



Ludwig-Maximilians-Universität
Fakultät für Mathematik, Informatik und Statistik
Institut für Statistik

Masterthesis

**Modellierung der Spielstärke von
Fußballmannschaften mit ordinalen
Paarvergleichsmodellen**

Bernhard Bauer

Betreuer: Prof. Dr. Gerhard Tutz,
Gunther Schauburger

29. Februar 2016

Abstract

Die vorliegende Arbeit zeigt verschiedene Ansätze zur Modellierung der Spielstärke von Bundesligisten mit Hilfe von ordinalen Paarvergleichsmodellen auf. Die Ansätze beruhen hauptsächlich auf den Arbeiten von Tutz und Schauburger (2015) und Cattelan et al. (2013). Zunächst wird ein allgemeines Paarvergleichsmodell vorgestellt, welches anschließend für ordinalen Response und einen globalen, bzw. teamspezifischen Heimeffekt erweitert wird. Es zeigt sich, dass sich die Modelle mit globalem und teamspezifischem Heimeffekt nicht signifikant voneinander unterscheiden.

Regularisierungstechniken werden angewandt, um Cluster für die einzelnen Koeffizientenschätzer der Mannschaften herauszufinden. Die stärksten Mannschaften bilden meist einzelne Cluster, wohingegen die restlichen Mannschaften sich nur auf wenige Cluster aufteilen. Anschließend wird das Modell so erweitert, dass teamspezifisch-erklärende Variablen mit aufgenommen werden können. Es wird festgestellt, dass die ausgewählten Variablen einen nicht unerheblichen Teil der ermittelten Spielstärken begründen.

Um eine dynamische Struktur der Spielstärken zu erhalten werden zwei weitere Ansätze betrachtet. Beim ersten Ansatz wird die Spielstärke durch einen exponential-gewichteten moving-average Prozess ermittelt. Der zweite Ansatz berechnet für jeden Spieltag das bereits bekannte erweiterte allgemeine Paarvergleichsmodell mit zusätzlicher Gewichtung der Likelihoods. Der rank probability score (rps) bietet eine Möglichkeit, die verschiedenen Modelle miteinander zu vergleichen. Das Modell mit den gewichteten Likelihoods liefert hierbei die besten Ergebnisse.

Abschließend werden die prädiktiven Fähigkeiten der Modelle überprüft. Der rps der verschiedenen Modelle unterscheidet sich dabei kaum. Bei der Prognose von Spielausgängen prognostiziert das beste Modell bei einer Simulationsstudie im Schnitt 42.3% der Ergebnisse richtig. Bei Tipp auf das wahrscheinlichste Ergebnis liegt die beste Trefferquote bei über 50%. Werden zusätzlich die Wettquoten für die einzelnen Spiele berücksichtigt, lässt sich feststellen, dass sich durch die Simulationsstudien keine rentablen Gewinne erzielen lassen. Tipps auf das wahrscheinlichste Ergebnis liefern teilweise gute Renditen. Unter Berücksichtigung der Rentabilität eines Tipps können diese noch weiter gesteigert werden. Die höchste Rendite mit 28.9% wird bei Berücksichtigung der Rentabilität und ausschließlichen Tipps auf Heimerfolge erreicht.

Inhaltsverzeichnis

Abbildungsverzeichnis	V
Tabellenverzeichnis	VI
1 Einleitung	1
2 Datenbeschreibung	4
3 Statisches Bradley-Terry Modell	7
3.1 Das binäre Bradley-Terry Modell	7
3.2 Erweiterung zum ordinalen Modell	8
3.3 Heimspieleffekt	9
3.3.1 Globaler Heimeffekt	9
3.3.2 Teamspezifischer Heimeffekt	11
3.4 Zusammenhang zum GLM	11
3.5 Ergebnisse	12
4 Clustering durch Regularisierung	16
4.1 Ranking LASSO	16
4.2 Auswahlkriterien	18
4.3 Ergebnisse	19
5 Modelle mit mannschaftsspezifischen Einflussgrößen	24
5.1 Modellerweiterung für erklärende Variablen	24
5.2 Ergebnisse	26
6 Dynamische Bradley-Terry Modellierung	35
6.1 Modellansatz nach Cattelan	35
6.2 Modellvalidierung	39
6.3 Dynamischer Ansatz über gewichtete Likelihoodfunktion	40
6.4 Anwendung	41
6.4.1 EWMA-Prozesse	41
6.4.2 Gewichtete Likelihood	47

Inhaltsverzeichnis

7	Prädiktiver Modellvergleich	52
7.1	Rank probability scores	53
7.2	Ergebnisvorhersage	56
7.2.1	Simulationsstudie	56
7.2.2	Tipp der höchsten Wahrscheinlichkeit	57
7.3	Wettquoten und Gewinnmöglichkeiten	58
7.3.1	Simulationsstudie	59
7.3.2	Tipp der höchsten Wahrscheinlichkeit	60
7.3.3	Tipp nach Rentabilität	61
8	Zusammenfassung und Diskussion	64
A	Verwendete Software	68
B	Anhang	69
C	Inhalt der CD	78
D	Eidesstattliche Erklärung	79
	Literaturverzeichnis	80

Abbildungsverzeichnis

3.1	Korrelation Punkte/Fähigkeiten	14
4.1	Koeffizientenpfade bei globalem Heimeffekt	20
4.2	Koeffizientenpfade bei teamspezifischem Heimeffekt	22
5.1	Koeffizientenpfade mit erklärender Variable Marktwert	27
5.2	Koeffizientenpfade mit erklärender Variable Vereinsbudget	29
5.3	Marktwert/Spielstärke	30
5.4	Koeffizientenpfade mit erklärender Variable Startaufstellung	32
5.5	Startaufstellung/Spielstärke	34
6.1	Spielstärkeentwicklung gemeinsame Startwerte	43
6.2	Spielstärkeentwicklung individuelle Startwerte	45
6.3	RPS-Mediane zur Koeffizientenbestimmung	47
6.4	Spielstärkeentwicklung gewichtete Likelihood	49
6.5	Vergleich verschiedener RPS	51
7.1	Vergleich RPS prädiktive Modelle	53
B.1	Koeffizientenpfade bei globalem Heimeffekt (k=3)	70
B.2	Koeffizientenpfade bei teamspezifischem Heimeffekt (k=3)	72
B.3	Koeffizientenpfade mit erklärender Variable Marktwert	74
B.4	Spielstärkeentwicklung gemeinsame Startwerte Rest	75
B.5	Spielstärkeentwicklung individuelle Startwerte Rest	76
B.6	Spielstärkeentwicklung gewichtete Likelihood Rest	77

Tabellenverzeichnis

2.1	Abschlusstabelle Saison 2014/2015	5
4.1	Cluster bei globalem Heimeffekt (BIC)	21
4.2	Cluster bei teamspezifischem Heimeffekt (BIC)	23
5.1	Cluster bei durchschnittlichem Marktwert (BIC)	28
5.2	Cluster bei durchschnittlichem Marktwert der Startaufstellungen (BIC)	32
6.1	Startwerte Heim- bzw. Auswärtsspielstärke	37
6.2	Vergleich statische und dynamische Spielstärken	44
6.3	Koeffizienten dynamisches Modell Gladbach	50
7.1	Kennwerte der einzelnen Modelle	54
7.2	Simulationsstudie richtige Ergebnisse	56
7.3	richtige Ergebnisse bei höchster Wahrscheinlichkeit	58
7.4	Simulationsstudie Gewinnberechnung	60
7.5	Gewinne bei höchster Wahrscheinlichkeit	61
7.6	Gewinne bei höchster Wahrscheinlichkeit und Rentabilität	62
7.7	Gewinne bei Tipp auf höchste Rentabilität	63
B.1	Cluster bei globalem Heimeffekt (AIC)	69
B.2	Cluster bei teamspezifischem Heimeffekt (AIC)	71
B.3	Lizenzspieleretats und Marktwerte der einzelnen Vereine	73

1 Einleitung

Verfolgt man verschiedene Sportveranstaltungen vor dem TV-Gerät, so fällt auf, dass hierbei der Zuschauer mit einer riesigen Anzahl an Statistiken überflutet wird. Mit Hilfe dieser Einblendungen versuchen die Kommentatoren oftmals die Überlegenheit eines Sportlers oder einer Mannschaft gegenüber ihren Gegnern zu belegen. Meist sind diese Statistiken jedoch nichtssagend und haben für die betroffenen Personen keinerlei Aussagekraft. Die statistische Modellierung hingegen nimmt im professionellen Sportbereich eine immer wichtigere Rolle ein. Im Baseball wurden Mitte des 20. Jahrhunderts erstmals statistische Daten verwendet, um die Stärke von Spielern zu ermitteln. Dieses entwickelte System wird mit dem Begriff *Sabermetrics* bezeichnet. Es wurde von den meisten Experten sehr kritisch gesehen und fand kaum Anwendung. Billy Beane, Manager des Baseballteams Oakland Athletics, verwendete ab dem Jahr 2000 dieses System, um Spieler zu finden, welche ein überdurchschnittliches Preis-/Leistungsverhältnis aufweisen. Das Team erreichte trotz geringer finanzieller Mittel viermal in Folge die Playoffs.

Auch im Fußball wird mittlerweile mit statistischen Modellen gearbeitet. Matthew Benham erkannte das Potential von statistischen Modellen und wurde so zum Millionär. Er gründete 2004 das Unternehmen Smartodds, welches professionell auf den Spielausgang bei Fußballspielen wettet. Mittlerweile ist er Eigentümer des englischen Vereins FC Brentford und des dänischen Vereins FC Midtjylland. Bei diesen Vereinen werden seitdem ausschließlich Spieler verpflichtet, welche von ihrem eigens entwickelten statistischen Modell als geeignet klassifiziert werden. Auch während des Spiels bekommen die Trainer der Mannschaften Informationen vom System zugesandt, anhand welcher sie ihre Taktik dem Spielverlauf anpassen. So kam das Modell beispielsweise zu der Erkenntnis, dass die führende Mannschaft beim Fußball ihren Vorsprung am besten verteidigen kann, indem sie weiterhin offensiv spielt und nicht versucht das Ergebnis zu verwalten. Das Modell scheint Erfolg zu haben. Der FC Midtjylland wurde 2015 dänischer Meister, obwohl die beiden größten Vereine des Landes, Brøndby Kopenhagen und der FC Kopenhagen, einen doppelt so hohen Etat zur Verfügung hatten. Auch die großen Fußballvereine Europas haben mittlerweile das immense Potential von statistischen Modellen erkannt und vereinseigene Statistikabteilungen gebildet. Der Einfluss statistischer Modelle im Profisportbereich wird daher in den nächsten Jahren noch weiter zunehmen.

Diese Arbeit beschäftigt sich mit der Modellierung der Spielstärke von Fußballmannschaften mit ordinalen Paarvergleichsmodellen. Nach einer kurzen Vorstellung der verwendeten

1 Einleitung

Daten in Kapitel 2 widmet sich Kapitel 3 mit dem klassischen Bradley-Terry Modell. Dieses wird zu einem ordinalen Modell erweitert. Zusätzlich werden verschiedene Annahmen für dieses ordinale Modell, wie zum Beispiel ein globaler oder auch ein teamspezifischer Heimspieleffekt, vorgestellt. Im Anwendungsteil dieses Kapitels finden sich dann die Ergebnisse einer solchen Betrachtungsweise wieder.

Kapitel 4 erweitert das vorgestellte Modell mit Hilfe von Regularisierungstechniken. Hierbei wird angenommen, dass Mannschaften teilweise gleiche Spielstärken besitzen und sich somit in verschiedene Cluster einteilen lassen. Diese Einteilung geschieht mittels Erweiterung der Likelihood um einen Strafterm für alle paarweisen Spielstärkekoeffizienten. Verschiedene Auswahlkriterien dienen zur Ermittlung eines Tuningparameters, durch welchen die Koeffizienten bestimmt werden können. Im Anwendungsteil werden wiederum die Ergebnisse dieses Modells betrachtet.

Im darauffolgenden Kapitel 5 wird das Modell nun um erklärende Variablen erweitert. Zum einen soll somit die Modellgüte festgestellt werden und zum anderen wird sich eine Verbesserung des Modells für spätere prädiktive Analysen erhofft. Zunächst wird in diesem Kapitel das Modell so erweitert, dass die durchschnittlichen Marktwerte der einzelnen Vereine als Variablen ins Modell aufgenommen werden. Durch Aufnahme dieser Variablen lassen sich die Spielstärken der Mannschaften bei Annahme von gleichen Marktwerten aller Vereine ermitteln und somit auch Aussagen über die Modellgüte treffen. Ein anderer Ansatz betrachtet die Marktwerte der einzelnen Startaufstellungen der Mannschaften. Da die Vereine oftmals ihre Startaufstellungen von Spiel zu Spiel ändern, ergibt sich bei dieser Betrachtung eine dynamische Variable für jedes Spiel. Außerdem erhofft man sich einen zusätzlichen Informationsgewinn für die prädiktiven Betrachtungen, weil die Startaufstellungen bereits vor Spielbeginn veröffentlicht werden.

Kapitel 6 erweitert das Modell zu einem dynamischen Modell, bei welchem sich die Spielstärken der Mannschaften für jedes Spiel verändern. Zunächst wird dabei ein von Cattelan et al. (2013) vorgeschlagener Ansatz, welcher die Spielstärken durch einen exponentialgewichteten moving-average Prozess modelliert, betrachtet. Bei diesem Ansatz gibt es eine strikte Unterscheidung zwischen Heimspielstärke und Auswärtsspielstärke eines Teams. Des Weiteren findet eine Unterscheidung zwischen teamspezifischen und gemeinsamen Startwerten statt. Eine weitere Möglichkeit der dynamischen Modellierung bietet die Betrachtung von gewichteten Likelihoods zur Ermittlung der Spielstärkekoeffizienten. Die Annahme hierbei besteht darin, dass weiter zurückliegende Spiele weniger Einfluss auf die aktuelle Spielstärke eines Teams haben als erst kürzlich gespielte Partien. Zusätzlich wird in diesem Kapitel noch eine Möglichkeit der Modellvalidierung über rank probability

1 Einleitung

scores vorgestellt. Der Anwendungsteil fasst wiederum die einzelnen Ergebnisse dieser Modelle zusammen.

Nach Vorstellung der verschiedenen Modellansätze sollen in Kapitel 7 die Ergebnisse der einzelnen Modelle miteinander verglichen werden. Es wird dabei hauptsächlich Augenmerk auf die prädiktive Aussagekraft der einzelnen Modelle gelegt. Mit Hilfe von Simulationsstudien wird untersucht, welches Modell die meisten Spiele richtig vorhersagt und zusätzlich wird überprüft, ob und wenn ja wie viel Geld sich mit diesen Ansätzen hätte in der Saison 2014/2015 verdienen lassen.

Abschließend werden die Ergebnisse nochmals kurz zusammengefasst und diskutiert. Ein Überblick über die verwendeten Pakete und Codes, sowie zusätzliche Grafiken und Tabellen sind im Appendix dieser Arbeit beigefügt.

2 Datenbeschreibung

Als Datengrundlage für die im Weiteren Verlauf dieser Arbeit vorgestellten Analysen dienen die Ergebnisse der 1. Fußballbundesliga in der Saison 2014/2015. Die Liga besteht aus 18 teilnehmenden Mannschaften, wobei die zwei schwächsten Teams in die 2. Bundesliga absteigen und in der nächsten Saison durch die beiden punktbesten Teams aus dieser Liga ersetzt werden. Zusätzlich muss der Tabellensechzehnte der abgelaufenen Saison gegen den Tabellendritten der 2. Liga in zwei Entscheidungsspielen um den Verbleib in der Liga spielen. Eine Saison besteht aus 34 Spieltagen mit jeweils neun Spielen. Insgesamt werden daher 306 Spiele ausgetragen, in denen jede Mannschaft je einmal zu Hause und einmal auswärts bei jeder anderen Mannschaft antreten muss. Die betrachtete Saison dauerte vom 22. August 2014 bis zum 23. Mai 2015. Insgesamt wurden dabei Spiele an 91 verschiedenen Tagen ausgetragen. Für einen Sieg erhält eine Mannschaft drei Punkte. Das unterlegene Team erhält keinen Punkt. Bei einem Unentschieden erhalten beide Mannschaften je einen Punkt. Die Mannschaft mit den meisten Punkten nach 34 Spieltagen gewinnt die deutsche Meisterschaft. In der betrachteten Saison war diese Mannschaft der FC Bayern München mit 79 Punkten. Der SC Paderborn 07 und der SC Freiburg mussten mit den wenigsten Punkten in die 2. Bundesliga absteigen. Der Hamburger SV konnte sich in den Relegationsspielen gegen den Karlsruher SC durchsetzen und ist auch in der Saison 2015/2016 Mitglied der 1. Bundesliga. Tabelle 2.1 zeigt unter anderem die Endplatzierungen der einzelnen Vereine nach Punkten sowie ihre Heim- und Auswärtspunkte. Die restlichen Angaben dieser Tabelle werden in Abschnitt 3.5 erläutert.

Insgesamt wurden in der Saison 2014/2015 47.4% (145 Spiele) der Spiele von der Heimmannschaft gewonnen, 26.8% (82 Spiele) der Spiele endeten Unentschieden und 25.8% (79 Spiele) der Spiele gewann die Gastmannschaft. Für die späteren Analysen ist nicht nur interessant, welches Team ein Spiel gewonnen hat, sondern auch, mit wie vielen Toren Unterschied der Sieg erzielt wurde. Hier ist 69 mal ein Heimsieg mit genau einem Tor Unterschied zu verzeichnen. Die anderen 76 Heimsiege endeten mit zwei oder mehr Toren Unterschied. Den höchsten Heimsieg erzielte dabei der FC Bayern München am 21. Spieltag mit einem 8:0 gegen den Hamburger SV. Bei den Auswärtssiegen betrug bei 42 Spielen der Unterschied genau ein Tor und bei 37 Spielen mindestens zwei Tore. Den höchsten Auswärtssieg erzielte ebenfalls der FC Bayern München mit einem 6:0 Sieg über den SC Paderborn 07 am 22. Spieltag.

Neben den Ergebnissen dieser Saison werden im Verlauf der Arbeit noch zusätzliche frei

2 Datenbeschreibung

Mannschaft	Punkte	zu Hause	auswärts	Spielstärke	QSE	Rang
FC Bayern München	79	43	36	0.602	0.331	1
VfL Wolfsburg	69	43	26	0.000	0.311	2
Borussia Mönchengladbach	66	39	27	-0.164	0.307	3
Bayer 04 Leverkusen	61	36	25	-0.380	0.304	4
FC Augsburg	49	31	18	-1.037	0.299	6
FC Schalke 04	48	35	13	-1.087	0.299	7
Borussia Dortmund	46	30	16	-0.939	0.299	5
TSG 1899 Hoffenheim	44	30	14	-1.398	0.300	12
Eintracht Frankfurt	43	32	11	-1.304	0.300	10
SV Werder Bremen	43	28	15	-1.379	0.300	11
1.FSV Mainz 05	40	24	16	-1.199	0.299	8
1.FC Köln	40	21	19	-1.238	0.299	9
Hannover 96	37	22	15	-1.576	0.302	15
VfB Stuttgart	36	19	17	-1.616	0.302	16
Hertha BSC Berlin	35	22	13	-1.531	0.301	14
Hamburger SV	35	23	12	-1.692	0.303	17
SC Freiburg	34	21	13	-1.440	0.300	13
SC Paderborn 07	31	18	13	-1.888	0.306	18

Tabelle 2.1: Abschlusstabelle der Bundesligasaison 2014/2015 mit Angabe der Gesamtpunktzahl, der gewonnenen Punkte bei Heim- und Auswärtsspielen sowie der modellierten Spielstärke, der dazugehörigen quasi-Standardabweichungen und der daraus folgenden theoretischen Endplatzierungen

2 Datenbeschreibung

zugängliche Daten verwendet. In Kapitel 5 werden die durchschnittlichen Marktwerte der einzelnen Bundesligavereine sowie die Marktwerte ihrer Startaufstellungen für jedes Spiel benötigt. Die genaue Datenherkunft und ihre Auflistung finden sich im dazugehörigen Abschnitt. Eine kurze Datenbeschreibung soll jedoch bereits hier gegeben werden.

