \\
 &= \text{CRPS}(G, G)
 \end{aligned}$$

Mit Hilfe dieser Umformungen wurde auch gleichzeitig die expected score function

$$\text{CRPS}(G, G) = - \int_{-\infty}^{\infty} G(x)(1 - G(x)) dx \tag{3.2.23}$$

hergeleitet und es lässt sich durch einfache Rechnung die Bregman divergence

$$\begin{aligned}
 d(F, G) &= \text{CRPS}(G, G) - \text{CRPS}(F, G) \\
 &= - \int_{-\infty}^{\infty} G(x)(1 - G(x)) dx \\
 &\quad - \left(- \int_{-\infty}^{\infty} (F(x) - G(x))^2 + G(x)(1 - G(x)) dx \right) \\
 &= \int_{-\infty}^{\infty} (F(x) - G(x))^2 dx
 \end{aligned}$$

bestimmen.

3.2.4. Energy score: Verallgemeinerung des CRPS

Bei den bisherigen Überlegungen zum CRPS haben wir uns bei dem Ereignisraum Ω immer auf den eindimensionalen Fall $\Omega = \mathbb{R}$ beschränkt. Wir wollen nun eine Verallgemeinerung des CRPS betrachten, welche die Bewertung von Wahrscheinlichkeitsvorhersagen auf dem \mathbb{R}^n ermöglicht. Dazu sei \mathcal{P} die Klasse der Borel Wahrscheinlichkeitsmaße auf dem \mathbb{R}^n und \mathcal{P}_β definiere für $\beta \in (0, 2)$ die Klasse der Wahrscheinlichkeitsmaße $P \in \mathcal{P}$, für die $\mathbb{E}_P \|X\|^\beta$ endlich ist. Hierbei bezeichne $\|\cdot\|$ die euklidische Norm.

Seien X und X' zwei unabhängige, identisch verteilte Zufallsvektoren aus dem \mathbb{R}^n , dann ist der *energy score* definiert als

$$\text{ES}(P, y) = \frac{1}{2} \mathbb{E}_P \|X - X'\|^\beta - \mathbb{E}_P \|X - y\|^\beta, \quad (3.2.24)$$

wobei P die Wahrscheinlichkeitsvorhersage im \mathbb{R}^n darstellt und $y \in \mathbb{R}^n$ das tatsächlich realisierte Ereignis (Gneiting & Raftery, 2007). Für den eindimensionalen Fall $m = 1$ und bei der Wahl von $\beta = 1$ entspricht die Darstellung (3.2.24) des energy scores der Erwartungswert Darstellung des CRPS. Somit kann man beim energy score von einer Verallgemeinerung des CRPS sprechen. Der energy score ist bezüglich der Klasse \mathcal{P}_β strictly proper. Der entsprechende Nachweis erfolgt über Theorem 1 aus Székely & Rizzo (2005).

4. Gewichtete scoring rules

Bei Dichtevorhersagen ist man oft an bestimmten Abschnitten der Dichte besonders interessiert. So kann es sein, dass man beispielsweise im Risikomanagement insbesondere Extremfälle, die sich an den Rändern der Dichte widerspiegeln, analysieren möchte oder dass man bei der strategischen Planung eher am Schwerpunkt, also am Zentrum der Dichte interessiert ist. Es werden folglich Vorhersagen benötigt, die speziell in diesen bestimmten Abschnitten gute Ergebnisse liefern. Um die Qualität solcher Vorhersagen zu bewerten liegt die Idee nahe, dass man die proper scoring rules durch nichtnegative Gewichtungsfunktionen erweitert. Zunächst werden wir in diesem Kapitel zeigen, dass eine einfache Konstruktion von gewichteten proper scoring rules durch Multiplikation einer proper scoring rule mit einer nichtnegativen Gewichtungsfunktion nicht möglich ist. Anschließend werden wir zwei von Gneiting & Ranjan (2011) entwickelte gewichtete Versionen des continuous ranked probability scores vorstellen, welche die Eigenschaft proper der ursprünglichen Bewertungsmethode beibehalten.

4.1. Konstruktion von gewichteten scoring rules

Einen intuitiven Ansatz, um gewichtete scoring rules zu konstruieren, stellt die Kombination einer scoring rule und einer Gewichtungsfunktion durch einfaches Multiplizieren da. Liegt uns eine proper scoring rule $S_0(f, y)$ vor, so ließe sich diese mit Hilfe einer geeigneten, nichtnegativen Gewichtungsfunktion $w(y)$ in die neue scoring rule

$$S(f, y) = w(y) S_0(f, y) \tag{4.1.1}$$

überführen. Damit die neue scoring rule aus (4.1.1) vernünftig angewendet werden kann, muss sichergestellt werden, dass diese die Eigenschaft proper der ursprünglichen scoring rule S_0 beibehält. Das Theorem 2.1 aus Gneiting & Ranjan (2011), welches wir im Anschluss betrachten wollen, zeigt jedoch, dass das Produkt aus einer proper scoring rule $S_0(f, y)$ und einer nichtnegativen Gewichtungsfunktion $w(y)$ im Allgemeinen nur dann wieder proper ist, falls es sich bei $w(y)$ um eine konstante Gewichtungsfunktion handelt. Hieraus lässt sich schließen, dass die Gewichtung einer proper scoring rule durch einfache Multiplikation mit einer Gewichtungsfunktion nicht sinnvoll ist.

Theorem 4.1.1 *Sei f die Dichte einer Zufallsvariable Y . Weiterhin sei S_0 eine proper scoring rule und w sei eine nichtnegative Gewichtungsfunktion, sodass*

$$0 < \int w(y) f(y) dy < \infty \tag{4.1.2}$$

gilt.

Dann wird der erwartete Wert der gewichteten scoring rule

$$S(g, Y) = w(Y) S_0(g, Y) \tag{4.1.3}$$

maximiert durch die Dichtevorhersage

$$g(y) = \frac{w(y) f(y)}{\int w(x) f(x) dx}. \tag{4.1.4}$$

Beweis: Sei h eine beliebige Dichtevorhersage. Dann folgt aufgrund der Eigenschaft proper von S_0 :

$$\begin{aligned} \mathbb{E}_f S(g, Y) &= \int w(y) f(y) S_0(g, y) dy \\ &= \int w(x) f(x) dx \int g(y) S_0(g, y) dy \\ &\geq \int w(x) f(x) dx \int g(y) S_0(h, y) dy \\ &= \int w(y) f(y) S_0(h, y) dy \\ &= \mathbb{E}_f S(h, Y). \end{aligned}$$

□

4.2. Schwellenwert und Quantil gewichtete Versionen des continuous ranked probability scores

Bisher ist kein allgemein gültiger Ansatz bekannt, um proper scoring rules geeignet zu gewichten. Eine individuelle Betrachtung einer jeden scoring rule ist deshalb nötig. Gneiting & Ranjan (2011) ist es gelungen, gewichtete Versionen des CRPS zu entwickeln, welche die Eigenschaft proper besitzen. Sie unterscheiden dabei zwischen der Schwellenwert gewichteten Version des CRPS, welche sich von der Brier Darstellung (3.2.16) des CRPS ableitet, und der Quantil gewichteten Version des CRPS, welche sich aus der Quantil Darstellung (3.2.18) des CRPS ableitet. Die Schwellenwert gewichtete Version des CRPS definieren Gneiting & Ranjan (2011) als

$$S(F, y) = - \int_{-\infty}^{\infty} \text{BS}(F(x), \mathbb{I}\{x \geq y\}) u(x) dx, \quad (4.2.1)$$

wobei u eine nichtnegative Gewichtungsfunktion auf den reellen Zahlen darstellt. Daran angelehnt definiert sich die Quantil gewichtete Version des CRPS als

$$S(F, y) = \int_0^1 \text{QS}_\alpha(F^{-1}(\alpha), y) v(\alpha) d\alpha, \quad (4.2.2)$$

wobei v einer nichtnegativen Gewichtungsfunktion auf dem Einheitsintervall $[0, 1]$ entspricht.

Die beiden gewichteten Versionen des CRPS behalten die Eigenschaft proper des CRPS bei. Dies lässt sich dadurch begründen, dass sowohl der negative Brier score, als auch der quantile score proper sind und eine Gewichtung dieser scores anhand eines beliebigen, nichtnegativen Maßes die Eigenschaft proper nicht beeinträchtigt (Gneiting & Ranjan, 2011). Für strikt positive Gewichtungsfunktionen sind die beiden gewichteten Versionen des CRPS sogar strictly proper. Diese Behauptungen wollen wir nun anhand der nachfolgenden Theoreme beweisen.

Theorem 4.2.1 *Sei S_b eine proper scoring rule für Wahrscheinlichkeitsvorhersagen auf dem binären Ereignisraum, u sei eine nichtnegative, integrierbare Funktion und F sei eine Verteilungsfunktion auf den reellen Zahlen. Dann ist auch die scoring rule*

$$S(F, y) = \int_{-\infty}^{\infty} S_b(F(x), \mathbb{I}\{x \geq y\}) u(x) dx \quad (4.2.3)$$

proper.

Falls es sich bei S_b sogar um eine strictly proper scoring rule handelt und die Gewichtungsfunktion u streng positiv ist, dann ist auch die scoring rules aus Darstellung (4.2.3) strictly proper.

Beweis: Sei G eine weitere beliebige Verteilung auf den reellen Zahlen mit $G \neq F$. Dann folgt mit dem Satz von Fubini (siehe z.B. Elstrodt (2011)) und der Eigenschaft proper

von S_b die erste Aussage des Theorems.

$$\begin{aligned}
 S(F, G) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} S_b(F(x), \mathbb{I}\{x \geq y\}) u(x) dx \right) dG(y) \\
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} S_b(F(x), \mathbb{I}\{x \geq y\}) dG(y) \right) u(x) dx \\
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^x S_b(F(x), 1) dG(y) + \int_x^{\infty} S_b(F(x), 0) dG(y) \right) u(x) dx \\
 &= \int_{-\infty}^{\infty} (G(x) S_b(F(x), 1) + (1 - G(x)) S_b(F(x), 0)) u(x) dx \\
 &\leq \int_{-\infty}^{\infty} (G(x) S_b(G(x), 1) + (1 - G(x)) S_b(G(x), 0)) u(x) dx \\
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} S_b(G(x), \mathbb{I}\{x \geq y\}) dG(y) \right) u(x) dx \\
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} S_b(G(x), \mathbb{I}\{x \geq y\}) u(x) dx \right) dG(y) \\
 &= S(G, G)
 \end{aligned}$$

Nimmt man zusätzlich an, dass die scoring rule S_b strictly proper und die Gewichtungsfunktion u streng positiv ist, dann ergibt sich in der Ungleichung ein strikt kleiner, da nun

$$\begin{aligned}
 &G(x) S_b(F(x), 1) + (1 - G(x)) S_b(F(x), 0) \\
 &< G(x) S_b(G(x), 1) + (1 - G(x)) S_b(G(x), 0)
 \end{aligned}$$

gilt und eine Gewichtung mit dem Faktor Null ausgeschlossen wird. Somit wäre auch die zweite Aussage des Theorems bewiesen. \square

Theorem 4.2.2 Sei QS_α eine proper scoring rule für Vorhersagen des α -Quantils auf den reellen Zahlen, v sei eine nichtnegative, integrierbare Funktion auf dem Intervall $[0, 1]$ und F sei eine Verteilungsfunktion auf den reellen Zahlen. Dann ist auch die scoring rule

$$S(F, y) = \int_0^1 QS_\alpha(F^{-1}(\alpha), y) v(\alpha) d\alpha \tag{4.2.4}$$

proper.

Falls es sich bei QS_α um eine strictly proper scoring rule handelt und die Gewichtungsfunktion v streng positiv ist, dann ist auch die scoring rule aus Darstellung (4.2.4) strictly proper.

Beweis: Sei G eine weitere beliebige Verteilung auf den reellen Zahlen mit $G \neq F$. Dann folgt mit dem Satz von Fubini (siehe z.B. Elstrodt (2011)) und der Eigenschaft proper

von QS_α die erste Aussage des Theorems.

$$\begin{aligned}
 S(F, G) &= \int_{-\infty}^{\infty} \left(\int_0^1 QS_\alpha(F^{-1}(\alpha), y) v(\alpha) d\alpha \right) dG(y) \\
 &= \int_0^1 \left(\int_{-\infty}^{\infty} QS_\alpha(F^{-1}(\alpha), y) dG(y) \right) v(\alpha) d\alpha \\
 &\leq \int_0^1 \left(\int_{-\infty}^{\infty} QS_\alpha(G^{-1}(\alpha), y) dG(y) \right) v(\alpha) d\alpha \\
 &= \int_{-\infty}^{\infty} \left(\int_0^1 QS_\alpha(G^{-1}(\alpha), y) v(\alpha) d\alpha \right) dG(y) \\
 &= S(G, G)
 \end{aligned}$$

Nimmt man zusätzlich an, dass die scoring rule QS_α strictly proper und v streng positiv ist, dann ergibt sich in der Ungleichung ein strikt kleiner, da nun

$$\int_{-\infty}^{\infty} QS_\alpha(F^{-1}(\alpha), y) dG(y) < \int_{-\infty}^{\infty} QS_\alpha(G^{-1}(\alpha), y) dG(y)$$

gilt und eine Gewichtung mit dem Faktor Null ausgeschlossen wird. Somit wäre auch die zweite Aussage des Theorems bewiesen. \square

Je nachdem, an welchen Bereichen der Dichte man besonders interessiert ist, kann man in den beiden gewichteten Versionen des CRPS die verschiedensten Gewichtungsfunktionen verwenden. Einige Beispiele für Gewichtungsfunktionen, die entsprechend das Zentrum oder beide bzw. die einzelnen Ränder der Dichte hervorheben, werden in Tabelle 4.2.1 aufgelistet. Sie wurden aus Gneiting & Ranjan (2011) entnommen und werden auch bei den späteren Berechnungen verwendet. Für eine starke Gewichtung auf das Zentrum der vorhergesagten Dichte wird dabei als Gewichtungsfunktion die Dichte der Normalverteilung mit Erwartungswert a und Standardabweichung b verwendet. Die Parameter a und b können individuell festgelegt werden. Beispielsweise könnte der Parameter a bei der Schätzung der erwarteten Inflation auf 0.02 festgelegt werden, da die Europäische Zentralbank eine Inflationsrate von knapp unter 2 Prozent als Richtwert festgelegt hat. Eine weitere Möglichkeit, welche wir auch bei unseren Berechnungen verfolgen werden, besteht darin, die Parameter durch Schätzung des Erwartungswertes und der Standardabweichung der Zielgröße unter Verwendung der historischen Daten zu bestimmen.

Bei der Bewertung einer vorhergesagten Verteilungsfunktion anhand von Formel (4.2.1) oder Formel (4.2.2) wird sich in der Praxis oftmals das Problem ergeben, dass man keine geschlossene Darstellung der Formeln ermitteln kann. In solchen Fällen greift man auf diskrete Approximationen der beiden Formeln zurück (Gneiting & Ranjan, 2011). Bei der Schwellenwert gewichteten Version legt man den für die Bewertung der Verteilungsfunktion interessanten Bereich durch das Intervall $[x_u, x_o] \subset \mathbb{R}$ fest und definiert sich I Bewertungspunkte

$$x_i = x_u + i \frac{x_o - x_u}{I}, \quad i = 1, \dots, I, \quad (4.2.5)$$

Gewichtung	Schwellenwert gewichtete Version	Quantil gewichtete Version
Zentrum	$u_1(y) = \phi_{a,b}(y)$	$v_1(\alpha) = \alpha(1 - \alpha)$
Ränder	$u_2(y) = 1 - \phi_{a,b}(y) / \phi_{a,b}(a)$	$v_2(\alpha) = (2\alpha - 1)^2$
rechter Rand	$u_3(y) = \Phi_{a,b}(y)$	$v_3(\alpha) = \alpha^2$
linker Rand	$u_4(y) = 1 - \Phi_{a,b}(y)$	$v_4(\alpha) = (1 - \alpha)^2$

Tabelle 4.2.1.: In obiger Tabelle sind mögliche Gewichtungsfunktionen für die Schwellenwert und die Quantil gewichtete Version des CRPS dargestellt. Dabei wurde bei der Schwellenwert gewichteten Version auf die Dichte $\phi_{a,b}$ und auf die Verteilungsfunktion $\Phi_{a,b}$ der Normalverteilung mit Erwartungswert a und Standardabweichung b zurückgegriffen. Die Tabelle stammt aus Gneiting & Ranjan (2011).

welche gleichmäßig auf dem Intervall verteilt sind. Die Schwellenwert gewichtete Version des CRPS lässt sich nun mithilfe der Formel

$$S(F, y) = -\frac{x_o - x_u}{I} \sum_{i=1}^I u(x_i) \text{BS}(F(x_i), \mathbb{I}\{x_i \geq y\}) \quad (4.2.6)$$

beliebig genau approximieren, wobei durch eine feinere Partitionierung eine genauere Approximation erreicht werden kann. Bei der Quantil gewichteten Version wählt man eine äquivalente Vorgehensweise. Man unterteilt das Intervall $[0, 1]$, welches von dem Parameter α durchlaufen wird, in J Abschnitte und approximiert die Quantil gewichtete Version des CRPS anhand der Formel

$$S(F, y) = \frac{1}{J-1} \sum_{j=1}^{J-1} v(\alpha_j) \text{QS}_{\alpha_j}(F^{-1}(\alpha_j), y), \quad \text{mit } \alpha_j = \frac{j}{J}. \quad (4.2.7)$$

Es sei noch erwähnt, dass auch die beiden diskreten Approximationen des CRPS proper scoring rules darstellen, da sie als Spezialfälle der Formeln (4.2.1) und (4.2.2) entstehen, wenn das Integral bezüglich eines geeigneten diskreten Stieltjes Maßes gebildet wird (Gneiting & Ranjan, 2011). Für nähere Erläuterungen zu Stieltjes Maßen siehe zum Beispiel Elstrodt (2011).

5. Tests und grafische Vergleichsmethoden

Mit dem bisher erlangten Wissen sind wir in der Lage, Wahrscheinlichkeitsvorhersagen anhand von proper scoring rules zu bewerten. Dieses Wissen wollen wir nutzen, um verschiedene Vorhersagemethoden bezüglich ihrer Vorhersagequalität zu vergleichen. Dazu greift man auf die historischen Daten einer Zeitreihe zurück, ermittelt mit jedem Vorhersageverfahren Prognosen für einen festen Zeitraum der Datenreihe und berechnet die mittleren scores der Vorhersagen. Diese scores ermöglichen es, ein Ranking der Vorhersagemethoden zu bestimmen. Da jedoch der absolute Wert eines mittleren scores oftmals wenig Aussagekraft besitzt und wir die scores deshalb nur nach ihrer Rangfolge klassifizieren können, werden wir nun einige weitere Möglichkeiten betrachten, mit denen man zwei Vorhersagemethoden vergleichen kann. Zunächst werden wir auf einen von Gneiting & Ranjan (2011) verwendeten Test auf gleiche Vorhersage Performance eingehen. Im Anschluss daran werden grafische Methoden vorgestellt, die zum Vergleich zweier Vorhersagemethoden herangezogen werden können.

5.1. Test auf gleiche Vorhersage Performance

In der Statistik spielen Hypothesentests eine wichtige Rolle, weil sie eine Möglichkeit darstellen, um den Wahrheitsgehalt einer Hypothese unter mathematisch fundierten Überlegungen zu testen. Gneiting & Ranjan (2011) verwenden einen Test, mit dessen Hilfe man Wahrscheinlichkeitsvorhersagen miteinander vergleichen kann. Die Grundidee für diesen Test geht auf einen Ansatz von Diebold & Mariano (1995) zurück, mit dem man zwei Vorhersageverfahren für Punktschätzungen auf gleiche Vorhersageeffizienz testen kann. Amisano & Giacomini (2007) übertrugen diese Überlegungen auf das Testen von Wahrscheinlichkeitsvorhersagen unter Verwendung des logarithmic scores und Gneiting & Ranjan (2011) nutzten wiederum diesen Test und zogen weitere scoring rules in ihre Überlegungen ein.

Für unseren Test definieren wir uns y als unsere Zielgröße, für die wir Wahrscheinlichkeitsvorhersagen vergleichen wollen. Wir gehen davon aus, dass uns als Datenbasis T tatsächlich eingetretene Werte y_1, \dots, y_T der Zielgröße zur Verfügung stehen. Außerdem bezeichnen f und g die zwei Wahrscheinlichkeitsvorhersagen, welche wir in unserem Test vergleichen wollen, und wir nehmen an, dass für die Schätzung der Vorhersagen f und g jeweils m Datensätze genutzt werden. Für eine Schätzung, die von einem Zeitpunkt t aus k Perioden in die Zukunft geht, ergeben sich folglich Wahrscheinlichkeitsvorhersagen $f_{t+k} = f_{t+k}(y_{t-m+1}, \dots, y_t)$ und $g_{t+k} = g_{t+k}(y_{t-m+1}, \dots, y_t)$. Wenn unsere Datenbasis aus $T = m + n + k$ Datensätzen besteht, können wir folglich für beide Vorhersagemethoden

$n + 1$ scores ermitteln, aus denen wir wiederum die mittleren scores

$$\bar{S}_{n+1}^f = \frac{1}{n+1} \sum_{t=m}^{m+n} S(f_{t+k}, y_{t+k}), \quad \bar{S}_{n+1}^g = \frac{1}{n+1} \sum_{t=m}^{m+n} S(g_{t+k}, y_{t+k}), \quad (5.1.1)$$

berechnen.

Der Ansatz, den Gneiting & Ranjan (2011) verfolgen, besteht darin, die Differenz aus den mittleren scores beider Vorhersagemethoden anhand einer geeigneten Teststatistik zu analysieren. Die von Gneiting & Ranjan (2011) dafür vorgeschlagene Teststatistik lautet

$$t_{n+1} = \sqrt{n+1} \frac{\bar{S}_{n+1}^f - \bar{S}_{n+1}^g}{\hat{\sigma}_{n+1}}, \quad (5.1.2)$$

mit

$$\hat{\sigma}_{n+1}^2 = \frac{1}{n+1} \sum_{j=-(k-1)}^{k-1} \sum_{t=m}^{m+n-|j|} \Delta_{t,k} \Delta_{t+|j|,k}$$

$$\Delta_{t,k} = S(f_{t+k}, y_{t+k}) - S(g_{t+k}, y_{t+k}).$$

Diese Teststatistik ist unter der Nullhypothese, dass beide Wahrscheinlichkeitsvorhersagen gleiche Vorhersage Performance besitzen, und unter drei weiteren Annahmen, die nachfolgend erläutert werden, asymptotisch standardnormalverteilt. Außerdem konvergiert die Wahrscheinlichkeit, dass der Test abgelehnt wird, unter der Gegenhypothese, dass die Performance der Vorhersagen unterschiedlich ist, gegen 1. Falls es bei dem Test zu einer Ablehnung der Nullhypothese kommt, wird man sich entsprechend für die Wahrscheinlichkeitsvorhersage mit dem höheren mittleren score entscheiden, d.h. man wähle f , falls die Teststatistik t_n positiv ist, und g , falls t_n negativ ist.

Damit die Teststatistik unter der Nullhypothese standardnormalverteilt ist, müssen noch die folgenden Bedingungen gelten:

1. Der Prozess $\{Y_t\}$ muss eine mixing condition erfüllen.
2. $\hat{\sigma}_n^2$ muss ein konsistenter Schätzer für

$$\sigma_n^2 = \text{var} \left(\sqrt{n} \left(\bar{S}_n^f - \bar{S}_n^g \right) \right) > 0 \quad (5.1.3)$$

sein.

3. Die Momentenbedingungen müssen erfüllt sein, d.h.

$$\mathbb{E}_{f_{t+k}|X|}, \quad \mathbb{E}_{g_{t+k}|X|} \quad \text{und} \quad \mathbb{E}_{t+k|Y_{t+k}|}^{2r} \quad (5.1.4)$$

müssen für alle t endlich sein, wobei $r \geq 2$ von der mixing condition an $\{Y_t\}$ abhängt.

Für weitere Informationen zu diesen Bedingungen und den entsprechenden Nachweisen, sowie zu dem Test im Allgemeinen sei auf Gneiting & Ranjan (2011), Amisano & Giacomini (2007) und Diebold & Mariano (1995) verwiesen.

5.2. Grafischer Vergleich durch Schwellenwert und Quantil Zerlegung des CRPS

Wenn man mit statistischen Verfahren versucht, zwei Vorhersagemethoden zu vergleichen, verwendet man im Allgemeinen Tests oder scoring rules und interpretiert die Testergebnisse bzw. die scores. Auf diese Weise erhält man zwar präzise Entscheidungshilfen, aber man hat meist nur eine eingeschränkte Vorstellung davon, inwieweit sich die Vorhersage Performance der Methoden unterscheidet und wo die Unterschiede liegen. Nach Möglichkeit zieht man deshalb auch grafische Hilfsmittel zur Analyse heran, da sie Unterschiede in der Vorhersage meist sehr gut offen legen.

Wir wollen nun zwei von Gneiting & Ranjan (2011) vorgeschlagene grafische Darstellungsformen für den CRPS betrachten, anhand derer wir die Stärken und Schwächen von Wahrscheinlichkeitsvorhersagen analysieren wollen. Dafür nehmen wir, wie schon im vorangegangenen Abschnitt über den Test auf gleiche Vorhersage Performance, an, dass uns von unserer Zielgröße y als Datenbasis $T = m + n + k$ Ausprägungen y_1, \dots, y_T zur Verfügung stehen und dass wir anhand von m Werten eine Wahrscheinlichkeitsvorhersage f erzeugen, die von einem Zeitpunkt $t \in \{m, \dots, m + n\}$ k Perioden in die Zukunft geht. Verwendet man nun den mittleren score $\overline{\text{CRPS}}_n^f$ und vertauscht Summe und Integral, dann erhält man für die Brier Darstellung (3.2.16) des CRPS den Ausdruck

$$\overline{\text{CRPS}}_n^f = - \int_{-\infty}^{\infty} \overline{\text{BS}}_n^f(x) dx \quad (5.2.1)$$

für den mittleren score, wobei

$$\overline{\text{BS}}_n^f(x) = \frac{1}{n+1} \sum_{t=m}^{m+n} \text{BS}(F_{t+k}(x), \mathbb{I}\{x \geq y_{t+k}\}) \quad (5.2.2)$$

den mittleren Brier score darstellt. Entsprechend lässt sich auch für die Quantil Darstellung (3.2.18) des CRPS die Darstellungsform

$$\overline{\text{CRPS}}_n^f = \int_{-\infty}^{\infty} \overline{\text{QS}}_n^f(\alpha) d\alpha \quad (5.2.3)$$

für den mittleren score ermitteln mit

$$\overline{\text{QS}}_n^f(\alpha) = \frac{1}{n+1} \sum_{t=m}^{m+n} \text{QS}_{\alpha}(F_{t+k}^{-1}(\alpha), y_{t+k}) \quad (5.2.4)$$

als mittleren quantile score.

Sowohl bei dem mittleren negativen Brier score $-\overline{\text{BS}}_n^f(x)$ als auch bei dem mittleren quantile score $\overline{\text{QS}}_n^f(\alpha)$ handelt es sich jeweils um eine Funktion, welche nur von dem Parameter x beziehungsweise α abhängt. Folglich können wir diese beiden Funktionen in Abhängigkeit von dem jeweiligen Parameter grafisch darstellen und mithilfe dieser Darstellung unsere Vorhersage f analysieren (siehe z.B. Abb. 7.4.4 und 7.4.5 auf Seite 75). In Anlehnung an Gneiting & Ranjan (2011) bezeichnen wir die Zerlegung der CRPS über den negativen Brier score als *Schwellenwert Zerlegung des CRPS* und verwenden für die Zerlegung über den quantile score den Begriff *Quantil Zerlegung des CRPS*.

6. Praktische Anwendung

Nachdem wir in den vorangegangenen Kapiteln die Theorie zur Bewertung verschiedener Wahrscheinlichkeitsvorhersagen mit proper scoring rules erarbeitet haben, wollen wir diese nun an praktischen Beispielen anwenden. Um dies zu tun, müssen wir uns zuerst mit den Vorhersagemethoden beschäftigen, welche wir vergleichen wollen. Da eine Vielzahl an Verfahren zur Vorhersage von makroökonomischen Kennzahlen zur Verfügung steht, müssen wir unsere Auswahl einschränken. Deshalb werden wir für die Analyse drei einfache Vorhersagemethoden, ein Zeitreihenmodell und drei komplexere Ansätze heranziehen. Die entsprechenden Verfahren werden in Abschnitt 6.2 vorgestellt. Außerdem müssen wir eine Datenbasis für die Berechnungen festlegen. Hierbei greifen wir auf wichtige makroökonomische Daten zurück und beschränken uns auf Daten zur Inflation, zum Bruttoinlandsprodukt und zur Arbeitslosenquote der Bundesrepublik Deutschland. Die verwendeten Daten werden im nächsten Abschnitt beschrieben.

6.1. Datengrundlage für die Berechnungen

Um bei unseren Berechnungen aussagekräftige Ergebnisse zu erhalten, müssen wir entsprechend lange Datenreihen verwenden. Das Statistische Amt der Europäischen Union (Eurostat) stellt uns dafür ein großes Angebot an makroökonomischen Daten zur Verfügung. Hieraus entnehmen wir die Vergangenheitswerte der Inflationsrate, des Bruttoinlandsproduktes und der Arbeitslosenquote der Bundesrepublik Deutschland. Diese liegen auf monatlicher Basis bzw. quartalsweise vor und gehören sicherlich zu den wichtigsten Kennzahlen, um die volkswirtschaftliche Entwicklung eines Landes beurteilen zu können. Der Datenabzug aus Eurostat erfolgte zum 15. Februar 2013 und enthält Daten aus 1991 bis einschließlich Dezember 2012. Nähere Einzelheiten werden im Folgenden erläutert.

6.1.1. Inflationsrate

Die Inflationsrate ist eine Kennzahl, welche die Preisniveauentwicklung von Gütern und Dienstleistungen beschreibt. Um die Entwicklung der Preise zu quantifizieren bestimmt man im Allgemeinen einen Warenkorb und ermittelt im Zeitverlauf den Wert des Warenkorbs anhand der entsprechenden Marktpreise. Natürlich gibt es verschiedene Möglichkeiten, wie man den Warenkorb festlegen kann. Außerdem existieren unterschiedliche Ansätze zur Anpassung des Warenkorbes an das Kaufverhalten der Verbraucher. Wir wollen bei unseren Analysen den harmonisierten Verbraucherpreisindex (HVPI) betrachten. Dieser wird EU-weit nach einheitlichen Regeln durch jedes EU-Mitgliedsland auf monatlicher Basis erhoben und dient als Preisindex für die Lebenshaltungskosten. Um

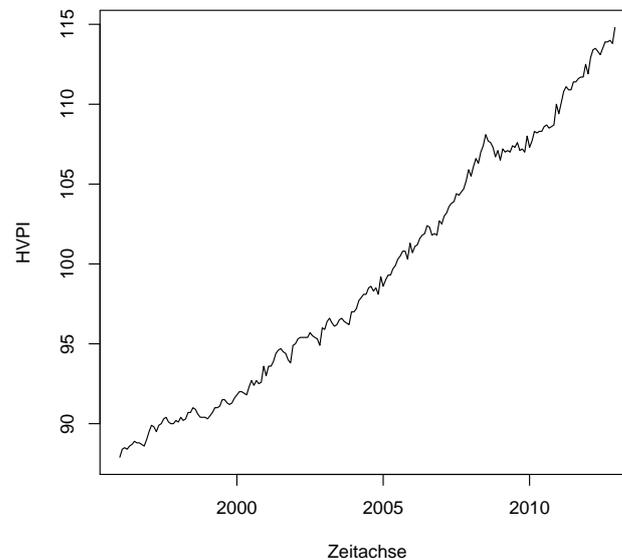


Abbildung 6.1.1.: Die Entwicklung des harmonisierten Verbraucherpreisindex im Zeitverlauf.

einen besseren Vergleich zu ermöglichen legte man als Referenzperiode das Jahr 2005 fest und definierte den Wert des Warenkorbs für dieses Jahr als 100. Aus den Werten des HVPI, welche in Abbildung 6.1.1 dargestellt werden, lassen sich die Wachstumsraten des Preisindex ermitteln, welche als Inflationsraten bezeichnet werden. Beispielsweise betrug der HVPI für Deutschland im November 2012 113.8 und im Dezember 114.8 Einheiten. Somit lässt sich von November auf Dezember eine Inflationsrate von

$$\pi = \frac{114.8 - 113.8}{113.8} \approx 0.9\%$$

bestimmen (siehe z.B. den Leitfaden für Datenutzer von der EU (2004) oder Blachard & Illing (2009)). Als Datenbasis für die Analyse werden die monatlichen HVPI-Werte von Januar 1996 (Beginn der Aufzeichnung in Eurostat) bis Dezember 2012 herangezogen und daraus die entsprechenden monatlichen Inflationsraten ermittelt. Diese dienen als Grundlage für die in Kapitel 7.2 folgenden Berechnungen. Bei den verwendeten Daten handelt es sich um nicht-saisonbereinigte Werte. Die von der Jahreszeit abhängigen Schwankungen lassen sich bereits in Abbildung 6.1.2, welche die monatlichen Inflationsraten darstellt, erkennen. Im Vergleich dazu zeigt Abbildung 6.1.3 die saisonbereinigten Daten. Diese saisonalen Effekte werden bei den späteren Berechnungen eine Rolle spielen und werden deshalb im Kapitel 7.2 nochmals genauer erläutert.

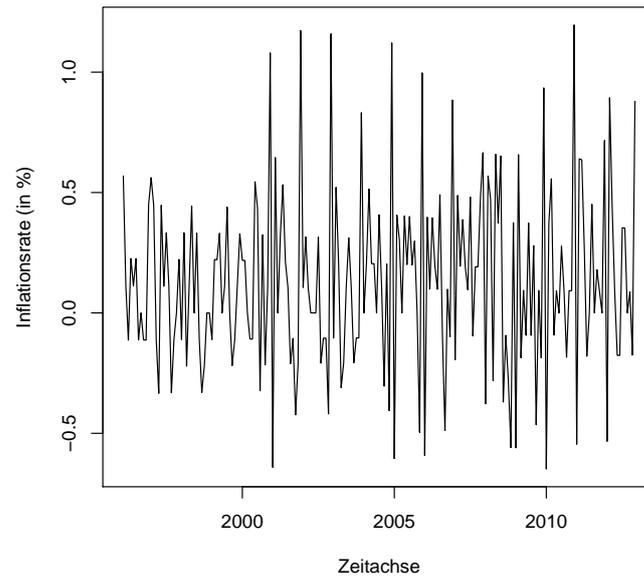


Abbildung 6.1.2.: Die Entwicklung der monatlichen Inflationsraten im Zeitverlauf.

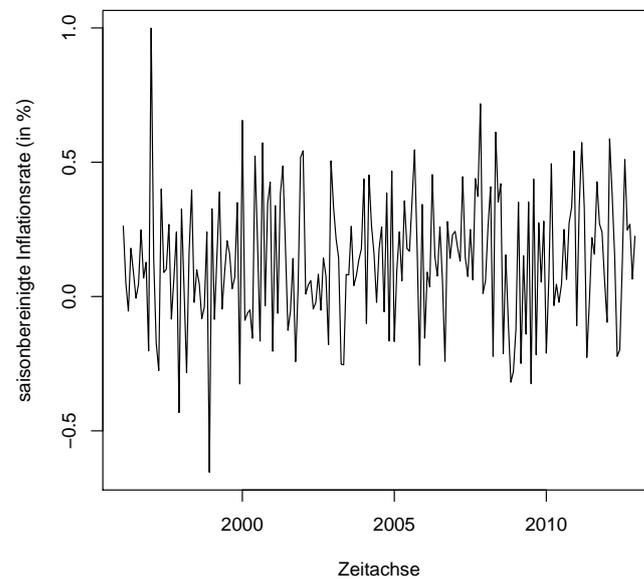


Abbildung 6.1.3.: Die Entwicklung der saisonbereinigten monatlichen Inflationsraten im Zeitverlauf.

6.1.2. Bruttoinlandsprodukt

Die zweite makroökonomische Kennzahl, für die wir Wahrscheinlichkeitsvorhersagen erzeugen und bewerten wollen, ist das Bruttoinlandsprodukt (BIP). Das Bruttoinlandsprodukt ist ein bedeutender Indikator für die wirtschaftliche Entwicklung eines Landes, denn es gibt den Wert der Waren und Dienstleistungen an, die innerhalb eines bestimmten Zeitraums für den Endverbrauch hergestellt bzw. erbracht wurden. Wir werden bei unseren Berechnungen die quartalsweisen Wachstumsraten des Bruttoinlandsproduktes beginnend mit dem zweiten Quartal 1991 (Beginn der Aufzeichnung in Eurostat) bis einschließlich dem vierten Quartal 2012 nutzen. Die Daten sind saison- und arbeitstäglich bereinigt, wurden zu den jeweiligen Marktpreisen der Waren und Dienstleistungen erhoben und entsprechen der prozentualen Veränderung gegenüber der Vorperiode. Die Entwicklung der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes wird in Abbildung 6.1.4 dargestellt. Die Wachstumsraten schwanken um einen leicht positiven Wert, aber man erkennt beispielsweise auch den starken Abschwung der Wirtschaft im 1. Quartal in 2009, welcher auf die Finanzkrise zurückzuführen ist (siehe z.B. Blachard & Illing (2009)).

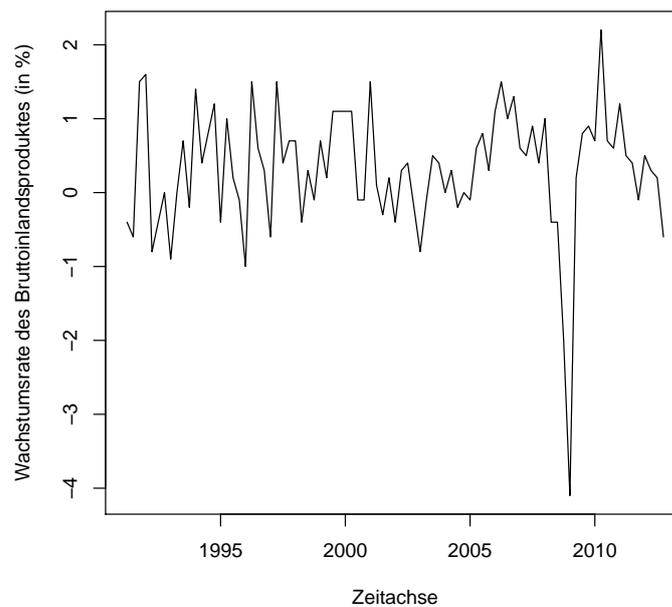


Abbildung 6.1.4.: Die Entwicklung der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes im Zeitverlauf.

6.1.3. Arbeitslosenquote

Schließlich wollen wir noch die Arbeitslosenquote von Deutschland analysieren. Die Arbeitslosenquote definiert sich als das Verhältnis zwischen der Zahl der Arbeitslosen und der Zahl der Erwerbspersonen (d.h. der Summe der Beschäftigten und Arbeitslosen) (siehe z.B. Blachard & Illing (2009)). Diese Quote unterliegt saisonalen Schwankungen und wird von der Bundesanstalt für Arbeit jeden Monat veröffentlicht. Wir greifen jedoch auf die von Eurostat zur Verfügung gestellten monatlichen Daten von Januar 1991 bis Dezember 2012 zurück, welche dort in saisonbereinigter Form vorliegen. Abbildung 6.1.5 zeigt den Verlauf der Arbeitslosenquote, wobei sich phasenweise kontinuierliche Anstiege und Rückgänge der Arbeitslosenquote erkennen lassen. Da wir für unsere spätere Zeitreihenanalyse Daten benötigen, die keinem steigenden oder fallenden Trend folgen, müssten wir die jeweiligen Trends erst schätzen und die anschließenden Berechnungen mit den trendbereinigten Daten fortsetzen. Um dies zu vermeiden ermitteln wir die prozentualen Wachstumsraten der Arbeitslosenquote. Deren Entwicklung wird in Abbildung 6.1.6 dargestellt. Es ist leicht ersichtlich, dass diese Werte nun um Null herum schwanken. Etwas problematisch ist jedoch der diskret wirkende Verlauf. Dieser ergibt sich zum Teil aufgrund der groben Angabe der Arbeitslosenquote auf nur eine Nachkommastelle. Da in Abbildung 6.1.5 allerdings auch lineare Anstiege und Rückgänge ersichtlich sind, scheint der diskret wirkende Verlauf der Wachstumsraten auch teil einer natürlichen Entwicklung zu sein. Eine genauere Erklärung, warum wir für die Vorhersage trendbereinigte Daten benötigen, erfolgt in Abschnitt 6.2.1.

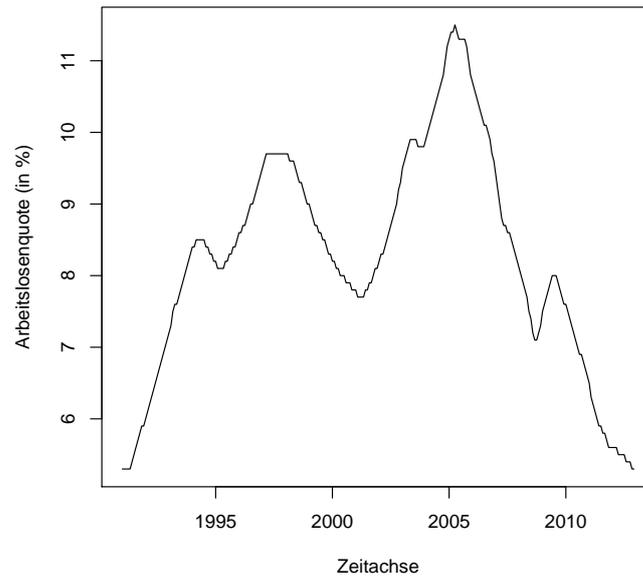


Abbildung 6.1.5.: Die Entwicklung der monatlichen Arbeitslosenquoten im Zeitverlauf.

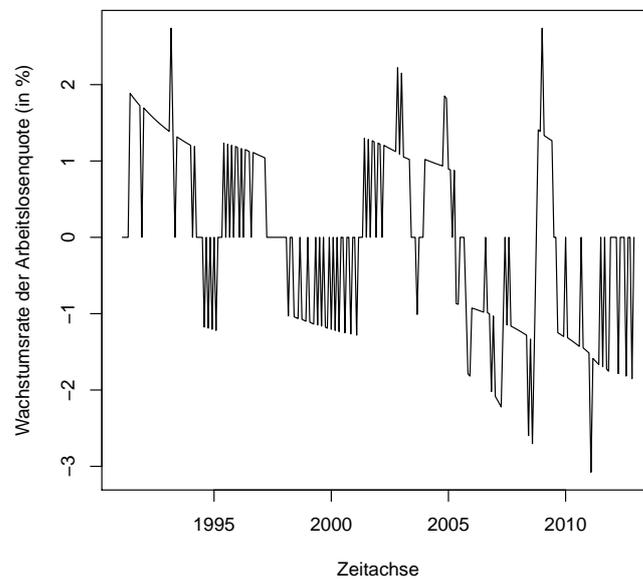


Abbildung 6.1.6.: Die Entwicklung der monatlichen Wachstumsraten der Arbeitslosenquote im Zeitverlauf.

6.2. Vorhersagemethoden

Vorhersagen von makroökonomischen Kennzahlen sind für die wirtschaftliche Planung und Steuerung in Unternehmen und Institutionen von besonderer Bedeutung. Dem zur Folge hat sich in der Vergangenheit eine umfangreiche Theorie über Zeitreihenvorhersagen entwickeln, welche innerhalb dieser Arbeit nicht in aller Ausführlichkeit dargestellt werden kann. Wir werden deshalb nur auf die wichtigsten Aspekte zur Zeitreihenanalyse eingehen. Im Anschluss daran werden die bei unseren Berechnungen herangezogenen Vorhersageverfahren vorgestellt. Hierbei werden aus der Vielzahl an Vorhersagemethoden zuerst drei einfache und intuitiv nachvollziehbare Methoden zur Erzeugung von Wahrscheinlichkeitsvorhersagen eingeführt, welche auf den Mittelwert und die empirische Varianz der historischen Daten zurückgreifen. Anschließend wollen wir ein autoregressives Zeitreihenmodell betrachten, anhand dessen wir eine Wahrscheinlichkeitsvorhersage erzeugen. Schließlich werden drei Modelle erläutert, deren Parameter durch eine Optimierung bestimmt werden.

6.2.1. Grundlagen der Zeitreihenanalyse

Für die Zeitreihenanalyse greift man im Allgemeinen auf die Datenhistorie y_1, \dots, y_T der Kenngröße y zurück, die man analysieren und vorhersagen möchte, sowie auf weitere historische Daten, welche man für die Vorhersage als relevant einschätzt. Damit man aus den Vergangenheitswerten sinnvolle Aussagen über die zukünftige Entwicklung der Zeitreihe treffen kann, sollten sich die Werte im Zeitverlauf ähnlich verhalten. Aus theoretischen Gründen hätte man deshalb gerne, dass die Verteilung der Daten zeitunabhängig ist. Diese Eigenschaft nennt man Stationarität und definiert sich wie folgt (siehe z.B. Kreiß & Neuhaus (2006) oder Schlittgen (2012)).

Definition 6.2.1 Sei $Y = (Y_t : t \in \mathbb{Z})$ eine reellwertige Zeitreihe. Die Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ heißt streng stationär, falls für alle $n \in \mathbb{N}$ und alle $t_1, \dots, t_n, h \in \mathbb{Z}$

$$P_{Y_{t_1}, \dots, Y_{t_n}} = P_{Y_{t_1+h}, \dots, Y_{t_n+h}} \quad (6.2.1)$$

gilt, wobei P_{Y_1, \dots, Y_n} die gemeinsame Verteilung der Y_1, \dots, Y_n bezeichne.

Die Eigenschaft der strengen Stationarität wird man bei Wirtschaftsdaten im Allgemein nicht vorfinden. Deshalb versucht man bei der Modellierung zumindest die schwache Stationarität sicherzustellen, welche lediglich Bedingungen an die ersten und zweiten Momente stellt.

Definition 6.2.2 Sei $Y = (Y_t : t \in \mathbb{Z})$ mit $\mathbb{E}|Y_t|^2 < \infty$ eine reellwertige Zeitreihe. Die Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ heißt schwach stationär, falls für alle $t \in \mathbb{Z}$ und alle $h \in \mathbb{Z}$ die Bedingungen

$$\mathbb{E}Y_t = \mu \quad (6.2.2)$$

$$\text{Cov}(Y_{t+h}, Y_t) = \gamma_h \quad (6.2.3)$$

gelten.

Die Forderung, dass der Erwartungswert im Zeitverlauf konstant bleibt, ist sicherlich bei Zeitreihen, die einem Trend folgen, nicht gegeben. Dies hat zur Folge, dass der Trend der Zeitreihe erst geschätzt und entfernt werden muss, bevor eine weitere sinnvolle Analyse möglich ist. Je nach Zeitreihe lässt sich eine Trendbereinigung aber auch durch die Betrachtung der Wachstumsraten der Zeitreihe umgehen. Wir werden diesen Ansatz verfolgen und haben deshalb bereits die Wachstumsraten des harmonisierten Verbraucherpreisindex (Inflationsrate), des Bruttoinlandsproduktes und der Arbeitslosenquote betrachtet, vgl. Abbildungen 6.1.2, 6.1.4 und 6.1.6.

Neben dem Vorliegen eines wachsenden oder fallenden Trends können auch saisonale Effekte dafür verantwortlich sein, dass sich der Erwartungswert im Zeitverlauf verändert. Solche periodischen Schwankungen lassen sich bei makroökonomischen Daten oftmals erkennen, da jahreszeitliche Entwicklungen die wirtschaftlichen Tätigkeiten der Unternehmen und das Konsumverhalten der Bürger beeinflussen. Hier sei beispielsweise das Weihnachtsgeschäft zu nennen, oder auch die niedrige Beschäftigungsrate im Baugewerbe und in der Landwirtschaft in den Wintermonaten. Bei den Daten zum Bruttoinlandsprodukt und zur Arbeitslosenquote wurden saisonale Effekte bereits entfernt, aber bei den Inflationsraten müssen wir erst saisonbereinigte Werte ermitteln. Natürlich gibt es auch zur Schätzung der saisonalen Anteile in Zeitreihen die unterschiedlichsten Methoden. Wir wollen uns in dieser Arbeit auf ein einfaches Verfahren beschränken, welches aus Kreiß & Neuhaus (2006) entnommen wurde und nun kurz vorgestellt wird.

Verfahren zur Ermittlung der saisonalen Anteile einer Zeitreihe

Seien y_1, \dots, y_T die realisierten Werte einer Zeitreihe, deren saisonalen Anteile wir bestimmen wollen. Weiterhin gehen wir davon aus, dass die Periodenlänge d der Zeitreihe bekannt ist. Nun berechnen wir den Mittelwert \bar{y} der gesamten Datenreihe, sowie den Mittelwert \bar{s}_i , $i = 1, \dots, d$, für jede Saison. Da wir im Allgemeinen nicht davon ausgehen können, dass die Anzahl T unserer Daten y_1, \dots, y_T immer einem Vielfachen der Periodenlänge d entspricht, bestimmen wir durch $\max\{k \in \mathbb{N} : (k-1) * d + i \leq T\}$, wie viele Werte der Saison i in der Datenreihe enthalten sind.

$$\bar{s}_i = \frac{1}{\max\{k \in \mathbb{N} : (k-1) * d + i \leq T\}} \sum_{j=1}^{\max\{k \in \mathbb{N} : (k-1) * d + i \leq T\}} y_{(j-1)*d+i} \quad (6.2.4)$$

Die Saisonanteile s_1, \dots, s_d definieren sich schließlich als

$$s_i = \bar{s}_i - \bar{y}, \quad i = 1, \dots, d, \quad (6.2.5)$$

wobei durch Abzug des Mittelwertes \bar{y} der gesamten Datenreihe dafür Sorge getragen werden soll, dass sich die saisonalen Effekte innerhalb einer Periode insgesamt wieder aufheben, d.h. $s_1 + \dots + s_d = 0$. Hierzu ist allerdings zu sagen, dass die Gleichheit im Allgemeinen nur eintritt, wenn die Datenanzahl T tatsächlich ein Vielfaches der Periodenlänge d ist, andernfalls erhalten wir eine kleine Ungenauigkeit, welche aber bei einer

großen Datenmenge zu vernachlässigen ist.