Der FC Bayern München besitzt mit deutlichem Abstand den größten Lizenzspieleretat aller 18 Bundesligisten mit einem Etat von 160 Millionen Euro. Dieser ist mehr als doppelt so hoch wie der zweitgrößte Lizenzspieleretat von 78 Millionen Euro. Dieser gehört zum FC Schalke 04. Auch der VfL Wolfsburg und Borussia Dortmund haben ähnlich hohe Etats wie Schalke zur Verfügung. Der SC Paderborn 07 startete mit dem geringsten Etat in Höhe von 17 Millionen Euro in die Saison 2014/2015. Die Höhe der zur Verfügung stehenden Gelder macht sich auch bei den Marktwerten der einzelnen Vereine bemerkbar. Zu Saisonbeginn lag der Marktwert aller Spieler des FC Bayern München bei über 564 Millionen Euro. Bei einer Spieleranzahl von 27 Spielern im Kader entspricht dies einem durchschnittlichen Marktwert von 20.9 Millionen Euro pro Spieler. Auch hier liegt der FC Bayern München weit vor allen anderen Mannschaften. Borussia Dortmund nimmt den zweiten Platz ein bei einem Gesamtmarktwert von gut 344 Millionen Euro und einer Kadergröße von 30 Spielern. Den geringsten Gesamtmarktwert besitzt der Kader des SC Paderborn 07 mit einem Wert von 22.35 Millionen Euro bei 28 Spielern im Kader. Bei den Mannschaftswerten der Startaufstellungen ergibt sich wiederum das gleiche Bild. Der FC Bayern München schickt an jedem Spieltag die wertvollste Mannschaft aller Bundesligisten auf den Platz. Am 23. Spieltag stellten sie die wertvollste Mannschaft der ganzen Saison mit einem durchschnittlichen Marktwert von 33.8 Millionen Euro pro Spieler. Der SC Paderborn 07 ist die Mannschaft mit dem geringsten Wert. Am 25. Spieltag erreichte die Startaufstellung der Mannschaft nur einen durchschnittlichen Marktwert von 895000 Euro pro Spieler.

Zusätzlich zu diesen Daten werden in Kapitel 7 noch die Wettquoten für jedes einzelne Spiel benötigt. Eine genaue Beschreibung wie sich diese Quoten ergeben und wie sie zu interpretieren sind, findet sich im dazugehörigen Abschnitt 7.3.

3 Statisches Bradley-Terry Modell

Im folgenden Kapitel soll das allgemeine statische Bradley-Terry Modell und seine Anpassungen an die zugrundeliegenden Daten vorgestellt werden. Das allgemeine Modell wurde von Bradley und Terry (1952) entwickelt. Teilweise wird es auch als Bradley-Terry-Luce Modell bezeichnet, da es einen engen Zusammenhang zu den Entscheidungsaxiomen von Luce (1959) aufweist. In Bezug auf Sportwettbewerbe lassen sich mit Hilfe dieses Modells nur binäre Ausgänge, sprich Sieg oder Niederlage, modellieren. Das Modell wurde unter anderem von Davidson (1970) und Rao und Kupper (1967) weiterentwickelt, so dass Unentschieden modelliert werden können. In dieser Arbeit spielt es später nicht nur eine Rolle, wer der Sieger ist, sondern auch, mit welcher Tordifferenz ein Spiel gewonnen wurde. Hierzu wird der Response um weitere Kategorien erweitert. Agresti (1992) und Tutz (1986) erweiterten das Modell so, dass ein geordneter ordinaler Response mit beliebig vielen Kategorien möglich ist. Weitere Erweiterungen des zuletzt genannten Modells wie das Hinzufügen eines globalen sowie eines teamspezifischen Heimvorteils finden sich in Tutz und Schauburger (2015). Die folgenden Notationen und der Aufbau des Kapitels sind größtenteils aus Tutz und Schauburger (2015) übernommen.

3.1 Das binäre Bradley-Terry Modell

Das Bradley-Terry Modell besitzt eine große Ähnlichkeit zum logistischen Regressionsmodell. Seien $\{a_1, \dots, a_m\}$ die einzelnen Teams, welche gegeneinander antreten können. Im binären Bradley-Terry Modell wird nur zwischen Sieg und Niederlage unterschieden. $Y_{rs} = 1$ bedeutet im Folgenden, dass Mannschaft r gegen Mannschaft s gewinnt. $Y_{rs} = 0$ bedeutet folglich eine Niederlage von Team r . Mit der Formel

$$P(Y_{rs} = 1) = \frac{\exp(\gamma_r - \gamma_s)}{1 + \exp(\gamma_r - \gamma_s)} \quad (3.1)$$

lässt sich nun die Wahrscheinlichkeit eines Sieges von Team r über Team s beschreiben. Die Parameter $\gamma_r, r = 1, \dots, m$ können dabei als die Spielstärke oder auch Fähigkeiten der einzelnen Mannschaften angesehen werden. Aus dieser Formel ist ersichtlich, dass die Wahrscheinlichkeit eines Sieges bei gleicher Spielstärke bei 50% liegt. Besitzt Mannschaft r eine höhere Spielstärke als Mannschaft s , so steigt auch die Gewinnwahrscheinlichkeit

3 Statisches Bradley-Terry Modell

von Team r .

Wird das Modell zu

$$\log \frac{P(Y_{rs} = 1)}{P(Y_{rs} = 0)} = \gamma_r - \gamma_s$$

umgeformt, so ist schnell erkennbar, dass es an eine Bedingung geknüpft werden muss. Würden den einzelnen Parametern γ_r Konstanten c hinzugefügt werden, so würde das Modell trotzdem die gleichen Gewinnwahrscheinlichkeiten ermitteln und damit wäre es nicht eindeutig identifizierbar. Die Identifizierbarkeit kann entweder durch die Wahl eines festen Parameters oder durch die Nullsummenbedingung $\sum_{i=1}^m \gamma_i = 0$ gewährleistet werden. Im Weiteren Verlauf wird zumeist erstere Bedingung eingesetzt und die Spielstärke des VfL Wolfsburg als Fixparameter gleich Null gesetzt [vgl. Tutz und Schaubberger (2015)].

3.2 Erweiterung zum ordinalen Modell

Wie bereits in der Einführung zu diesem Kapitel erwähnt, gilt das Interesse dieser Arbeit nicht nur dem Ereignis, welches Team das Spiel gewonnen hat, sondern der Response soll auch die Dominanz einer Mannschaft widerspiegeln. Hierfür wird der Ausgang Y_{rs} eines Spiels zwischen zwei Mannschaften r und s mit einer geordneten ordinalen Skala mit Wertebereich $Y_{rs} \in \{1, \dots, k\}$ dargestellt. Wählt man $k = 3$, so lässt sich der Ausgang „Unentschieden“ ins Modell mit aufnehmen. $k = 1$ würde in diesem Fall einen Sieg von Team r bedeuten, $k = 2$ „Unentschieden“ und $k = 3$ „Sieg für Team s “. Wird die Anzahl der Responsemöglichkeiten auf $k = 5$ erhöht, so können die einzelnen Responses als „starke Dominanz von Team r “, „schwache Dominanz von Team r “, „Unentschieden“, „schwache Dominanz von Team s “ und „starke Dominanz von Team s “ interpretiert werden. Im weiteren Verlauf werden die Analysen hauptsächlich für den Fall $k = 5$ durchgeführt. Als „starke Dominanz eines Teams“ wird definiert, dass dieses Ereignis gleichbedeutend mit einem Sieg einer Mannschaft mit mindestens 2 Toren Differenz ist. Folglich wird ein Sieg mit genau einem Tor Unterschied als „schwache Dominanz eines Teams“ festgelegt. Um diesen geordneten ordinalen Response anwenden zu können, wird ein kumulatives Modell verwendet. Die Wahrscheinlichkeit eines Ereignisses lässt sich hier als kumulative Wahrscheinlichkeit folgendermaßen modellieren:

3 Statisches Bradley-Terry Modell

$$P(Y_{rs} \leq j) = F(\eta_{rsj}) \quad (3.2)$$

$F(\cdot)$ ist hier eine symmetrische Verteilungsfunktion, welcher das Bradley-Terry Modell als logistische Verteilungsfunktion zugrunde liegt. Der lineare Prädiktor η_{rsj} ergibt sich als

$$\eta_{rsj} = \theta_j + \gamma_r - \gamma_s,$$

d.h. er setzt sich aus der Differenz $\gamma_r - \gamma_s$ der Spielstärken der einzelnen Mannschaften und einem Schwellenparameter θ_j zusammen. Erst dieser Schwellenparameter ermöglicht die Darstellung als ordinalskaliertes Modell. Da es sich bei $F(\cdot)$ um eine symmetrische Verteilungsfunktion handelt, folgt daraus, dass der Schwellenparameter θ die Einschränkung $\theta_j = -\theta_{k-j}$ für $t = 1, \dots, [k/2]$ besitzt. In der Praxis bedeutet dies, dass bei gleichen Spielstärken $\gamma_r = \gamma_s$ die Wahrscheinlichkeiten das gegnerische Team schwach bzw. stark zu dominieren für beide Teams identisch sind, sprich $P(Y_{rs} = j) = P(Y_{rs} = k + 1 - j)$. Der Beweis für diese Äquivalenz in den Wahrscheinlichkeiten findet sich in Tutz (1986), Seite 308. Für den Fall $k = 5$ folgt dadurch bei gleicher Spielstärke zweier Mannschaften $P(Y_{rs} = 1) = P(Y_{rs} = 5)$ und $P(Y_{rs} = 2) = P(Y_{rs} = 4)$ [vgl. Tutz und Schauburger (2015)].

3.3 Heimspieleffekt

In Kapitel 2 wurde bereits erwähnt, dass in der Saison 2014/2015 47,4% aller Spiele von der gastgebenden Mannschaft gewonnen wurden und nur 25,8% der Spiele durch den Gast. Daher liegt die Vermutung nahe, dass eine Mannschaft einen Vorteil daraus zieht, auf eigenem Platz anzutreten. Um dies zu untersuchen, werden zwei unterschiedliche Ansätze betrachtet, zum einen die Annahme eines globalen Heimeffektes und zum anderen die eines teamspezifischen Heimeffektes. Beide Ansätze seien im Weiteren kurz beschrieben.

3.3.1 Globaler Heimeffekt

Soll ein globaler Effekt in das Modell mit aufgenommen werden, so muss der lineare Prädiktor zu der Form

3 Statisches Bradley-Terry Modell

$$\eta_{rsj} = \alpha + \theta_j + \gamma_r - \gamma_s$$

erweitert werden. α repräsentiert dann den globalen Heimeffekt. Da von einem Heimvorteil ausgegangen wird, sollte α positiv sein und dadurch sollten sich die Wahrscheinlichkeiten für einen kleinen Response, also einen Heimsieg, erhöhen. Dieser Parameter kann unter bestimmten Umständen folgendermaßen interpretiert werden: Er stellt für den Fall $k = 3$ und identische Spielstärken die anteiligen Gewinnchancen für einen Sieg von Mannschaft r bzw. Mannschaft s dar [vgl. Tutz und Schaubberger (2015)].

Formell ausgedrückt:

$$\alpha = \frac{1}{2} \log \frac{P(Y_{rs} = 1)/(1 - P(Y_{rs} = 1))(*)}{P(Y_{rs} = 3)/(1 - P(Y_{rs} = 3))(\#)} \quad (3.3)$$

Nachfolgend sei kurz gezeigt, dass der rechte Teil der Gleichung wirklich den Parameter α liefert.

$$P(Y_{rs} = 1) = \frac{\exp(\alpha + \theta_1 + \gamma_r - \gamma_s)}{1 + \exp(\alpha + \theta_1 + \gamma_r - \gamma_s)}$$

$$P(Y_{rs} = 3) = 1 - P(Y_{rs} \leq 2) = 1 - \frac{\exp(\alpha + \theta_2 + \gamma_r - \gamma_s)}{1 + \exp(\alpha + \theta_2 + \gamma_r - \gamma_s)} = \frac{1}{1 + \exp(\alpha + \theta_2 + \gamma_r - \gamma_s)}$$

$$\gamma_r - \gamma_s = 0$$

$$\theta_1 = -\theta_2$$

In (*) eingesetzt

$$\Rightarrow \frac{\exp(\alpha + \theta_1)}{1 + \exp(\alpha + \theta_1)} / \frac{1}{1 + \exp(\alpha + \theta_1)} = \exp(\alpha + \theta_1)$$

In (#) eingesetzt

$$\Rightarrow \frac{1}{1 + \exp(\alpha - \theta_1)} / \frac{1 + \exp(\alpha - \theta_1) - 1}{1 + \exp(\alpha - \theta_1)} = \frac{1}{\exp(\alpha - \theta_1)}$$

Beides in (3.3) eingesetzt

$$\Rightarrow \frac{1}{2} \log \frac{\exp(\alpha + \theta_1)}{\frac{1}{\exp(\alpha - \theta_1)}} = \frac{1}{2} \log(\exp(2\alpha)) = \alpha$$

Für den Fall $k = 5$ drückt α bei identischer Spielstärke die anteiligen Gewinnchancen eines schwach bzw. klar dominierten Spieles von Mannschaft r im Verhältnis zu einem schwach dominierten bzw. stark dominierten Sieg von Mannschaft s aus.

3.3.2 Teamspezifischer Heimeffekt

Eine Betrachtung der Ergebnisse zeigt eindeutig, dass ein Heimvorteil existiert. Jede Mannschaft holte in der Spielzeit 2014/2015 mehr Punkte zu Hause als auf fremdem Platz. Es stellt sich jedoch die Frage, ob die Annahme eines globalen Heimvorteils ausreicht. Daher wird zusätzlich auch ein teamspezifischer Heimvorteil betrachtet. Hierfür muss lediglich der Parameter α zu einem spezifischen Parameter α_r verändert werden. Der Parameter besitzt die selben Indizes wie der γ -Parameter der Heimmannschaft. Der lineare Prädiktor lautet dann

$$\eta_{rsj} = \alpha_r + \theta_j + \gamma_r - \gamma_s.$$

Durch diese Anpassungen eines Heimeffektes verändert sich auch die Interpretation der γ -Parameter. Diese lassen sich jetzt nicht mehr sofort als Spielstärke einer Mannschaft interpretieren, da nun unterschieden werden muss, ob eine Mannschaft zu Hause oder auswärts spielt. Die Parameter können jedoch als Spielstärke eines Teams bei Auswärtsspielen angesehen werden. Soll die Spielstärke bei Heimspielen ermittelt werden, so ergibt sich diese aus der Summe des Heimeffektes und des γ -Koeffizienten der jeweiligen Mannschaft. Der teamspezifische α -Koeffizient lässt sich äquivalent zum globalen Heimkoeffizienten interpretieren [vgl. Tutz und Schauburger (2015)].

3.4 Zusammenhang zum GLM

Wie bereits erwähnt, handelt es sich bei unserem Modellansatz um ein kumulatives Modell. Im Folgenden soll dieser Ansatz in die Struktur eines multivariaten generalisierten linearen Modells (GLM) eingebettet werden. Diese Zusammenführung der Modelle wurde von McCullagh und Nelder (1989) eingeführt. Bei unserem Modell kann der lineare Prädiktor mit $Y_{rs} \in \{1, \dots, k\}$ und $r, s \in \{1, \dots, m\}$ als

$$\eta_{rsj} = \alpha_r + \theta_j + \gamma_r - \gamma_s = \alpha_r + \theta_j + x_2^{(r,s)} \gamma_2 + \dots + x_m^{(r,s)} \gamma_m = \alpha_r + \theta_j + (\mathbf{x}^{(r,s)})^T \boldsymbol{\gamma}$$

3 Statisches Bradley-Terry Modell

geschrieben werden, wobei die $(m-1)$ Komponenten des $\mathbf{x}^{(r,s)}$ -Vektors durch

$$x_l^{(r,s)} = \begin{cases} 1, & l = r \\ -1, & l = s \\ 0, & \text{sonst} \end{cases}$$

gegeben sind. Bei dieser Gleichung handelt es sich also um ein kumuliertes Modell mit Schwellenparameter θ_j , teamspezifischem Heimeffekt α_r und dem Prädiktorenvektor $\mathbf{x}^{(r,s)}$. Fahrmeir und Tutz (2001) und auch Tutz (2012) verwenden die Einbettung des kumulierten Modells in den Rahmen der multivariaten generalisierten Modelle zur Parameterschätzung. Diese Maßnahme erlaubt es, die bereits bekannte Inferenz der GLMs auf die vorgestellten Modelle anzuwenden und somit sind Hypothesentests in unserer Arbeit mit der bekannten goodness-of-fit-Statistik sowie der Likelihood-ratio-Statistik durchführbar. Einzige Bedingung hierbei ist, dass die einzelnen Beobachtungen bei gegebenen Spielstärken als unabhängig angenommen werden müssen [vgl. Tutz und Schaubberger (2015)].

3.5 Ergebnisse

Wendet man die oben genannten Methodiken auf die Daten der Bundesligasaison 2014/2015 mit $k = 5$ verschiedenen Responsemöglichkeiten und globalen Heimeffekt an, so erhält man folgende Ergebnisse:

Der geschätzte globale Heimeffekt beträgt $\hat{\alpha} = 0.528$. Für die Schwellenparameter ergeben sich die Werte $\hat{\theta}_1 = -\hat{\theta}_4 = -1.794$ und $\hat{\theta}_2 = -\hat{\theta}_3 = -0.671$. Mit diesen Schätzern lassen sich jetzt die allgemeinen Wahrscheinlichkeiten für ein bestimmtes Ereignis modellieren. Die Wahrscheinlichkeit eines Heimsieges mit mindestens zwei Toren Unterschied bei gleicher Spielstärke liegt demnach bei 22% und für einen Heimsieg mit genau einem Tor Differenz bei 24.4%. In 46.4% aller Spiele sollte laut Modell daher die Heimmannschaft das Spielfeld als Sieger verlassen. Für Unentschieden schätzt das Modell einen Wert von 30.4%. Eine knappe Niederlage tritt in 14.3% Prozent aller Heimspiele auf und eine klare Niederlage gibt es in 8.9% der Spiele. Die Wahrscheinlichkeit einer Heimmiederlage bei gleicher Spielstärke liegt im Allgemeinen also bei 23.4%. Es gilt natürlich zu beachten, dass sich diese Werte als Mittelwerte aller Spiele ergeben und die Mannschaften unterschiedliche Spielstärken besitzen. Hier zeigt sich ein großer Vorteil der latent trait Modelle.

3 Statisches Bradley-Terry Modell

Diese Modelle beachten die Unterschiede in den Spielstärken zwischen den einzelnen Mannschaften und beziehen diese in die Schätzung des Heimeffektes mit ein.

In Tabelle 2.1 sind die modellierten Spielstärken und ihre theoretischen Platzierungen durch diese Werte für $k = 5$ angegeben. Für die vier besten Mannschaften stimmt die modellierte Platzierung mit der wahren Platzierung überein. Bei den restlichen Mannschaften kommt es zu einigen Veränderungen im Vergleich zur wahren Platzierung. Diese Veränderungen können jedoch durch die Betrachtung der quasi-Standardabweichungen relativiert werden. Die Berechnung dieser erfolgt wie in Firth und De Menezes (2004) beschrieben. Mit Hilfe des quasi-Standardfehlers kann überprüft werden, ob sich die Spielstärken zweier Mannschaften signifikant voneinander unterscheiden oder nicht. Dies ist möglich, da sich nun die Standardabweichungen der Differenzen zweier Mannschaften berechnen lassen. Soll beispielsweise überprüft werden, ob sich die Spielstärke von Schalke 04 und Borussia Dortmund signifikant unterscheidet, so ergibt sich als Standardfehler der Differenz der Spielstärken der Mannschaften ein Wert von $\sqrt{(0.299^2 + 0.299^2)} = 0.423$. Das dazugehörige 95%-Konfidenzintervall ergibt sich dann mit $\mu = -1.087 - (-0.939) = -0.148$ als $[-0.148 \pm 1.96 \cdot 0.423] = [-0.987, 0.690]$. Die Spielstärken der beiden Mannschaften unterscheiden sich daher nicht signifikant. Vergleicht man die Spielstärken des FC Bayern München und des VfL Wolfsburg, so lässt sich auch zwischen diesen beiden Mannschaften kein signifikanter Unterschied in den Spielstärken nachweisen, obwohl Bayern München die Meisterschaft mit 10 Punkten Vorsprung gewann. Dies zeigt, dass sich anhand der errechneten Spielstärken keine konkreten Aussagen über die Endplatzierungen treffen lassen.