Die saisonbereinigten Daten y_1^s, \dots, y_T^s erhält man durch Abzug des entsprechenden Saisonanteils (Kreiß & Neuhaus, 2006).

$$y_i^s = y_i - s_{h+1}, \quad \text{mit } h = i \bmod d, \quad i = 1, \dots, T. \quad (6.2.6)$$

Besonderheiten bei der Vorhersage von makroökonomischen Daten

Mithilfe des beschriebenen Verfahrens ermittelten wir saisonbereinigte Inflationsraten, welche bereits in Abbildung 6.1.3 dargestellt wurden. Zur Bestimmung der saisonalen Effekte wurde auf die gesamte Datenreihe zurückgegriffen. Da aber bei den späteren Berechnungen auch Vorhersagen erzeugt werden, welche zeitlich gesehen in der Vergangenheit liegen, werden bei den Berechnungen immer nur die Daten für die Ermittlung der Saisonanteile berücksichtigt, die zum Zeitpunkt der Vorhersage schon vorlagen. Dies hat zwar zur Folge, dass die Schätzung der saisonalen Effekte für Vorhersagen, die weit in der Vergangenheit liegen, wegen der entsprechend geringeren Datenbasis schlechter ausfallen. Aber dafür ist diese Vorgehensweise praxisorientiert.

Weiterhin sei noch zu erwähnen, dass man aus einer Vorhersage, die auf Grundlage der saisonbereinigten Daten erzeugt wurde, durch Addition der entsprechenden Saisonkomponente eine nicht-saisonbereinigte Vorhersage erhält. Diesen Ansatz werden wir bei der Analyse der Inflationsraten verfolgen, sodass wir Vorhersagen für die tatsächlichen, nicht-saisonbereinigten Inflationsraten erhalten.

Da unsere Daten aufgrund des Übergangs zu Wachstumsraten und nach durchgeführter Saisonbereinigung gleichmäßig um einen uns unbekanntem konstanten Wert schwanken (vgl. Abb. 6.1.3, 6.1.4 und 6.1.6), könnte man aus theoretischer Sicht vermuten, dass unsere Datenreihe die Bedingungen der schwachen Stationarität erfüllen. Bei genauerem Betrachten wird man jedoch feststellen, dass dies nicht der Fall ist. Besonders auffällig ist beispielsweise, dass die Wachstumsraten der Arbeitslosenquote in manchen Zeitabschnitten ausschließlich größer oder gleich Null sind und in anderen Phasen kleiner oder gleich Null (vgl. Abbildung 6.1.6). Außerdem sind die Schwankungen der Wachstumsraten im Zeitverlauf unterschiedlich stark ausgeprägt, was auf eine Verletzung der zweiten Bedingung der schwachen Stationarität (6.2.3) hindeutet (vgl. Definition 6.2.2).

Derartige Veränderungen der Zeitreihenverläufe sind bei makroökonomischen Kennzahlen normal und entstehen unter anderem dadurch, dass die Wirtschaft durch Schocks beeinträchtigt wird, welche die Entwicklung über längere Zeit oder auch dauerhaft beeinflussen. Dabei kann es sich zum Beispiel um das Platzen einer Immobilienblase handeln, welche die Finanzkrise ausgelöst hat, oder um Gesetzesänderungen wie die Durchführungen der Hartz-IV-Reformen. Auch der technologische Fortschritt, die Einführung des Euros, die terroristischen Anschläge vom 11. September 2001 oder der Ausbruch des Irakkriegs sorgten für eine mehr oder weniger starke Beeinflussung der wirtschaftlichen

Entwicklung. Wir müssen folglich davon ausgehen, dass sich die Verteilungen unserer Zeitreihen mit der Zeit immer wieder verändern werden.

Um diesem Problem Rechnung zu tragen, verwenden wir bei der Bestimmung unserer Vorhersagen nicht die gesamte Datenhistorie, sondern beschränken uns zur Schätzung der Modellparameter auf die letzten m Datenpunkte, die uns zum Zeitpunkt der Vorhersage zur Verfügung stehen. Dieser Datenbereich, welcher bei der jeweiligen Schätzung berücksichtigt wird, nennt man auch *training window* (siehe z.B. Gneiting & Ranjan (2011)). Der Ansatz, den wir zur Bestimmung der training windows verfolgen werden, sowie die allgemeine Vorgehensweise bei der Analyse wird in Abschnitt 7.1 beschrieben.

6.2.2. Einfache Vergleichsvorhersagen

Im Folgenden werden konkrete Methoden zur Erzeugung von Wahrscheinlichkeitsvorhersagen vorgestellt. Dazu gehen wir davon aus, dass wir uns im Zeitpunkt t befinden und aus den letzten m saisonbereinigten Ausprägungen y_t, \dots, y_{t-m+1} einer Zeitreihe y eine Wahrscheinlichkeitsvorhersage für den nächsten Wert y_{t+1} der Zeitreihe erzeugen wollen. Unsere Wahrscheinlichkeitsvorhersage geben wir in Form einer Dichte an. Zur besseren Veranschaulichung sind die von den Vorhersagemethoden erzeugten Dichten mit beispielhaft gewählten Parametern in Abbildung 6.2.1 dargestellt. Zunächst wollen wir drei einfache Vorhersagemethoden erläutern, welche uns bei der späteren Analyse als Vergleichsvorhersagen zur Verfügung stehen sollen.

Ein intuitiver Ansatz zur Ermittlung einer Vorhersage stellt die Annahme dar, dass der Wert von y_{t+1} in etwa im Bereich des Mittelwertes $\hat{\mu}$ der m vorherigen Ausprägungen liegen dürfte und dass dessen mögliche Abweichung von diesem Wert durch die empirische Varianz $\hat{\sigma}^2$ der letzten m Werte beschrieben werden kann. Somit bietet sich als erste Vorhersage für y_{t+1} eine Normalverteilung $N(\mu, \sigma^2)$ an mit $\mu = \hat{\mu}$ und $\sigma^2 = \hat{\sigma}^2$, deren Dichte man mithilfe einer Transformation aus der Dichte der Standardnormalverteilung ϕ erhält.

1. Vorhersage: $N(\hat{\mu}, \hat{\sigma}^2)$

$$f_1(y) = \frac{1}{\hat{\sigma}} \phi\left(\frac{y - \hat{\mu}}{\hat{\sigma}}\right) \quad (6.2.7)$$

Nachdem wir bei unserem ersten Ansatz den Mittelwert als Erwartungswert angenommen haben und somit keine zeitliche Struktur der Zeitreihe unterstellt bzw. berücksichtigt haben, wollen wir für unsere zweite Vorhersage die Vermutung aufstellen, dass die zukünftige Entwicklung der Zeitreihe sehr stark von deren letzter Ausprägung abhängt. Setzt man eine solch hohe zeitliche Abhängigkeit voraus, dann dürfte y_{t+1} in etwa dem Wert seines Vorgängers y_t entsprechen und eine Normalverteilung mit y_t als Erwartungswert liegt nahe. Bei der Schätzung der Varianz vertrauen wir erneut auf die empirische

Varianz $\hat{\sigma}^2$ der vorangegangenen m Ausprägungen.

2. Vorhersage: $N(y_t, \hat{\sigma}^2)$

$$f_2(y) = \frac{1}{\hat{\sigma}} \phi\left(\frac{y - y_t}{\hat{\sigma}}\right) \quad (6.2.8)$$

Bisher haben wir jeweils eine Normalverteilung zugrunde gelegt, da diese gerne als Verteilung angenommen wird, wenn man keine genaueren Informationen über die Zeitreihenentwicklung hat und die Ausprägungen leicht um einen erwarteten Wert schwanken. Bei einem Blick auf die Entwicklung der saisonbereinigten Inflationsraten (siehe Abb. 6.1.3) liegt eine solche Annahme nahe. Wir müssen allerdings bedenken, dass unsere makroökonomischen Kennzahlen immer mal wieder durch starke Schocks beeinflusst werden. Ein deutlicher Einbruch der Wirtschaftsleistung wie beispielsweise während der Finanzkrise ist immer möglich und sollte uns davor warnen, extreme Ereignisse als sehr unwahrscheinlich einzuschätzen. Deshalb bietet sich als weitere Verteilung die Studentische t -Verteilung an, da diese der Normalverteilung sehr nahe kommt, aber den Ausreißern eine etwas höhere Wahrscheinlichkeit zuordnet. Unsere dritte Vorhersagemethode soll daher eine t -Verteilung erzeugen, deren Freiheitsgrad wir der Einfachheit halber im Vorhinein als 10 festlegen, sodass im Vergleich zur Normalverteilung die Ränder der Dichte etwas stärker, aber nicht zu stark ausgeprägt sind. Zusätzlich wird die Dichte der t -Verteilung d_t unter Verwendung des Mittelwertes $\hat{\mu}$ und der empirischen Varianz $\hat{\sigma}^2$ der m letzten Ausprägungen entsprechend lokalisiert und skaliert, um eine Anpassung des Erwartungswertes und der Varianz der Verteilung an die tatsächlichen Ausprägungen zu erzeugen.

3. Vorhersage: $t(\hat{\mu}, \hat{\sigma}^2)$

$$f_3(y) = \frac{1}{\hat{\sigma}} d_t\left(\frac{y - \hat{\mu}}{\hat{\sigma}}\right) \quad (6.2.9)$$

6.2.3. Autoregressives Zeitreihenmodell

Nachdem wir bei den Vorhersagemethoden 1 und 3 angenommen haben, dass die Entwicklung der Daten keine Zeitreihenstruktur aufweist, und bei Vorhersage 2 eine sehr starke zeitliche Abhängigkeit unterstellt haben, wollen wir nun einen Ansatz verfolgen, bei dem speziell die zeitliche Abhängigkeit gemessen und ausgenutzt werden soll. Dazu werden wir ein autoregressives Zeitreihenmodell verwenden.

Bei einem autoregressiven Modell geht man von einer stationären Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ aus, bei der der Wert Y_t der Zeitreihe zu einem beliebigen Zeitpunkt t linear abhängt von den vorangegangenen Werten $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$. Der Parameter p gibt dabei die Anzahl der vorangegangenen Zeitreihenglieder (auch *Lags* genannt) an, die einen Einfluss auf Y_t ausüben. Ein Modell der Ordnung p wird durch die Gleichung

$$Y_t = a_0 + a_1 Y_{t-1} + \dots + a_p Y_{t-p} + \epsilon_t \quad (6.2.10)$$

beschrieben, wobei a_0, \dots, a_p Parameter des Modells sind und ϵ_t einem *Weißes Rauschen* entspricht, d.h. einer Folge unabhängiger, identisch verteilter Zufallsvariablen mit Erwartungswert Null ($\mathbb{E}(\epsilon_t) = 0$) und fester Varianz ($\text{Var}(\epsilon_t) = \sigma_\epsilon^2$) (siehe z.B. Schlittgen (2012)).

Für unsere Vorhersage unterstellen wir die Annahme, dass die makroökonomische Kennzahl der Struktur eines autoregressiven Zeitreihenmodells folgt. Da unsere historischen Daten den Bedingungen der Stationarität nicht genügen und wir deshalb für die Vorhersage jeweils nur eine noch nicht festgelegte Anzahl von m Datenpunkten verwenden, ist die Schätzung eines autoregressiven Modells mit hoher Ordnung problematisch. Wir werden deshalb standardmäßig ein Modell mit Ordnung 1 verwenden. Bei den Berechnungen erfolgt die Schätzungen der Parameter mit der Methode *ar*, welche in der Statistik-Software R zur Verfügung steht und eine Yule-Walker-Schätzung durchführt. Mit den berechneten Parametern bestimmen wir die Punktvorhersage $\hat{y}_{t+1} = \mathbb{E}(y_{t+1})$ für y_{t+1} , welche wir als Erwartungswert für unsere Wahrscheinlichkeitsvorhersage annehmen. Außerdem stellt uns die Methode *ar* eine Schätzung $\hat{\sigma}_{AR}^2$ für den Teil der Varianz unserer Datenreihe zur Verfügung, der nicht durch das autoregressive Zeitreihenmodell erklärt werden kann. Schließlich definieren wir unsere vierte Wahrscheinlichkeitsvorhersage als eine Normalverteilung mit Erwartungswert \hat{y}_{t+1} und Varianz $\hat{\sigma}_{AR}^2$.

4. Vorhersage: AR1

$$f_{AR}(y) = \frac{1}{\hat{\sigma}_{AR}} \phi \left(\frac{y - \hat{y}_{t+1}}{\hat{\sigma}_{AR}} \right) \quad (6.2.11)$$

Während wir bei den ersten drei vorgestellten Vorhersagemethoden keine Möglichkeit haben die Dichtevorhersage im Falle einer Zweischrittvorhersage für y_{t+2} sinnvoll anzupassen, können wir bei der Autoregression y_{t+2} besser prognostizieren. Dies tun wir, indem wir auf die bereits geschätzten Parameter \hat{a}_0 und \hat{a}_1 zurückgreifen und den Schätzer für y_{t+2} definieren als $\hat{y}_{t+2} = \hat{a}_0 + \hat{a}_1 \hat{y}_{t+1}$. Die Dichtevorhersage ergibt sich anschließend aus den Parametern \hat{y}_{t+2} und $\hat{\sigma}_{AR}^2$. Bei allen anderen Methoden, die nachfolgenden drei Verfahren eingeschlossen, liefern die einperiodische und die mehrperiodische Wahrscheinlichkeitsvorhersage identische Resultate.

6.2.4. Bank-of-England-Vorhersage

Die Bank of England veröffentlicht alle drei Monate einen Quarterly Inflation Report. In diesem Bericht gibt sie unter anderem ihre Erwartungen an die Entwicklung der Inflation in England in Form einer Wahrscheinlichkeitsvorhersage an. Die Bank of England verwendet für ihre Vorhersage eine asymmetrische zweiteilige Normalverteilung. Diese wird durch die drei Parameter μ , σ_1 und σ_2 charakterisiert, wobei μ die Verteilung in zwei Hälften teilt, welche für sich gesehen jeweils einer Normalverteilung mit Erwartungswert μ und Varianz σ_1^2 beziehungsweise σ_2^2 entsprechen, jedoch durch Multiplikation mit einem konstanten Faktor skaliert wurden (zur Veranschaulichung siehe Abb. 6.2.1). Die Dichte einer asymmetrischen zweiteiligen Normalverteilung, welche wir auch für unsere

Vorhersage nutzen werden, ist durch nachfolgende Formel gegeben (Gneiting & Thorarinsdottir, 2010).

5. Vorhersage: Bank of England (BoE)

$$f_{BoE}(y) = \begin{cases} \left(\frac{2}{\pi}\right)^{1/2} (\sigma_1 + \sigma_2)^{-1} \exp\left(-\frac{(y-\mu)^2}{2\sigma_1^2}\right), & \text{falls } y \leq \mu \\ \left(\frac{2}{\pi}\right)^{1/2} (\sigma_1 + \sigma_2)^{-1} \exp\left(-\frac{(y-\mu)^2}{2\sigma_2^2}\right), & \text{falls } y \geq \mu \end{cases} \quad (6.2.12)$$

Für $\sigma_1 = \sigma_2$ entspricht diese Funktion der Dichte einer Normalverteilung. Sollte σ_1 jedoch größer sein als σ_2 , so erhält man eine linksschiefe Glockenform, für σ_1 kleiner als σ_2 entsprechend eine rechtsschiefe Glockenform. Die Verteilungsfunktion, welche wir für die Berechnung des CRPS benötigen, lautet (Gneiting & Thorarinsdottir, 2010)

$$F_{BoE}(y) = \begin{cases} \frac{2\sigma_1}{\sigma_1 + \sigma_2} \Phi\left(\frac{y-\mu}{\sigma_1}\right), & \text{falls } y \leq \mu \\ \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2} + \frac{2\sigma_2}{\sigma_1 + \sigma_2} \Phi\left(\frac{y-\mu}{\sigma_2}\right), & \text{falls } y \geq \mu \end{cases}. \quad (6.2.13)$$

Um für unsere Zielgröße y_{t+1} eine spezifische Vorhersage erzeugen zu können, müssen die drei Parameter unseres Modells geeignet geschätzt werden. Dies tun wir, indem wir auf die m vorangegangenen Ausprägungen unserer Datenreihe zurückgreifen und die Parameter durch Maximierung der Log-Likelihood-Funktion bestimmen.

Maximum-Likelihood-Schätzer

Der *Maximum-Likelihood-Schätzer* ist eine in der Statistik genutzte Methode, um Parameter einer Verteilung zu schätzen. Sei im Folgenden $Y : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable, deren Verteilung die Dichte f_θ besitzt, wobei θ einen oder mehrere unbekannte Parameter repräsentiert. Dann ordnet die Dichte f_θ jeder Ausprägung y von Y einen Wert $f_\theta(y)$ zu.

$$y \mapsto f_\theta(y) \quad (6.2.14)$$

Wenn die Ausprägung y der Zufallsvariable Y bekannt ist, dann lässt sich für dieses feste y die Likelihood-Funktion

$$L : \Theta \rightarrow \mathbb{R}, \quad \theta \mapsto f_\theta(y) \quad (6.2.15)$$

definieren, wobei Θ dem Raum aller möglichen Parameterwerte entspricht. Der Maximum-Likelihood-Schätzer befolgt nun den Ansatz, dass er die Parameter θ so wählt, dass $f_\theta(y)$ maximal ist.

$$ML : \mathbb{R} \rightarrow \Theta, \quad y \mapsto \operatorname{argmax}_\theta f_\theta(y) \quad (6.2.16)$$

Dieses Vorgehen funktioniert auch im mehrdimensionalen Fall, wobei dann die Parameter θ durch Maximierung der gemeinsamen Dichte f_θ^n der Zufallsvariable $Y = (Y_1, \dots, Y_n)$ ermittelt werden. Falls es sich bei den Zufallsvariablen Y_1, \dots, Y_n um unabhängige, identisch verteilte Zufallsvariablen handelt, die jeweils die Dichte f_θ besitzen, dann erhält man für die Ausprägungen y_1, \dots, y_n die Formel

$$\hat{\theta}_{ML} = \operatorname{argmax}_\theta \prod_{i=1}^n f_\theta(y_i) \quad (6.2.17)$$

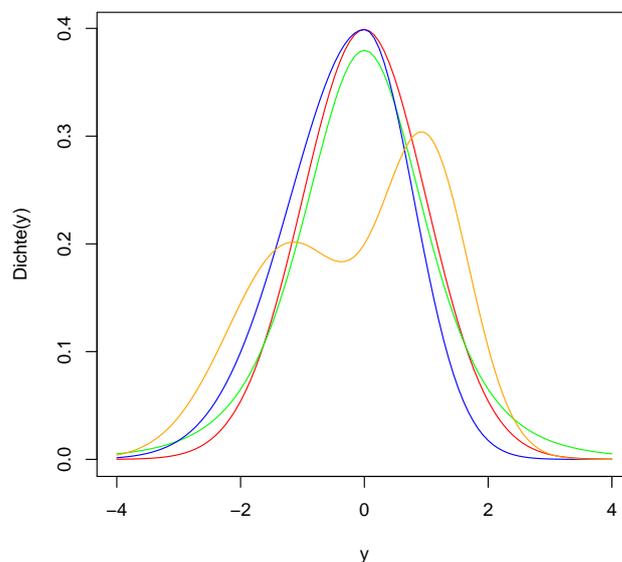


Abbildung 6.2.1.: Veranschaulichung der Dichtefunktionen: Normalverteilung (rot, mit $\mu = 0, \sigma^2 = 1$), t-Verteilung (grün, mit Freiheitsgrad=5), BoE (blau, mit $\mu = 0, \sigma_1 = 1.2, \sigma_2 = 0.8$), GMM (orange, mit $\alpha = 0.5, \mu_1 = -1.2, \mu_2 = 1, \sigma_1 = 1, \sigma_2 = 0.7$).

für den Maximum-Likelihood-Schätzer. Um aus technischer Sicht das Optimierungsproblem besser lösen zu können, wendet man auf die Likelihood-Funktion zusätzlich die streng monoton wachsende Logarithmus-Funktion an, sodass man bei unabhängig, identisch verteilten Zufallsvariablen die *Log-Likelihood-Funktion*

$$\log(L(\theta)) = \sum_{i=1}^n \log(f_{\theta}(y_i)) \quad (6.2.18)$$

maximiert (siehe z.B. Fahrmeir et al. (2007)).

Bei der Erzeugung der Wahrscheinlichkeitsvorhersagen verwenden wir für die Bestimmung der Log-Likelihood-Funktion jeweils die m letzten Ausprägungen unserer Zeitreihe und unterstellen die Unabhängigkeit der Zeitreihenpunkte. Die anschließende Optimierung und somit die Bestimmung der Parameter erfolgt jeweils mit der in R zur Verfügung stehenden Methode *optim*.

6.2.5. Gaussian-Mixture-Model

Gneiting & Thorarinsdottir (2010) schlagen für die Wahrscheinlichkeitsvorhersage ein *Gaussian-Mixture-Model* vor. Bei den erzeugten Vorhersagen handelt es sich um Ver-

teilungen, die durch die gewichtete Kombination zweier Normalverteilungen entstehen. Die zwei Normalverteilungen werden durch die Erwartungswerte μ_1 beziehungsweise μ_2 und die Varianzen σ_1^2 beziehungsweise σ_2^2 festgelegt und mit den Faktoren α beziehungsweise $(1 - \alpha)$, mit $\alpha \in [0, 1]$, gewichtet. Wenn man die Verteilungsfunktion der Standardnormalverteilung $N(0, 1)$ mit Φ bezeichnet, dann ergibt sich für das Gaussian-Mixture-Model die Verteilungsfunktion

$$F_{GMM}(y) = \alpha \Phi\left(\frac{y - \mu_1}{\sigma_1}\right) + (1 - \alpha) \Phi\left(\frac{y - \mu_2}{\sigma_2}\right). \quad (6.2.19)$$

Die entsprechende Dichte lautet

6. Vorhersage: Gaussian-Mixture-Model (GMM)

$$f_{GMM}(y) = \frac{\alpha}{\sigma_1} \phi\left(\frac{y - \mu_1}{\sigma_1}\right) + \frac{(1 - \alpha)}{\sigma_2} \phi\left(\frac{y - \mu_2}{\sigma_2}\right), \quad (6.2.20)$$

wobei ϕ der Dichte der Standardnormalverteilung entspricht. Für unsere Vorhersage müssen wir die fünf Parameter α , μ_1 , μ_2 , σ_1 und σ_2 geeignet schätzen. Dies tun wir, indem wir wie bei der Bank-of-England-Vorhersage die Log-Likelihood-Funktion ermitteln und den Maximum-Likelihood-Schätzer anwenden. Dies erfolgt ebenfalls unter Verwendung der letzten m historischen Datenpunkte und unter der Annahme der Unabhängigkeit der Zeitreihenpunkte.

6.2.6. t-Verteilung mit Maximum-Likelihood-Schätzung

Um eine Aussage darüber treffen zu können, ob und in welchen Fällen sich eine aufwendige Parameterschätzung gegenüber einfachen Schätzungen als vorteilhaft erweist, wollen wir mit der letzten Vorhersagemethode einen Ansatz verfolgen, der die Vorhersage einer einfachen Verteilung und die Maximum-Likelihood-Schätzung kombiniert. Dazu betrachten wir erneut die transformierte t-Verteilung $t(\mu, \sigma^2, df)$ mit den Parametern μ (Erwartungswert), σ^2 (Skalierungsfaktor) und df (Freiheitsgrade), vgl. Formel (6.2.9). Die Maximum-Likelihood-Schätzung erfolgt erneut auf Grundlage der letzten m Ausprägungen der Zeitreihe und es wird die Unabhängigkeit der Zeitreihenpunkte unterstellt. Wir erhalten schließlich für unsere Vorhersage die Dichte

7. Vorhersage: t_{MLS}

$$f(y) = \frac{1}{\hat{\sigma}_{MLS}} d_t\left(\frac{y - \hat{\mu}_{MLS}}{\hat{\sigma}_{MLS}}, df_{MLS}\right), \quad (6.2.21)$$

wobei $d_t(\cdot, df)$ der Dichte der t-Verteilung mit df Freiheitsgraden entspricht.

7. Berechnungen und Analysen

In diesem Kapitel wollen wir die makroökonomischen Daten (Inflationsrate, Wachstumsrate des Bruttoinlandsproduktes und Wachstumsrate der Arbeitslosenquote) analysieren und die vorgestellten Vorhersagemethoden bewerten. Alle drei Kennzahlen werden mit den gleichen Analysemethoden untersucht, sodass wir die Ergebnisse unserer Berechnungen gut vergleichen können. Dies ermöglicht es, Unterschiede in der Zeitreihenstruktur der Daten und deren Auswirkungen auf die Vorhersagequalität der Methoden ersichtlich zu machen. Die genaue Vorgehensweise, die wir bei der Analyse der historischen Daten verfolgen werden, wird im nächsten Abschnitt beschrieben. Anschließend werden die Resultate der Untersuchungen für jede Kennzahl in einem separaten Unterkapitel dargelegt.

7.1. Vorgehensweise bei der Analyse

Nachdem wir bereits die historischen Verläufe der Kennzahlen betrachtet haben, wollen wir uns zu Beginn der Analyse einen Überblick über die Verteilung und die Struktur der jeweiligen Datenreihe verschaffen. Dazu verwenden wir einen Kerndichteschätzer, der uns aus allen zur Verfügung stehenden Vergangenheitswerten eine Dichte der Datenpunkte ermittelt (für nähere Informationen zu Kerndichteschätzern siehe z.B. Fahrmeir et al. (2007)). Die Bestimmung der Dichte erfolgt mit der R-Funktion *density* und die berechnete Dichte wird in einer Abbildung dargestellt. Um ein Gefühl dafür zu bekommen, wie gut die in den Vorhersagemethoden unterstellten Verteilungen (Normalverteilung, transformierte t-Verteilung, Verteilung der Bank of England und Verteilung des Gaussian-Mixture-Models) sich dieser Dichte anpassen, haben wir auf Grundlage der gesamten Datenreihe die Dichtevorhersagen ermittelt und ebenfalls in diese Abbildung integriert. Auf das Abbilden der Dichten, welche durch die Vorhersagemethoden $N(y_t, \hat{\sigma}^2)$ und AR1 erzeugt werden, wurde verzichtet, da diese sehr stark von dem letzten Datenpunkt abhängen und deshalb im Vergleich zu den anderen Dichten leicht nach rechts oder links versetzt abgebildet wären.

Als weiteren Analyseschritt schauen wir uns die Autokorrelationen der jeweiligen Zeitreihe an, da wir auf diese Weise Erkenntnisse über zeitliche Abhängigkeiten in der Datenreihe gewinnen können. Die Ergebnisse der Autokorrelationsfunktion, welche für eine Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ definiert ist als (Schlittgen, 2012)

$$\text{Cor}(Y_{t+h}, Y_t) = \gamma_h, \quad h \in \mathbb{Z}, \quad (7.1.1)$$

lassen wir uns zur besseren Veranschaulichung grafisch darstellen.

Wie bereits in Abschnitt 6.2.1 angesprochen wurde, müssen wir, bevor wir mit der Erzeugung der Wahrscheinlichkeitsvorhersagen beginnen können, die Länge des training windows festlegen. Um diese Entscheidung zu treffen schauen wir uns die mittleren scores der Vorhersagemethoden in Abhängigkeit von der Länge des training windows an. Zur Sicherstellung der Vergleichbarkeit der scores wurden jeweils Vorhersagen für einen festen Zeitraum erzeugt, sodass sich Unterschiede lediglich aufgrund der Länge des training windows ergeben. Die Höhe der mittleren scores sollte uns ein Gefühl dafür geben können, wie sich die Qualität der Vorhersagen in Abhängigkeit vom training window verhält. Eine Veränderung der Verteilung unserer makroökonomischen Kennzahl im Zeitverlauf sollte sich bei der Wahl eines zu großen Datenbereiches für die Schätzung auf die Vorhersagequalität negativ auswirken. Wir können somit aus unserer Analyse entsprechende Rückschlüsse auf die Entwicklung der Verteilung ziehen und ein angemessenes training window bestimmen.

Im Gegensatz zu den bereinigten Daten des Bruttoinlandsproduktes und der Arbeitslosenquote liegen uns für die Analyse der Inflationsraten nur die nicht-saisonbereinigten Daten vor. Da die Saisoneffekte einen erheblichen Einfluss auf die Verteilung und die Struktur der Zeitreihe haben und wir unsere eigentliche Schätzung der Verteilung mit den saisonbereinigten Daten durchführen, betrachten wir bei den beschriebenen Voruntersuchungen im Falle der Inflationsraten ebenfalls die saisonbereinigten Daten. Bei den Berechnungen gehen wir dann wie folgt vor: Zuerst bestimmen wir, wie bereits in Abschnitt 6.2.1 beschrieben, für jede Vorhersage die Saisonanteile aus den bis dato bekannten Daten. Anschließend erfolgt die Wahrscheinlichkeitsvorhersage auf Grundlage der saisonbereinigten Daten und schließlich wird durch Addition der entsprechenden Saisonkomponente eine Vorhersage für die zukünftige, nicht-saisonbereinigte Inflationsrate ermittelt. Als Inputvariablen der Vorhersagemethoden sind somit die saisonbereinigten Daten relevant, weshalb wir diesen besondere Beachtung schenken.

Im folgenden Part unserer Untersuchungen analysieren wir die mittleren scores für Einschrittvorhersagen, welche für einen vorgegebenen Zeitraum erstellt wurden. Es werden die in Abschnitt 3.2 vorgestellten scoring rules *quadratic score*, *pseudospherical score* (mit $\alpha = 2$), *logarithmic score* und *continuous ranked probability score* (CRPS), sowie die Schwellenwert gewichtete scoring rule des CRPS aus Abschnitt 4.2 mit den Gewichtungsfunktionen aus Tabelle 4.2.1 verwendet, wobei der CRPS mithilfe der Brier-Darstellung (3.2.16) ermittelt wurde und die Parameter a und b der Gewichtungsfunktionen jeweils durch den Mittelwert bzw. die empirische Varianz der m letzten historischen Daten festgelegt wurden. Neben der Betrachtung der mittleren scores führen wir für die zwei Vorhersagemethoden mit dem jeweils höchsten bzw. zweihöchsten score pro scoring rule den Test auf gleiche Vorhersage Performance durch und interpretieren die Testergebnisse. Außerdem werten wir für alle Vorhersagemethoden die Schwellenwert und die Quantil Zerlegungen des CRPS aus.

Schließlich wollen wir noch analysieren wie sich die Vorhersage Performance verhält, wenn wir nicht nur eine, sondern zwei Perioden in die Zukunft schauen. Dazu wurden

mit jeder Vorhersagemethode Zweischrittvorhersagen für den gleichen Zeitraum erzeugt, den wir auch zur Bestimmung der mittleren scores bei den Einschrittvorhersagen verwendet haben. Wir untersuchen auch hierfür die mittleren scores und betrachten die Testergebnisse auf gleiche Vorhersage Performance.

Es sei noch zu erwähnen, dass wir bei unserer Analyse lediglich Dichtevorhersagen für die Wachstumsraten des harmonisierten Verbraucherpreisindex, des Bruttoinlandsproduktes und der Arbeitslosenquote analysieren. Durch entsprechendes Lokalisieren und Skalieren dieser Dichten könnte man aber auch Wahrscheinlichkeitsvorhersagen für den harmonisierten Verbraucherpreisindex, das Bruttoinlandsprodukt und die Arbeitslosenquote erzeugen und diese untersuchen.

7.2. Analyse der Inflationsrate

Um einen ersten Eindruck davon zu bekommen, welche Vorhersagemethode für die Prognose der monatlichen Inflationsraten geeignet sein könnte, schauen wir uns in Abbildung 7.2.1 die durch einen Kerndichteschätzer ermittelte Dichte der saisonbereinigten Inflationsraten von 1996 bis 2012 an. Die Dichte besitzt annähernd die für eine Normalverteilung typische Glockenform, sie ordnet jedoch den Werten an den Rändern eine etwas höhere Wahrscheinlichkeit zu. Diese ergeben sich aufgrund von Ausreißern und werden als *fat tails* bezeichnet. Die durch unsere Vorhersagemethoden erzeugten Dichten, welche ebenfalls in Abbildung 7.2.1 eingezeichnet wurden und als training window den kompletten Datensatz der saisonbereinigten Inflationsdaten (1996-2012) berücksichtigen, besitzen alle eine fast identische Form. Leichte Unterschiede ergeben sich an den Rändern, hier ordnen die Vorhersagen mit t-Verteilung ($t(\hat{\mu}, \hat{\sigma}^2)$) und t_{MLS} den Ausreißern eine vergleichsweise höhere Wahrscheinlichkeit zu. Man könnte deshalb vermuten, dass die mittleren scores der Vorhersagemethoden nahe bei einander liegen, sich aber Vorteile für die Vorhersagemethoden ergeben, die fat tails besser berücksichtigen können. Da die Verteilung der Inflationsraten eine typische Glockenform aufweist, werden wir bei der Vorhersagemethode Gaussian-Mixture-Model (GMM) die Parameter μ_1 und μ_2 , welche die Erwartungswerte der zwei Normalverteilungen festlegen, gleichsetzen ($\mu_1 = \mu_2$). Dadurch hängt die Verteilung nur noch von vier Parametern ab und wir können bei der Maximum-Likelihood-Schätzung bessere Ergebnisse erwarten.

Als nächstes wollen wir den Verlauf der Inflationsrate auf mögliche zeitliche Abhängigkeiten hin untersuchen. Die Autokorrelationen für die saisonbereinigten Daten, welche in Abbildung 7.2.2 zu sehen sind, weisen nur sehr geringe Werte auf. Es lässt sich schlussfolgern, dass die saisonbereinigten Inflationsraten keiner oder einer nur sehr geringen zeitlichen Abhängigkeit unterliegen. Da ein Zeitreihenmodell, wie wir es in der Vorhersagemethode AR1 verwenden, seine Vorteile nur beim Vorhandensein einer Zeitreihenstruktur ausnutzen kann, dürften sich durch dieses Modell keine besseren Ergebnisse erzielen lassen als durch die einfache Vorhersagemethode $N(\hat{\mu}, \hat{\sigma}^2)$. Vielmehr könnten sich innerhalb der Berechnung durch den Versuch, eine nicht vorhandene Zeitreihenstruktur zu berücksichtigen, Nachteile für die AR1-Vorhersage ergeben. Noch schlechtere Resultate dürften wir mit unserer Vorhersagemethode $N(y_t, \hat{\sigma}^2)$ erhalten, da diese eine besonders starke zeitliche Abhängigkeit unterstellt.

Die angesprochene geringe Autokorrelation innerhalb der saisonbereinigten Inflationsraten deutet außerdem darauf hin, dass sich die Verteilung der saisonbereinigten Inflationsraten im Zeitverlauf nicht sonderlich schnell verändert. Um diese Vermutung genauer zu hinterfragen, schauen wir uns die mittleren scores der Vorhersagemethoden für die Inflationsraten von 2002 bis 2012 in Abhängigkeit von der Länge des training windows an (siehe Abbildung 7.2.3). Bei allen vier scoring rules (quadratic score, pseudospherical score, logarithmic score und CRPS) lässt sich erkennen, dass die mittleren scores zunächst mit der Länge des training windows ansteigen. Besonders deutlich wird die Steigerung der scores bei den komplexeren Vorhersagemethoden AR1, BoE und GMM,

da diese für eine gute Parameterschätzung auf eine größere Datenbasis angewiesen sind. Ab einer Länge des training windows von etwa 54 Datenpunkten erreichen die scores ein Niveau, auf dem sie sich einpendeln und sich nicht mehr großartig verändern. Aufgrund dieser Informationen nehmen wir an, dass die Verteilung der Inflationsraten über längere Zeit relativ konstant bleibt, und verwenden für die Schätzung der Parameter in unseren Vorhersagemethoden jeweils die 54 letzten Inflationsraten.

Nachdem wir uns einen Überblick über die Entwicklung der Inflationsraten verschafft haben und bereits einige Vermutungen über die Vorhersagequalität unserer Methoden geäußert haben, wollen wir diese Vermutungen mithilfe der scoring rules überprüfen. Zunächst werfen wir einen Blick auf Vorhersagen mit einem Vorhersagehorizont von einem Monat und analysieren deren mittlere scores für die Inflationsraten von Januar 2001 bis Dezember 2012. Wie aus Tabelle 7.2.1, in der die mittleren scores nach Vorhersagemethode und scoring rule aufgelistet sind, ersichtlich ist, schneidet die Vorhersagemethode $t(\hat{\mu}, \hat{\sigma}^2)$ beim quadratic score, beim logarithmic score und auch beim CRPS am besten ab. Dies bekräftigt unsere Vermutung, dass eine Verteilung mit fat tails extreme Inflationsraten besser berücksichtigt. Auffällig ist, dass die Vorhersagemethode t_{MLS} , welche ebenfalls eine transformierte t-Verteilung vorhersagt und bei der Bestimmung der Parameter mehr Freiheit besitzt, nicht mit der einfacheren Vorhersagemethode $t(\hat{\mu}, \hat{\sigma}^2)$ mithalten kann. Dies macht deutlich, dass die Nutzung des Mittelwertes als Erwartungswert und der empirischen Varianz als Skalierungsfaktor offensichtlich schon sehr gute Prognosen liefert. Die Vorhersagemethode BoE erzielt den höchsten Wert bei Verwendung des pseudospherical scores und liegt beim quadratic score und beim CRPS auf Rang zwei. Es lässt sich deshalb vermuten, dass die BoE-Dichte mit ihrer Schiefe punkten kann, indem sie den etwas häufiger eintretenden Ausreißern im negativen Bereich einen größeren tail zuweist und im rechten Part der Dichte die Wahrscheinlichkeit schneller abfallen lässt. Die Vorhersagemethode $N(y_t, \hat{\sigma}^2)$ schneidet bei allen scores mit Abstand am schlechtesten ab und die AR1-Vorhersage erhält etwas niedrigere scores als die einfache Vorhersagemethode $N(\hat{\mu}, \hat{\sigma}^2)$. Die schlechte Bewertung der beiden Methoden $N(y_t, \hat{\sigma}^2)$ und AR1 hatten wir bereits aufgrund der geringen Autokorrelation erwartet. Die scoring rules bestrafen somit den Versuch, eine nicht vorhandene zeitliche Abhängigkeit in der Modellierung ausnutzen zu wollen. Schließlich sei noch zu erwähnen, dass das Gaussian-Mixture-Modell seine Gestaltungsfreiräume durch die Anpassung von insgesamt vier Parametern nicht ausreichend nutzen kann. Dies könnte eventuell an Ungenauigkeiten bei der Schätzung liegen. Allerdings deuten die Ergebnisse unserer Tests auf gleiche Vorhersage Performance, welche in Tabelle 7.2.2 zu sehen sind, auch darauf hin, dass keine der Vorhersagemethoden nachweislich eine bessere Vorhersage Performance besitzt. Bei den vier Tests konnte jeweils die Nullhypothese, dass die Vorhersagemethoden mit den beiden besten scores eine gleiche Vorhersagequalität vorweisen, nicht verworfen werden.

Auch bei den gewichteten Versionen des CRPS bekommt die Vorhersagemethode $t(\hat{\mu}, \hat{\sigma}^2)$ vergleichsweise gute mittlere scores (siehe Tabelle 7.2.3). Es fällt jedoch auf, dass die AR1-Vorhersage bei der Gewichtung auf die Ränder und auf den rechten Rand minimal bessere Ergebnisse erzielt. Dies könnte damit zu tun haben, dass es besonders im positi-

ven Bereich weniger Ausreißer gibt und somit die t-Verteilung die Wahrscheinlichkeiten am rechten Rand zu hoch einschätzt. Außerdem kann es sein, dass gerade in wirtschaftlichen Phasen, in denen es zu Ausreißern kommt, leichte zeitliche Abhängigkeiten in der Entwicklung der Inflationsrate vorliegen, sodass die AR1 Vorhersage hier punkten kann. Die Tests auf gleiche Vorhersage Performance weisen allerdings erneut p-Werte auf, die allesamt über 5% liegen (siehe Tabelle 7.2.4). Die Nullhypothese kann folglich in keinem der vier Tests verworfen werden.

Sowohl die Schwellenwert Zerlegung, als auch die Quantil Zerlegung des continuous ranked probability scores (CRPS) bestätigen unsere bisherigen Beobachtungen (siehe Abbildungen 7.2.4 und 7.2.5). Die Vorhersagemethode $N(y_t, \hat{\sigma}^2)$ erhält mit Abstand die schlechtesten scores, alle anderen Vorhersagemethoden zeigen jedoch bei den beiden Zerlegungen ein sehr ähnliches Bild auf.

Bei dem Übergang von einer Einschrittvorhersage zu einer Zweischrittvorhersage können wir im Bezug auf die mittleren scores keine großen Unterschiede erkennen (siehe Tabelle 7.2.5). Die Vorhersagemethode $t(\hat{\mu}, \hat{\sigma}^2)$ kann erneut beim quadratic score, beim logarithmic score und beim CRPS mit den höchsten mittleren scores glänzen. Die Bank-of-England-Vorhersage liegt beim pseudospherical score vorne und erhält die zweitbeste Bewertung beim quadratic score und beim CRPS. Außerdem liegen die scores der Vorhersagemethode $N(y_t, \hat{\sigma}^2)$ weiterhin weit unterhalb der Bewertungen, die von den anderen Methoden erzielt werden. Schließlich kann auch keiner der Tests die Nullhypothese auf gleiche Vorhersage Performance ablehnen, wenn man das Signifikanzniveau bei 5% ansetzt (siehe Tabelle 7.2.6).

Als Resultat unserer Analysen können wir festhalten, dass sich zur Vorhersage der Inflationsraten keine Vorhersagemethode besonders auszeichnet hat. Die Vorhersagemethode $t(\hat{\mu}, \hat{\sigma}^2)$ erhielt zwar in den meisten Fällen die höchsten mittleren scores, aber sie konnte sich nicht deutlich von den anderen Verfahren absetzen. Da nur eine sehr geringe zeitliche Abhängigkeit in der Zeitreihe vorliegt, sind Zeitreihenmodelle nur bedingt zu empfehlen. Vielmehr ist für eine gute Wahrscheinlichkeitsvorhersage die Wahl einer angemessenen Verteilung sehr wichtig. Diese sollte insbesondere am linken Rand eine etwas höhere Wahrscheinlichkeit vorweisen, um Ausreißer im negativen Bereich zu berücksichtigen.

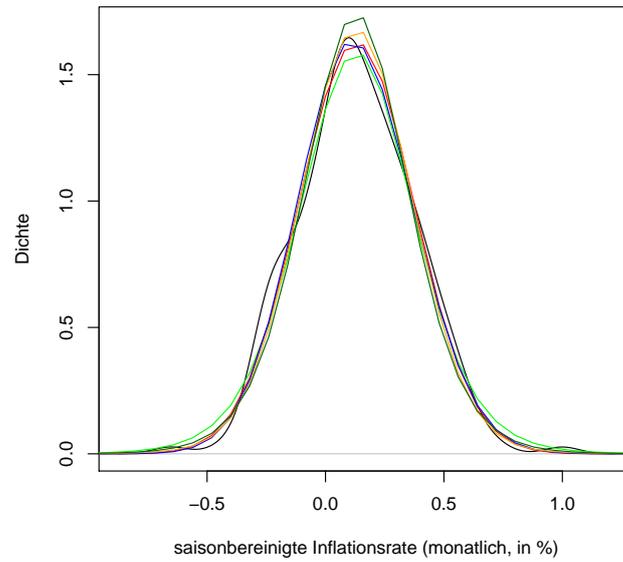


Abbildung 7.2.1.: Abbildung der mit einem Kerndichteschätzer ermittelten Dichte der saisonbereinigten monatlichen Inflationsraten (schwarz) und die geschätzten Dichtevorhersagen auf Grundlage der monatlichen Inflationsraten aus den Jahren 1996 bis 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

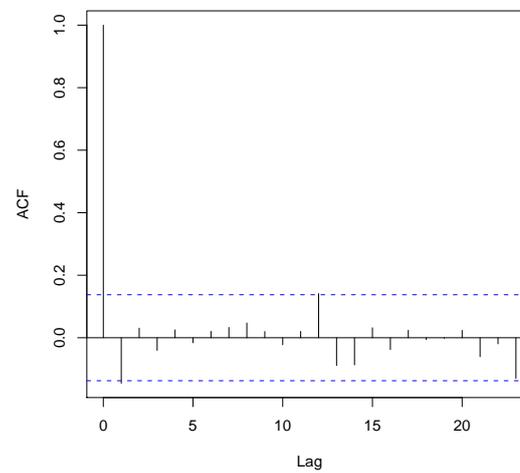


Abbildung 7.2.2.: Abbildung der Autokorrelationen für die saisonbereinigten monatlichen Inflationsraten in Abhängigkeit von den Lags.

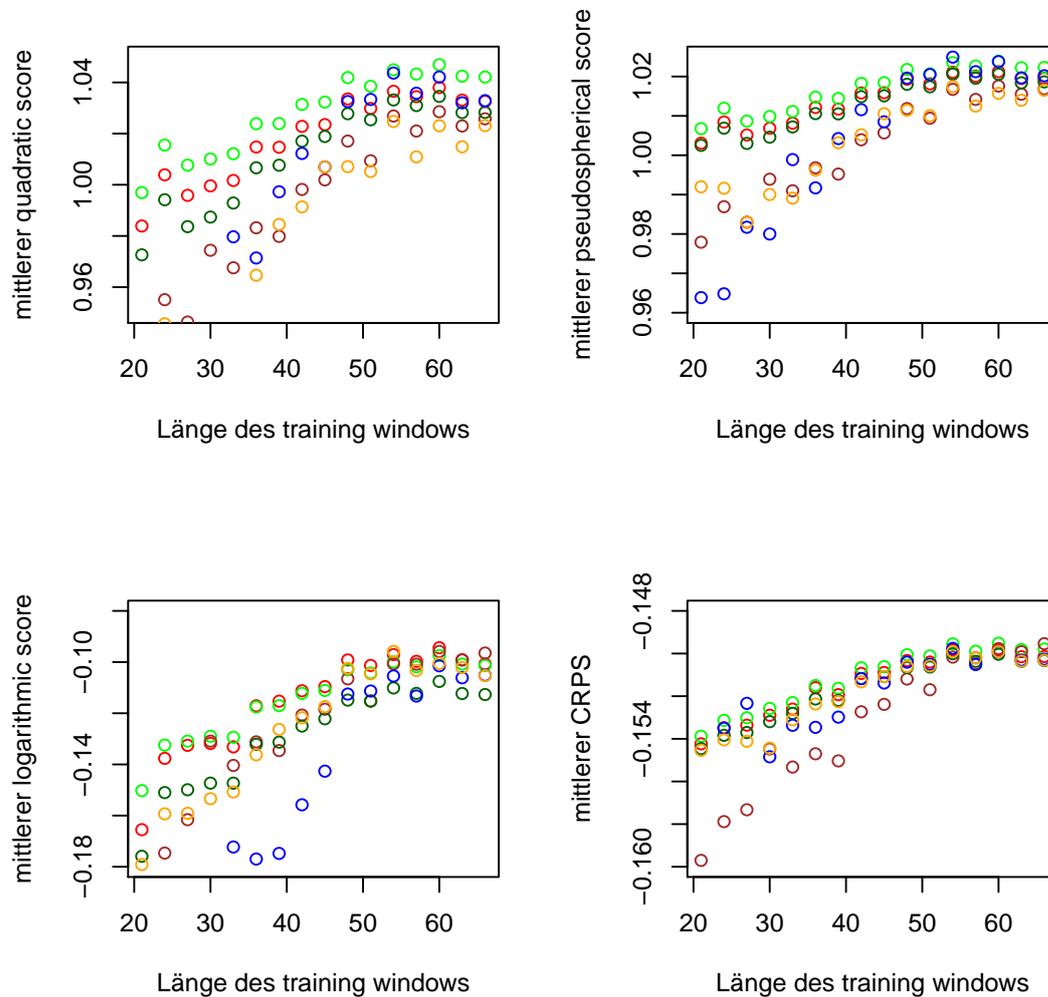


Abbildung 7.2.3.: Die vier Abbildungen zeigen für jede scoring rule separat die Entwicklung der mittleren scores der Vorhersagemethoden für die Vorhersagen der monatlichen Inflationsraten von 2002 bis 2012 in Abhängigkeit von der Länge des training windows ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

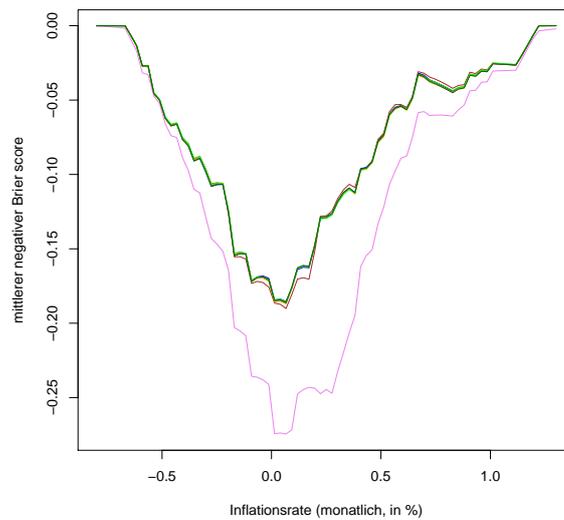


Abbildung 7.2.4.: Schwellenwert Zerlegung des CRPS der verschiedenen Vorhersagen unter Verwendung der monatlichen Inflationsraten von 2001 bis 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

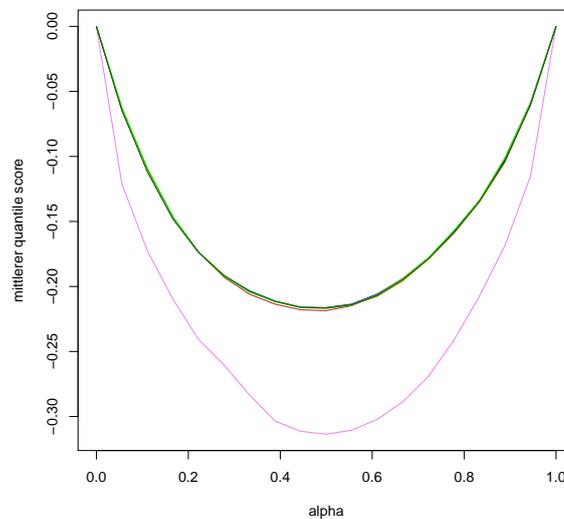


Abbildung 7.2.5.: Quantil Zerlegungen des CRPS der verschiedenen Vorhersagen unter Verwendung der monatlichen Inflationsraten von 2001 bis 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

Scoring rule	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
quadratic	1.0081	0.6125	1.0172	0.9946	1.0145	0.9845	0.9985
pseudosph.	1.0088	0.8256	1.0111	1.0029	1.0126	1.0015	1.0064
logarithmic	-0.1644	-0.9394	-0.1438	-0.1642	-0.2212	-0.1502	-0.1644
CRPS	-0.1555	-0.2294	-0.1551	-0.1558	-0.1555	-0.1560	-0.1558

Tabelle 7.2.1.: Mittlere scores nach scoring rule und Vorhersagemethode für die Inflationsraten von 2001 bis 2012 bei einer Einschrittvorhersage. Die Vorhersage erfolgte auf Grundlage der monatlichen Inflationsraten von Februar 1996 bis Dezember 2012. Hierbei wurden in jedem Vorhersageschritt die Saisonkomponenten aus allen bis zu diesem Zeitpunkt bekannten Inflationsraten neu berechnet. Bei den weiteren Schätzungen mit den saisonbereinigten Werten wurde ein fortlaufendes Datenfenster von 54 Datenpunkten verwendet.