In Kapitel 2 wurde bereits erwähnt, dass eine Mannschaft für ein gewonnenes Spiel drei Punkte erhält, für ein Unentschieden einen Punkt und bei einer Niederlage keine Punkte. Diese Punktwertung wurde im Fußball bei der Weltmeisterschaft 1994 in den USA eingeführt und findet seit der Saison 1995/1996 auch in der deutschen Bundesliga Anwendung. Zuvor gab es für einen Sieg nur zwei anstatt drei Punkte. In diversen Artikeln, wie zum Beispiel auf ran.de (2015) oder von Nedo (2015), ist nachzulesen, dass diese Einführung damals sehr umstritten war. So wäre zum Beispiel mit der alten Zwei-Punkte-Regel in der Saison 1999/2000 nicht der FC Bayern München, sondern Bayer 04 Leverkusen deutscher Meister geworden. Vor allem im Kampf um den Klassenerhalt würde sich oft ein anderes Tabellenbild ergeben. Daher ist es interessant zu überprüfen, wie hoch die Korrelation zwischen den geschätzten Spielstärken der Teams und den gesammelten Punkten im Verlauf der Saison ist und bei welcher Punktwertung diese maximiert werden würde. Die Korrelation zwischen den beiden Ergebnissen für die Saison 2014/2015 liegt bei 0.983 und

3 Statisches Bradley-Terry Modell

ist damit sehr hoch.

Abbildung 3.1 stellt dar, wie sich die Korrelationen entwickeln, wenn man die Punktein-

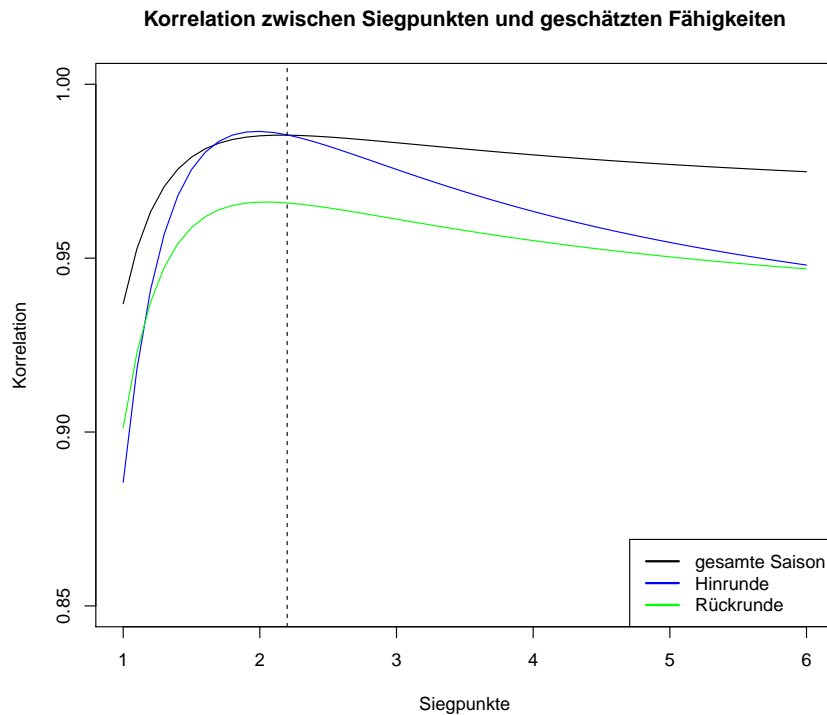


Abbildung 3.1: Korrelation zwischen den geschätzten Spielstärken und den zu erhaltenden Punkten für einen Sieg.

zahl für einen Sieg nun verändert. Die maximale Korrelation erhält man mit einem Wert von 0.985 bei einer Wertung von 2.2 Punkten für einen Sieg. Zu demselben Ergebnis kamen auch Tutz und Schauburger (2015) für die Daten der Saison 2012/2013. Es fällt auf, dass die Korrelation selbst bei starker Veränderung der Punktwertung relativ konstant bleibt. Dies heißt, dass es für ein Ligensystem, bei dem jeder gegen jeden antritt, eine relativ geringe Rolle spielt, wie viele Punkte ein Sieg einbringt. Wird die Korrelation getrennt für Hin- und Rückrunde betrachtet, so ist für höhere Punktwertungen der Abfall der Korrelation wesentlich größer.

Abschließend sollen noch das Modell mit globalen Heimeffekt und das Modell mit teamspezifischen Heimeffekt miteinander verglichen werden. Der Likelihood-ratio Test mit der Nullhypothese, alle Effekte seien gleich, gibt für $k = 5$ einen Wert von 12.58 bei 17 Freiheitsgraden an, was einem p-Wert von 0.764 entspricht. Die Nullhypothese kann daher eindeutig nicht abgelehnt werden, so dass der teamspezifische Heimeffekt keine signi-

3 Statisches Bradley-Terry Modell

fikante Verbesserung des Modells bringt. Bei einem Test derselben Hypothese für beide Modelle und $k = 3$, liegt die Devianz bei 12.62, was einem p-Wert von 0.761 entspricht. Demzufolge kann auch hier die Nullhypothese nicht verworfen werden. Ein Vergleich der gesammelten Punkte der Mannschaften in Heim- und Auswärtsspielen zeigt, dass dieses Ergebnis wenig überraschend ist. Die ersten vier Mannschaften der Tabelle haben sowohl zu Hause als auch auswärts die meisten Punkte geholt. Bei den anderen 14 Teams ergeben sich zwar teils große Unterschiede zwischen Heim- und Auswärtstabelle - so belegt der 1.FC Köln zum Beispiel in der Heimtabelle Platz 15 und in der Auswärtstabelle Platz 5 -, da die Mannschaften aber in der Auswärtstabelle alle zwischen 11 und 19 Punkte gesammelt haben, also innerhalb von 8 Punkten liegen, hat dies eine relativ geringe Aussagekraft. Auf selbiges Ergebnis kommt auch Knorr-Held (2000) bei der Betrachtung von Bundesligadaten der Saison 1996/1997. Die Aussage muss jedoch für jede Saison einzeln überprüft werden. So spielt der teamspezifische Heimvorteil bei anderen Analysen, wie z.B. Glickman und Stern (1998) oder auch Clark und Norman (1995), eine bedeutende Rolle. Auch Tutz und Schauburger (2015) erweiterten ihre Analysen auf Basis des teamspezifischen Heimvorteils, obwohl die Verbesserung dieses Modells anhand des p-Wertes nicht signifikant gegeben ist.

4 Clustering durch Regularisierung

Die Betrachtungen der quasi-Standardabweichungen in Abschnitt 3.5 zeigten, dass sich mit diesem Kriterium die einzelnen Spielstärken der Mannschaften nicht signifikant voneinander unterscheiden lassen. Im Folgenden soll nun ein weiterer Ansatz zur Unterscheidung der Spielstärken der Mannschaften betrachtet werden. Mit Hilfe von Regularisierungstechniken sollen Cluster von Teams mit gleicher Spielstärke herausgefiltert werden. Gertheiss und Tutz (2010) stellen solche Regularisierungstechniken vor. Diese wurden von Masarotto und Varin (2012) benutzt, um genau solche Cluster in verschiedenen Sportarten - mit und ohne mögliche Unentschieden - zu finden. Tutz und Schaubberger (2015) erweiterten diese Techniken zusätzlich für den Fall eines teamspezifischen Heimvorteils.

4.1 Ranking LASSO

Im unpenalisierten Modell ergeben sich die einzelnen Koeffizienten durch die Maximierung der Bradley-Terry-Likelihood, welche sich unter der Annahme der Unabhängigkeit der einzelnen Spiele als

$$l(\gamma, \alpha, \theta) = \sum_{r \neq s} \sum_{u=1}^k 1_{\{Y_{rs}=u\}} \cdot \log(P(Y_{rs} = u)) \quad (4.1)$$

bestimmen lässt. n gibt die Anzahl der Spiele und k die unterschiedlichen ordinalen Responses an. Um nun eine Gruppierung der einzelnen Spielstärkekoeffizienten der Mannschaften zu erhalten, wird die Bradley-Terry-Likelihood mit einem Strafterm für alle paarweisen Spielstärkedifferenzen belegt und muss dann wiederum maximiert werden. Der Strafterm ergibt sich als

$$J(\gamma, \alpha, \theta) = \sum_{r < s} w_{rs} |\gamma_r - \gamma_s| \quad (4.2)$$

und somit muss der Term

$$l_p(\gamma, \alpha, \theta) = l(\gamma, \alpha, \theta) - \lambda J(\gamma, \alpha, \theta) = l(\gamma, \alpha, \theta) - \lambda \sum_{r < s} w_{rs} |\gamma_r - \gamma_s| \quad (4.3)$$

4 Clustering durch Regularisierung

maximiert werden. λ wird hierbei als Tuning-Parameter bezeichnet und w_{rs} sind spezifische Gewichte für die einzelnen Differenzen. Der hier verwendete Strafterm ist eine Verallgemeinerung des fusionierten Lasso-Strafterms von Tibshirani et al. (2005). Durch die Verwendung der L_1 -Norm für die Differenzen folgt, dass für einen wachsenden Tuning-Parameter λ sich die Fähigkeiten der einzelnen Mannschaften ähnlicher werden. Für den Extremfall $\lambda \rightarrow \infty$ sind alle Stärkeparameter gleich.

Die Lasso-Methode kann jedoch zu inkonsistenten Schätzern für die Parameter führen, da die Möglichkeit besteht, dass durch die L_1 -Norm zu viele Parameter gleich Null geschätzt werden. Für den in dieser Arbeit betrachteten Fall mit der Anwendung des Ranking Lasso bedeutet dies, dass der Bias der Schätzer der paarweisen Fähigkeitsdifferenzen ungleich Null bei steigender Spielanzahl nicht gegen Null konvergiert. Eine Möglichkeit, dies zu verhindern, ist die von Fan und Li (2001) vorgeschlagene SCAD (smoothly clipped absolute deviation)-Methode, in welcher der L_1 -Strafterm durch einen stetigen Strafterm, welcher starke Effekte stärker bestraft, ersetzt wird. In dieser Arbeit wird jedoch ein anderer Ansatz verfolgt. Die Inhalte des L_1 -Strafterms sollen stärker gewichtet werden als die Abnahme der Effekte. Dies kann durch den von Zou (2006) entwickelten adaptiven Lasso erreicht werden. Die Gewichte ergeben sich hier als inverse der Maximum-Likelihood Schätzer, formell ausgedrückt:

$$w_{rs} = |\hat{\gamma}_r^{(ML)} - \hat{\gamma}_s^{(ML)}|^{-1}. \quad (4.4)$$

Diese Herangehensweise ist an eine Bedingung geknüpft. Eine Anwendung ist nur möglich, wenn kein Verein alle Spiele gewinnt oder alle Spiele verliert. Würde dies der Fall sein, so würde der Maximum-Likelihood-Schätzer divergieren. Aus diesem Grund erhält der Maximum-Likelihood-Schätzer einen Ridge-Strafterm mit der Festlegung $\epsilon = 10^{-4}$, um diesem Problem entgegenzuwirken [vgl. Masarotto und Varin (2012)].

Diese oben genannte Form des Strafterms bezieht sich auf den Fall eines globalen Heimeffektes. Soll ein Clustering mit den teamspezifischen Heimeffekten stattfinden, so gestaltet sich dies schwieriger. Grund dafür ist die nun nötige Unterscheidung zwischen der Spielstärke eines Teams bei Heimspielen und bei Auswärtsspielen. Nach Tutz und Schauburger (2015) ändert sich dadurch der Strafterm zu

$$J(\gamma, \alpha, \theta) = \sum_{r < s} w_{rs} |\gamma_r - \gamma_s| + \sum_{r < s} u_{rs} |\gamma_r - \gamma_s + \alpha_r - \alpha_s| + \sum_{r < s} v_{rs} |\alpha_r - \alpha_s|. \quad (4.5)$$

4 Clustering durch Regularisierung

In diesem Strafterm werden die Spielfähigkeiten für ein Auswärtsspiel γ_r , für ein Heimspiel $\gamma_r + \alpha_r$ und der teamspezifische Heimeffekt α_r als eigene Terme mit den Gewichten $w_{rs} = |\hat{\gamma}_r^{(ML)} - \hat{\gamma}_s^{(ML)}|^{-1}$, $u_{rs} = |\hat{\gamma}_r^{(ML)} - \hat{\gamma}_s^{(ML)} + \hat{\alpha}_r^{(ML)} - \hat{\alpha}_s^{(ML)}|^{-1}$ und $v_{rs} = |\hat{\alpha}_r^{(ML)} - \hat{\alpha}_s^{(ML)}|^{-1}$ berücksichtigt. Dadurch können beide Spielfähigkeiten einer Mannschaft - für Heimspiele und für Auswärtsspiele - getrennt geclustert werden.

4.2 Auswahlkriterien

Durch die gerade vorgestellte Methodik können nun die einzelnen Koeffizientenschätzer für unterschiedliche Wahl des Tuning-Parameters λ ermittelt werden. Doch welche Wahl von λ liefert die besten Ergebnisse für die Daten? Bei generalisierten linearen Modellen wird in solchen Fällen meist auf das Akaike-Kriterium (AIC) oder das Schwarz-Bayes-Kriterium (BIC) zurückgegriffen. Da sich das Bradley-Terry-Modell in die Gruppe der GLMs einbetten lässt, verwenden wir nun ebenfalls diese beiden Kriterien, obwohl es noch keinen Nachweis für die Effizienz bei einer solchen Anwendung gibt. Zunächst seien beide Kriterien kurz erläutert.

AIC

Der AIC wurde von Akaike (1974) entwickelt und ist das älteste Kriterium zur Auswahl eines Modells. Seine Formel lautet

$$AIC(\lambda) = -2 \cdot l(\beta) + 2 \cdot df(\lambda), \quad (4.6)$$

d.h es berechnet sich aus der Summe der doppelten negativen log-Likelihood-Funktion und der doppelten Anzahl der Freiheitsgrade des betrachteten Modells. Das Kriterium ist zu minimieren. Als Kritikpunkt am AIC wird häufig angeführt, dass die Zahl der Beobachtungen für die Modellwahl bei diesem Kriterium keine Rolle spielt. Das Modell neigt deshalb oft dazu, zu viele Koeffizienten in ein Modell mit aufzunehmen.

4 Clustering durch Regularisierung

BIC

Ein weiteres anerkanntes Kriterium, welches restriktivere Resultate durch die Berücksichtigung der Stichprobengröße zur Folge hat, ist das von Schwarz (1978) entwickelte Bayessche Informationskriterium.

$$BIC(\lambda) = -2 \cdot l(\beta) + \log(n) \cdot df(\lambda) \quad (4.7)$$

In diesem wird die Zahl der Freiheitsgrade mit der logarithmierten Stichprobengröße multipliziert. Dies hat zur Folge, dass bereits ab einer Stichprobengröße von acht Beobachtungen die Zahl der Parameter stärker bestraft wird als beim AIC.

Es gilt zu beachten, dass bei den Analysen die Zahl der Parameter nicht der Anzahl der Parameter des Modells gleicht. Grund hierfür sind die Effekte der Variablenselektion und des Shrinkages, welche sich bei penalisierten Modellen ergeben. Die Zahl der Freiheitsgrade ergibt sich nach Buja et al. (1989) als $tr(2\mathbf{H} - \mathbf{H}^T \mathbf{H})$, wobei \mathbf{H} die hat-Matrix des letzten Fischer-Scoring-Schritts des PIRLS-Algorithmus, welcher hier verwendet wurde, darstellt. Der Algorithmus ist in Oelker und Tutz (2013) näher beschrieben [vgl. Tutz und Schauburger (2015)].

4.3 Ergebnisse

Abschließend sollen in diesem Kapitel noch die Ergebnisse des Clusterings vorgestellt werden. Abbildung 4.1 zeigt die Koeffizientenentwicklungen der einzelnen Teams für unterschiedliche Tuning-Parameter λ und fünf Responsekategorien.

Die x-Achse ist äquivalent zu den Analysen von Tutz und Schauburger (2015) gewählt. Mit der gewählten Darstellung lassen sich die Pfade in einem Wertebereich zwischen 0 und 1 fixieren. Wie in Abschnitt 4.1 beschrieben, werden sich die Koeffizientenschätzer bei größer werdendem Tuning-Parameter λ immer ähnlicher und die gewählte x-Achsen-Funktion wird dadurch immer kleiner. Bei einem gewählten $\lambda = 0$ entsprechen die Koeffizientenschätzer denen des unpenalisierten Modells. In Abbildung 4.1 wurden die Schätzer für $\lambda \in [0, 3]$ für 100 verschiedene Tuning-Parameter in diesem Wertebereich modelliert. Es gilt jedoch zu beachten, dass durch die gewählte Darstellung die Beobachtungspunkte nicht gleichmäßig verteilt sind. So sind beispielsweise die ersten 71 gemessenen Modelle

4 Clustering durch Regularisierung

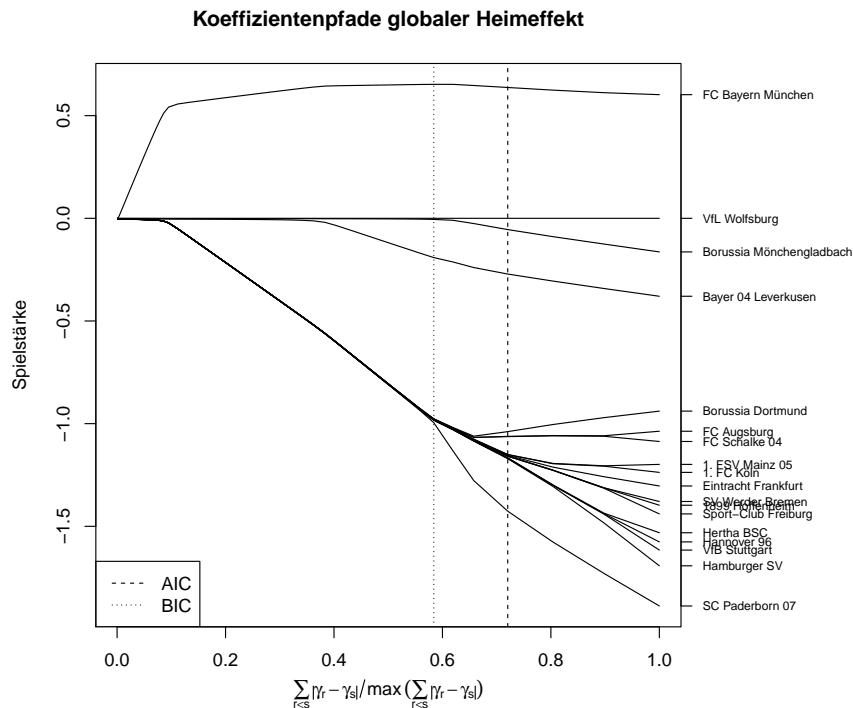


Abbildung 4.1: Koeffizientenpfade bei globalem Heimeffekt unter Verwendung eines adaptiven L_1 -Strafterms

($\lambda \in [0.88, 3]$) in einem Bereich kleiner 0.1 zu finden. Die in der Grafik mit angegebenen Werte bei einer Optimierung des AIC bzw. BIC liegen bei $\lambda = 0.09$, bzw $\lambda = 0.18$. Bei Betrachtung des AIC lassen sich die 18 Vereine in zehn verschiedene Cluster einteilen, beim restriktiveren BIC sind es sechs unterschiedliche Cluster. Tabelle 4.1 gibt diese durch den BIC erhaltenen Cluster an.

Es fällt auf, dass fünf der sechs Cluster nur aus jeweils einer Mannschaft bestehen. Die restlichen 13 Mannschaften bilden den sechsten Cluster. Würden die Ergebnisse auf nur eine Nachkommastelle analysiert werden, so würden sich die Spielstärken sogar nur auf vier unterschiedliche Cluster verteilen. Dieses Ergebnis spiegelt auch den tatsächlichen Verlauf dieser Bundesligasaison wider. Die ersten vier Mannschaften setzten sich am Ende deutlich von den restlichen Mannschaften ab, welche sich auf relativ ähnlichem Niveau bewegten. Für $k = 3$ bilden der AIC und BIC acht bzw. fünf Cluster. Die dazugehörige Grafik findet sich im Anhang.

In Abschnitt 3.5 wurde zwar gezeigt, dass das Modell bei teamspezifischem Heimvorteil keine signifikante Verbesserung aufweist, nichtsdestotrotz sollen die Ergebnisse bei einer

4 Clustering durch Regularisierung

Cluster	Mannschaften	Spielstärke
1	FC Bayern München	0.65
2	VfL Wolfsburg	0
3	Borussia Mönchengladbach	-0.01
4	Bayer 04 Leverkusen	-0.19
5	1.FC Köln, 1.FSV Mainz, TSG 1899 Hoffenheim Borussia Dortmund, Eintracht Frankfurt, FC Augsburg FC Schalke 04, Hamburger SV, Hannover 96 SC Freiburg, Hertha BSC Berlin, SV Werder Bremen, VfB Stuttgart	-0.98
6	SC Paderborn	-1.00

Tabelle 4.1: Cluster der Spielstärken bei globalem Heimeffekt unter Verwendung des BIC-Kriteriums

solchen Betrachtung hier vorgestellt werden um diesen Sachverhalt nochmals zu erläutern. Abbildung 4.2 zeigt die ermittelten Koeffizientenpfade für die einzelnen Teams bei Unterscheidung zwischen Heimvorteil, Heimstärke und Auswärtsstärke. Wiederum sind auch der AIC und BIC mit angegeben. Die Ergebnisse für den BIC finden sich ebenfalls in Tabelle 4.2.