Scoring rule	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
quadratic	$t(\hat{\mu}, \hat{\sigma}^2)$	BoE	0.1910	0.424
pseudospherical	BoE	$t(\hat{\mu}, \hat{\sigma}^2)$	0.2416	0.405
logarithmic	$t(\hat{\mu}, \hat{\sigma}^2)$	GMM	0.5438	0.293
CRPS	$t(\hat{\mu}, \hat{\sigma}^2)$	BoE	0.7138	0.238

Tabelle 7.2.2.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach scoring rule bei einer Einschrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Inflationsraten von 2001 bis 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der monatlichen Inflationsraten von Februar 1996 bis Dezember 2012. Hierbei wurden in jedem Vorhersageschritt die Saisonkomponenten aus allen bis zu diesem Zeitpunkt bekannten Inflationsraten neu berechnet. Bei den weiteren Schätzungen mit den saisonbereinigten Werten wurde ein fortlaufendes Datenfenster von 54 Datenpunkten verwendet.

Gewichtung	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
Zentrum	-0.1165	-0.1793	-0.1164	-0.1176	-0.1165	-0.1168	-0.1166
Ränder	-0.0485	-0.0674	-0.0482	-0.0479	-0.0485	-0.0488	-0.0488
rechts	-0.0750	-0.1180	-0.0749	-0.0747	-0.0749	-0.0754	-0.0752
links	-0.0805	-0.1141	-0.0802	-0.0811	-0.0806	-0.0806	-0.0807

Tabelle 7.2.3.: Mittlere scores der gewichteten Version des CRPS nach Gewichtung und Vorhersagemethode für die Inflationsraten von 2001 bis 2012 bei einer Einschrittvorhersage. Die Vorhersage erfolgte auf Grundlage der monatlichen Inflationsraten von Februar 1996 bis Dezember 2012. Hierbei wurden in jedem Vorhersageschritt die Saisonkomponenten aus allen bis zu diesem Zeitpunkt bekannten Inflationsraten neu berechnet. Bei den weiteren Schätzungen mit den saisonbereinigten Werten wurde ein fortlaufendes Datenfenster von 54 Datenpunkten verwendet.

Gewichtung	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
Zentrum	$t(\hat{\mu}, \hat{\sigma}^2)$	BoE	0.2555	0.399
Ränder	AR1	$t(\hat{\mu}, \hat{\sigma}^2)$	0.4356	0.332
rechts	AR1	$t(\hat{\mu}, \hat{\sigma}^2)$	0.1566	0.438
links	$t(\hat{\mu}, \hat{\sigma}^2)$	$N(\hat{\mu}, \hat{\sigma}^2)$	1.3464	0.089

Tabelle 7.2.4.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach Gewichtung des CRPS bei einer Einschrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Inflationsraten von 2001 bis 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der monatlichen Inflationsraten von Februar 1996 bis Dezember 2012. Hierbei wurden in jedem Vorhersageschritt die Saisonkomponenten aus allen bis zu diesem Zeitpunkt bekannten Inflationsraten neu berechnet. Bei den weiteren Schätzungen mit den saisonbereinigten Werten wurde ein fortlaufendes Datenfenster von 54 Datenpunkten verwendet.

Scoring rule	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
quadratic	1.0035	0.6645	1.0132	0.9959	1.0105	0.9703	0.9941
pseudosph.	1.0076	0.8510	1.0101	1.0047	1.0119	0.9965	1.0054
logarithmic	-0.1796	-0.6396	-0.1504	-0.1832	-0.2560	-0.1904	-0.1715
CRPS	-0.1557	-0.2075	-0.1553	-0.1560	-0.1557	-0.1564	-0.2320

Tabelle 7.2.5.: Mittlere scores nach scoring rule und Vorhersagemethode für die Inflationsraten von 2001 bis 2012 bei einer Zweischrittvorhersage. Die Vorhersage erfolgte auf Grundlage der monatlichen Inflationsraten von Februar 1996 bis Dezember 2012. Hierbei wurden in jedem Vorhersageschritt die Saisonkomponenten aus allen bis zu diesem Zeitpunkt bekannten Inflationsraten neu berechnet. Bei den weiteren Schätzungen mit den saisonbereinigten Werten wurde ein fortlaufendes Datenfenster von 54 Datenpunkten verwendet.

Scoring rule	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
quadratic	$t(\hat{\mu}, \hat{\sigma}^2)$	BoE	0.2051	0.419
pseudospherical	BoE	$t(\hat{\mu}, \hat{\sigma}^2)$	0.3439	0.365
logarithmic	$t(\hat{\mu}, \hat{\sigma}^2)$	t_{MLS}	1.5223	0.064
CRPS	$t(\hat{\mu}, \hat{\sigma}^2)$	BoE	0.6715	0.251

Tabelle 7.2.6.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach scoring rule bei einer Zweischrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Inflationsraten von 2001 bis 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der monatlichen Inflationsraten von Februar 1996 bis Dezember 2012. Hierbei wurden in jedem Vorhersageschritt die Saisonkomponenten aus allen bis zu diesem Zeitpunkt bekannten Inflationsraten neu berechnet. Bei den weiteren Schätzungen mit den saisonbereinigten Werten wurde ein fortlaufendes Datenfenster von 54 Datenpunkten verwendet.

7.3. Analyse des Bruttoinlandsproduktes

Aufgrund der Finanzkrise ging das Bruttoinlandsprodukt der Bundesrepublik Deutschland im 4. Quartal 2008 um 2.0% und im 1. Quartal 2009 sogar im 4.1% zurück. Diese starken Rückgänge der Wirtschaftsleistung spiegeln sich auch in der durch einen Kerndichteschätzer erzeugten Dichte der Wachstumsraten des Bruttoinlandsproduktes wider, welche auf Grundlage der Wachstumsraten aus Q2 1991 bis Q4 2012 ermittelt wurde (siehe Abbildung 7.3.1). Abgesehen von den deutlich sichtbaren Ausreißern im negativen Bereich zeigt die Dichte eine Form, die im Zentrum etwas schmaler und spitzer verläuft, als die typische Glockenform der Normalverteilung. Anhand der ebenfalls in Abbildung 7.3.1 eingezeichneten Dichtevorhersagen lässt sich erkennen, dass das Gaussian-Mixture-Model und die Vorhersagemethode t_{MLS} dem schmäleren Verlauf im Zentrum am ehesten Rechnung tragen können. Hierbei könnte sich insbesondere das Gaussian-Mixture-Model zur Vorhersage eignen, da es zwei Dichten kombinieren kann, eine mit einer geringen Varianz zur Abbildung des Zentrums und eine Dichte mit entsprechend höherer Varianz, um die fat tails abzubilden. Da für die eingezeichneten Dichtevorhersagen die gesamte Datenhistorie (Q2 1991 bis Q4 2012) verwendet wurde, wird bei den späteren Vorhersagen fraglich sein, ob die Vorhersage des Gaussian-Mixture-Models auch bei einer kleineren, für die Schätzung der Parameter zur Verfügung stehenden Datenmenge ausreichend gut sein wird. Um bessere Ergebnisse bei der Parameterschätzung zu erhalten, reduzieren wir die Parameteranzahl beim Gaussian-Mixture-Model, indem wir wie bei der Analyse der Inflationsraten die Parameter μ_1 und μ_2 , welche die Erwartungswerte der zwei Normalverteilungen festlegen, gleichsetzen ($\mu_1 = \mu_2$). Die Wahrscheinlichkeitsvorhersagen der einfachen Vorhersagemethoden $N(\hat{\mu}, \hat{\sigma}^2)$ und $t(\hat{\mu}, \hat{\sigma}^2)$ passen sich nicht so gut an die Kerndichteschätzung an, da die Ausreißer im negativen Bereich die empirische Varianz stark beeinflussen und die Dichten deshalb im Zentrum zu breit sind.

In Abbildung 7.3.2 werden die Autokorrelationen der Wachstumsraten des Bruttoinlandsproduktes dargestellt. Hier lässt sich eine leichte Korrelation mit dem ersten Lag erkennen. Dies dürfte zur Folge haben, dass unser Zeitreihenmodell AR1, welches bei der Vorhersage der Inflationsrate seine Vorteile nicht ausspielen konnte, nun etwas bessere Vorhersagen ermitteln sollte als das einfachere Verfahren $N(\hat{\mu}, \hat{\sigma}^2)$. Da die Zeitreihe keine besonders starken zeitlichen Abhängigkeiten vorweist, erwarten wir von dem Vorhersagemodell $N(y_t, \hat{\sigma}^2)$ erneut relativ schlechte Resultate.

Als Inputvariablen für unsere Vorhersagemethoden liegen uns nur die quartalsweisen Wachstumsraten des BIP vor. Es wird deshalb von entscheidender Rolle sein, wie stabil die Verteilung der Wachstumsraten im Zeitverlauf ist, da wir unser training window daran ausrichten müssen. Die Analyse der mittleren scores für die Wachstumsraten von Q3 2004 bis Q4 2012 in Abhängigkeit von der Länge des training windows zeigt, dass die mittleren scores beim quadratic score und beim pseudospherical score zunächst stark ansteigen, der Anstieg aber ab der Länge des training windows von etwa 32 Datenpunkten flacher verläuft (siehe Abbildung 7.3.3). Beim CRPS bleiben die mittleren scores ab einer Vorhersage mit ungefähr 28 Wachstumsraten auf gleichem Niveau und beim logarithmic

score verschlechtern sich die scores je nach Vorhersagemethode sogar ab einer Länge des training windows von etwa 32 Datenpunkten. Es ist davon auszugehen, dass sich die Verteilung des Bruttoinlandsproduktes im Zeitverlauf verändert. Wir ziehen daraus die Konsequenz, dass unser training window nicht zu groß gewählt werden sollte, und entscheiden uns für eine Länge von 32 Datenpunkten, was einem Zeitraum von 8 Jahren entspricht.

Für die weitere Analyse erzeugen wir mit jeder Vorhersagemethode Einschrittvorhersagen für die quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012. Die mittleren scores, welche sich für diese Vorhersagen ergeben, sind in Tabelle 7.3.1 zu sehen. Das Gaussian-Mixture-Model erhält beim quadratic score, beim pseudospherical score und beim logarithmic score jeweils den vergleichsweise höchsten Wert, gefolgt von der Vorhersagemethode t_{MLS} mit dem jeweils zweitbesten Ergebnis. Dieses Resultat bestätigt unsere vorangegangene Vermutung, dass diese beiden Verfahren dem schmalen Verlauf im Zentrum der in Abbildung 7.3.1 dargestellten Dichte der Wachstumsraten am besten Rechnung tragen können und zusätzlich die Ausreißer durch höhere Wahrscheinlichkeiten an den Rändern berücksichtigen. Bei der Bewertung mit dem CRPS erzielt die AR1-Vorhersage jedoch ein minimal besseres Ergebnis als das Gaussian-Mixture-Model, was auf die leichte Autokorrelation der Wachstumsraten zurückzuführen sein könnte. Die einfachen Vorhersagemethoden $N(\hat{\mu}, \hat{\sigma}^2)$, $N(y_t, \hat{\sigma}^2)$ und $t(\hat{\mu}, \hat{\sigma}^2)$, sowie die Bank-of-England-Vorhersage erhalten bei allen vier scoring rules etwas schlechtere mittlere Werte.

Die Testergebnisse auf gleiche Vorhersage Performance zeigen, dass bei einer Bewertung mit dem quadratic score oder mit dem pseudospherical score das Gaussian-Mixture-Model bessere Vorhersagen erzeugt. Beide Tests lehnen die Nullhypothese bei zugrunde legen eines Signifikanzniveaus von 5% ab (siehe Tabelle 7.3.2). Wählt man allerdings ein Signifikanzniveau von nur 2.5%, dann erfolgt keine Ablehnung. Bei Verwendung des logarithmic scores oder des CRPS kann die Nullhypothese, dass die zwei Vorhersagemethoden mit den besten scores gleiche Vorhersage Performance besitzen, nicht verworfen werden.

Bei der Bewertung der Vorhersagen mit der gewichteten Version des CRPS schneidet das Gaussian-Mixture-Model sowohl bei Gewichtung auf beide Ränder, als auch bei Gewichtung auf den linken Rand am besten ab (siehe Tabelle 7.3.4). Die t_{MLS} -Vorhersage folgt jeweils an zweiter Stelle. Anders sieht es bei der Betonung des Zentrums und des rechten Randes aus. Bei ersterem erzielt die Vorhersagemethode $N(y_t, \hat{\sigma}^2)$ das beste Ergebnis, gefolgt von der AR1-Vorhersage, bei letzterem liegt die Vorhersage mit dem autoregressiven Modell AR1 vor dem Gaussian-Mixture-Model. Diese Resultate lassen einerseits darauf schließen, dass das Gaussian-Mixture-Model und die Vorhersagemethode t_{MLS} die Ausreißer im negativen Bereich aufgrund von fat tails besser berücksichtigen, andererseits zeigen sie aber auch, dass Zeitreihenmodelle bei der Schätzung des Zentrum und des rechten Randes der Dichte ihre Vorteile haben. Die Auswertung der entsprechenden Tests offenbart, dass nur bei der Gewichtung auf die Ränder die Nullhypothese abgelehnt

werden kann, und dies auch nur bei einem Signifikanzniveau von 5% (p-Wert = 4.2%). Bei den anderen Tests kann die Nullhypothese jeweils nicht abgelehnt werden.

Die Schwellenwert Zerlegung des CRPS, welche in Abbildung 7.3.4 zu sehen ist, bestätigt die leichten Vorteile der Vorhersagemethode $N(y_t, \hat{\sigma}^2)$ im Zentrum der Dichte. Eine bessere Vorhersagequalität des Gaussian-Mixture-Models am linken Rand lässt sich anhand der Grafik nicht erkennen. Die Quantil Zerlegung liefert uns eher unerwartete Resultate (siehe Abbildung 7.3.5). Zwar zeigt sie auch auf, dass die Vorhersagemethoden $N(\hat{\mu}, \hat{\sigma}^2)$, $t(\hat{\mu}, \hat{\sigma}^2)$ und BoE etwas schlechtere Vorhersagen erzeugen, aber wir können weder einen Vorteil für die $N(y_t, \hat{\sigma}^2)$ -Vorhersage im Zentrum sehen, noch ein besseres Abschneiden des Gaussian-Mixture-Models im linken Bereich.

Die Analyse der Zweischrittvorhersage lässt uns erkennen, dass die Verfahren $N(y_t, \hat{\sigma}^2)$ und AR1, welche zeitliche Abhängigkeiten bei der Vorhersage berücksichtigen, keine vergleichsweise besseren mittleren scores erzielen können (siehe Tabelle 7.3.5). Begründen kann man dies durch den abnehmenden zeitlichen Bezug bei Vergrößerung des Vorhersagehorizonts. Das Gaussian-Mixture-Model erlangt beim quadratic score, beim pseudospherical score und beim CRPS die beste Bewertung. Das Verfahren t_{MLS} liegt beim logarithmic score vorne und nimmt bei den anderen scoring rules jeweils Platz zwei ein. Die Nullhypothese des Tests auf gleiche Vorhersage Performance kann bei Verwendung des quadratic scores oder des pseudospherical scores erneut bei zugrunde legen eines 5%-Signifikanzniveaus verworfen werden. Bei den anderen zwei Tests ist eine Ablehnung nicht möglich (siehe Tabelle 7.3.6).

Zusammenfassend kann man sagen, dass sich auch zur Vorhersage der Wachstumsraten des Bruttoinlandsproduktes keine unserer Vorhersagemethoden als 'das beste' Verfahren herausgestellt hat. Das Gaussian-Mixture-Model erhielt zwar oftmals den vergleichsweise höchsten score, war aber nicht in allen Bereichen überzeugend. Insbesondere beim ungewichteten CRPS und bei der Gewichtung auf das Zentrum und den rechten Rand der Dichte erlangte das autoregressive Modell AR1 eine bessere Bewertung. Dies deutet darauf hin, dass zum einen die Berücksichtigung der Zeitreihenstruktur sinnvoll sein kann und dass zum anderen eine Dichte vorteilhaft sein könnte, die zwar am linken Rand einen fat tail vorweist, am rechten Rand jedoch nicht. Außerdem lässt sich festhalten, dass die einfachen Vorhersagemethoden $N(\hat{\mu}, \hat{\sigma}^2)$, $N(y_t, \hat{\sigma}^2)$ und $t(\hat{\mu}, \hat{\sigma}^2)$, sowie die Bank-of-England-Vorhersage bei unserer Analyse der Wachstumsraten des Bruttoinlandsproduktes nicht überzeugen konnten.

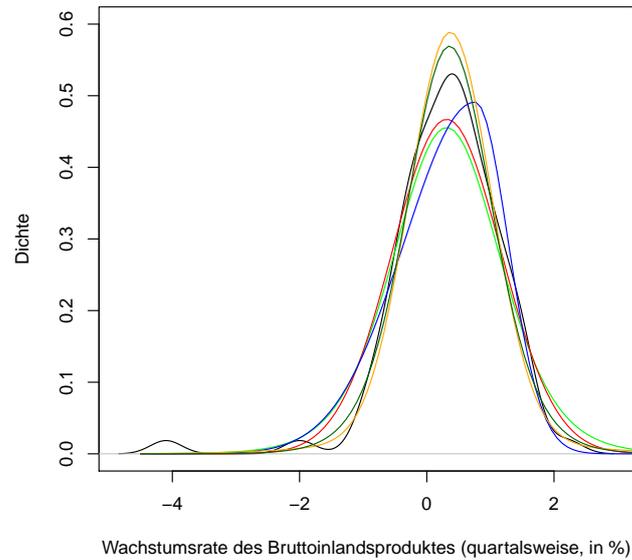


Abbildung 7.3.1.: Abbildung der mit einem Kerndichteschätzer ermittelten Dichte der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes (schwarz) und die geschätzten Dichtevorhersagen auf Grundlage der Daten aus den Jahren 1991 bis 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

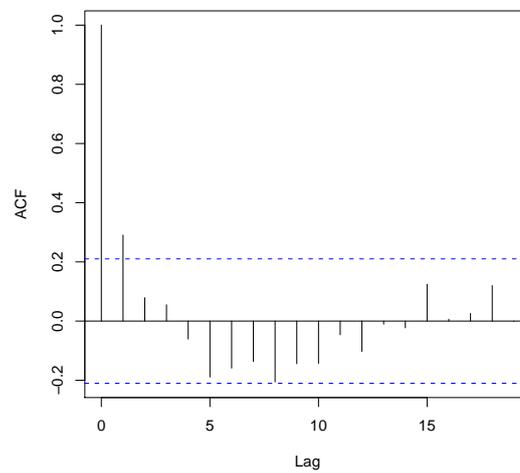


Abbildung 7.3.2.: Abbildung der Autokorrelationen für die quartalsweisen Wachstumsraten des Bruttoinlandsproduktes in Abhängigkeit von den Lags.

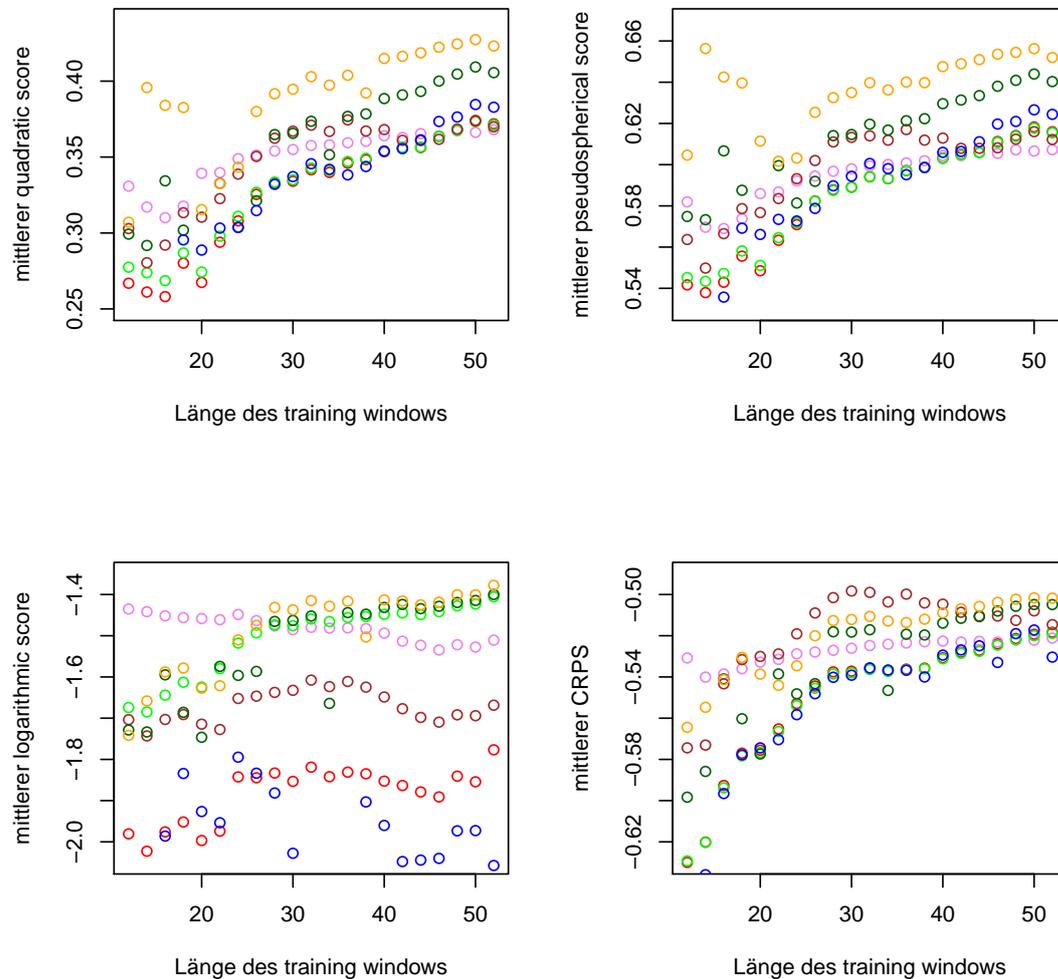


Abbildung 7.3.3.: Die vier Abbildungen zeigen für jede scoring rule separat die Entwicklung der mittleren scores der Vorhersagemethoden für die Vorhersagen der Wachstumsraten des Bruttoinlandsproduktes von Q3 2004 bis Q4 2012 in Abhängigkeit von der Länge des training windows ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

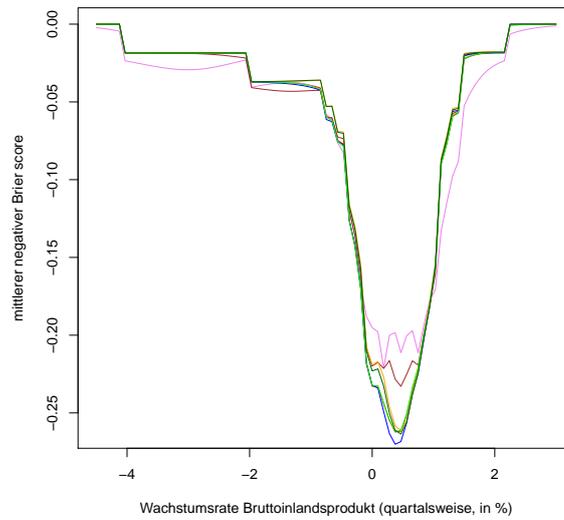


Abbildung 7.3.4.: Schwellenwert Zerlegung des CRPS der verschiedenen Vorhersagen unter Verwendung der quart. Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

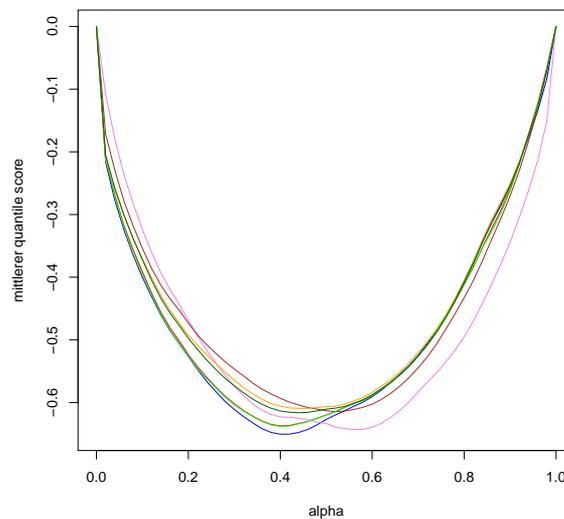


Abbildung 7.3.5.: Quantil Zerlegungen des CRPS der verschiedenen Vorhersagen unter Verwendung der quart. Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

Scoring rule	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
quadratic	0.3757	0.3755	0.3750	0.3875	0.3580	0.4141	0.3940
pseudosph.	0.6173	0.6134	0.6165	0.6247	0.6061	0.6459	0.6319
logarithmic	-1.4916	-1.3265	-1.2773	-1.3709	-1.6964	-1.2362	-1.2659
CRPS	-0.4687	-0.4770	-0.4696	-0.4506	-0.4730	-0.4528	-0.4571

Tabelle 7.3.1.: Mittlere scores nach scoring rule und Vorhersagemethode für die quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012 bei einer Einschrittvorhersage. Die Vorhersage erfolgte auf Grundlage der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q2 1991 bis Q4 2012. Bei den Schätzungen für die Vorhersagen wurde ein fortlaufendes Datenfenster von 32 Datenpunkten verwendet.

Scoring rule	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
quadratic	GMM	t_{MLS}	1.7007	0.044
pseudospherical	GMM	t_{MLS}	1.8739	0.030
logarithmic	GMM	t_{MLS}	0.4408	0.330
CRPS	AR1	GMM	0.0708	0.472

Tabelle 7.3.2.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach scoring rule bei einer Einschrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q2 1991 bis Q4 2012. Bei den Schätzungen für die Vorhersagen wurde ein fortlaufendes Datenfenster von 32 Datenpunkten verwendet.

Scoring rule	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
Zentrum	-0.1524	-0.1409	-0.1524	-0.1430	-0.1540	-0.1470	-0.1488
Ränder	-0.1424	-0.1790	-0.1432	-0.1480	-0.1433	-0.1380	-0.1385
rechts	-0.2077	-0.2145	-0.2082	-0.1982	-0.2094	-0.2026	-0.2048
links	-0.2609	-0.2661	-0.2613	-0.2560	-0.2635	-0.2502	-0.2523

Tabelle 7.3.3.: Mittlere scores der gewichteten Version des CRPS nach Gewichtung und Vorhersagemethode für die Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012 bei einer Einschrittvorhersage. Die Vorhersage erfolgte auf Grundlage der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q2 1991 bis Q4 2012. Bei den Schätzungen für die Vorhersagen wurde ein fortlaufendes Datenfenster von 32 Datenpunkten verwendet.

Gewichtung	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
Zentrum	$N(y_t, \hat{\sigma}^2)$	AR1	0.1628	0.435
Ränder	GMM	t_{MLS}	1.7324	0.042
rechts	AR1	GMM	0.5089	0.3054
links	GMM	t_{MLS}	1.6067	0.054

Tabelle 7.3.4.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach Gewichtung des CRPS bei einer Einschrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q2 1991 bis Q4 2012. Bei den Schätzungen für die Vorhersagen wurde ein fortlaufendes Datenfenster von 32 Datenpunkten verwendet.

Scoring rule	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
quadratic	0.3731	0.3296	0.3727	0.3782	0.3526	0.4067	0.3883
pseudosph.	0.6165	0.5786	0.6158	0.6201	0.6035	0.6409	0.6282
logarithmic	-1.6845	-1.8678	-1.3169	-1.7081	-2.6334	-1.6540	-1.3089
CRPS	-0.4747	-0.5879	-0.4755	-0.4737	-0.4779	-0.4588	-0.4633

Tabelle 7.3.5.: Mittlere scores nach scoring rule und Vorhersagemethode für die quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012 bei einer Zweischrittvorhersage. Die Vorhersage erfolgte auf Grundlage der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q2 1991 bis Q4 2012. Bei den Schätzungen für die Vorhersagen wurde ein fortlaufendes Datenfenster von 32 Datenpunkten verwendet.

Scoring rule	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
quadratic	GMM	t_{MLS}	1.8828	0.030
pseudosph.	GMM	t_{MLS}	2.0297	0.021
logarithmic	t_{MLS}	$t(\hat{\mu}, \hat{\sigma}^2)$	0.1443	0.443
CRPS	GMM	t_{MLS}	1.6280	0.052

Tabelle 7.3.6.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach scoring rule bei einer Zweischrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q2 1991 bis Q4 2012. Bei den Schätzungen für die Vorhersagen wurde ein fortlaufendes Datenfenster von 32 Datenpunkten verwendet.

7.4. Analyse der Arbeitslosenquote

Wie schon bei der Analyse der Inflationsraten und der Wachstumsraten des Bruttoinlandsproduktes wollen wir auch bei der Untersuchung der Wachstumsraten der Arbeitslosenquote zunächst die Verteilung der historischen Daten analysieren. Die mit einem Kerndichteschätzer erzeugte Dichte der Wachstumsraten aus 1991 bis 2012, welche in Abbildung 7.4.1 zu sehen ist, besitzt drei Häufungspunkte. Die Entstehung dieser drei Häufungspunkte ist, wie bereits in Abschnitt 6.1.3 angesprochen wurde, zum einen auf die grobe Angabe der Arbeitslosenquote auf eine Nachkommastelle zurückzuführen und zum anderen auf die gleichmäßigen Anstiege und Rückgänge der Arbeitslosenquote (vgl. Abbildung 6.1.5). Bei der Betrachtung der vorhergesagten Dichten, die ebenfalls in Abbildung 7.4.1 dargestellt sind, ergeben sich auf den ersten Blick Vorteile für das Gaussian-Mixture-Model, da es durch die Mischung zweier Dichten die Massepunkte besser abbilden kann als die anderen Verteilungen, welche nur einen Häufungspunkt vorweisen können. Um herauszufinden, ob sich diese besondere Charakteristik des Gaussian-Mixture-Models bei der anschließenden Bewertung der Vorhersagen mit den proper scoring rules tatsächlich auszahlt, berücksichtigen wir bei der Vorhersage alle fünf Parameter der Vorhersagemethode und nehmen Schätzungenauigkeiten in Kauf. Im Gegensatz zu den Inflationsraten und den Wachstumsraten des Bruttoinlandsproduktes können wir bei den Wachstumsraten der Arbeitslosenquote keine extremen Ausreißer feststellen. Die Dichte zeigt deshalb keine fat tails auf, was besonders bei der Vorhersagemethode $t(\hat{\mu}, \hat{\sigma}^2)$ zu einer Überschätzung der Randbereiche führt. Ansonsten lassen sich bei den Dichten, die durch die Verfahren $N(\hat{\mu}, \hat{\sigma}^2)$, $t(\hat{\mu}, \hat{\sigma}^2)$, BoE und t_{MLS} erzeugt wurden, anhand von Abbildung 7.4.1 keine großen Unterschiede feststellen.

Aus der Autokorrelationsfunktion für die Wachstumsraten der Arbeitslosenquote, welche in Abbildung 7.4.2 zu sehen ist, lässt sich eine klare zeitliche Abhängigkeit ableiten. Dies ist typisch für die Arbeitslosenquote, da Unternehmen ihren Personalbestand in Krisenzeiten aufgrund von Kündigungsfristen nur mit einem großen finanziellen Aufwand kurzfristig stark reduzieren können. Außerdem versuchen Unternehmen qualifizierte und eingearbeitete Fachkräfte auch in wirtschaftlich angespannten Zeiten zu halten, da die Einstellung und Einarbeitung neuer Mitarbeiter in der nächsten Aufschwungphase sehr kostspielig ist. Anhand des Verlaufs der Arbeitslosenquote, den wir in Abbildung 6.1.5 betrachtet haben, wurden diese trendartigen Entwicklungen deutlich. Die Wachstumsraten der Arbeitslosenquote sind entsprechend weniger volatil und weisen eine Zeitreihenstruktur auf. Dies dürfte zur Folge haben, dass unsere Vorhersagemethode $N(y_t, \hat{\sigma}^2)$, welche bei der Vorhersage der Inflationsrate und der Wachstumsraten des Bruttoinlandsproduktes sehr schlecht abgeschnitten hat, nun im Verhältnis bessere scores erzielen sollte. Das autoregressive Modell AR1 könnte die $N(y_t, \hat{\sigma}^2)$ -Vorhersage allerdings noch übertreffen, da sie den Grad der zeitlichen Abhängigkeit in jedem Vorhersageschritt neu ermittelt. Die Autokorrelationsfunktion lässt auch darauf schließen, dass ein autoregressives Modell mit höherer Ordnung sinnvoll sein könnte. Zur besseren Vergleichbarkeit mit den Ergebnissen aus der Analyse der anderen beiden makroökonomischen Kennzahlen werden wir jedoch auf die Verwendung eines autoregressiven Modells mit höherer

Ordnung verzichten.

Die zeitliche Abhängigkeit innerhalb der Zeitreihe wirkt sich auch auf die Verteilung der Daten aus. Ein Blick auf Abbildung 7.4.3 zeigt, dass die mittleren scores, welche auf Grundlage der Wachstumsraten aus 1997 bis 2012 ermittelt wurden, stark von der Länge des training windows abhängen. Die Vorhersagen erlangen im Allgemeinen bei Nutzung eines kleineren training windows bessere Ergebnisse. Lediglich bei der Vorhersagemethode $N(y_t, \hat{\sigma}^2)$ kann durch eine Verkleinerung des training windows keine höhere Vorhersagequalität erzielt werden. Für die Schätzung der Modellparameter legen wir deshalb für alle Verfahren ein Datenfenster von 12 Monatswerten fest. Eine noch kleinere Anzahl an Daten können wir für die Schätzungen nicht heranziehen, da sich ansonsten beispielsweise wegen sehr kleiner empirischer Varianzen Probleme bei der Berechnung ergeben. Anhand von Abbildung 7.4.3 lässt sich schon erkennen, dass die Vorhersagemethode AR1 verhältnismäßig höhere scores erhält. Hierbei muss man allerdings erwähnen, dass für die Ermittlung der Bank-of-England-Vorhersagen und der Gaussian-Mixture-Model-Vorhersagen der bei der Maximum-Likelihood-Schätzung zulässige Parameterbereich eingeschränkt werden musste. Für die im weiteren Verlauf durchgeführten Schätzung mit einem training window von 12 Datenpunkten konnte der zulässige Parameterbereich etwas vergrößert werden, sodass wir für die Verfahren BoE und GMM bessere Resultate erwarten dürfen.

Wir wollen nun die mittleren scores der Einschrittvorhersagen für die Wachstumsraten der Arbeitslosenquote von 1997 bis 2012, welche in Tabelle 7.4.1 zu sehen sind, betrachten. Besonders auffällig sind die hohen scores des Gaussian-Mixture-Modells beim quadratic score und beim pseudospherical score. Die Tests auf gleiche Vorhersage Performance bestätigen die bessere Vorhersagequalität gegenüber der jeweils zweitbesten Methode BoE, indem sie die Nullhypothese in beiden Fällen ablehnen (siehe Tabelle 7.4.2). Beim logarithmic score und beim CRPS erzielt das autoregressive Modell AR1 die besten Ergebnisse und verweist das Gaussian-Mixture-Modell bei ersterem und das Verfahren $N(y_t, \hat{\sigma}^2)$ bei letzterem auf Rang zwei. Die Methode $N(y_t, \hat{\sigma}^2)$ schneidet auch bei allen anderen scores schlechter ab als die AR1-Vorhersage. Die Überlegenheit des autoregressiven Modells wird durch die Ablehnung der Nullhypothese beim Test auf gleiche Vorhersage Performance unter Verwendung des CRPS untermauert. Beim logarithmic score kann die Nullhypothese nicht abgelehnt werden.

Bei den gewichteten Versionen des CRPS erhält die AR1-Vorhersage bei allen Gewichtungen den höchsten score (siehe Tabelle 7.4.3). Die Methode $N(y_t, \hat{\sigma}^2)$ erlangt bei der Betonung des Zentrums, sowie des rechten Randes den zweitbesten Wert. Diesmal können bei den entsprechenden Tests die Nullhypothesen nicht verworfen werden (siehe Tabelle 7.4.4). Ein anderes Bild zeigt sich bei der Gewichtung auf beide Ränder beziehungsweise auf den linken Rand. Hier liegt das Gaussian-Mixture-Modell auf dem zweiten Rang und der Test auf gleiche Vorhersage Performance lehnt die Nullhypothese bei beiden scoring rules ab.

Sowohl die Schwellenwert Zerlegung des CRPS, als die Quantil Zerlegung, welche in Abbildung 7.4.4 bzw. 7.4.5 dargestellt werden, bestätigen die bessere Qualität der AR1-Vorhersage. Dem Gaussian-Mixture-Model wurde zwar beim quadratic score und beim pseudospherical score eine gute Vorhersage bescheinigt, die Zerlegungen des CRPS zeigen jedoch keine Vorteile dieses Verfahrens auf. Außerdem legen die Zerlegungen sowohl der Bank-of-England-Vorhersage, als auch den Vorhersagemethoden $N(\hat{\mu}, \hat{\sigma}^2)$, $t(\hat{\mu}, \hat{\sigma}^2)$ und t_{MLS} , welche bei den mittleren scores bisher unterdurchschnittliche Werte erzielten, ebenfalls ein schlechtes Ergebnis dar.

Für den Zeitraum von 1997 bis 2012 wurden ebenfalls Zweischrittvorhersagen erzeugt. Die Analyse der mittleren scores, welche in Tabelle 7.4.5 zu sehen sind, zeigt, dass das Gaussian-Mixture-Model bei den scoring rules quadratic score und pseudospherical score weiterhin die mit Abstand höchsten Werte erlangt. Auch hier lehnen die beiden Tests die Nullhypothese auf gleiche Vorhersage Performance ab und bestätigen die Dominanz des Gaussian-Mixture-Modells (siehe Tabelle 7.4.6). Es sollte jedoch erwähnt werden, dass nun die Vorhersagemethode $N(y_t, \hat{\sigma}^2)$ die zweitbesten scores erzielt. Bei den anderen beiden scoring rules gibt es leichte Veränderungen im Vergleich zu den Resultaten bei den Einschrittvorhersagen. Die AR1-Vorhersage schneidet schlechter ab, was sich, wie schon bei den Zweischrittvorhersagen für die Wachstumsraten des Bruttoinlandsproduktes, durch den abnehmenden zeitlichen Bezug bei Vergrößerung des Vorhersagehorizontes erklären lässt. Die Bank-of-England-Vorhersage liegt beim logarithmic score an erster Stelle, gefolgt von der $t(\hat{\mu}, \hat{\sigma}^2)$ -Vorhersage. Beim CRPS erzielt die Methode $N(y_t, \hat{\sigma}^2)$ den höchsten score und verweist die AR1-Vorhersage auf Platz zwei. Die Tests können in den letzten beiden Fällen die Nullhypothese jeweils nicht verwerfen, sodass wir dem jeweiligen Verfahren keine bessere Vorhersagequalität bescheinigen können.

Zum Abschluss der Analyse können wir sagen, dass das Gaussian-Mixture-Model zur Vorhersage der Wachstumsraten der Arbeitslosenquote den Vorteil ausnutzen kann, eine Dichte mit zwei Häufungspunkten erzeugen zu können. Dadurch passt sie sich der Dichte der wahren Verteilung anscheinend besser an und erhält gerade beim quadratic score und beim pseudospherical score überdurchschnittliche mittlere scores. Dem muss man entgegenhalten, dass bei der Einschrittvorhersage das autoregressive Modell AR1 bei allen anderen scoring rules die höchsten scores erzielt. Dieses Verfahren kann die vorliegende Zeitreihenstruktur ausnutzen und ist ebenfalls zur Vorhersage zu empfehlen. Vergleichsweise schlechte Resultate werden den Vorhersagemethoden $t(\hat{\mu}, \hat{\sigma}^2)$ und t_{MLS} ausgestellt. Dies lässt vermuten, dass die transformierte t-Verteilung der wahren Verteilung nicht sehr nahe kommt. Da außerdem das einfache Verfahren $N(\hat{\mu}, \hat{\sigma}^2)$ schlecht abschneidet, trifft die Annahme, dass der Mittelwert zur Schätzung des Erwartungswertes der Verteilung gut geeignet ist, eher nicht zu. Zusammenfassend lässt sich sagen, dass sowohl das Gaussian-Mixture-Model, als auch das autoregressive Model ihre Vorteile bei der Vorhersage der Wachstumsraten der Arbeitslosenquote haben, die Vorhersagequalität sich aber sicherlich noch verbessern lässt. Eine Kombination der beiden Verfahren könnte sich als sinnvoll erweisen. Einige weitere Ansätze zur Verfeinerung der Vorhersage Performance werden im nächsten Kapitel vorgestellt.

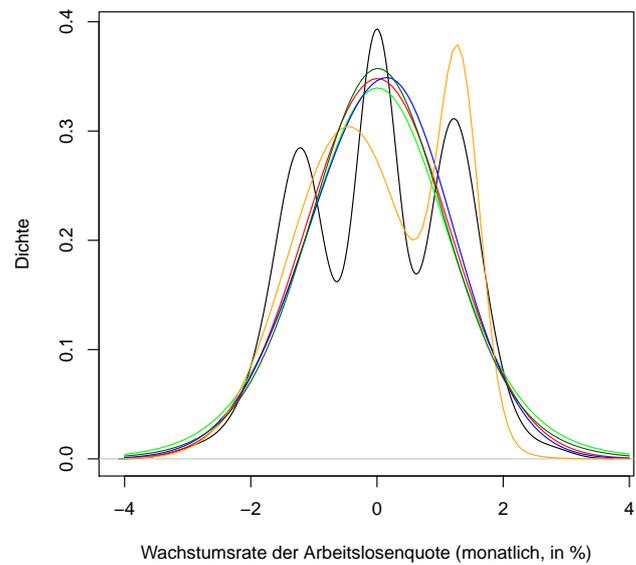


Abbildung 7.4.1.: Abbildung der mit einem Kerndichteschätzer ermittelten Dichte der monatlichen Wachstumsraten der Arbeitslosenquote (schwarz) und die geschätzten Dichtevorhersagen auf Grundlage der Daten zur Arbeitslosenquote aus den Jahren 1991 bis 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

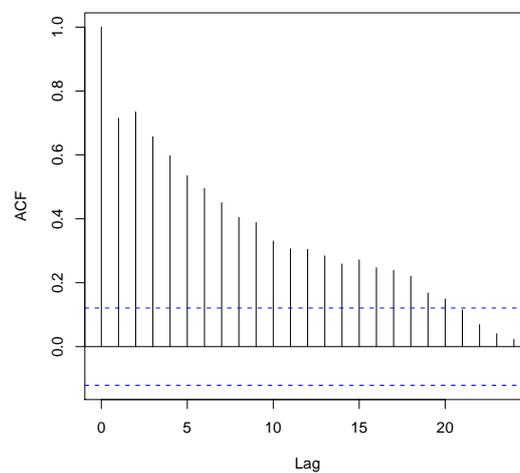


Abbildung 7.4.2.: Abbildung der Autokorrelationen für die quartalsweisen Wachstumsraten der Arbeitslosenquote in Abhängigkeit von den Lags.

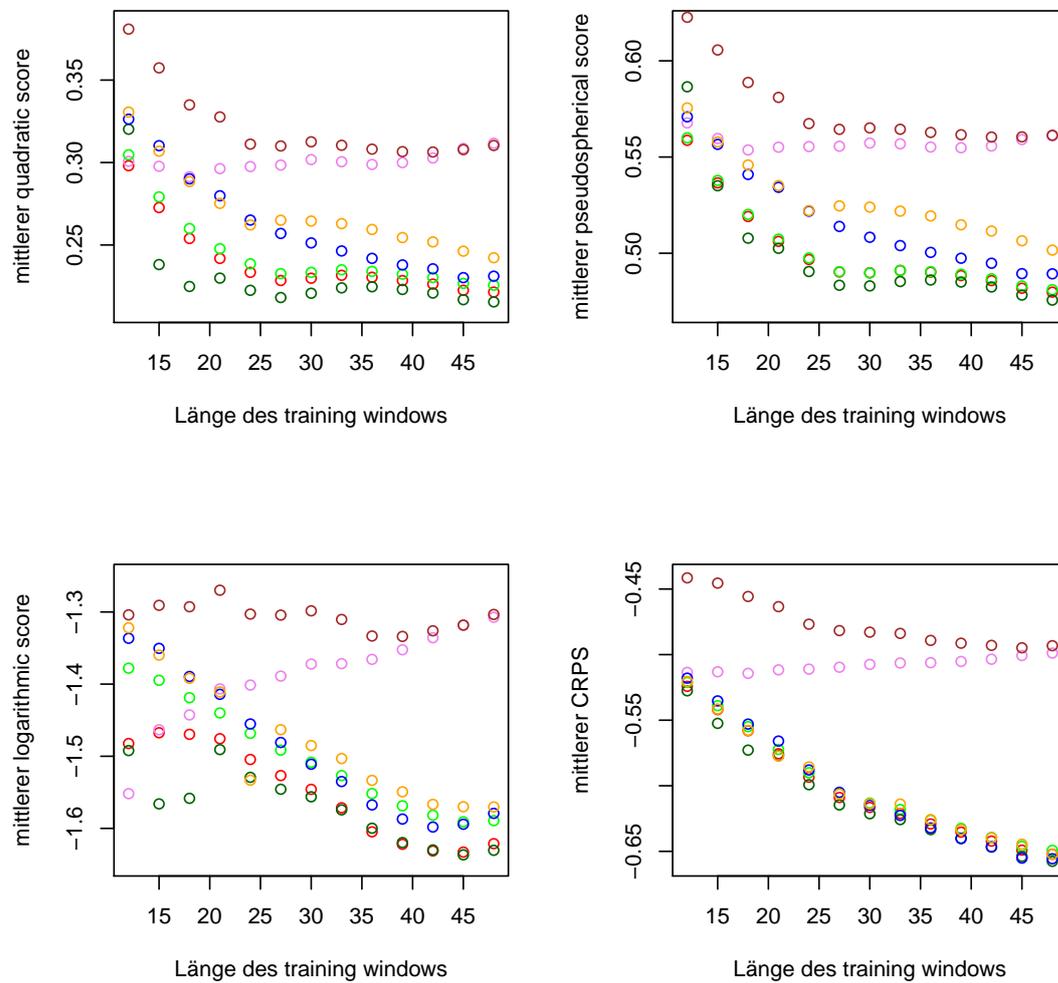


Abbildung 7.4.3.: Die vier Abbildungen zeigen für jede scoring rule separat die Entwicklung der mittleren scores der Vorhersagemethoden für die Vorhersagen der monatlichen Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 in Abhängigkeit von der Länge des training windows ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

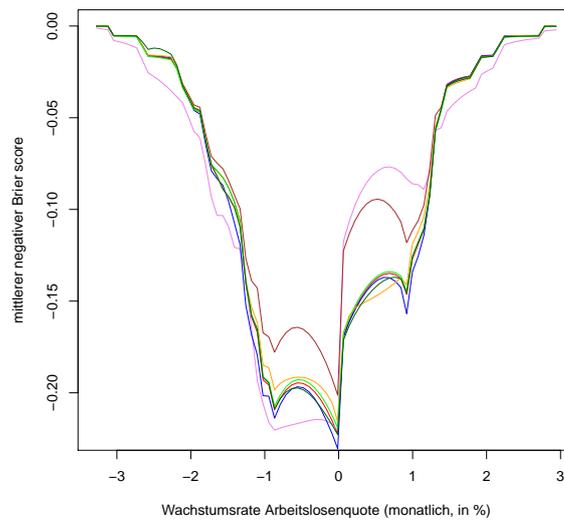


Abbildung 7.4.4.: Schwellenwert Zerlegung des CRPS der verschiedenen Vorhersagen unter Verwendung der monatl. Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

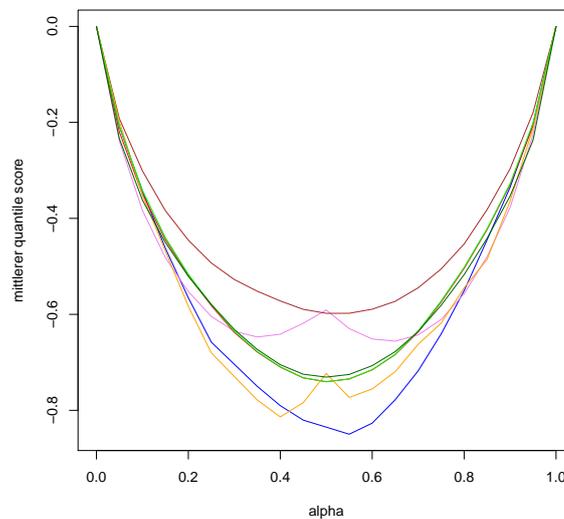


Abbildung 7.4.5.: Quantil Zerlegungen des CRPS der verschiedenen Vorhersagen unter Verwendung der monatlichen Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 ($N(\hat{\mu}, \hat{\sigma}^2)$ =rot, $N(y_t, \hat{\sigma}^2)$ =violett, $t(\hat{\mu}, \hat{\sigma}^2)$ =grün, AR1=braun, BoE=blau, GMM=orange, t_{MLS} =dunkelgrün).