Es fällt auf, dass sich für die verschiedenen Betrachtungen bis zu elf verschiedene Cluster ergeben. Ein Blick auf die Spielstärken der einzelnen Cluster lässt erkennen, dass sich diese teils kaum unterscheiden. Sowohl bei der Heimstärke als auch bei der Auswärtsstärke bilden die vier besten Mannschaften der Spielzeit 2014/2015 wie schon bei Betrachtung der Spielstärke bei globalem Heimeffekt eigene Cluster. Für die restlichen Mannschaften ändern sich durch den teamspezifischen Heimvorteil die Zuordnungen zu den Clustern. So ist beispielsweise Eintracht Frankfurt dem Modell zufolge eine besonders auswärtschwache Mannschaft, zu Hause gehört sie durch ihren Heimvorteil jedoch zum vorderen Drittel der Mannschaften. Bei näherer Betrachtung des Heimvorteils fällt auf, dass sich die neun Cluster in einem Wertebereich von $[0.3, 0.79]$ befinden, d.h. der größte gemessene Unterschied bei den Heimvorteilen bei 0.49 liegt. Hierdurch erklärt sich zum einen, warum bei den Heim- und Auswärtsspielstärken die Cluster sehr ähnlich sind, und zum anderen, warum dieses Modell keine signifikante Verbesserung zum Modell mit globalem Heimeffekt darstellt.

4 Clustering durch Regularisierung

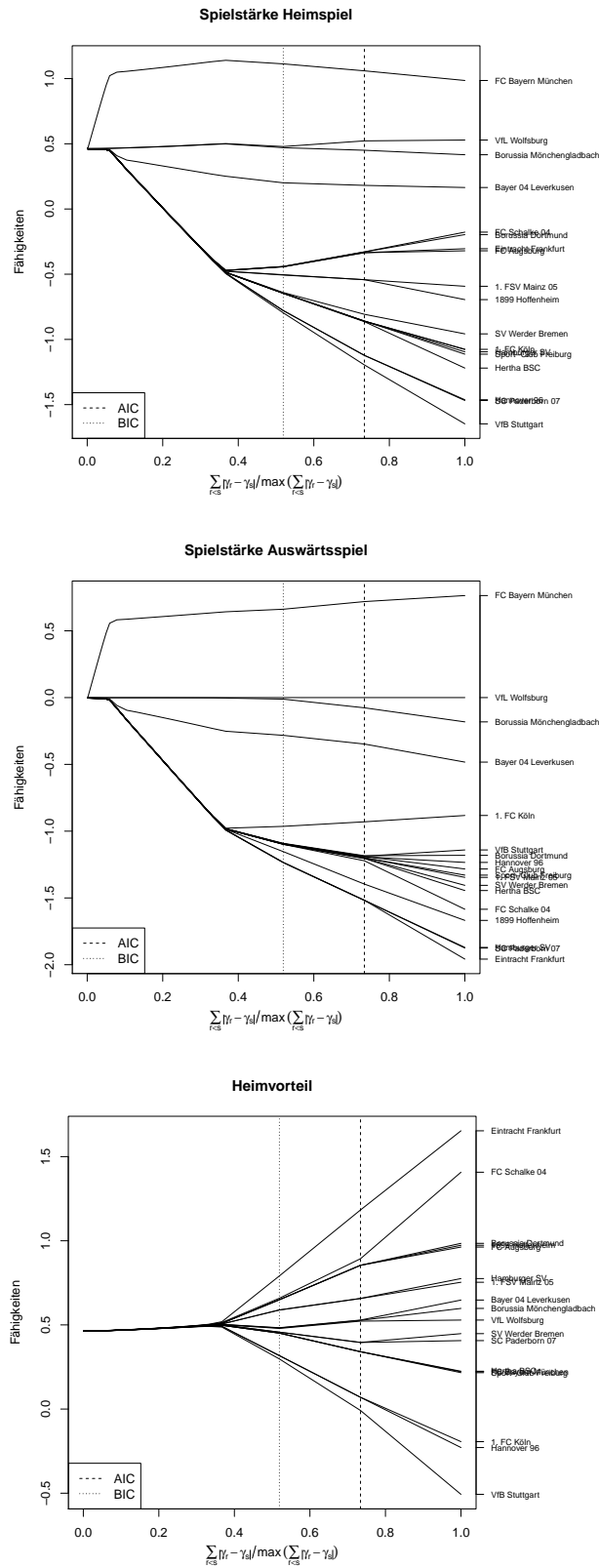


Abbildung 4.2: Koeffizientenpfade bei teamspezifischem Heimeffekt unter Verwendung eines adaptiven L_1 -Strafterms

4 Clustering durch Regularisierung

Cluster	Mannschaften	Spielstärke
Heimstärke		
1	FC Bayern München	1.11
2	VfL Wolfsburg	0.48
3	Borussia Mönchengladbach	0.47
4	Bayer 04 Leverkusen	0.20
5	Eintracht Frankfurt, Borussia Dortmund, FC Augsburg, Schalke 04	-0.44
6	TSG Hoffenheim	-0.50
7	1.FSV Mainz 05	-0.51
8	1. FC Köln, SV Werder Bremen	-0.64
9	Hamburger SV, Hertha BSC Berlin, SC Freiburg	-0.65
10	Hannover 96, SC Paderborn 07	-0.78
11	VfB Stuttgart	-0.79
Auswärtsstärke		
1	FC Bayern München	0.66
2	VfL Wolfsburg	0.00
3	Borussia Mönchengladbach	-0.01
4	Bayer Leverkusen	-0.28
5	1.FC Köln	-0.96
6	Borussia Dortmund, FC Augsburg, VfB Stuttgart	-1.09
7	1.FSV Mainz 05, FC Schalke 04, Hannover 96	-1.10
	Hertha BSC Berlin, SC Freiburg, SV Werder Bremen	
8	TSG 1899 Hoffenheim	-1.15
9	Eintracht Frankfurt	-1.23
10	Hamburger SV, SC Paderborn 07	-1.24
Heimvorteil		
1	Eintracht Frankfurt	0.79
2	FC Schalke 04	0.66
3	TSG 1899 Hoffenheim, Borussia Dortmund, FC Augsburg	0.65
4	1.FSV Mainz 05, Hamburger SV	0.59
5	Bayer 04 Leverkusen, Borussia Mönchengladbach, VfL Wolfsburg	0.48
6	SC Paderborn 07, SV Werder Bremen	0.46
7	FC Bayern München, Hertha BSC Berlin, SC Freiburg	0.45
8	1.FC Köln, Hannover 96	0.32
9	VfB Stuttgart	0.30

Tabelle 4.2: Cluster der Spielstärken und des Heimvorteils bei teamspezifischem Heimeffekt unter Verwendung des BIC-Kriteriums

5 Modelle mit mannschaftsspezifischen Einflussgrößen

Die letzten beiden Kapitel beschäftigten sich mit der Bestimmung der Spielstärke der einzelnen Mannschaften. Zur Modellierung dienten dabei die Resultate der Mannschaften über die 34 Spieltage. Dabei wurde die Frage nach den Gründen für die erzielten Ergebnisse bisher völlig außer Acht gelassen. Im Folgenden soll daher versucht werden, die Streuung der einzelnen Spielstärken durch die Hinzunahme von Kovariablen in das Modell zu erklären. Dabei beschränken wir uns auf Variablen, welche eindeutig einem Verein zugeschrieben werden können, wie zum Beispiel das Spielerbudget oder der Mannschaftswert eines Vereins. Variablen, welche von keinem der Vereine beeinflusst werden können, wie zum Beispiel das Wetter am Spieltag, werden im Weiteren nicht betrachtet.

5.1 Modellerweiterung für erklärende Variablen

Im Folgenden seien die Daten gegeben durch $(Y_{rs} \in \{1, \dots, k\}, x_1, \dots, x_m)$, wobei k der Anzahl der Responsemöglichkeiten und m der Zahl der Mannschaften entspricht. \mathbf{x}_r stellt den zu Mannschaft r gehörenden Vektor der erklärenden Variablen dar. Durch den Zusatz solcher erklärenden Variablen lässt sich die Spielstärke einer Mannschaft bei der Annahme eines globalen Heimeffektes im Folgenden nicht mehr durch γ_r , sondern durch $\gamma_r + \mathbf{x}_r^T \boldsymbol{\beta}$ beschreiben. Für den linearen Prädiktor η_{rsj} ergibt sich nun

$$\eta_{rsj} = \alpha + \theta_j + \gamma_r - \gamma_s + (\mathbf{x}_r - \mathbf{x}_s)^T \boldsymbol{\beta}. \quad (5.1)$$

Für den Fall eines teamspezifischen Heimeffektes würde α einfach durch α_r ersetzt werden. Dieses Modell ist jedoch nicht eindeutig identifizierbar, weshalb eine Restriktion getroffen werden muss, um die Identifizierbarkeit zu gewährleisten. Eine von Springall (1973) vorgeschlagene Restriktion gewährleistet Identifizierbarkeit durch die Bedingung $\gamma_r = 0$ für alle $r = 1, \dots, m$. Dies führt zu einem sehr restriktiven Modell, da diese Bedingung die Annahme zur Folge hat, dass die modellierten Spielstärken vollständig durch die zusätzlichen Variablen erklärt werden können. Durch die Beschränkung auf wenige Variablen in dieser Arbeit erscheint diese Annahme hier nicht angemessen.

5 Modelle mit mannschaftsspezifischen Einflussgrößen

Ein weiterer Ansatz löst das Problem der Identifizierbarkeit durch die Annahme, die Spielstärken seien Zufallseffekte und können daher durch Random-Effects-Modelle geschätzt werden. Beschrieben wird dieser Ansatz von Turner und Firth (2012).

Betrachtet wird jedoch ein dritter Ansatz, welcher das Problem durch den Einsatz penalisierter Schätzprozesse innerhalb eines Modells mit festen Effekten löst. Angenommen die Spielstärken sind wie in Kapitel 4 geclustert, so kann Gleichung 4.2 verwendet werden. Dadurch erhält der nicht durch die Kovariablen erklärte Teil des Modells γ_r für $r = 1, \dots, m$ eine Penalisierung, der Parameter β jedoch nicht. Wiederum gilt hier: Konvergiert $\lambda \rightarrow \infty$, sind alle Koeffizienten γ_r identisch. Die Spielstärke ergibt sich dann ausschließlich durch $\mathbf{x}_r^T \beta$. Dies würde zum Modell von Springall (1973) führen. Für die Wahl eines positiven Tuning-Parameters wurde die zuletzt beschriebene Methode von Friedman et al. (2010) für überschätzte multinomiale Regressionsmodelle verwendet. Solch ein Modell mit festen Effekten hat den Vorteil, dass keine Annahme einer Unkorreliertheit zwischen Zufallseffekten und den Kovariablen nötig ist. Eine solche Annahme ist in unseren Betrachtungen aber sowieso nicht gefragt. Viel wichtiger ist der zweite Vorteil, den diese Methodik mit sich bringt: Durch die Verwendung von bereits bekannten Straftermen lässt sich die Methodik relativ problemlos in die bereits verwendete Schätzprozedur einfügen [vgl. Tutz und Schauburger (2015)].

Nun stellt sich noch die Frage, welche Kovariablen für das Modell verwendet werden sollen. Tutz und Schauburger (2015) berücksichtigten in ihrer Arbeit die Lizenzspieleretats der einzelnen Mannschaften, was zu guten Ergebnissen führte. Der Lizenzspieleretat gibt an, welches Budget einem Verein für die Gehälter all seiner Profis zur Verfügung steht. Da der Verein nicht verpflichtet ist, diese Zahlen zu veröffentlichen, können sie oft nur geschätzt werden. Ein weiteres Problem dieser Wahl als Kovariable ist, dass Vereine mit einem hohen Spieleretat meist für ähnlich starke Spieler grundsätzlich höhere Gehälter zahlen als Vereine mit niedrigerem Etat. Eine alternative Möglichkeit wäre daher die Betrachtung des durchschnittlichen Marktwertes der einzelnen Mannschaften. Dadurch wird das gerade beschriebene Problem des Spieleretats gelöst. Noch dazu kann durch die Betrachtung des Marktwertes zusätzliche Information gewonnen werden. Im Fußball besitzen Spieler meist mehrjährige Verträge. Es ist möglich, dass mehrere Spieler in kurzer Zeit einen Leistungssprung in ihrer Entwicklung machen. Dadurch steigt der Marktwert dieser Spieler an. Da diese Spieler durch Verträge jedoch an ihren Verein gebunden sind, können sich immer wieder Vereine mit einem relativ kleinen Spieleretat mittelfristig in der Bundesliga etablieren. So liefert der Marktwert in gewisser Weise Informationen über die jüngste Entwicklung eines Vereins.

5 Modelle mit mannschaftsspezifischen Einflussgrößen

Als weitere Alternative zur vorgestellten Wahl der Kovariablen können auch die durchschnittlichen Marktwerte der eingesetzten Spieler pro Partie betrachtet werden. Dadurch fließen in das Modell Informationen mit ein, ob eine Mannschaft seine besten Spieler auf das Feld schickte oder ob sie geschwächt antrat. Dies könnte das Modell unter Umständen noch weiter verbessern. Um solche spieltagsabhängigen Kovariablen ins Modell aufnehmen zu können, muss der lineare Prädiktor zu

$$\eta_{rsji} = \alpha + \theta_j + \gamma_r - \gamma_s + (\mathbf{x}_{r,i} - \mathbf{x}_{s,i})\boldsymbol{\beta} \quad (5.2)$$

erweitert werden. $i \in \{1, \dots, 34\}$ gibt dabei den jeweiligen Spieltag an. Diese Erweiterung erscheint vor allem in Hinblick auf die späteren prädiktiven Analysen in Kapitel 7 sinnvoll, da die Startaufstellungen der Mannschaften bereits vor dem Spiel bekannt gegeben werden und somit zusätzliche Informationen liefern.

5.2 Ergebnisse

Zum Vergleich werden zunächst die geschätzten Spielstärken der Mannschaften sowohl für den Lizenzspieleretat als auch für die Marktwerte der einzelnen Vereine betrachtet. Die Betrachtung der Ergebnisse bei ausschließlicher Berücksichtigung der zu Spielbeginn eingesetzten Spieler erfolgt anschließend. Die Spieleretats erschienen in einem Artikel von Reich und Fritzen (2014). Die Marktwerte können auf der Internetseite transfermarkt.de (2016) eingesehen werden. Sie werden regelmäßig aktualisiert. Die Analysen wurden mit den Marktwerten am 23.10.2014, 01.02.2015 und dem Mittel der beiden Werte durchgeführt. Die jeweiligen Startaufstellungen der Vereine finden sich ebenfalls auf transfermarkt.de (2016).

Abbildung 5.1 zeigt die Spielstärken der Mannschaften unter Berücksichtigung der durchschnittlichen Marktwerte der Vereine. Die einzelnen Vereinsmarktwerte sowie die Lizenzspieleretats der Vereine finden sich im Anhang.

Bei Betrachtung des BIC lassen sich die Teams in vier unterschiedliche Cluster einteilen. Wie auch schon bei den früheren Ergebnissen bilden nur sehr wenige Mannschaften die besten Cluster. Die vier Mannschaften der ersten beiden Cluster sind der VfL Wolfsburg, Borussia Mönchengladbach, Bayer 04 Leverkusen und der FC Augsburg. Es ist interessant zu sehen, dass diese vier Mannschaften auch in der realen Tabelle unter den besten fünf zu finden sind, obwohl die γ -Parameter nun die Spielstärke, welche nicht durch den

5 Modelle mit mannschaftsspezifischen Einflussgrößen

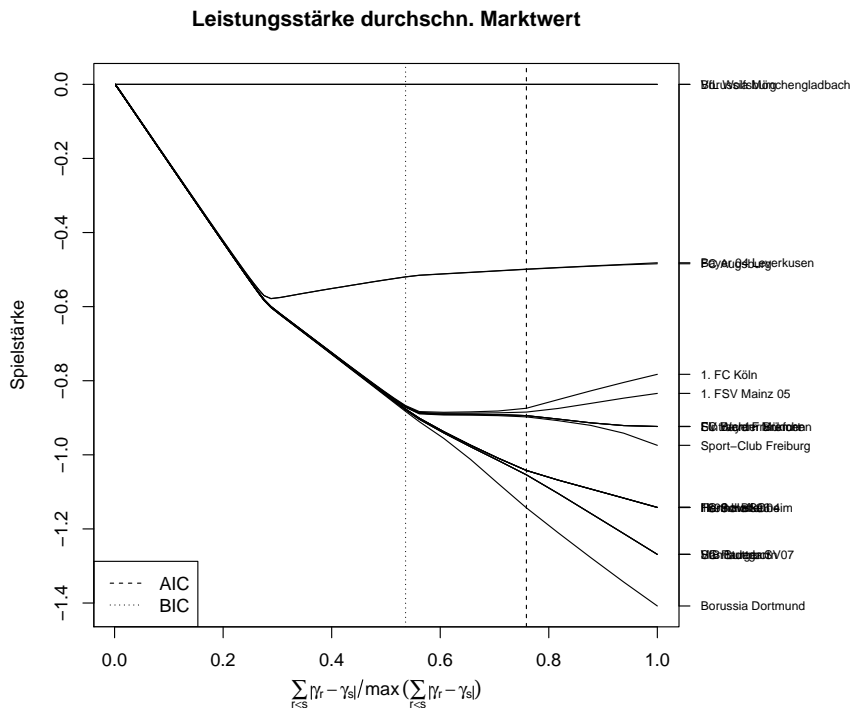


Abbildung 5.1: Koeffizientenpfade der Spielstärke bei globalem Heimeffekt und dem durchschnittlichen Marktwert als erklärende Variable

Marktwert des Teams erklärt werden kann, abbilden. Nur der FC Bayern München fehlt bei diesen Mannschaften. Dies ist wenig überraschend, da der FC Bayern den mit Abstand größten durchschnittlichen Marktwert besitzt und dadurch bereits deutliche Überlegenheit zu erwarten ist. Tabelle 5.1 zeigt die Einteilung der Vereine in die vier unterschiedlichen Cluster.

Zusätzlich fällt auf, dass sich die beiden großen Cluster nur um 0.01 Einheiten unterscheiden. Es kann daher unter Umständen auch von einer Einteilung in nur drei verschiedene Cluster gesprochen werden.

Der geschätzte Parameter für die erklärenden Variablen liegt bei $\beta = 0.1$ für das vom BIC gewählte $\lambda = 0.1$. β ist nicht in der Grafik aufgeführt, es sei aber gesagt, dass der Koeffizient für unterschiedliche λ relativ konstant bleibt. Da die Differenzen der durchschnittlichen Marktwerte zweier Teams teilweise sehr groß sind, kann suggeriert werden, dass die erklärenden Variablen einen relativ großen Einfluss auf das Modell haben. Die Betrachtung der durchschnittlichen Mannschaftswerte zu Saisonbeginn und zum Beginn der Rückrunde sind fast identisch zu den vorgestellten Ergebnissen. Die dazugehörigen

5 Modelle mit mannschaftsspezifischen Einflussgrößen

Cluster	Mannschaften	Spielstärke
1	Borussia Mönchengladbach, VfL Wolfsburg	0.00
2	Bayer 04 Leverkusen, FC Augsburg	-0.52
3	1.FC Köln, 1.FSV Mainz 05, Eintracht Frankfurt, FC Bayern München	-0.87
4	1899 Hoffenheim, Borussia Dortmund, FC Schalke 04, Hamburger SV Hannover 96, Hertha BSC Berlin, SC Paderborn 07, SC Freiburg Werder Bremen, VfB Stuttgart	-0.88

Tabelle 5.1: Cluster der Spielstärken bei globalem Heimeffekt und dem durchschnittlichen Marktwert als erklärende Variable unter Verwendung des BIC-Kriteriums

Grafiken finden sich im Anhang.

Bei Betrachtung des Lizenzspieleretats als Kovariable sieht die ermittelte Spielstärke etwas anders aus.