Scoring rule	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
quadratic	0.2981	0.3008	0.3047	0.3809	0.4297	0.8040	0.3202
pseudosph.	0.5587	0.5678	0.5600	0.6227	0.6479	0.8864	0.5865
logarithmic	-1.4823	-1.5519	-1.3779	-1.3040	-1.3365	-1.3219	-1.4921
CRPS	-0.5242	-0.5135	-0.5211	-0.4415	-0.5380	-0.5180	-0.5276

Tabelle 7.4.1.: Mittlere scores nach scoring rule und Vorhersagemethode für die Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 bei einer Einschrittvorhersage. Die Vorhersage erfolgte auf Grundlage der monatlichen Arbeitslosenquoten von Januar 1991 bis Dezember 2012. Bei Schätzungen für die Vorhersage wurde ein fortlaufendes Datenfenster von 12 Datenpunkten verwendet.

Scoring rule	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
quadratic	GMM	BoE	6.4759	<0.001
pseudospherical	GMM	BoE	8.8243	<0.001
logarithmic	AR1	GMM	0.3209	0.374
CRPS	AR1	$N(y_t, \hat{\sigma}^2)$	2.2487	0.012

Tabelle 7.4.2.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach scoring rule bei einer Einschrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der monatlichen Arbeitslosenquoten von Januar 1991 bis Dezember 2012. Bei Schätzungen für die Vorhersage wurde ein fortlaufendes Datenfenster von 12 Datenpunkten verwendet.

Scoring rule	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
Zentrum	-0.1355	-0.1247	-0.1344	-0.1116	-0.1378	-0.1337	-0.1370
Ränder	-0.1352	-0.1553	-0.1353	-0.1214	-0.1390	-0.1335	-0.1342
rechts	-0.2273	-0.2013	-0.2260	-0.1869	-0.2317	-0.2253	-0.2297
links	-0.2972	-0.3127	-0.2954	-0.2550	-0.3062	-0.2926	-0.2981

Tabelle 7.4.3.: Mittlere scores der gewichteten Version des CRPS nach Gewichtung und Vorhersagemethode für die Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 bei einer Einschrittvorhersage. Die Vorhersage erfolgte auf Grundlage der monatlichen Arbeitslosenquoten von Januar 1991 bis Dezember 2012. Bei Schätzungen für die Vorhersage wurde ein fortlaufendes Datenfenster von 12 Datenpunkten verwendet.

Gewichtung	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
Zentrum	AR1	$N(y_t, \hat{\sigma}^2)$	1.5585	0.060
Ränder	AR1	GMM	3.2148	<0.001
rechts	AR1	$N(y_t, \hat{\sigma}^2)$	1.0363	0.150
links	AR1	GMM	3.8995	<0.001

Tabelle 7.4.4.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach Gewichtung des CRPS bei einer Einschrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der monatlichen Arbeitslosenquoten von Januar 1991 bis Dezember 2012. Bei Schätzungen für die Vorhersage wurde ein fortlaufendes Datenfenster von 12 Datenpunkten verwendet.

Scoring rule	$N(\hat{\mu}, \hat{\sigma}^2)$	$N(y_t, \hat{\sigma}^2)$	$t(\hat{\mu}, \hat{\sigma}^2)$	AR1	BoE	GMM	t_{MLS}
quadratic	0.2709	0.3725	0.2786	0.3083	0.3487	0.6834	0.2826
pseudosph.	0.5395	0.6204	0.5409	0.5706	0.6147	0.8288	0.5576
logarithmic	-1.6255	-1.5028	-1.4701	-1.5309	-1.4504	-1.4832	-1.6344
CRPS	-0.5574	-0.4814	-0.5537	-0.5125	-0.5766	-0.5566	-0.5625

Tabelle 7.4.5.: Mittlere scores nach scoring rule und Vorhersagemethode für die Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 bei einer Zweischrittvorhersage. Die Vorhersage erfolgte auf Grundlage der monatlichen Arbeitslosenquoten von Januar 1991 bis Dezember 2012. Bei Schätzungen für die Vorhersage wurde ein fortlaufendes Datenfenster von 12 Datenpunkten verwendet.

Scoring rule	Vorhersage 1	Vorhersage 2	t-Wert	p-Wert
quadratic	GMM	$N(y_t, \hat{\sigma}^2)$	2.9746	0.001
pseudospherical	GMM	$N(y_t, \hat{\sigma}^2)$	3.7539	<0.001
logarithmic	BoE	$t(\hat{\mu}, \hat{\sigma}^2)$	0.4291	0.334
CRPS	$N(y_t, \hat{\sigma}^2)$	AR1	0.8557	0.196

Tabelle 7.4.6.: t-Werte und p-Werte der Tests auf gleiche Vorhersage Performance nach scoring rule bei einer Zweischrittvorhersage. Es wurden jeweils die Vorhersagemethoden mit den beiden besten mittleren scores für die Wachstumsraten der Arbeitslosenquote von 1997 bis 2012 verglichen. Die Vorhersage erfolgte auf Grundlage der monatlichen Arbeitslosenquoten von Januar 1991 bis Dezember 2012. Bei Schätzungen für die Vorhersage wurde ein fortlaufendes Datenfenster von 12 Datenpunkten verwendet.

8. Ansätze zur Verfeinerung der Vorhersage

Bei unserer Analyse haben wir bereits 7 Methoden untersucht, mit denen man makroökonomische Kennzahlen vorhersagen kann. Keine dieser Methoden stellte sich als 'die beste' Methode heraus, aber die unterschiedlichen Ansätze der Verfahren und deren Auswirkungen auf die Vorhersagequalität wurden deutlich. Es existiert eine Vielzahl an weiteren Ansätzen, die in der Praxis zur Vorhersage herangezogen werden. Eine Auswahl an Verfahren, die unter anderem speziell für die Vorhersage unserer drei makroökonomischen Kennzahlen Inflationsrate, Bruttoinlandsprodukt und Arbeitslosenquote verwendet werden, wollen wir in diesem Kapitel vorstellen. Beginnen wollen wir mit einem Ansatz, der unsere Vorhersagemethode AR1 erweitert.

8.1. Autoregression mit exogenen Variablen

Die bisher vorgestellten Methoden haben immer nur die entsprechenden historischen Daten der Zielgröße als Inputvariablen verwendet. Vernachlässigt haben wir dabei, dass es eventuell noch weitere Informationen gibt, die wir in unsere Vorhersage einfließen lassen sollten. Eine Möglichkeit um bessere Wahrscheinlichkeitsvorhersagen zu erhalten ist folglich die Einbindung exogener Variablen in das Modell. Dabei bezeichnen exogene Variablen Kenngrößen, die zusätzlich zu den historischen Daten der Zielgröße für die Vorhersage herangezogen werden. Für unser autoregressives Modell (6.2.10) aus Abschnitt 6.2.3 bedeutet dies, dass man die Lags einer zweiten Zeitreihe $X = (X_t : t \in \mathbb{Z})$ in das Modell einbezieht und die erweiterte Version

$$Y_t = a_0 + a_1 Y_{t-1} + \dots + a_p Y_{t-p} + b_1 X_{t-1} + \dots + b_q X_{t-q} + \epsilon_t \quad (8.1.1)$$

erhält, wobei a_0, \dots, a_p und b_1, \dots, b_q Parameter des Modells sind und ϵ_t einem *Weißes Rauschen* entspricht (siehe z.B. Shumway & Stoffer (2011)).

Eine Aussage über die zeitliche Abhängigkeit der beiden Zeitreihen untereinander kann man treffen, wenn man sich die Autokorrelationen

$$\text{Cor}(Y_{t+h}, X_t) = \gamma_h, \quad h \in \mathbb{Z}, \quad (8.1.2)$$

anschaut. Eine hohe Autokorrelation spricht dafür, dass man durch die zusätzliche Information, die man aus der Zeitreihe X gewinnt, die zeitliche Entwicklung der Zeitreihe Y besser erklären kann (siehe z.B. Shumway & Stoffer (2011)).

Wir wollen nun dieses erweiterte autoregressive Modell auf unsere Daten anwenden. Zuerst wollen wir versuchen, ob wir für die Wachstumsraten des Bruttoinlandsproduktes

bessere Vorhersagen erzeugen können, wenn wir den ifo-Geschäftsklimaindex als exogene Variable heranziehen. Anschließend wollen wir die Wachstumsraten des Bruttoinlandsproduktes nutzen, um bessere Prognosen für die Wachstumsraten der Arbeitslosenquote zu erhalten. Die Schätzung der Parameter und die Vorhersage erfolgte jeweils mit den in R zur Verfügung stehenden Methoden *arima* und *predict*.

8.1.1. Vorhersage des Bruttoinlandsproduktes unter Verwendung des ifo-Geschäftsklimaindexes

Der ifo-Geschäftsklimaindex ist ein von dem ifo Institut für Wirtschaftsforschung in monatlichen Abständen veröffentlichter Index. Die Erhebung erfolgt durch die Auswertung eines Fragebogens, der an verschiedenste Unternehmen versandt wird. Da der ifo-Geschäftsklimaindex als Frühindikator für die wirtschaftliche Entwicklung gilt, dürfte er als zusätzliche Information für die Vorhersage des Bruttoinlandsproduktes geeignet sein. Für die Modellierung ist es nötig, dass wir die quartalsweisen Mittelwerte des ifo-Geschäftsklimaindexes bilden, damit sowohl die Wachstumsraten des Bruttoinlandsproduktes, als auch unsere exogene Variable in quartalsweisen Werten vorliegen. Um eine bessere Vorstellung über den Zusammenhang zwischen den Wachstumsraten des Bruttoinlandsproduktes und dem ifo-Geschäftsklimaindex zu erhalten, schauen wir uns die Autokorrelationen in Abbildung 8.1.1 an. Es lässt sich eine hohe Korrelation von 0.58 zwischen der Wachstumsrate des Bruttoinlandsproduktes und dem ifo-Geschäftsklimaindex aus dem vergangenen Quartal erkennen. Wir werden deshalb den ifo-Geschäftsklimaindex mit Lag 1 als exogene Variable in unser autoregressives Modell AR1 aufnehmen.

Mit diesem neuen Modell erzeugen wir ebenfalls sowohl Ein-, als auch Zweischrittvorhersagen für den Zeitraum von Q3 1999 bis Q4 2012 unter Verwendung eines training windows von 32 Datenpunkten. Die mittleren scores, die sich für diese Vorhersagen ergeben, sind in Tabelle 8.1.1 aufgeführt. Ein Vergleich mit den anderen Vorhersagemethoden zeigt, dass unser neues Modell bei der Einschrittvorhersage beim logarithmic score und beim CRPS besser abschneidet als alle anderen Verfahren. Außerdem erlangt es beim quadratic score und beim pseudospherical score jeweils höhere mittlere scores als die AR1-Vorhersage, aber es kommt nicht an die Werte des Gaussian-Mixture-Models heran. Bei der Zweischrittvorhersage sind die Abstände zwischen den mittleren scores des erweiterten autoregressiven Modells und des Gaussian-Mixture-Models im Falle des quadratic scores und des pseudospherical scores deutlich geringer. Das Gaussian-Mixture-Model erzielt aber auch hier eine höhere Bewertung. Wenn es um die zwei-periodische Vorhersage geht, dann beurteilen der logarithmic score und der CRPS die Vorhersagequalität des neuen Verfahrens als äußerst schlecht. Das Verfahren landet bei den mittleren scores jeweils auf dem vorletzten Rang.

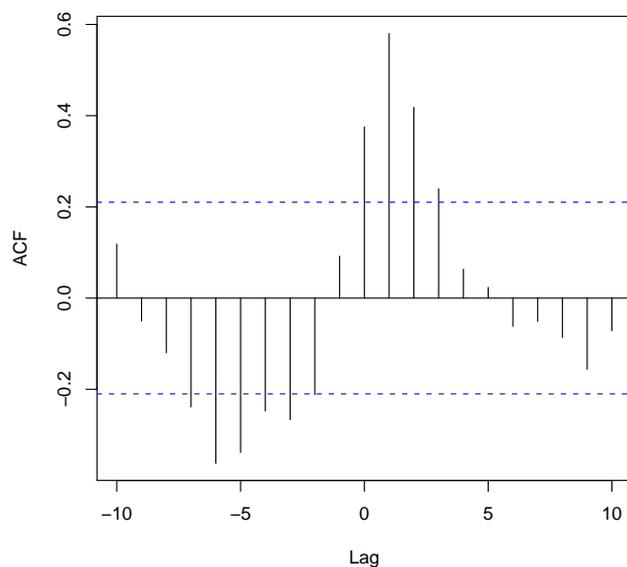


Abbildung 8.1.1.: Abbildung der Autokorrelationen zwischen den quartalsweisen Wachstumsraten des Bruttoinlandsproduktes und den quartalsweisen Ifo-Geschäftsklimaindizes (Mittelwerte der drei monatlichen Indizes) in Abhängigkeit von den Lags.

Scoring rule	Einschritt	Zweischritt
quadratic	0.3878	0.3914
pseudosph.	0.6285	0.6300
logarithmic	-1.2165	-2.1266
CRPS	-0.4074	-0.5151

Tabelle 8.1.1.: Mittlere scores des erweiterten AR1-Modells für die Wachstumsraten des Bruttoinlandsproduktes von Q3 1999 bis Q4 2012. Die Vorhersage erfolgte auf Grundlage der quartalsweisen Wachstumsraten des Bruttoinlandsproduktes von Q2 1991 bis Q4 2012 und dem quartalsweisen Ifo-Geschäftsklimaindex (Mittelwert der drei monatlichen Indizes) von Q1 1991 bis Q4 2012. Bei Schätzungen für die Vorhersage wurde ein fortlaufendes Datenfenster von 32 Datenpunkten verwendet.

Dieses Resultat lässt darauf schließen, dass es bei der mehrperiodischen Vorhersage dazu kommen kann, dass das Modell die Entwicklung der Zielgröße in manchen Fällen überschätzt. Besonders in volatilen Phasen kann sich die Verwendung des ifo-Geschäftsklimaindexes, der als Frühindikator für das Wirtschaftsgeschehen gilt, bei der Mehrschrittvorhersage des Bruttoinlandsproduktes als nachteilig erweisen, da durch Fehleinschätzungen der Befragten die Vorhersage weit von der tatsächlichen Ausprägung entfernt liegen kann. Die entsprechenden Ausreißer führen speziell beim logarithmic score zu schlechten Bewertungen und machen sich auch beim CRPS bemerkbar.

8.1.2. Vorhersage der Arbeitslosenquote unter Verwendung des Bruttoinlandsproduktes

Ein wirtschaftlicher Aufschwung bzw. Abschwung wirkt sich auf den Umsatz der Unternehmen aus. Um eine höhere Nachfrage der Kunden zu bedienen steigern die Unternehmen ihre Produktion und benötigen dazu mehr Personal. In wirtschaftlich schlechten Zeiten reduziert sich die Nachfrage und die Unternehmen können ihre Produkte nicht veräußern. Sie besitzen ein Überangebot an Gütern und Arbeitern und reagieren mit Personalabbau. Diese Überlegungen lassen schlussfolgern, dass sich die Entwicklung des Bruttoinlandsproduktes zeitverzögert auf die Arbeitslosenquote auswirken sollte und wir somit die Wachstumsrate des Bruttoinlandsproduktes als einen Indikator für die Wachstumsrate der Arbeitslosenquote verwenden können. Damit wir für beide Zeitreihen monatliche Datenpunkte zur Verfügung haben, gehen wir vereinfachend davon aus, dass sich das quartalsweise Wachstum des Bruttoinlandsproduktes jeweils gleichmäßig auf die drei Monate verteilt. Die zeitliche Abhängigkeit der beiden Kennzahlen lässt sich in Abbildung 8.1.2 erkennen, welche die Autokorrelationen zwischen den Kenngrößen darstellt. Die höchste Korrelation liegt mit einem Wert von 0.39 bei Lag 2 vor. Wir werden deshalb die Wachstumsrate des Bruttoinlandsproduktes mit Lag 2 in das autoregressive Modell aufnehmen.

Bei den bisherigen Schätzungen zur Arbeitslosenquote haben wir aufgrund der ausgeprägten Zeitreihenstruktur jeweils ein sehr kurzes training window von 12 Datenpunkten verwendet. Unser neues Modell benötigt für eine vernünftige Vorhersage mehr Daten, sodass nach einer Analyse der mittleren scores die Länge des training windows auf 48 festgelegt wurde. Die Einschrittvorhersagen erhalten bei allen vier scoring rules schlechtere mittlere scores als das einfache autoregressive Modell AR1 (siehe Tabelle 8.1.2). Anders sieht die Situation bei der Zweischrittvorhersage aus. Hier erzielt die erweiterte Methode jeweils bessere Ergebnisse als das einfache AR1-Verfahren und erlangt in Vergleich zu den anderen Verfahren beim logarithmic score den höchsten und beim CRPS den zweithöchsten Wert.

Aufgrund dieser Resultate liegt die Vermutung nahe, dass die Korrelation zwischen den Kennzahlen in Höhe von 0.39 im Verhältnis zu der hohen Autokorrelation bei den Wachstumsraten der Arbeitslosenquote (vgl. Abbildung 7.4.2) zu gering ist um die Einschrittvorhersage des AR1-Verfahrens zu verbessern. Da die zeitliche Abhängigkeit bei

der Zweischrittvorhersage nachlässt und die Wachstumsrate des Bruttoinlandsproduktes mit Lag 2 in die Vorhersage einfließt, kann das erweiterte Modell durch den Zusatz an Information eine höhere Vorhersagequalität liefern.

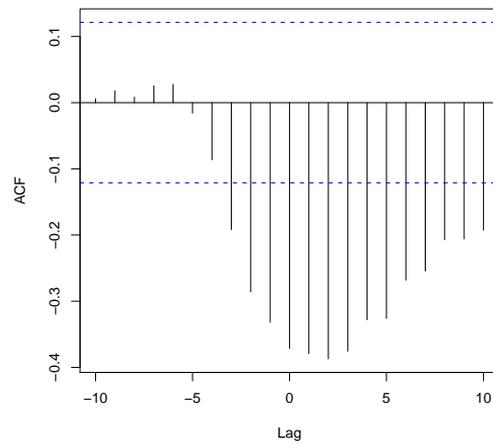


Abbildung 8.1.2.: Abbildung der Autokorrelationen zwischen den monatlichen Wachstumsraten der Arbeitslosenquote und den monatlichen Wachstumsraten des Bruttoinlandsproduktes (Quartalswachstum des BIP auf die drei Monate gleichmäßig verteilt) in Abhängigkeit von den Lags.

Scoring rule	Einschritt	Zweischritt
quadratic	0.3084	0.3300
pseudosph.	0.5621	0.5774
logarithmic	-1.3140	-1.3722
CRPS	-0.4877	-0.5007

Tabelle 8.1.2.: Mittlere scores des erweiterten AR1-Modells für die Wachstumsraten der Arbeitslosenquote von 1997 bis 2012. Die Vorhersage erfolgte auf Grundlage der monatlichen Wachstumsraten der Arbeitslosenquote von 1991 bis 2012 und des Bruttoinlandsproduktes (Quartalswachstum des BIP auf die drei Monate gleichmäßig verteilt) von Q2 1991 bis Q4 2012. Bei den Schätzungen für die Vorhersage wurde ein fortlaufendes Datenfenster von 48 Datenpunkten verwendet.

8.2. Weitere Ansätze im Überblick

Die vorangegangene Analyse der Arbeitslosenquote hat gezeigt, dass wir aus der Entwicklung des Bruttoinlandsproduktes Rückschlüsse auf die Entwicklung des Arbeitsmarktes ziehen können. Clements & Smith (2000) verfolgen die Annahme, dass auch die Geschehnisse am Arbeitsmarkt die wirtschaftliche Entwicklung beeinflussen. Diese wechselseitige Abhängigkeit zwischen den beiden Kenngrößen modellieren sie mit einem *Vektorautoregressiven Modell* (VAR). Unter einem VAR-Modell versteht man eine Autoregression, bei der die zu erklärende Variable einer mehrdimensionalen Zufallsvariable entspricht. Es handelt sich somit um die mehrdimensionale Variante der einfachen Autoregression, wie wir sie aus Abschnitt 6.2.3 kennen. Bei ihren Analysen verwenden Clements & Smith (2000) sowohl ein lineares, als auch ein nicht-lineares VAR-Modell. Der Ansatz, auch nicht-lineare autoregressive Modelle für die Vorhersage zu verwenden, wurde bereits von Rothman (1998) analysiert. Er beschränkte sich dabei auf eine einfache Autoregression und untersuchte die Arbeitslosenquote der USA. Zum Vergleich der Vorhersagen zog Rothman (1998) zusätzlich ein *Regime-Switching Model* heran. Bei diesem wechselt das Modell in Abhängigkeit von der Höhe der Arbeitslosenquote.

Zur Vorhersage makroökonomischer Kennzahlen greift man auch auf makroökonomische Theorien und Modelle zurück. So existieren beispielsweise Überlegungen, nach denen die Höhe der Zinssätze und die Zinsdifferenz für als sicher geltende Anlagen/Kredite mit unterschiedlicher Laufzeit als Indikatoren für die Entwicklung der Inflationsrate und des Bruttoinlandsproduktes herangezogen werden können. In wirtschaftlich schwachen Phasen ist der Zinssatz im Allgemeinen niedrig, da die Nachfrage nach Geld zum Konsumieren oder Investieren gering ist. Im Gegensatz dazu steigt der Zinssatz bei einem Aufschwung aufgrund der erhöhten Nachfrage nach Zahlungsmitteln. Eine große Differenz zwischen den Zinssätzen für kurzfristige und langfristige Anlagen/Kredite deutet, darauf hin, dass der Anbieter des Finanzproduktes mit steigenden Zinssätzen rechnet und somit ein höheres Wirtschaftswachstum oder eine höhere Inflation erwartet (siehe z.B. Blachard & Illing (2009)) (Liquiditätsrisiken und Kreditrisiken wurden bei diesen Überlegungen vernachlässigt). Ang et al. (2006) versuchen diesen Zusammenhang auszunutzen und verwenden deshalb zur Vorhersage des Bruttoinlandsproduktes eine Vektorautoregression, welche als Variablen die Wachstumsrate des Bruttoinlandsproduktes, einen Zinssatz für kurze Laufzeiten und die Zinsdifferenz zu einem Zinssatz für Anlagen mit langer Laufzeit beinhaltet. Die Informationen der *Zinsstrukturkurve*, welche die Höhe des Zinssatzes in Abhängigkeit von der Laufzeit des Finanzproduktes darstellt, werden auch in dem von Kozicki (1997) geschriebenen Economics Review der Federal Reserve Bank of Kansas City genutzt, um die Entwicklung der Inflationsrate und des Bruttoinlandsproduktes in den USA zu prognostizieren. In diesem Bericht wird für die Vorhersage ein autoregressives Modell erläutert, das als exogene Variablen Kennzahlen hinzuzieht, die aus der Zinsstrukturkurve ermittelt wurden.

Ein weiterer theoretischer Ansatz, der bei der Vorhersage der Inflationsrate eine Rolle

spielt, ist die *Quantitätsgleichung*

$$\text{Geldmenge} * \text{Umlaufgeschwindigkeit} = \text{Preisniveau} * \text{realesBIP}.$$

Zur Analyse unterstellt man die Annahme, dass die Kennzahlen sich von einer Periode auf die nächste nur leicht verändern, sodass man durch Betrachtung der logarithmierten Werte die Approximation

$$\pi \approx \Delta \log(\text{Preisniv.}) = \Delta \log(\text{Geldmenge}) + \Delta \log(\text{Umlaufgeschw.}) - \Delta \log(\text{realesBIP})$$

für die Inflationsrate erhält, wobei Δ die Änderung der Kennzahl bezeichnet. Scheide & Trabandt (2000) greifen diesen Zusammenhang auf und analysieren mit einem autoregressiven Modell mit entsprechenden exogenen Variablen die Inflationsrate im Euroraum. Außerdem verwenden sie ein *Fehlerkorrekturmodell*, welches davon ausgeht, dass die Inflationsrate auf lange Sicht auf einem gleichmäßigen Niveau bleibt. Kurzfristige Abweichungen von diesem Niveau werden durch das Fehlerkorrekturmodell korrigiert, sodass eine Annäherung an das langfristige Niveau erfolgt.

Auch Bachmeier & Swanson (2005) wenden bei der Analyse der Inflationsrate ein Fehlerkorrekturmodell an, welches auf die Quantitätsgleichung zurückzuführen ist, und vergleichen dieses mit autoregressiven Modellen mit und ohne zusätzliche exogene Variablen. Darüber hinaus betrachten Bachmeier & Swanson (2005) auch ein autoregressives Modell, welches durch die *Phillips-Kurve* motiviert wird. Bei der Phillips-Kurve handelt es sich um ein Modell aus der Wirtschaftstheorie, welches einen Bezug zwischen der Arbeitslosenquote und der Inflationsrate herstellt (siehe z.B. Blachard & Illing (2009)). Dabei geht man davon aus, dass bei geringer Arbeitslosigkeit Unternehmen ihren Mitarbeitern aufgrund der erhöhten Nachfrage nach Fachkräften höhere Löhne zahlen müssen. Um die zusätzlichen Kosten zu kompensieren, passen die Unternehmen die Preise für ihre Dienstleistungen entsprechend an und es kommt zu einer steigenden Inflation. Im Gegensatz dazu ist der Wettbewerb um Mitarbeiter bei hoher Arbeitslosigkeit geringer und die Lohn- und Preissteigerungen fallen schwächer aus.