In Abbildung 5.2 ist bei Betrachtung der Ergebnisse des BIC zu erkennen, dass Borussia Mönchengladbach klar die größte Spielstärke besitzt. Von den ersten vier Mannschaften der Abschlusstabelle hat Borussia Mönchengladbach mit 38 Millionen Euro den mit Abstand kleinsten Lizenzspieleretat. Der VfL Wolfsburg und Bayer 04 Leverkusen bilden das zweite Cluster, dicht gefolgt vom FC Augsburg. Die Augsburger wurden in der Saison 2014/2015 Tabellenfünfter, hatten jedoch zwölf Punkte Rückstand auf die vor ihnen platzierten Leverkusener. Da den Augsburgern aber nur einen Etat von 19 Millionen Euro zur Verfügung stand, ist dies für sie ein sehr gutes Abschneiden und daher ist auch zu erklären, warum sie bei dieser Betrachtung gut abschneiden. Die restlichen Mannschaften bilden noch weitere zwei Cluster, welche sich nur um 0.01 Einheiten unterscheiden. Der geschätzte Parameter für die erklärenden Variablen beträgt $\beta = 1.44$ für das vom BIC gewählte $\lambda = 0.15$. Im Vergleich zur vorherigen Betrachtung ist dieses β sehr hoch. Dies liegt an den erheblichen Unterschieden in der Größe der einzelnen Variablen. Der Einfluss der Kovariablen kann aber auch hier als relativ stark angesehen werden.

Eine Möglichkeit, den Einfluss der erklärenden Variablen auf die Spielstärke zu ermitteln, besteht darin, die Spielstärke als abhängige Variable und die gewählten Kovariablen für die Teams als unabhängige Variable zu modellieren. Im Folgenden sei der durchschnittliche Marktwert der Teams die gewählte unabhängige Variable. Es besteht hier die Frage, ob der Marktwert als linearer Effekt oder als nicht-linearer Effekt angesehen werden soll. Abbildung 5.3 bildet die in diesem Fall zentrierten Spielstärken und durchschnittlichen Marktwerte der Teams ab. Zusätzlich ist der oberen Grafik die lineare und der unteren Grafik die additive Regressionsgleichung des Modelles angefügt.

5 Modelle mit mannschaftsspezifischen Einflussgrößen

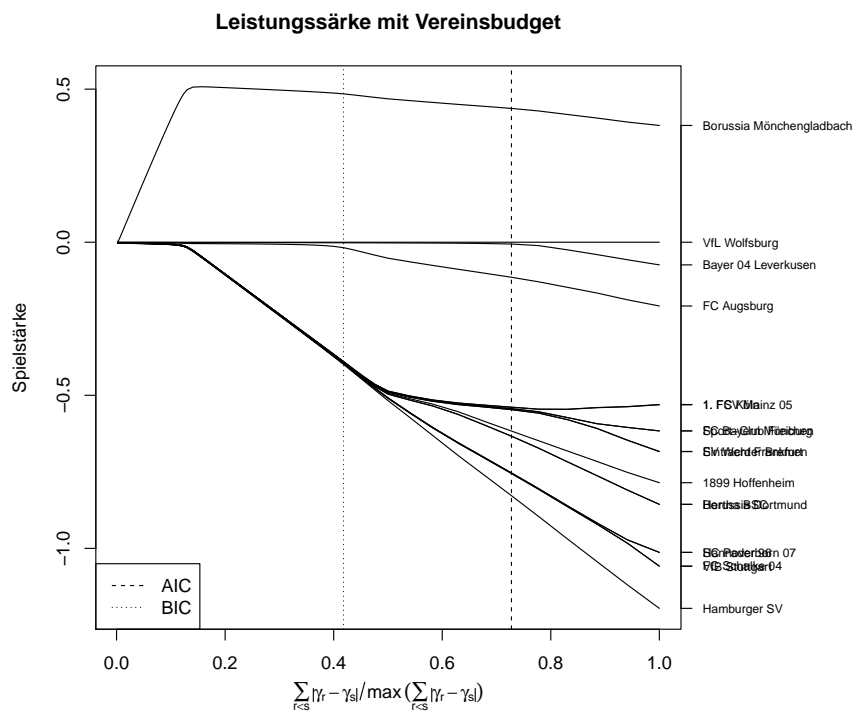


Abbildung 5.2: Koeffizientenpfade der Spielstärke bei globalem Heimeffekt und dem Vereinsbudget als erklärende Variable

5 Modelle mit mannschaftsspezifischen Einflussgrößen

Es ist zu erkennen, dass das lineare Modell bereits eine gute Annäherung an die Daten

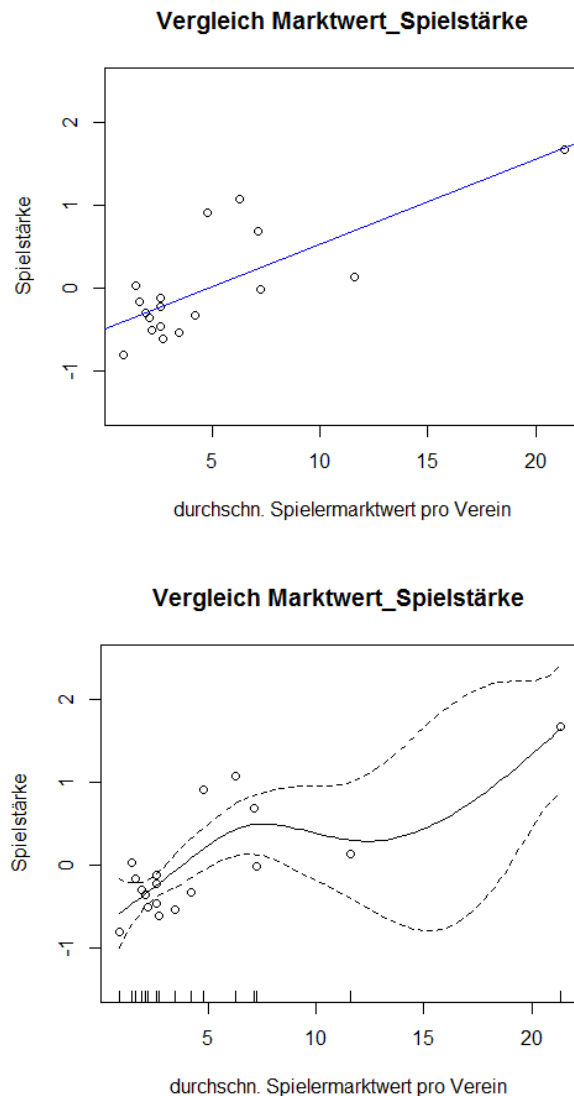


Abbildung 5.3: durchschnittlicher Marktwert der Spieler der Vereine im Vergleich zu den Spielstärken der Mannschaften mit zusätzlicher Angabe des linearen und additiven Modellfits

fittet. Das adjustierte Bestimmtheitsmaß beim linearen Modell liegt bei 0.557, d.h mehr als 55% der Streuung werden durch die Marktwerte der Mannschaften erklärt. Bei einem nicht-linearen Modell steigt dieser Wert sogar auf 0.628. Durch die Verteilung der einzelnen durchschnittlichen Marktwerte der Vereine ist ein overfitting zu erwarten, da bei einem durchschnittlichen Marktwert von über 10 Millionen Euro pro Spieler das Modell nur

5 Modelle mit mannschaftsspezifischen Einflussgrößen

durch zwei Beobachtungen erklärt wird. Dies ist an dem sehr breiten Konfidenzintervall des zusätzlich angegebenen Glättungseffektes zu erkennen. Ein weiteres Indiz für das overfitting liefern die Veränderungen des adjustierten Bestimmtheitsmaßes bei Betrachtung der Marktwerte zu unterschiedlichen Zeitpunkten. So liegt der Wert bei der Betrachtung der Marktwerte zum 23.10.2014 bei einem Wert von 0.79, am 01.02.2015 sogar bei 0.82. Der Wert von 0.628 ergibt sich bei Betrachtung der Mittelwerte der beiden Zeitpunkte. Wird nochmals das lineare Modell zu Grunde gelegt, so lassen sich mit diesem die erhaltenen Cluster gut erklären. Im Bereich zwischen 5 und 8 Millionen Euro Marktwert pro Spieler befinden sich drei Mannschaften, welche sich deutlich von der Regressionsgeraden abheben. Dies sind Borussia Mönchengladbach, Bayer 04 Leverkusen und der VfL Wolfsburg. Bei den Teams mit kleinem Marktwert hebt sich ebenso eine Mannschaft etwas heraus. Hierbei handelt es sich um den FC Augsburg. Dies sind auch die Teams, welche sich in Abbildung 5.3 als stärkste Mannschaften zeigen. Der FC Bayern schneidet trotz seiner Ausnahmestellung beim Marktwert vergleichbar mit den restlichen Mannschaften ab. Bei Betrachtung des Lizenzspieleretats als unabhängige Variable kann man konstatieren, dass das Modell keinen nicht-linearen Effekt feststellt. Es kann daher nur über die lineare Regressionsgleichung beschrieben werden. Hierbei ergibt sich ein Wert für das adjustierte Bestimmtheitsmaß von 0.529. Wie bereits vermutet, liefert der Marktwert bessere Informationen zur Modellerklärung als das Vereinsbudget.

Abschließend werden in diesem Kapitel nun die Koeffizientenpfade der Spielstärke der Mannschaften unter Berücksichtigung der durchschnittlichen Marktwerte der Startaufstellungen betrachtet. Abbildung 5.4 stellt diese dar.

Bei Betrachtung des BIC-Kriteriums zur Wahl der Koeffizientenschätzer lassen sich die Spielstärken der Mannschaften in sechs verschiedene Cluster einteilen, wie Tabelle 5.2 zeigt.

Borussia Mönchengladbach und der VfL Wolfsburg bilden wie bei der Betrachtung der durchschnittlichen Marktwerte den ersten Cluster. Der zweite Cluster ist mit dem FC Bayern München besetzt. Dies ist sehr überraschend, da Bayern München mit erheblichem Abstand den größten durchschnittlichen Marktwert der Spieler bei den Startaufstellungen aufweist. Er liegt bei einem Wert von über 27 Millionen Euro. Dies sind fast 12 Millionen mehr als der durchschnittliche Wert der Spieler von Borussia Dortmund, welche bei den Startaufstellungen den zweitstärksten Kader stellen, und liegt über 20 Millionen Euro höher als der durchschnittliche Wert aller Vereine. Von daher ist eher zu erwarten, dass sich die Spielstärke von Bayern München hauptsächlich durch diesen Wert erklären lässt. Bayer Leverkusen bildet den nächsten Cluster. Mit einer Abweichung von nur 0.01 Einheiten zu

5 Modelle mit mannschaftsspezifischen Einflussgrößen

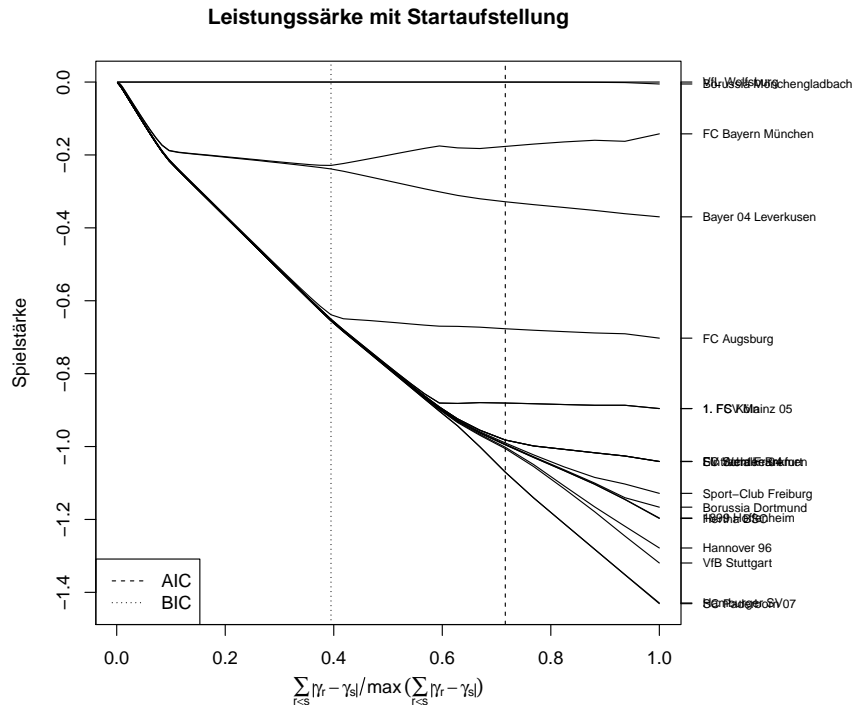


Abbildung 5.4: Koeffizientenpfade der Spielstärke bei globalen Heimeffekt und dem durchschnittlichen Marktwert der Startaufstellung als erklärende Variable

Cluster	Mannschaften	Spielstärke
1	Borussia Mönchengladbach, VfL Wolfsburg	0.00
2	FC Bayern München	-0.23
3	Bayer 04 Leverkusen	-0.24
4	FC Augsburg	-0.64
5	1.FC Köln, 1.FSV Mainz 05, Eintracht Frankfurt, 1899 Hoffenheim, Borussia Dortmund, FC Schalke 04, Hannover 96, Hertha BSC Berlin, SC Freiburg, Werder Bremen, VfB Stuttgart	-0.65
6	Hamburger SV, SC Paderborn 07	-0.66

Tabelle 5.2: Cluster der Spielstärken bei globalem Heimeffekt und dem durchschnittlichen Marktwert der Startaufstellungen als erklärende Variable unter Verwendung des BIC-Kriteriums

5 Modelle mit mannschaftsspezifischen Einflussgrößen

Bayern München könnten diese beiden Vereine auch in einen Cluster zusammengefasst werden. Der FC Augsburg bildet ebenfalls einen eigenen Cluster. Die Differenz zu den nächsten beiden Clustern, in denen sich alle restlichen Mannschaften befinden, beträgt jedoch nur 0.01, bzw. 0.02 Einheiten, so dass theoretisch alle diese Mannschaften als ein Cluster angesehen werden können.

Der geschätzte Parameter für die erklärenden Variablen liegt bei $\beta = 0.59$ für den durch das BIC-Kriterium bestimmten Parameter $\lambda = 0.17$. Der Einfluss der erklärenden Variablen scheint auch hier nicht vernachlässigbar zu sein.

Um dies zu überprüfen, werden auch hier die Spielstärke als abhängige und die durchschnittlichen Marktwerte als unabhängige Variablen betrachtet. Dargestellt wird dies in Abbildung 5.5.

Da die Startaufstellungen der Teams von Spiel zu Spiel meistens variieren, gibt es in diesem Fall wesentlich mehr unterschiedliche Datenpunkte. Die Spielstärken jeder Mannschaft sind für alle Spiele jedoch gleich, so dass pro Spielstärke 34 Datenpunkte existieren. Betrachtet man das Modell mit einem additiven Effekt, so fällt auf, dass die Konfidenzbänder eine wesentlich geringere Bandbreite aufweisen als bei Betrachtung der durchschnittlichen Marktwerte. Für das lineare Regressionsmodell ergibt sich ein Wert für das adjustierte Bestimmtheitsmaß von 0.567, beim additiven Modell sogar von 0.694. Es lassen sich also 69.4% der Streuung durch den durchschnittlichen Wert der Startaufstellungen erklären. Dieser Wert stellt nochmals eine deutliche Steigerung zu obigen Betrachtungen dar. Vor allem die Senkung der Varianz durch die Variabilität der unabhängigen Variablen bietet eine starke Verbesserung des Modells.

5 Modelle mit mannschaftsspezifischen Einflussgrößen

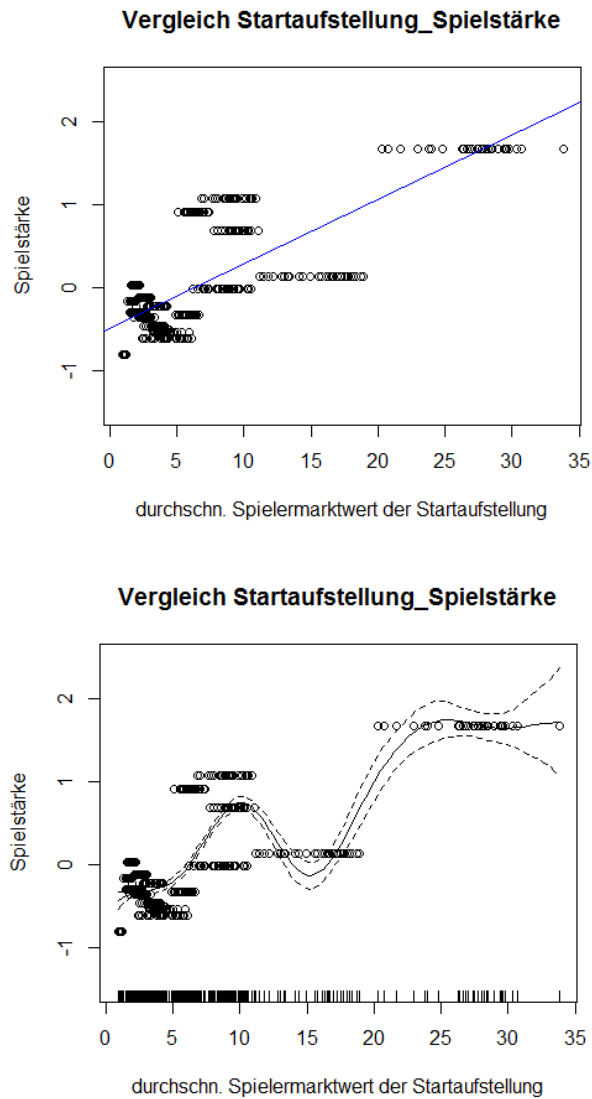


Abbildung 5.5: durchschnittlicher Marktwert der Startaufstellungen im Vergleich zu den Spielstärken der Mannschaften mit zusätzlicher Angabe des linearen und additiven Modellfits

6 Dynamische Bradley-Terry Modellierung

Nachdem in den vorangegangenen Kapiteln die Modellierung auf statische Modelle beschränkt war, sollen nun dynamische Ansätze für die Spielstärkemodellierung betrachtet werden. Die Bundesligasaison ist in 34 Spieltage aufgeteilt, wobei die Spiele bis auf eine längere Pause im Winter meist wöchentlich an den Wochenenden ausgetragen werden. Aufgrund der langen Dauer einer Saison ist es unwahrscheinlich, dass eine Mannschaft über die gesamte Saison eine konstante Leistung bringt. Wahrscheinlicher ist es, dass ein Team Leistungsschwankungen über die Saison aufweist, sei es aufgrund von Verletzungen wichtiger Spieler oder einfach weil einige Spieler eine Formschwäche aufweisen. Harville (1980) nahm für eine dynamische Modellierung der Stärke von Mannschaften einen autoregressiven Prozess an, welcher die Mannschaftsstärken saisonweise modellierte. Eine Analyse der wöchentlichen Stärkeparameter der Mannschaften führten Glickman und Stern (1998) unter Annahme eines autoregressiven Prozesses erster Ordnung für die wöchentlichen Veränderungen durch. Dieses Kapitel widmet sich zunächst einem Ansatz von Cattelan et al. (2013), in welchem die Spielstärke einer Mannschaft durch einen exponential-gewichteten moving-average Prozess beeinflusst wird. Zunächst wird hierfür das Bradley-Terry-Modell zu einem dynamischen Modell erweitert und die Modellierung der Spielstärken beschrieben. Des Weiteren wird ein Score zur Validierung des Modells vorgestellt. Als zusätzlichen Ansatz zur dynamischen Modellierung der Spielstärke wird anschließend versucht, die Likelihood der einzelnen Spiele mit Hilfe einer Kernfunktion zu gewichten, je nachdem, wie lange die Spiele bereits zurückliegen. Anschließend folgt wiederum die Anwendung der beschriebenen Modelle auf die Daten.

6.1 Modellansatz nach Cattelan

In den vorangegangenen Kapiteln wurde ersichtlich, dass die Spielstärke davon abhängt ob eine Mannschaft zu Hause oder auswärts spielt. Hierfür wurde dann sowohl ein globaler als auch ein teamspezifischer Heimeffekt mit modelliert. Beim dynamischen Ansatz von Cattelan et al. (2013) werden nun für jede Mannschaft eigene Spielstärken für Heim- und

6 Dynamische Bradley-Terry Modellierung

Auswärtspartien bestimmt, so dass der Heimeffekt α nicht benötigt wird. Das dynamische Bradley-Terry Modell lautet daher:

$$P(Y_{rs,i} \leq j | Y_{i-1} = y_{i-1}, \dots, Y_1 = y_1) = \frac{\exp(\theta_j + \gamma_r(t_i) - \gamma_s(t_i))}{1 + \exp(\theta_j + \gamma_r(t_i) - \gamma_s(t_i))}. \quad (6.1)$$

$Y_i, i = 1, \dots, 34$ stellt einen Sammelvektor dar, welcher für jeden Spieltag die einzelnen Ergebnisse der neun stattfindenden Partien enthält. $Y_{rs,i}$ bezeichnet eine bestimmte Partie zwischen Mannschaft r und Mannschaft s am i -ten Spieltag. Das Modell berechnet die kumulierte Wahrscheinlichkeit des Spielausganges eines Spiels zwischen zwei Mannschaften r und s unter der Bedingung, dass die Ergebnisse der zurückliegenden Spiele beider Teams bekannt sind. $\gamma_r(t_i)$ gibt die Spielstärke der Heimmannschaft zum Zeitpunkt t_i an. $\gamma_s(t_i)$ beschreibt folglich die Spielstärke der Gastmannschaft zu diesem Zeitpunkt. Im Folgenden wird angenommen, dass sich die Entwicklung der Spielstärke der Heimmannschaft nur aus den Ergebnissen der vergangenen Heimspiele, nicht aber aus den Auswärtsspielen ergibt. Die Leistungsänderung der Gastmannschaft soll hingegen nur von den vergangenen Auswärtsspielen abhängen. Des Weiteren sei $t_i^{(-1)}$ der Zeitpunkt des letzten Heimspiels von Mannschaft r . Die Heimspielstärke von Team r lässt sich nach Cattelan et al. (2013) nun durch den exponential-gewichteten moving-average (EWMA) Prozess

$$\gamma_r(t_i) = \lambda_1 \mu_r(t_i) + (1 - \lambda_1) \gamma_r(t_i^{(-1)}) \quad (6.2)$$

berechnen. Bei $\lambda_1 \in [0, 1]$ handelt es sich dabei um einen heimspezifischen Glättungsparameter und bei

$$\mu_r(t_i) = \beta_1 r_r(t_i^{(-1)}) \quad (6.3)$$

um die gemittelte Heimstärke von Team r , welche nur vom Ausgang des letzten Heimspiels $r_r(t_i^{(-1)})$ und einem heimspezifischen Parameter β_1 abhängt. Für $r_r(t_i^{(-1)})$ wird festgelegt:

$$r_r(t_i^{(-1)}) = \begin{array}{ll} -1, & \text{Niederlage mit mindestens 2 Toren Differenz} \\ 0, & \text{Niederlage mit genau einem Tor Differenz} \\ 1, & \text{Unentschieden} \\ 3, & \text{Sieg mit genau einem Tor Differenz} \\ 4, & \text{Sieg mit mindestens 2 Toren Differenz} \end{array} \quad (6.4)$$

6 Dynamische Bradley-Terry Modellierung

Da am ersten Spieltag noch kein Spiel vorher stattgefunden hat, muss für Gleichung 6.2 noch eine Startbedingung festgelegt werden. Im Folgenden werden hierfür zwei verschiedene Möglichkeiten betrachtet, zum einen die Grundannahme, dass alle Teams mit gleichen Heim- bzw. Auswärtsspielstärken in die Saison starten, und zum anderen, dass jede Mannschaft die Saison mit einer individuellen Spielstärke beginnt. Als Startwerte werden bei beiden Ansätzen jeweils die durchschnittlich pro Spiel gesammelten Punkte aller Mannschaften in der Vorsaison festgelegt, wobei hierfür nicht das klassische Punktesystem, sondern die Verteilung aus Gleichung 6.4 verwendet wird. In der Saison 2013/2014 erzielten die Mannschaften hierbei durchschnittlich $\bar{r}_r = 1.761$ Punkte pro Heimspiel. Die Startwerte bei individueller Heim- bzw. Auswärtsspielstärke sind in Tabelle 6.1 gegeben. Bei Betrachtung von Gleichung 6.2 und Gleichung 6.3 fällt auf, die Berechnung der ak-

Mannschaft	Startwert Heimstärke	Startwert Auswärtsstärke
1. FC Köln	0.471	0.353
1. FSV Mainz	2.176	0.765
1899 Hoffenheim	1.882	0.824
Bayer 04 Leverkusen	2.294	1.765
Borussia Dortmund	2.353	2.412
Borussia Mönchengladbach	2.588	1.000
Eintracht Frankfurt	1.176	0.765
FC Augsburg	1.882	1.176
FC Bayern München	3.294	3.235
FC Schalke 04	2.765	1.530
Hamburger SV	0.824	0.235
Hannover 96	1.765	0.412
Hertha BSC	1.118	1.059
SC Paderborn 07	1.000	-0.118
SC Freiburg	1.176	0.588
SV Werder Bremen	1.294	0.588
VfB Stuttgart	1.235	0.588
VfL Wolfsburg	2.412	1.353

Tabelle 6.1: Startwerte Heim- bzw. Auswärtsspielstärke für $k=5$

tuellen Spielstärke einer Mannschaft hängt von der Spielstärke und dem Ergebnis des vorangegangenen Spiels ab. Nun sei angenommen, Mannschaft r hat bisher N Heimspiele bestritten. Die Spielstärke für die aktuelle Partie kann nun iterativ durch

$$\gamma_r(t_i) = \beta_1 \left\{ \lambda_1 \sum_{n=0}^{N-1} (1 - \lambda_1)^n r_r(t_i^{(-n-1)}) + (1 - \lambda_1)^N \bar{r}_r \right\} = \beta_1 x_r(t_i; \lambda_1) \quad (6.5)$$

6 Dynamische Bradley-Terry Modellierung

modelliert werden. Die Heimspielstärke $\gamma_r(t_i)$ ist also eine Funktion, welche sich aus den Ergebnissen aller vergangenen Heimspiele $r_r(t_i^{(-1)}), r_r(t_i^{(-2)}), \dots, r_r(t_i^{(-N)})$ ermitteln lässt. Dabei wird ein gewichtetes Mittel der vergangenen Ergebnisse berechnet, wobei der Einfluss der länger zurückliegenden Ergebnisse immer weiter abnimmt und geometrisch gegen Null wandert. Durch den Glättungsparameter λ_1 wird die Abhängigkeit der Spielstärke von den vergangenen Ergebnissen bestimmt. Ein λ_1 nahe Null bedeutet, die Spielstärke der Mannschaft bleibt relativ konstant, da der Einfluss der vergangenen Spiele nur langsam abnimmt. Somit würde sich wie in Abschnitt 3.3 quasi ein teamspezifischer Heimspieleffekt ermitteln lassen. Bei einem Glättungsparameter nahe eins hängt γ_r hauptsächlich nur vom Ausgang der letzten Heimpartie ab.

Die eben beschriebenen Modelle dienen zur Modellierung der Heimspielstärke der einzelnen Mannschaften. Äquivalent hierzu lässt sich auch die Auswärtsspielstärke der Mannschaften modellieren. Hierzu wird ein zweiter EWMA-Prozess

$$\gamma_s(t_i) = \lambda_2 \mu_r(t_i) + (1 - \lambda_2) \gamma_s(t_i^{(-1)})$$

verwendet. Anstatt der Heimspiele werden nun die vergangenen Auswärtsspiele der einzelnen Mannschaften betrachtet. Die Bedeutung und Interpretation der einzelnen Parameter gleicht dem Modell für die Heimspielstärke. Als Startwerte für die Auswärtsspielstärke werden nun die durchschnittlich gesammelten Punkte pro Auswärtsspiel - im ersten Ansatz gesamt, im zweiten Ansatz für jedes Team individuell - verwendet. Für die Saison 2013/2014 waren dies $\bar{r}_s = 1.029$ Punkte pro Auswärtspartie bei Betrachtung aller Spiele. Die individuellen Startwerte finden sich in Tabelle 6.1.

Da sich die Spielstärken iterativ berechnen lassen, sind bei diesem Ansatz von Cattelán et al. (2013) nur sechs Parameter zu schätzen. Zum Vergleich: Beim statischen Bradley-Terry Modell müssen 20 Parameter geschätzt werden. Dabei werden zuerst die beiden λ 's mit Hilfe einer Profile-Likelihood-Funktion bestimmt und anschließend werden die restlichen vier Parameter ermittelt. Es sei aber erwähnt, dass diese vier restlichen Parameter bereits indirekt bei der Bestimmung der λ -Parameter mit geschätzt werden.

6.2 Modellvalidierung

Es stellt sich nun die Frage, ob dieses Modell valide ist oder nicht. Hierzu sollen die Ergebnisse des dynamischen Modells von Cattelan et al. (2013) denen des statischen Bradley-Terry Modells von Tutz und Schauburger (2015) entgegengestellt werden, d.h. es werden die geschätzten Wahrscheinlichkeiten des dynamischen Modells mit denen des statischen Modells verglichen. Hier gibt es zu beachten, dass sich die zeitabhängigen Spielstärken bei einem möglichen Response mit $k = 5$ verschiedenen Spielausgängen mit nur sechs Parametern bestimmen lassen, wohingegen beim statischen Modell $m + 2$ freie Parameter nötig sind. m ist dabei die Anzahl der Mannschaften. Bei Daten der Fußballbundesliga werden also 20 Parameter geschätzt. Dadurch ist zu erwarten, dass das statische Modell die Daten besser fitet. Das statische Modell soll daher einen Maßstab für die Güte des dynamischen Modells bilden. Je ähnlicher sich die Wahrscheinlichkeiten der beiden Modelle sind, desto besser ist der Fit des dynamischen Modells. Für den Vergleich beider Modelle bietet sich an, einen Score zu verwenden. Cattelan et al. (2013) greifen in ihren Ausführungen auf den sogenannten rank probability score zurück. Dieser wird von Czado et al. (2009) für den Fall von mehr als zwei Responsekategorien vorgeschlagen und auch in dieser Arbeit verwendet.

$$RPS_{rs,i} = \sum_{j=0}^{k-1} \{P(Y_{rs,i} \leq j | Y_{i-1} = y_{i-1}, \dots, Y_1 = y_1) - 1_{\{y_i \leq j\}}\}^2 \quad (6.6)$$

Der Score berechnet sich aus der Summe der quadrierten Differenzen der modellierten kumulierten Wahrscheinlichkeiten des Spielausganges und einer Indikatorvariable, welche angibt, ob das Ergebnis bereits eingetreten ist. Der Score kann bei der Wahl von $k = 5$ einen Wert zwischen Null und vier annehmen, wobei bei einem Wert von Null das Modell die tatsächlichen Ergebnisse perfekt widerspiegeln würde. Ein Wert von vier würde bedeuten, es gab nur Ergebnisse mit klarer Dominanz einer Mannschaft, das Modell legt seine komplette Wahrscheinlichkeitsmasse jedoch auf die klare Dominanz der jeweils anderen Mannschaft. Der Score kann auch auf alle anderen vorgestellten Modelle angewendet werden und bietet daher eine gute Möglichkeit zum Modellvergleich. Ein ausführlicher Vergleich, welcher auch die Stärken und Schwächen dieses Scores beleuchtet, folgt in Kapitel 7.

6.3 Dynamischer Ansatz über gewichtete Likelihoodfunktion

Beim Modellansatz von Cattelan et al. (2013) spielt zwar die Reihenfolge der einzelnen Spieltage eine Rolle, jedoch fließt in dieses Modell nicht mit ein, welcher zeitliche Abstand zwischen den einzelnen Spielen vorherrscht. Da die Spiele der Bundesliga aufgrund von Länderspielpausen, der Winterpause oder auch durch einzelne Spieltage unter der Woche jedoch nicht in regelmäßigen Abständen stattfinden, sollte dies nicht außer Acht gelassen werden. Herrscht zum Beispiel eine Länderspielpause, so haben die Mannschaften meist zwei Wochen kein Bundesligaspiel und können diese Zeit nutzen, um ihre aktuelle Form zu verbessern. Daher scheint es realistisch, dass länger zurückliegende Spiele einen geringeren Einfluss haben. Ein Ansatz für eine solche dynamische Modellierung der Spielstärke beschäftigt sich mit der Gewichtung der Likelihood der einzelnen Spiele. Hierzu soll die Likelihoodfunktion mit einer Kernfunktion kombiniert werden, um eine zeitliche Gewichtung der einzelnen Spiele zu erhalten.

Als Ausgangspunkt für dieses Modell wird zunächst das statische Bradley-Terry Modell zu einem dynamischen Modell erweitert. Im Gegensatz zum Ansatz von Cattelan et al. (2013) sollen Heim- und Auswärtsspiele nicht getrennt betrachtet werden, so dass ein globaler bzw. teamspezifischer Heimkoeffizient α_r verwendet werden muss. Die kumulierten Wahrscheinlichkeiten des Modells ergeben sich daher als

$$P(Y_{rs,i} \leq j | Y_{i-1} = y_{i-1}, \dots, Y_1) = \frac{\exp(\alpha_r(t_i) + \theta_j(t_i) + \gamma_r(t_i) - \gamma_s(t_i))}{1 + \exp(\alpha_r(t_i) + \theta_j(t_i) + \gamma_r(t_i) - \gamma_s(t_i))}. \quad (6.7)$$

Die Gleichung erinnert auf den ersten Blick sehr stark an Gleichung 6.1 aus Abschnitt 6.1, jedoch sind nun alle Koeffizienten vom gewählten Zeitpunkt abhängig und die Interpretation der einzelnen Koeffizienten ist hier etwas anders. Y_i stellt einen Sammelvektor der vergangenen Spieltage dar und auch $Y_{rs,i}$ beschreibt wie in Gleichung 6.1 ein Spiel zwischen den Mannschaften r und s am i -ten Spieltag. $\gamma_r(t_i)$ ist hier jedoch nicht die Heimspielstärke von Mannschaft r zum Zeitpunkt t_i , sondern ebenso wie $\gamma_s(t_i)$ die Auswärtsspielstärke des Teams zum Zeitpunkt t_i . Die Heimstärke ergibt sich durch den zusätzlichen zeitpunktabhängigen Koeffizienten $\alpha_r(t_i)$, welcher diesem Modell hinzugefügt ist. Die Schwellenparameter θ hängen ebenfalls vom gewählten Zeitpunkt ab. Bei konstanten Spielstärken, Heimeffekt und Schwellenparametern über die Zeit würde t_i keinen Einfluss auf das Modell haben und man erhielte das statische Bradley-Terry-Modell

6 Dynamische Bradley-Terry Modellierung

aus Kapitel 3. Es lässt sich somit feststellen, dass sich ein einfaches dynamisches Modell aus dem statischen Modell entwickeln lässt, indem die Koeffizientenschätzer für jeden Spieltag einzeln ermittelt werden. So ergibt sich das dynamische Modell aus dem Zusammenfügen mehrerer statischer Modelle.

Wie bereits erwähnt, sollen die Koeffizienten in diesem Modell nun zusätzlich durch einen gewichteten Log-Likelihood-Ansatz geschätzt werden. Als Gewichtungsfaktor wird eine Kernfunktion mit einem Gaußkern verwendet. Die Log-Likelihoodfunktion ergibt sich als

$$l(t_0|\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \sum_{r,s,t} \sum_{u=1}^k 1_{\{Y_{rs}=u\}} \cdot \log(P(Y_{rs} = u)) \cdot w(t_0, t) \quad (6.8)$$

mit

$$w(t_0, t) \propto K\left(\frac{t - t_0}{\sigma}\right).$$

Im Gegensatz zur Log-Likelihoodfunktion aus Gleichung 4.1 spielen nun die Zeitpunkte der einzelnen Spiele eine Rolle. Die Log-Likelihoodfunktion zum Zeitpunkt t_0 ergibt sich also aus der Summe der Log-Likelihoodfunktionen aller Spiele zu den unterschiedlichen Zeitpunkten $t < t_0$, der Gewichtungsfaktor $w(t_0, t)$ aus der Kernfunktion $K(\frac{t-t_0}{\sigma})$. Der Parameter σ gibt die Bandbreite der Kernfunktion an. Mit Hilfe dieses Parameters kann die Log-Likelihoodfunktion optimiert werden. Der Parameter ist allerdings unbekannt und muss deshalb geschätzt werden. Dies geschieht mittels prädiktiver Simulation vergangener Bundesligadaten aus dem Jahr 2013/14. Die genaue Vorgehensweise und Wahl des Parameters werden in Unterabschnitt 6.4.2 beschrieben.

6.4 Anwendung

6.4.1 EWMA-Prozesse

Gemeinsame Startwerte

Wird die beschriebene Methodik nun auf die Daten angewandt, so erhält man als Schätzer für die beiden Glättungsparameter die Werte $\lambda_1 = 0.198$ und $\lambda_2 = 0.142$. Da die Werte nahe 0 sind, bedeutet dies, dass die Spielstärken auch von den schon länger vergangenen

6 Dynamische Bradley-Terry Modellierung

Spielen abhängen und nicht nur vom Ergebnis der letzten Partie. Wären beide Parameter $\lambda_1 = \lambda_2 = 0$, so würde dies eine Variante eines statischen Bradley-Terry Modells mit globalem Heimeffekt darstellen. Diese Variante wird vom Modell jedoch eindeutig nicht unterstützt. Die maximierte Profile-Log-Likelihood für das geschätzte Modell liegt bei -466.47, für das Modell mit $\lambda_1 = \lambda_2 = 0$ ist sie wesentlich niedriger bei -478.28. Schätzt man nun die beiden Heim- bzw. Auswärtskoeffizienten β_1 und β_2 , sowie die Schwellenparameter θ mit Hilfe der erhaltenen Glättungsparameter, so ergeben sich Werte von $\beta_1 = 0.809$ und $\beta_2 = 0.862$ mit Standardfehlern von 0.146 bzw. 0.217, sowie $\theta_1 = -\theta_4 = -1.637$ und $\theta_2 = -\theta_3 = -0.613$ mit Standardfehlern von 0.062 bzw. 0.100. Wird der rank-probability score für jedes Spiel der Saison berechnet, so erhält man einen Mittelwert von 0.700 und einen Median von 0.533 für das dynamische Modell. Beim statischen Modell aus Kapitel 3 beträgt der Score 0.635 für den Mittelwert und 0.482 für den Median. Der Score ist für das dynamische Modell daher um 10.2%, bzw. 10.6% höher als beim statischen Modell. Die Korrelation der Scores der beiden Modelle beträgt 0.785.

Mit Hilfe dieser geschätzten Parameter können nun die Spielstärken der einzelnen Mannschaften unterschieden nach Heim- und Auswärtsstärke berechnet werden. Abbildung 6.1 zeigt beispielhaft die Spielstärkeentwicklung für einige Mannschaften.

Durch die erste gewählte Anfangsbedingung der Spielstärke starten alle Mannschaften mit gleichen Spielstärken in die Saison. Die durchgezogene Linie zeigt die Heimspielstärke zum jeweiligen Heimspiel, die gestrichelte Linie die Auswärtsspielstärke. Es werden je 17 Heim- und Auswärtsspiele pro Saison bestritten. Die Überlegenheit des FC Bayern München in der Saison 2014/2015 ist durch seine rapide Verbesserung der Spielstärke sofort erkennbar. Vor allem bei der Auswärtsspielstärke ist diese Entwicklung im Vergleich zu den anderen Mannschaften hervorzuheben. Borussia Mönchengladbach verbesserte ihre Spielstärke vor allem in der zweiten Saisonhälfte. In der Rückrunde holte sie tatsächlich die meisten Punkte aller Mannschaften. Auch bei Borussia Dortmund ist in der zweiten Saisonhälfte ein Anstieg der Spielstärke zu verzeichnen. Holte sie in der Hinrunde gerade einmal 15 Punkte und war auf Position 17 platziert, so waren es in der Rückrunde 31 Punkte und sie beendete die Saison noch auf Platz sieben. Eintracht Frankfurt konnte in der Rückrunde kein einziges Auswärtsspiel gewinnen und erreichte nur zwei Unentschieden auf fremdem Platz. Dementsprechend verschlechtert sich auch ihre Auswärtsspielstärke. Die Spielstärkeentwicklungen der restlichen Mannschaften finden sich im Anhang.