Neben dem Einbinden zusätzlicher makroökonomischer Kennzahlen in ein Modell kann man auch die Ergebnisse von Umfragen als weitere Informationsquelle für die Vorhersage verwenden. Gneiting & Thorarinsdottir (2010) analysieren beispielsweise Experteneinschätzungen aus dem *Survey of Professional Forecasters*. In diesem werden durch eine vierteljährliche Befragung, die an Unternehmen und Institutionen gerichtet ist, Einschätzungen zur Entwicklung von makroökonomischen Kennzahlen ermittelt. Gneiting & Thorarinsdottir (2010) greifen auf die Prognosen zur Inflationsrate zurück und vergleichen deren Vorhersage Performance mit Vorhersagen, die durch einfache Vorhersagemodelle erzeugt wurden. Als weiteren Schritt nutzen die beiden Autoren in ihrem Artikel die Mittelwerte und die empirischen Varianzen der Experteneinschätzungen und erzeugen mit deren Hilfe Wahrscheinlichkeitsvorhersagen.

Ang et al. (2007) greifen einige der bisher beschriebenen Ansätze zur Vorhersage makroökonomischer Kennzahlen auf und führen eine Vergleichsstudie durch. Dazu erzeugen

sie für die Inflationsraten der USA verschiedenste Vorhersagen, unter anderem mit einem *autoregressive integrated moving average model* (ARIMA), siehe zum Beispiel Shumway & Stoffer (2011). Außerdem verwenden sie autoregressive Modelle unter Einbeziehung exogener Variablen. Bei den zusätzlichen Variablen handelt es sich um makroökonomische Daten, die durch die Phillips-Kurve motiviert werden. Weiterhin wird die Zinsstrukturkurve in linearer, nicht-linearer und Arbitrage-freier Variante genutzt und in vektorautoregressive Modelle eingebunden. Als weitere Vorhersagemethode verwenden Ang et al. (2007) ein Regressionsmodell, in das Umfrageergebnisse als erklärende Variablen eingehen. Dabei wird sowohl der Survey of Professional Forecasters berücksichtigt, als auch der Livingston und der Michigan Survey. Schließlich untersuchen Ang et al. (2007) auch Regime-Switching Models und die Kombination verschiedener Vorhersagemethoden.

Neben den bereits angesprochenen Ansätzen zur Vorhersage makroökonomischer Kennzahlen sei noch das *Faktorenmodell* zu nennen. Bei dieser Methode analysiert man mehrere Zeitreihen gleichzeitig und versucht gemeinsame Faktoren zu bestimmen, welche die Entwicklung der Zeitreihen beschreiben (siehe z.B. Shumway & Stoffer (2011)). Banbura & Rünstler (2011) nutzen dieses Verfahren zur Prognose des Bruttoinlandsproduktes der Eurozone und lassen in ihre Berechnungen sowohl Wirtschaftsdaten, wie beispielsweise den Index der Industrieproduktion nach verschiedenen Sektoren, als auch Umfrageergebnisse und Finanzdaten, wie beispielsweise Wechselkurse und Zinssätze, einfließen. Da die einbezogenen Zeitreihen teilweise auf monatlicher oder sogar auf täglicher Basis vorliegen, können Banbura & Rünstler (2011) in jedem Monat neue, präzisere Vorhersagen für das Bruttoinlandsprodukt erzeugen, obwohl dieses nur quartalsweise veröffentlicht wird. Außerdem sei noch zu erwähnen, dass Banbura & Rünstler (2011) in ihr Modell eine Autoregression integrieren, sodass auch entsprechende Lags der Faktoren in die Vorhersage einfließen. Antipa et al. (2012) greifen bei ihren Analysen des Bruttoinlandsproduktes der Bundesrepublik Deutschland auf ein ähnliches Faktorenmodell zurück und nutzen Wirtschaftsdaten, Umfrageergebnisse und Finanzdaten für die Bestimmung der Faktoren. Die ermittelten Prognosen vergleichen sie mit den Vorhersagen eines Bridge-Modells, welches einem autoregressiven Modell mit exogenen Variablen entspricht, das ebenfalls auf monatlicher Basis Vorhersagen erzeugt und somit monatliche Inputvariablen und quartalsweise Daten des Bruttoinlandsproduktes verknüpft. Schließlich seien noch Schumacher & Breitung (2008) zu nennen, die ebenfalls Prognosen für das Bruttoinlandsprodukt der Bundesrepublik Deutschland mit Hilfe von Faktorenmodellen erzeugen und dabei verschiedene Vorhersagemethoden vergleichen, welche die ermittelten Faktoren in unterschiedlicher Art und Weise in die Vorhersage eingehen lassen.

9. Zusammenfassung und Fazit

Zu Beginn der Diplomarbeit haben wir uns mit *scoring rules* auseinandergesetzt. In diesem theoretischen Teil der Arbeit wurde deutlich, dass für eine sinnvolle Bewertung von Wahrscheinlichkeitsvorhersagen *proper scoring rules* verwendet werden sollten. Hervorzuheben wäre hier der *continuous ranked probability score* (CRPS) aus Abschnitt 3.2.3, da für diesen score die in Kapitel 4 vorgestellten gewichteten Versionen existieren. Diese können durch Einbeziehung von Gewichtungsfunktionen spezielle Bereiche der Dichtevorhersage besonders kritisch betrachten. Darüber hinaus ist mit einer speziellen Zerlegung des CRPS eine grafische Begutachtung der Vorhersage Performance von Vorhersagemethoden möglich, sodass sich der CRPS als ein vielfältiges Analyseinstrument erweist. Man sollte allerdings auch beachten, dass neben den von uns verwendeten *scoring rules* *quadratic score*, *pseudospherical score*, *logarithmic score* und CRPS noch viele weitere Bewertungsmethoden existieren und dass uns mit den erläuterten Ansätzen zur Konstruktion von *proper scoring rules* noch viele Möglichkeiten offen stehen.

Im praktischen Teil der Arbeit, in dem wir Dichtevorhersagen für die makroökonomischen Kennzahlen Inflationsrate, Bruttoinlandsprodukt und Arbeitslosenquote der Bundesrepublik Deutschland untersucht haben, stellten sich die *scoring rules* als sinnvolle Hilfsmittel heraus. Etliche unserer Vermutungen wurden durch die *scoring rules* bestätigt und wir konnten unter anderem durch den Test auf gleiche Vorhersage Performance und durch die Schwellenwert Zerlegung des CRPS zusätzliche Erkenntnisse gewinnen. Allerdings wurde bei unserer Analyse auch deutlich, dass sich die *scoring rules* bei den Bewertungen unterscheiden und teilweise unterschiedliche Vorhersageverfahren favorisierten. Man sollte deshalb niemals nur auf eine Bewertungsmethode vertrauen, sondern mehrere *scoring rules* für die Beurteilung von Wahrscheinlichkeitsvorhersagen heranziehen. Außerdem würde sich eine genauere Untersuchung der einzelnen *scoring rules* anbieten, um die unterschiedlichen Bewertungen der *scoring rules* besser interpretieren zu können.

Über unsere Vorhersagemethoden lässt sich sagen, dass sich keines dieser Verfahren bezüglich seiner Vorhersagequalitäten von den anderen Verfahren besonders absetzen konnte. Vielmehr ist es bei der Vorhersage von makroökonomischen Kennzahlen sehr wichtig, jede Kennzahl individuell zu betrachten. So wurde beispielsweise bei der Analyse der Arbeitslosenquote deutlich, dass ein Zeitreihenmodell die vorhandenen zeitlichen Abhängigkeiten innerhalb der Zeitreihe zu seinem Vorteil ausnutzen konnte. Im Gegensatz dazu spielte bei der Vorhersage der Inflationsrate und des Bruttoinlandsproduktes die Form der vorhergesagten Dichten eine besondere Rolle, da entsprechende Ausreißer berücksichtigt werden mussten.

Außerdem konnten wir feststellen, dass neben der Datenhistorie einer Zeitreihe die Aufnahme zusätzlicher Informationen in das Vorhersagemodell sinnvoll sein kann. Man sollte sich deshalb darüber Gedanken machen, welche zusätzlichen Informationen aus theoretischer Sicht zur Verbesserung der Vorhersagequalität verwendet werden können. Die in Kapitel 8 vorgestellten Ansätze bieten uns hierfür entsprechende Möglichkeiten und zeigen auch alternative Modellierungsansätze auf.

Schließlich sollte einem aber auch bewusst sein, dass mathematische Vorhersagemethoden immer auf theoretischen Modellen beruhen. Diese unterstellen gewisse Annahmen an ihre Inputvariablen und können auch nur eine begrenzte Anzahl an Informationen berücksichtigen. Die perfekte Vorhersage einer makroökonomischen Kennzahl wird es deshalb nie geben und man sollte immer hinterfragen, ob die Annahmen und die mathematische Vorhersage aus wirtschaftlicher Sicht sinnvoll erscheinen.

A. Anhang

A.1. Herleitungen zu den Beispielen aus Abschnitt 3.1.3

Im Folgenden wird jeweils die Herleitung der scoring rule und der Bregman divergence aus der expected score function beschrieben. Bei dem Nachweis, dass die expected score functions konvex sind, sei auf den nächsten Abschnitt über Beispiele in stetigen Ereignisräumen verwiesen, da die Nachweise äquivalent erfolgen können.

Seien $p = (p_1, \dots, p_m)$ und $q = (q_1, \dots, q_m)$ Wahrscheinlichkeitsmaße aus \mathcal{P}_m .

A.1.1. Quadratic score

$$G(p) = \sum_{j=1}^m p_j^2 \quad G'(p) = (2p_1, \dots, 2p_m)$$

$$\begin{aligned} S(p, i) &= G(p) - \langle G'(p), p \rangle + G'_i(p) \\ &= \sum_{j=1}^m p_j^2 - \langle (2p_1, \dots, 2p_m), (p_1, \dots, p_m) \rangle + 2p_i \\ &= \sum_{j=1}^m p_j^2 - 2 \sum_{j=1}^m p_j^2 + 2p_i \\ &= 2p_i - \sum_{j=1}^m p_j^2 \end{aligned}$$

$$\begin{aligned} d(p, q) &= S(q, q) - S(p, q) \\ &= \sum_{i=1}^m q_i \left(2q_i - \sum_{j=1}^m q_j^2 \right) - \sum_{i=1}^m q_i \left(2p_i - \sum_{j=1}^m p_j^2 \right) \\ &= \sum_{i=1}^m q_i \left(\sum_{j=1}^m (p_j^2 - q_j^2) + 2(q_i - p_i) \right) \\ &= 1 \sum_{j=1}^m (p_j^2 - q_j^2) + 2 \sum_{i=1}^m q_i^2 - 2 \sum_{i=1}^m q_i p_i \\ &= \sum_{j=1}^m (p_j - q_j)^2 \end{aligned}$$

A.1.2. Pseudospherical score

$$G(p) = \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}} \quad G'(p) = \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} (p_1^{\alpha-1}, \dots, p_m^{\alpha-1}) \quad , \alpha > 1$$

$$\begin{aligned} S(p, i) &= G(p) - \langle G'(p), p \rangle + G'_i(p) \\ &= \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}} - \left\langle \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} (p_1^{\alpha-1}, \dots, p_m^{\alpha-1}), (p_1, \dots, p_m) \right\rangle \\ &\quad + \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} p_i^{\alpha-1} \\ &= \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}} - \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} \sum_{j=1}^m p_j^\alpha + \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} p_i^{\alpha-1} \\ &= \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} p_i^{\alpha-1} \end{aligned}$$

$$\begin{aligned} d(p, q) &= S(q, q) - S(p, q) \\ &= \sum_{i=1}^m q_i \left(\sum_{j=1}^m q_j^\alpha \right)^{\frac{1}{\alpha}-1} q_i^{\alpha-1} - \sum_{i=1}^m q_i \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} p_i^{\alpha-1} \\ &= \left(\sum_{j=1}^m q_j^\alpha \right)^{\frac{1}{\alpha}} - \left(\sum_{j=1}^m p_j^\alpha \right)^{\frac{1}{\alpha}-1} \sum_{i=1}^m q_i p_i^{\alpha-1} \end{aligned}$$

A.1.3. Logarithmic score

$$G(p) = \sum_{j=1}^m p_j \log p_j \quad G'(p) = (\log p_1 + 1, \dots, \log p_m + 1)$$

$$\begin{aligned} S(p, i) &= G(p) - \langle G'(p), p \rangle + G'_i(p) \\ &= \sum_{j=1}^m p_j \log p_j - \langle (\log p_1 + 1, \dots, \log p_m + 1), (p_1, \dots, p_m) \rangle + \log p_i + 1 \\ &= \sum_{j=1}^m p_j \log p_j - \sum_{j=1}^m (\log p_j + 1) p_j + \log p_i + 1 \\ &= \sum_{j=1}^m p_j \log p_j - \sum_{j=1}^m p_j \log p_j - \sum_{j=1}^m p_j + \log p_i + 1 \\ &= \log p_i \end{aligned}$$

$$\begin{aligned} d(p, q) &= S(q, q) - S(p, q) \\ &= \sum_{i=1}^m q_i \log q_i - \sum_{j=1}^m q_j \log p_j \\ &= \sum_{i=1}^m q_i \log \frac{q_i}{p_i} \end{aligned}$$

A.2. Herleitung zu den Beispielen aus Abschnitt 3.2.2

Im Folgenden wird jeweils die Herleitung der expected score function und der Bregman divergence beschrieben. Außerdem erfolgt der Nachweis, dass die jeweilige expected score function konvex ist. Seien dazu p und q die Dichten der Wahrscheinlichkeitsmaße P bzw. Q bezüglich des Maßes μ .

A.2.1. Quadratic score

$$\begin{aligned}
 G(p) &= \int_{\Omega} \text{QuadS}(p, \omega) dP(\omega) & d(p, q) &= \text{QuadS}(q, q) - \text{QuadS}(p, q) \\
 &= \int_{\Omega} 2p(\omega) - \|p\|_2^2 dP(\omega) & &= \|q\|_2^2 - \left(\int_{\Omega} 2p(\omega) - \|p\|_2^2 dQ(\omega) \right) \\
 &= 2 \int_{\Omega} p(\omega) p(\omega) d\mu(\omega) - \|p\|_2^2 & &= \int_{\Omega} q(\omega)^2 - 2p(\omega) q(\omega) + p(\omega)^2 d\mu(\omega) \\
 &= \|p\|_2^2 & &= \|p - q\|_2^2
 \end{aligned}$$

Sei $\lambda \in (0, 1)$, dann folgt die Konvexität aufgrund der Minkowski Ungleichung (siehe z.B. Elstrodt (2011)):

$$\begin{aligned}
 G(\lambda p + (1 - \lambda) q) &= \|\lambda p + (1 - \lambda) q\|_2^2 \\
 &= \int (\lambda p + (1 - \lambda) q)^2 d\mu \\
 &\leq \lambda^2 \int p^2 d\mu + (1 - \lambda)^2 \int q^2 d\mu \\
 &< \lambda \int p^2 d\mu + (1 - \lambda) \int q^2 d\mu \\
 &= \lambda G(p) + (1 - \lambda) G(q)
 \end{aligned}$$

A.2.2. Pseudospherical score

$$\begin{aligned}
 G(p) &= \int_{\Omega} \text{PseudoS}(p, \omega) dP(\omega) & d(p, q) &= \text{PseudoS}(q, q) - \text{PseudoS}(p, q) \\
 &= \int_{\Omega} \frac{p(\omega)^{\alpha-1}}{\|p\|_{\alpha}^{\alpha-1}} p(\omega) d\mu(\omega) & &= \|q\|_{\alpha} - \int_{\Omega} \frac{p(\omega)^{\alpha-1}}{\|p\|_{\alpha}^{\alpha-1}} q(\omega) d\mu(\omega) \\
 &= \frac{\|p\|_{\alpha}^{\alpha}}{\|p\|_{\alpha}^{\alpha-1}} \\
 &= \|p\|_{\alpha}
 \end{aligned}$$

Der Nachweis der Konvexität von G erfolgt ebenfalls, wie beim quadratic score, über die Minkowski Ungleichung und wird deshalb nicht aufgeführt.

A.2.3. Logarithmic score

$$\begin{aligned}
 G(p) &= \int_{\Omega} \text{LogS}(p, \omega) \, dP(\omega) & d(p, q) &= \text{LogS}(q, q) - \text{LogS}(p, q) \\
 &= \int_{\Omega} \log p(\omega) p(\omega) \, d\mu(\omega) & &= \int_{\Omega} \log q(\omega) q(\omega) \, d\mu(\omega) \\
 & & &\quad - \int_{\Omega} \log p(\omega) q(\omega) \, d\mu(\omega) \\
 & & &= \int_{\Omega} \log \left(\frac{q(\omega)}{p(\omega)} \right) q(\omega) \, d\mu(\omega)
 \end{aligned}$$

Um die Konvexität der expected score function zu beweisen, zeigen wir, dass die Kullback-Leibler divergence stets größer oder gleich null ist. Daraus folgt, dass der logarithmic score (strictly) proper ist. Aus Theorem 3.2.1 folgern wir dann die Konvexität der expected score function.

Behauptung:

$$d(p, q) = \int_{\Omega} \log \left(\frac{q(\omega)}{p(\omega)} \right) q(\omega) \, d\mu(\omega) \geq 0 \quad (\text{A.2.1})$$

Der folgende Beweis stammt aus Decker (2005).

Beweis:

Die Jensensche Ungleichung besagt für eine konvexe Funktion ϕ (Elstrodt, 2011):

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)), \quad X \in \mathcal{L}_1$$

Da $-\log$ konvex ist, folgt mit der Jensenschen Ungleichung und den Logarithmischen Gesetzen:

$$\begin{aligned}
 d(p, q) &= \int_{\Omega} \log \left(\frac{q(\omega)}{p(\omega)} \right) q(\omega) \, d\mu(\omega) \\
 &= \mathbb{E}_Q \left(-\log \left(\frac{p(\omega)}{q(\omega)} \right) \right) \\
 &\geq -\log \left(\mathbb{E}_Q \left(\frac{p(\omega)}{q(\omega)} \right) \right) \\
 &= -\log \left(\int_{\Omega} \frac{p(\omega)}{q(\omega)} q(\omega) \, d\mu(\omega) \right) \\
 &= -\log(1) \\
 &= 0
 \end{aligned}$$

□

Nun zeigen wir noch, dass in der Ungleichung (A.2.1) genau dann Gleichheit herrscht, wenn $p(\omega) = q(\omega)$ für alle $\omega \in \Omega$ gilt.

Beweis :

„ \Leftarrow “: Dies folgt direkt aus der Definition der Kullback-Leibler divergence.

„ \Rightarrow “:

Es gilt folgende Ungleichung, die wir im weiteren Verlauf verwenden werden:

$$x - 1 \geq \log(x), \quad x > 0.$$

Dabei gilt Gleichheit genau dann, wenn $x = 1$ ist. Es folgt:

$$d(p, q) = - \int_{\Omega} \log \left(\frac{p(\omega)}{q(\omega)} \right) q(\omega) \, d\mu(\omega) \geq - \int_{\Omega} \left(\frac{p(\omega)}{q(\omega)} - 1 \right) q(\omega) \, d\mu(\omega)$$

Die Gleichheit

$$\int_{\Omega} \log \left(\frac{p(\omega)}{q(\omega)} \right) q(\omega) \, d\mu(\omega) = \int_{\Omega} \left(\frac{p(\omega)}{q(\omega)} - 1 \right) q(\omega) \, d\mu(\omega) = 0$$

tritt folglich nur ein, wenn

$$\log \left(\frac{p(\omega)}{q(\omega)} \right) = \frac{p(\omega)}{q(\omega)} - 1$$

und damit

$$\begin{aligned} \frac{p(\omega)}{q(\omega)} &= 1 \\ p(\omega) &= q(\omega) \end{aligned}$$

gilt. □

Literaturverzeichnis

- Amisano, G. & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, 25(2), 177–190.
- Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4), 1163–1212.
- Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about gdp growth? *Journal of Econometrics*, 131(1), 359–403.
- Antipa, P., Barhoumi, K., Brunhes-Lesage, V., & Darné, O. (2012). Nowcasting german gdp: A comparison of bridge and factor models. *Journal of Policy Modeling*, 34(6), 864–878.
- Bachmeier, L. J. & Swanson, N. R. (2005). Predicting inflation: does the quantity theory help? *Economic Inquiry*, 43(3), 570–585.
- Banbura, M. & Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting gdp. *International Journal of Forecasting*, 27(2), 333–346.
- Baringhaus, L. & Franz, C. (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 88(1), 190–206.
- Blachard, O. & Illing, G. (2009). *Makroökonomie*. Pearson Studium.
- Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3), 200–217.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3.
- Bröcker, J. & Smith, L. A. (2007). Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, 22(2), 382–388.
- Buja, A., Stuetzle, W., & Shen, Y. (2005). Loss functions for binary class probability estimation and classification: Structure and applications. unveröffentlichtes manuskript vom 3. november 2005. erhältlich unter <http://www-stat.wharton.upenn.edu/~buja/>.
- Cervera, J. L. & Muñoz, J. (1996). Proper scoring rules for fractiles. in *Bayesian Statistics 5*, Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., eds., pp. 513-519. Oxford University Press, Oxford.

- Clements, M. P. & Smith, J. (2000). Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. *Journal of Forecasting*, 19(4), 255–276.
- Decker, T. (2005). *Datenklassifikation mittels Bayestechniken*. Diplomarbeit, angefertigt am Fraunhofer ITWM in Kaiserslautern, Technische Fachhochschule Berlin, Germany.
- Diebold, F. X. & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Elliott, G. & Timmermann, A. (2008). Economic forecasting. *Journal of Economic Literature*, 46(1), 3–56.
- Elstrodt, J. (2011). *Maß- und Integrationstheorie*. Springer-Lehrbuch. Springer.
- Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8(6), 985–987.
- EU (2004). *Harmonisierte Verbraucherpreisindizes (HVPI) - Ein kurzer Leitfaden für Datennutzer*. Europäische Gemeinschaft - Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften.
- Fahrmeir, L., Künstler, R., Pigeot, I., & Tutz, G. (2007). *Statistik*. Springer-Lehrbuch. Springer London, Limited.
- Gerds, T. (2002). *Nonparametric efficient estimation of prediction error for incomplete data models*. Ph.D. Thesis, Mathematische Fakultät, Albert-Ludwigs-Universität Freiburg, Germany.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746–762.
- Gneiting, T. & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Gneiting, T. & Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3), 411–422.
- Gneiting, T. & Thorarinsdottir, T. L. (2010). Predicting inflation: Professional experts versus no-change forecasts. *ArXiv e-prints 1010.2318*.
- Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14(1), 107–114.
- Good, I. J. (1971). Comment on “Measuring Information and Uncertainty“ by Robert J. Buehler. In *Foundations of Statistical Inference*, Godambe, V. P. and Sprott, D. A., eds., pp. 337-339. Holt, Rinehart and Winston, Toronto.

- Grünwald, P. D. & Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust bayesian decision theory. *Annals of Statistics*, 32(4), 1367–1433.
- Hendrickson, A. D. & Buehler, R. J. (1971). Proper scores for probability forecasters. *The Annals of Mathematical Statistics*, 42(6), 1916–1921.
- Kozicki, S. (1997). Predicting real growth and inflation with the yield spread. *Economic Review - Fourth Quarter 1997 - Federal Reserve Bank of Kansas City*, (pp. 39–57).
- Kreiß, J. P. & Neuhaus, G. (2006). *Einführung in Die Zeitreihenanalyse*. Statistik und ihre Anwendungen. Springer London, Limited.
- Laio, F. & Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences*, 11(4), 1267–1277.
- Matheson, J. E. & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, 22(10), 1087–1096.
- McCarthy, J. (1956). Measures of the value of information. *Proceedings of the National Academy of Sciences*, 42(9), 654–655.
- Rockafellar, R. T. (1970). *Convex Analysis*. Princeton and NJ: Princeton University Press.
- Rothman, P. (1998). Forecasting asymmetric unemployment rates. *The Review of Economics and Statistics*, 80(1), 164–168.
- Roulston, M. S. & Smith, L. A. (2002). Evaluating probabilistic forecasts using information theory. *Monthly Weather Review*, 130, 1653–1660.
- Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336), 783–801.
- Scheide, J. & Trabandt, M. (2000). Predicting inflation in euroland : the pstar approach. *Kiel Working Papers - Kiel Institute for the World Economy (IfW)*, (1019).
- Schlittgen, R. (2012). *Angewandte Zeitreihenanalyse mit R*. Lehr- und Handbücher der Statistik. Oldenbourg Verlag.
- Schumacher, C. & Breitung, J. (2008). Real-time forecasting of german gdp based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, 24(3), 386–398.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423 and 623–656.
- Shumway, R. H. & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*. Springer Texts in Statistics. Springer.

Székel, G. J. & Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1), 58–80.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Diplomarbeit eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in dieser oder in einer ähnlichen Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Marburg, den 27. März 2013

Martin Reuter

Danksagung

An dieser Stelle möchte ich Herrn Prof. Dr. Hajo Holzmann für die Vergabe des sehr interessanten Themas und für die ausgezeichnete Betreuung danken. Weiterhin danke ich meiner Familie und meinen Freunden, die mich während meines Studiums begleitet und unterstützt haben. Ein ganz besonderer Dank gilt meinen Eltern, ohne deren Hilfe dieses Studium nicht möglich gewesen wäre. Vielen Dank!