Da die Spielstärken nur indirekt geschätzt werden ist im Modell keine Bedingung wie in Kapitel 3 nötig, um die Identifizierbarkeit des Modells gewährleisten zu können. Vielmehr hängen sie von der Anfangsbedingung der gewählten Heim- und Auswärtsspielstärken

6 Dynamische Bradley-Terry Modellierung

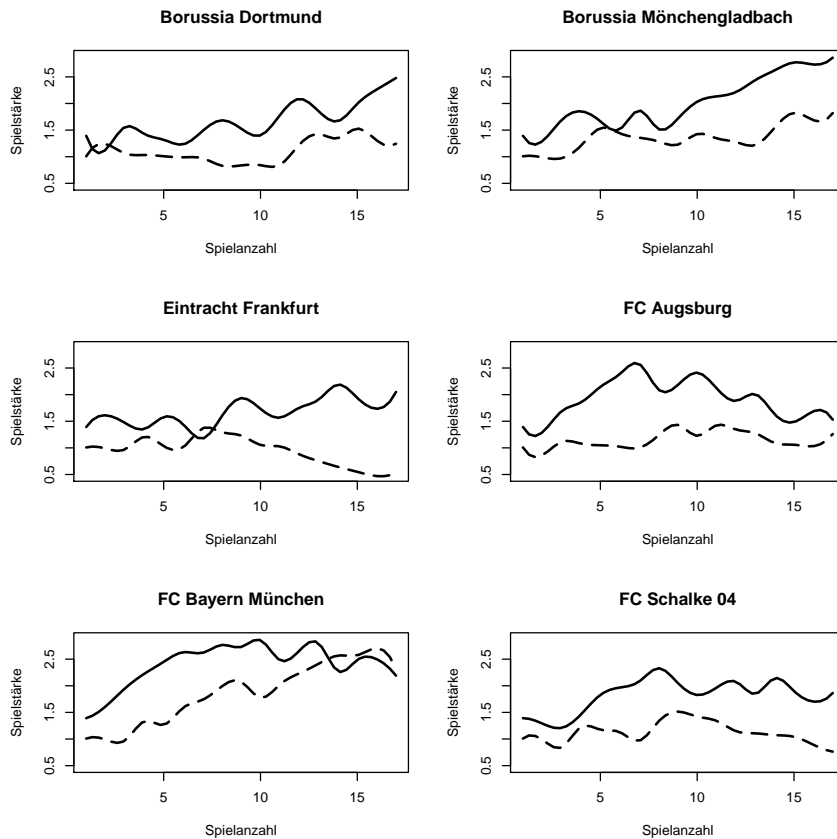


Abbildung 6.1: Spielstärkeentwicklung einzelner Mannschaften mit Unterscheidung zwischen Heim- (durchgezogene Linie) und Auswärtsspielstärke (gestrichelte Linie).

ab. Um trotzdem eine gewisse Vergleichbarkeit des statischen und dynamischen Bradley-Terry-Modells herzustellen wird nun im Nachhinein festgelegt, dass der Mittelwert der dynamischen Spielstärken des VfL Wolfsburg den Referenzwert bezeichnen soll und dieser daher als Null gewählt wird. Tabelle 6.2 zeigt nochmals die Spielstärken des statischen Modells, sowie zusätzlich die Mittelwerte der dynamischen Spielstärken - für gemeinsame und individuelle Startwerte - und die durch die Modelle erhaltenen Platzierungen der Mannschaften.

Die mittleren Spielstärkeunterschiede zwischen den Mannschaften sind beim dynamischen Modell wesentlich geringer als beim statischen Modell. Die Unterschiede zu den tatsächlichen Platzierungen sind nur marginal. Nur der 1.FC Köln und der SC Paderborn 07 verbessern sich im Vergleich zu ihren tatsächlichen Endplatzierungen. Bei den restlichen Mannschaften gibt es dadurch maximal Verschlechterungen um einen einzigen Tabellenplatz. Die letzten beiden Spalten beziehen sich auf das Modell mit individuellen

6 Dynamische Bradley-Terry Modellierung

Mannschaft	Punkte	stat.Stärke	QSE	Rang	gem.dyn	Rang	ind.dyn	Rang
FC Bayern München	79	0.602	0.331	1	0.253	1	0.831	1
VfL Wolfsburg	69	0.000	0.311	2	0	2	0	5
Borussia Mgladbach	66	-0.164	0.307	3	-0.179	3	-0.082	6
Bayer 04 Leverkusen	61	-0.380	0.304	4	-0.252	4	0.025	4
FC Augsburg	49	-1.037	0.299	6	-0.343	5	-0.291	7
FC Schalke 04	48	-1.087	0.299	7	-0.402	6	0.058	3
Bor. Dortmund	46	-0.939	0.299	5	-0.486	7	0.122	2
1899 Hoffenheim	44	-1.398	0.300	12	-0.496	8	-0.42	9
Eintracht Frankfurt	43	-1.304	0.300	10	-0.562	9	-0.676	12
SV Werder Bremen	43	-1.379	0.300	11	-0.612	11	-0.705	13
1.FSV Mainz 05	40	-1.199	0.299	8	-0.616	12	-0.367	8
1.FC Köln	40	-1.238	0.299	9	-0.579	10	-0.968	17
Hannover 96	37	-1.576	0.302	15	-0.664	13	-0.585	10
VfB Stuttgart	36	-1.616	0.302	16	-0.810	18	-0.759	14
Hertha BSC Berlin	35	-1.531	0.301	14	-0.684	14	-0.648	11
Hamburger SV	35	-1.692	0.303	17	-0.763	17	-0.966	16
SC Freiburg	34	-1.440	0.300	13	-0.732	15	-0.759	15
SC Paderborn 07	31	-1.888	0.306	18	-0.744	16	-0.983	18

Tabelle 6.2: Vergleich der statischen und dynamischen Spielstärken der 18 Bundesligavereine in der Saison 2014/2015

Startwerten und werden im nächsten Abschnitt genauer beschrieben.

Individuelle Startwerte

Nun werden dieselben Analysen wie in Abschnitt 6.4.1, jedoch mit individuellen Startwerten, welche in Tabelle 6.1 gegeben sind, durchgeführt. Als Startwerte der beiden Aufsteiger aus Köln und Paderborn wurden die Punktzahlen der beiden Absteiger der Vorsaison, Nürnberg und Braunschweig, übernommen. Für die Glättungsparameter erhält man nun Werte von $\lambda_1 = 0.019$ und $\lambda_2 = 0.043$. Die Werte sind nochmals wesentlich kleiner als bei den gemeinsamen Startwerten. Der Einfluss schon länger zurückliegender Spiele sowie der Startwerte ist deshalb nochmals wesentlich höher. Es kann stark angezweifelt werden, ob sich die ermittelten Werte überhaupt noch signifikant von der Annahme $\lambda_1 = \lambda_2 = 0$ unterscheiden. Bei Betrachtung der jeweiligen Profile-Likelihoods ergeben sich Werte von -456.77 für die Nullannahme und -456.26 für die optimierten λ -Werte. Für die Heim- bzw.

6 Dynamische Bradley-Terry Modellierung

Auswärtsspielstärkekoeffizienten ergeben sich Werte von $\beta_1 = 0.730$ und $\beta_2 = 0.700$ mit Standardfehlern von 0.098 und 0.138. Die Schwellenparameter lauten $\theta_1 = -\theta_4 = -1.707$ und $\theta_2 = -\theta_3 = -0.638$ mit Standardfehlern von 0.062 und 0.100. Für den rank probability score ergibt sich für das dynamische Modell ein Mittelwert von 0.672 und ein Median von 0.514. Dieser Wert ist um 5.8%, bzw. 6.6% höher als der Mittelwert bzw. Median des statischen Modells und um 4%, bzw. 3.6% niedriger als bei Annahme gemeinsamer Startwerte. Die Korrelation der Scores des dynamischen und statischen Modells liegt hier bei 0.867 und ist damit wesentlich höher als im vorherigen Abschnitt.

In Abbildung 6.2 sind die ermittelten Spielstärken beispielhaft für dieselben sechs Mannschaften wie in Abbildung 6.1 abgebildet.

Durch die sehr kleinen Werte für die beiden Glättungsparameter verändern sich die jewei-

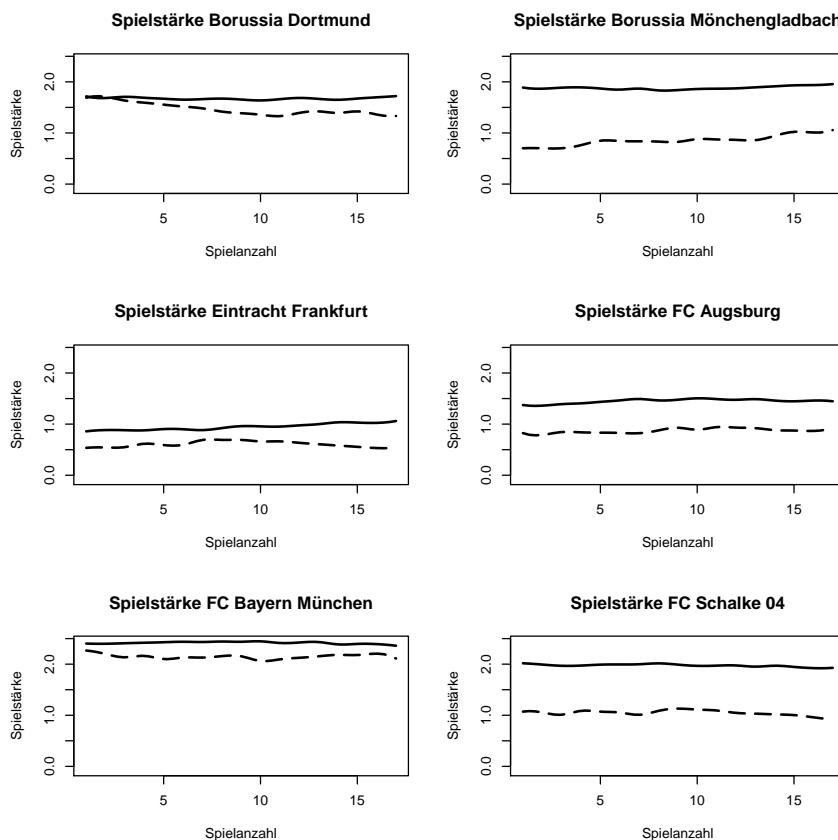


Abbildung 6.2: Spielstärkeentwicklung einzelner Mannschaften mit Unterscheidung zwischen Heim- (durchgezogene Linie) und Auswärtsspielstärke (gestrichelte Linie) und individuellen Startwerten.

ligen Heim- bzw. Auswärtsspielstärken der einzelnen Mannschaften kaum. Dadurch ist

6 Dynamische Bradley-Terry Modellierung

es hier nicht möglich, Aussagen über die aktuelle Formentwicklung einer Mannschaft zu treffen. Borussia Dortmund zum Beispiel erzielte in der Rückrunde, wie bereits erwähnt, wesentlich mehr Punkte als in der Hinrunde. Dies lässt sich aber aus den ermittelten Spielstärken nicht erkennen. Die Spielstärke einer Mannschaft ergibt sich in diesem Modell fast ausschließlich durch die Startwerte. Daraus lässt sich folgern, dass sich die Spielstärken der Mannschaften über die Saison hinaus nicht stark verändern. Die Verläufe der restlichen Mannschaften sind im Anhang angefügt. Werden die gemittelten Spielstärken der Mannschaften über die gesamte Saison, welche in Tabelle 6.2 gegeben sind, betrachtet, so ist zwar wiederum der FC Bayern München die stärkste Mannschaft, jedoch gibt es auf den restlichen Positionen größere Unterschiede. Borussia Dortmund wird nun als zweitstärkste Mannschaft klassifiziert, wohingegen der VfL Wolfsburg nur auf Platz fünf geführt wird. Bei einem Vergleich mit dem Ausgang der Vorsaison 2013/14 stellt man fest, dass die ermittelten Platzierungen des Modells mit diesen fast vollständig übereinstimmen. Besonders deutlich wird diese Eigenschaft des Modells bei der Betrachtung des 1. FC Köln. Für diesen Verein wurden die Werte der Vorsaison des 1. FC Nürnberg übernommen, da Köln in der Vorsaison nicht in der Bundesliga vertreten und Nürnberg einer von zwei Absteigern war. Obwohl die Kölner die Saison im sicheren Mittelfeld abgeschlossen haben, prognostiziert das Modell sie auf Platz 17. Dies entspricht der Platzierung des 1. FC Nürnberg in der Vorsaison. Warum wird dieses Modell trotz dieser starken Einschränkung trotzdem betrachtet? Die Grundannahme, dass immer dieselben Mannschaften in ähnlichen Tabellenbereichen landen, ist nicht so unrealistisch, wie sie im ersten Moment klingt. Vereine werden sich kaum innerhalb eines Jahres auf Dauer im oberen Tabellenbereich festsetzen können, wenn sie die Jahre vorher nur im Tabellenmittelfeld platziert waren. Ebenso wird der FC Bayern München mit seinen finanziellen Mitteln wohl kaum mehrere Jahre in Folge nur einen Platz im Tabellenmittelfeld belegen. Ein weiterer Punkt, der diesen Modellansatz plausibel wirken lässt, sind die quasi-Standardabweichungen, welche sich im statischen Modell ergeben. Bereits bei diesem Modell hat sich gezeigt, dass sich allein durch die Spielstärken der meisten Mannschaften keine genauen Aussagen über ihre Endplatzierungen treffen lassen. Als Letztes sei noch angeführt, dass der mittlere rank probability score bei diesem Modellansatz niedriger ist als bei der Wahl gemeinsamer Startwerte. Deshalb wird das Modell auch in die später noch folgenden Analysen miteinbezogen.

6.4.2 Gewichtete Likelihood

Um die einzelnen Parameterschätzer für das dynamische Modell zu erhalten, muss zunächst der Parameter σ geschätzt werden. Für die Schätzung dieses Parameters wird eine prädiktive Simulation der Bundesligadaten der Saison 2013/14 mit verschiedenen σ -Parametern durchgeführt. Als Optimierungskriterium wird dann der rank probability score betrachtet und mit diesem der optimale Parameter ausgewählt. Als Auswahlkriterium innerhalb dieses Scores fiel die Entscheidung auf den Median. In späteren Betrachtungen, genauer in Kapitel 7, wird diese Wahl näher begründet.

Die prädiktive Simulation wurde für σ -Werte zwischen 15 und 100 durchgeführt. Abbildung 6.3 zeigt die Mediane der einzelnen Simulationen.

Die Mediane bewegen sich etwa in einem Bereich zwischen 0.6 und 0.85. Für steigende

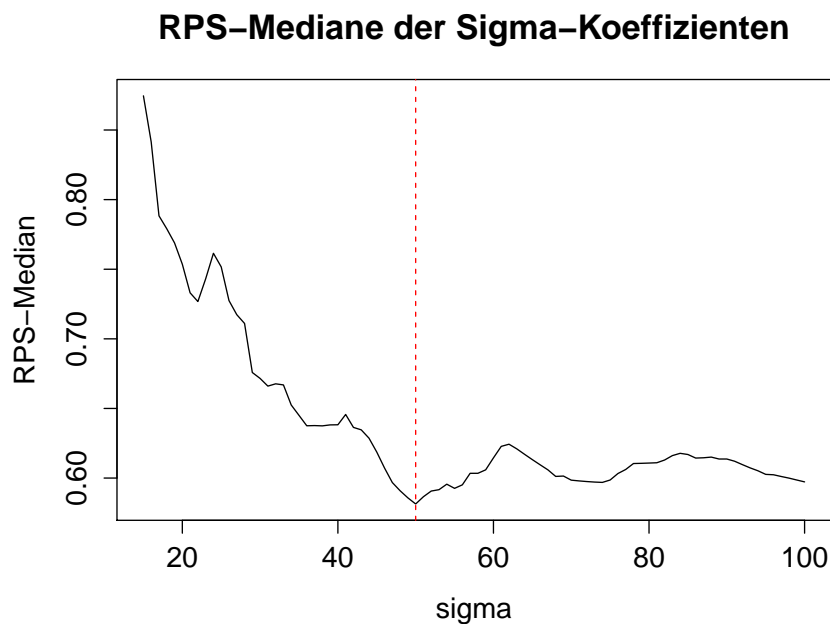


Abbildung 6.3: RPS-Mediane für verschiedene Sigma-Werte anhand der Daten aus der Saison 2013/14

σ -Werte sinkt der Median zu Beginn relativ rasch. Bei $\sigma = 50$ wird das Minimum im betrachteten Bereich mit einem Wert von 0.581 erreicht. Für höhere σ -Werte bewegen sich die Mediane in einem relativ ähnlichen Bereich.

Nach der Bestimmung dieses Koeffizienten ergibt sich aber bei diesem Modellansatz noch ein weiteres Problem. Da für die dynamische Modellierung nur die bereits vergangenen

6 Dynamische Bradley-Terry Modellierung

und der aktuelle Spieltag als Datengrundlage betrachtet werden, ist die Schätzung, vor allem zu Saisonbeginn, sehr unsicher. Ebenso ist die Schätzung zu Beginn der Rückrunde instabil, da durch die Winterpause, welche ca. 40 Tage dauert, der aktuelle Spieltag eine sehr große Gewichtung im Vergleich zu den anderen Partien erfährt. Das zuletzt genannte Problem lässt sich dadurch etwas entschärfen, dass der Zeitraum der Winterpause künstlich verkürzt wird und zwar zwischen dem 17. und 18. Spieltag um 21 Tage, so dass zwischen diesen beiden eine Zeitspanne von 19 Tagen liegt. Dies ist damit immer noch der größte Zeitraum zwischen zwei Spielen. Ein weiteres Problem besteht darin, dass sich die Modellierung erst ab dem Zeitpunkt durchführen lässt, ab welchem jede Mannschaft nicht mehr jedes Spiel gewonnen oder verloren hat. Solange dies der Fall ist, würde das Modell die Spielstärke dieser Mannschaften auf $+\infty$ oder $-\infty$ setzen. In der Saison 2014/15 ist dies bereits ab dem 3. Spieltag der Fall, so dass ab diesem Spieltag dynamische Spielstärken berechnet werden können. Abbildung 6.4 zeigt für einige Vereine die geschätzten Spielstärkekoeffizienten.

Im Gegensatz zum Ansatz von Cattelan et al. (2013) besitzt nun jede Mannschaft nur einen einzigen Spielstärkekoeffizientenpfad. Es ist zu erkennen, dass erst ab dem dritten Spieltag Koeffizientenschätzer existieren. Die teilweise großen Schwankungen bis zum fünften Spieltag sind durch die große Unsicherheit aufgrund der geringen Datenmenge zu erklären. Bei Borussia Dortmund fällt auf, dass ihre Spielstärke zu Saisonbeginn bis etwa zum 10. Spieltag sinkt, jedoch danach im Laufe der Zeit wieder zunimmt. Borussia Mönchengladbach ist vor allem zu Saisonbeginn und noch auffälliger zum Saisonende hin sehr stark. Diese Auffälligkeiten stimmen auch mit der tatsächlichen Entwicklung der Mannschaften überein, wie sie in Abschnitt 6.4.1 bereits beschrieben wurden. Die Koeffizientenpfade für die restlichen Mannschaften sind ergänzend im Anhang beigefügt. Durch die Modellierung jedes einzelnen Spieltags darf jedoch nicht außer Acht gelassen werden, dass für jeden Spieltag individuelle Heimstärkeparameter α und Schwellenparameter θ existieren. In Tabelle 6.3 sind nachfolgend die einzelnen Heimstärkeparameter sowie die Spielstärken und quasi-Standardabweichungen für Borussia Mönchengladbach abgebildet.

Die Heimstärkekoeffizienten schwanken an den einzelnen Spieltagen zwischen -0.136 und 0.793. Im Vergleich zu den Schwankungen der Spielstärken der einzelnen Mannschaften ist diese Spanne jedoch relativ gering. Die Spielstärke von Borussia Mönchengladbach schwankt zum Beispiel zwischen -1.994 am 3. Spieltag und 2.580 am 33. Spieltag. Am wichtigsten sind in der Tabelle jedoch die angegebenen quasi-Standardabweichungen. Vor allem am 3. Spieltag ist diese mit einem Wert von 3.126 sehr hoch. Je weiter die Saison

6 Dynamische Bradley-Terry Modellierung

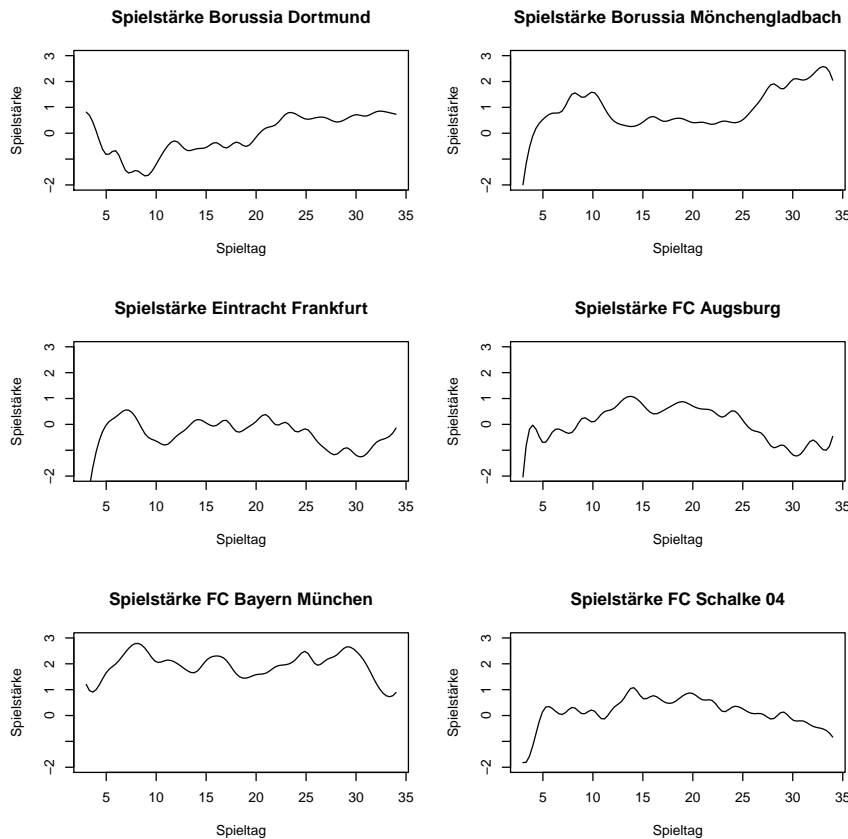


Abbildung 6.4: Geschätzte Spielstärkekoeffizienten einzelner Mannschaften bei gewichteter Likelihood vom 3. bis 34. Spieltag

voranschreitet, desto sicherer werden auch die einzelnen Schätzer. Am 24. Spieltag nimmt die quasi-Standardabweichung der Spielstärke von Borussia Mönchengladbach ihren geringsten Wert mit 0.846 an. Jedoch lassen sich auch hier wie bereits beim statischen Modell die Spielstärken vieler Vereine nicht signifikant voneinander unterscheiden, weshalb die Spielstärken alleine den Tabellenplatz nicht exakt widerspiegeln können.

Aus diesem Grund stellt sich nun wieder die Frage nach der Güte des Modells. Daher sollen die Ergebnisse wie bereits beim Ansatz von Cattelan et al. (2013) mit dem statischen Modell verglichen werden. Während bei Cattelan et al. (2013) nur sechs verschiedene Parameter geschätzt wurden und somit das Ziel des Modells nur darin bestand, möglichst nah an die Ergebnisse des statischen Modells heranzukommen, werden bei diesem Modell für jeden Spieltag dieselben 20 Koeffizienten geschätzt wie beim statischen Modell. Als Gütekriterium dient der rank probability score. Beim statischen Modell beträgt der Mittelwert dieses Scores 0.635 und der Median nimmt einen Wert von 0.482 an. Beim gewichteten

6 Dynamische Bradley-Terry Modellierung

Spieltag	3	4	5	6	7	8	9	10	11	12	13
α	-0.136	0.389	0.483	0.437	0.530	0.609	0.715	0.614	0.661	0.544	0.632
γ	-1.994	-0.088	0.535	0.779	0.891	1.565	1.380	1.593	1.083	0.504	0.318
QSE	3.126	1.862	1.471	1.322	1.200	1.264	1.165	1.147	1.067	1.113	1.038
Spieltag	14	15	16	17	18	19	20	21	22	23	24
α	0.619	0.469	0.499	0.438	0.485	0.373	0.272	0.291	0.268	0.379	0.450
γ	0.260	0.428	0.648	0.465	0.545	0.556	0.409	0.424	0.344	0.467	0.410
QSE	1.001	0.979	0.960	0.918	0.972	0.921	0.873	0.855	0.850	0.862	0.846
Spieltag	25	26	27	28	29	30	31	32	33	34	
α	0.470	0.501	0.553	0.675	0.674	0.787	0.686	0.738	0.764	0.793	
γ	0.519	0.930	1.420	1.914	1.709	2.080	2.054	2.266	2.580	2.051	
QSE	0.862	0.943	1.107	1.205	1.165	1.191	1.115	1.114	1.130	1.105	

Tabelle 6.3: Heimstärkekoeffizienten der einzelnen Spieltage sowie die Spielstärken und die dazugehörigen quasi-Standardabweichungen von Borussia Mönchengladbach

dynamischen Modell mit einem σ -Wert von 50 erhält man für den Mittelwert des Scores 0.508 und für den Median 0.358. Die Werte sind also um 20% bzw. 25.7% niedriger als beim statischen Modell. Es ist daher davon auszugehen, dass das gewichtete dynamische Modell bessere Ergebnisse als das statische Bradley-Terry-Modell liefert. Die Korrelation zwischen den beiden Modellen liegt bei 0.729. Würde man ein dynamisches Modell mit identischen Gewichten, sprich einer Gleichverteilung als Kernfunktion verwenden, so würde der rank probability score Werte von 0.586 für den Mittelwert und 0.416 für den Median liefern. Diese Werte sind 7.7% bzw. 13.7% niedriger als die des statischen Modells, aber auch 15.4% bzw. 16.2% höher als die des dynamischen Modells mit Gauß-gewichteter Likelihood. Eine solche Gewichtung über einen Gaußkern scheint daher gerechtfertigt. Zum besseren Vergleich der Scores zeigt Abbildung 6.5 die Boxplots der einzelnen Scores für das statische Modell, die beiden exponential-gewichteten moving-average Prozesse mit gemeinsamen und individuellen Startwerten sowie die Boxplots für die beiden dynamischen Modelle mit unterschiedlicher und ohne unterschiedliche Gewichtung.

Vorneweg sei gesagt, dass für die ersten drei Boxplots Scores für alle Spiele - 306 Werte - einfließen und für die beiden dynamischen Modelle nur 288 Werte, da erst ab dem 3. Spieltag Ergebnisse modelliert werden können. Es ist gut zu erkennen, dass die beiden EWMA-Prozesse im Schnitt höhere Werte liefern als das statische Modell. Ebenso sieht man, dass bei individuellen Startwerten leicht bessere Ergebnisse erzielt werden als bei gemeinsamen Startwerten. Die Ausreißer sind bei allen drei Modellen in etwa gleich

6 Dynamische Bradley-Terry Modellierung

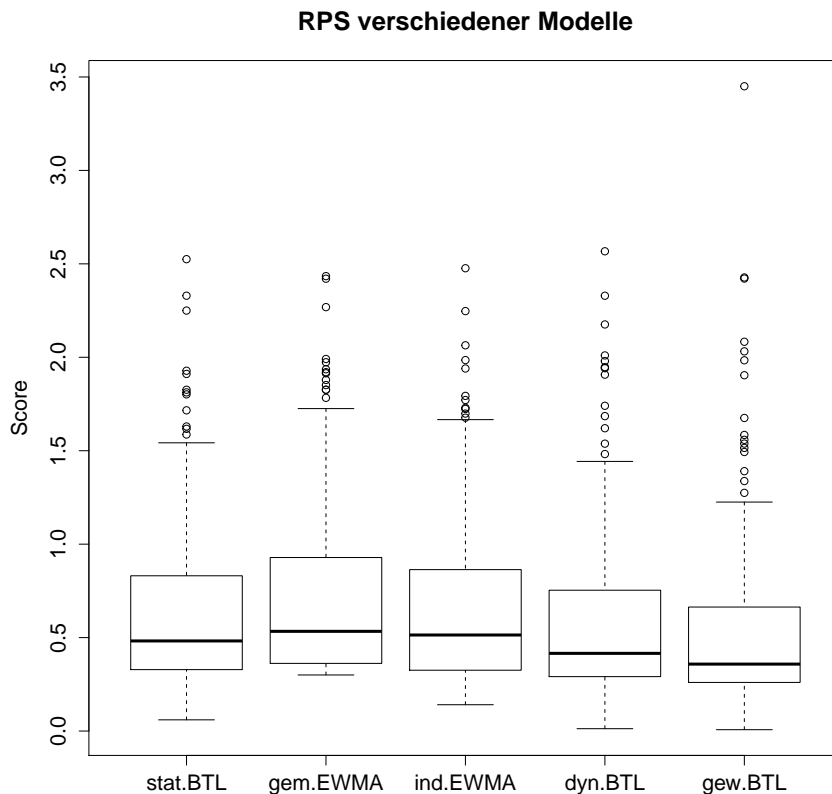


Abbildung 6.5: Vergleich der rank probability scores für das statische Modell, die beiden exponential-gewichteten moving-average-Prozesse, sowie das gewichtete dynamische Modell

und nehmen maximal einen Scorewert von etwa 2.5 an. Das dynamische Modell ohne Gewichtung sieht dem statischen Modell relativ ähnlich, vor allem bei den Ausreißern. Alle gekennzeichneten Quantile liegen etwas niedriger als beim statischen Modell. Das gewichtete dynamische Modell liefert im Schnitt sichtbar geringere Werte als die anderen Modelle. Nicht nur der Median, sondern auch das 75%-Quantil sind deutlich niedriger bei diesem Modell. Die Ausreißer sind ähnlich, bis auf eine Ausnahme: Der maximale Score erreicht einen Wert von 3.45 und hebt sich somit deutlich von den anderen ab. Dieser Wert ergibt sich beim Spiel zwischen Borussia Mönchengladbach und dem FC Augsburg. Der Borussia wird in diesem Spiel eine über 93%-ige Siegeschance modelliert, jedoch gewann der FC Augsburg dieses Spiel mit 3:1, was sogar einen Sieg mit zwei Toren Unterschied bedeutet. Diesem Ereignis wird nur eine Wahrscheinlichkeit von 0.5% zugerechnet.

7 Prädiktiver Modellvergleich

In den vergangenen Kapiteln wurden Modelle vorgestellt, mit denen sich die Spielstärke der Mannschaften mit Hilfe der erzielten Ergebnisse und zusätzlicher Kovariablen modellieren lassen. Dabei wurde immer angenommen, dass die Ergebnisse bekannt sind, und aufgrund dieser wurde die Spielstärke modelliert. Diese Modellergebnisse wurden danach angewandt und ausgewertet. Ein wichtiger und gleichzeitig interessanter Punkt wurde bei all diesen Vorgehensweisen jedoch bisher noch nicht beleuchtet: die prädiktive Aussagekraft der Modelle. Die Spielstärke im Nachhinein zu modellieren liefert zwar bessere Ergebnisse als eine prädiktive Analyse, jedoch hat dies für einen Anwender relativ geringen Nutzen. Viel wichtiger ist es einem solchen, anhand vergangener Daten zukünftige Ergebnisse prognostizieren zu können, um sich so einen Vorteil gegenüber Anderen zu verschaffen. Dieses Kapitel widmet sich daher der prädiktiven Aussagekraft der vorgestellten Modelle. Das bedeutet, es werden nur die Daten bis zu einem bestimmten Spieltag betrachtet. Mit diesen werden dann die einzelnen Koeffizienten geschätzt und auf den folgenden Spieltag angewandt. Dieses Vorgehen wird für jeden Spieltag wiederholt. Da die Modelle zu Beginn bereits einige Daten benötigen, um „stabile“ Ergebnisse zu erhalten, wird dieses Vorgehen am 7. Spieltag der Saison 2014/15 gestartet. Es werden also zu Beginn nur die Ergebnisse der ersten sechs Spieltage herangezogen und mit diesen die Koeffizienten geschätzt. Mit diesen Schätzern werden anschließend die Siegeswahrscheinlichkeiten der einzelnen Mannschaften für den 7. Spieltag berechnet. Anschließend werden die Daten der ersten sieben Spieltage betrachtet und wiederum die Koeffizientenschätzer modelliert. Die somit erhaltenen Ergebnisse werden im Folgenden unter drei verschiedenen Gesichtspunkten betrachtet. Zuerst sollen die einzelnen rank probability scores der Modelle miteinander verglichen werden. Anschließend wird überprüft, wie viele Spiele sich, mit Hilfe der Modelle, richtig vorhersagen lassen. Hierfür wird zum einen eine Simulationsstudie durchgeführt, welche die Ergebnisse anhand der durch das Modell erhaltenen Wahrscheinlichkeiten simuliert, und zum anderen werden die Spiele anhand der wahrscheinlichsten Ergebnisse getippt. Zuletzt soll noch mit verschiedenen Ansätzen überprüft werden, ob sich mit Hilfe dieser Modelle in der Saison 2014/15 Gewinne durch gezielte Wetten erzielen lassen.

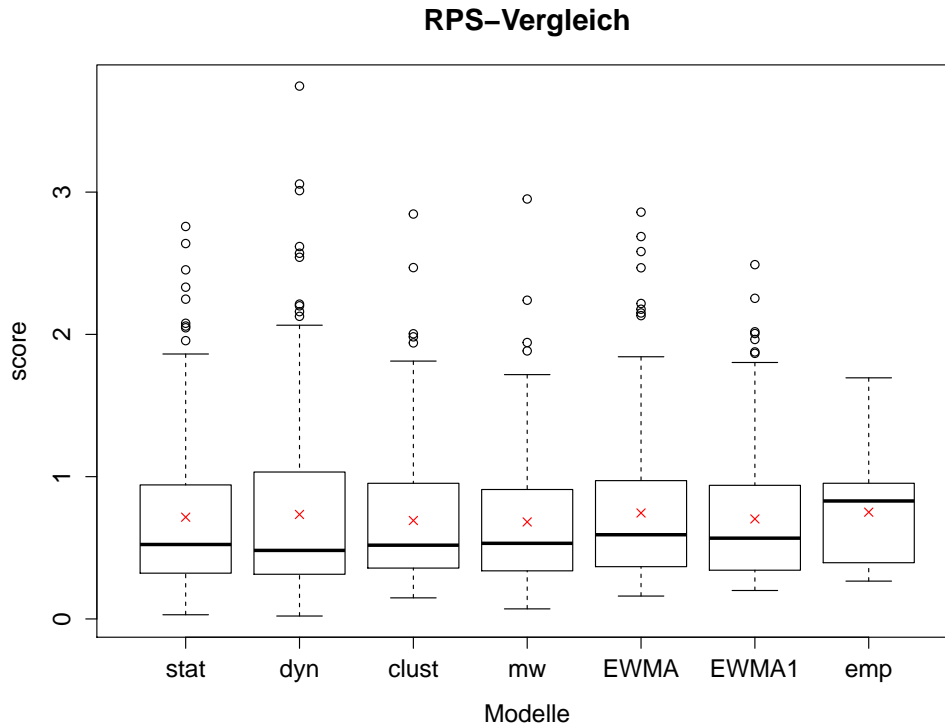


Abbildung 7.1: Vergleich der verschiedenen RPS für sieben verschiedene Modelle: Einfaches BTL (stat), gewichtetes BTL (dyn), Modell mit Penalisierung und globalem Heimeffekt (clust), Modell mit Penalisierung, globalem Heimeffekt und Berücksichtigung der Startaufstellung (mw), exponential-gewichteter moving-average Prozess mit gemeinsamen Startwerten (EWMA), exponential-gewichteter moving-average Prozess mit individuellen Startwerten (EWMA1) und ein Modell mit gemittelten Ergebniswahrscheinlichkeiten (emp)

7.1 Rank probability scores

Nachfolgende Abbildung 7.1 zeigt die ermittelten rank probability scores für insgesamt sieben verschiedene Modellansätze.

Bei den ersten sechs Modellen handelt es sich um einige der im bisherigen Verlauf dieser Arbeit vorgestellten Modelle. Da bei der prädiktiven Analyse, wie bereits beschrieben, die Modelle für jeden Spieltag neu modelliert werden, können nun alle als dynamische Modelle angesehen werden. Das siebte Modell betrachtet einen sehr allgemeinen Ansatz. Bei diesem wird die Anzahl der Heimsiege, Unentschieden und Auswärtssiege jeweils durch die Anzahl aller bisher stattgefundenen Spiele geteilt. Dies ergibt dann die jeweiligen Wahrscheinlichkeiten. Folglich sind diese für jedes Heimteam bzw. jedes Auswärtsteam

7 Prädiktiver Modellvergleich

gleich. Bei Betrachtung des rps für dieses Modell fällt auf, dass keine Ausreißer existieren. Der Mittelwert ist ähnlich dem der anderen Modelle. Das 25%-Quantil und noch deutlicher der Median sind klar höher als bei den anderen Modellen. Anhand dieses Boxplots lassen sich gut die Stärken und Schwächen des rps erkennen. Aufgrund der Unabhängigkeit der einzelnen Wahrscheinlichkeiten von den Mannschaften ergeben sich bei einem allgemeinen Ansatz keine eindeutigen Siegeswahrscheinlichkeiten. Deshalb ist es folglich auch nicht möglich, eine völlig falsche Einschätzung abzugeben, und daher existieren keine großen Ausreißer. Das Problem dabei ist nun jedoch, dass dadurch auch keine besonders kleinen Werte für den rps erzielt werden können. Deshalb muss sich ein Anwender je nach Ziel seiner Auswertung entscheiden, wie er sein Modell wählen will. Ein Modell, welches klare Aussagen darüber trifft, welche Mannschaft ein Spiel gewinnt wird viele kleine Werte, aber auch einige große Ausreißer produzieren. Ein Modell, welches potenziell den Teams eher ausgeglichene Siegeswahrscheinlichkeiten zuweist, wird alle seine Scores etwa im selben Bereich erhalten. Wie Abbildung 7.1 zeigt, führt dies dazu, dass die Mittelwerte der Modelle kaum unterscheidbar sind. Deshalb wird in dieser Arbeit hauptsächlich der Median betrachtet. Dieser ist beim Modell mit gemittelten Siegeswahrscheinlichkeiten, wie erwähnt, deutlich höher, weshalb dieses Modell als schwächstes bezeichnet werden kann. Bei den anderen Modellen ist für eine Unterscheidung eine genauere Betrachtung der Werte nötig. In Tabelle 7.1 sind daher das 25%-Quantil, der Mittelwert, der Median und das 75%-Quantil für die verschiedenen Modelle aufgeführt.

Die Tabelle ist dabei in vier verschiedene Gruppen aufgeteilt. Die erste Gruppe betrachtet

	stat	dyn	clust	mw	EWMA	EWMA1	emp
0.25-Quantil	0.323	0.314	0.358	0.338	0.368	0.344	0.395
Median	0.523	0.482	0.518	0.531	0.591	0.568	0.829
Mittelwert	0.715	0.734	0.692	0.682	0.745	0.703	0.750
0.75-Quantil	0.940	1.029	0.953	0.908	0.971	0.934	0.954

Tabelle 7.1: Kennwerte der sieben zu vergleichenden Modelle: Einfaches BTL (stat), gewichtetes BTL (dyn), Modell mit Penalisierung und globalem Heimeffekt (clust), Modell mit Penalisierung, globalem Heimeffekt und Berücksichtigung der Startaufstellung (mw), exponential-gewichteter moving-average Prozess mit gemeinsamen Startwerten (EWMA), exponential-gewichteter moving-average Prozess mit individuellen Startwerten (EWMA1) und ein Modell mit gemittelten Ergebniswahrscheinlichkeiten (emp)

die dynamischen Modelle basierend auf dem klassischen Bradley-Terry Modell, mit und ohne Gewichtung der Likelihood. Die zweite Gruppe umfasst die geclusterten Modelle,

7 Prädiktiver Modellvergleich

welche sich durch das BIC-Kriterium ergeben, zum einen ohne Kovariablen und zum anderen mit den durchschnittlichen Marktwerten der Startaufstellungen der Mannschaften. Die dritte Gruppe enthält die Modelle, welche über einen exponential-gewichteten moving-average Prozess, wie in Cattelan et al. (2013) beschrieben, hergeleitet werden können, einmal mit gemeinsamen und einmal mit individuellen Startwerten für den Prozess. Die vierte und letzte Gruppe besteht schließlich aus dem allgemeinen Ansatz, wie oben beschrieben.

Das 25%-Quantil ist beim gewichteten dynamischen Modell am geringsten mit einem Wert von 0.314. Bei den geclusterten Modellen sind die Werte des Modells ohne den Einfluss von Kovariablen höher als beim Modell mit Berücksichtigung der Marktwerte der Startaufstellungen. Beide Werte sind höher als bei den nicht-geclusterten Modellen. Die EWMA-Prozesse weisen ähnlich hohe Werte auf, wobei der Prozess mit individuellen Startwerten besser abschneidet als der Prozess mit gemeinsamen Startwerten. Warum die von Tutz und Schauburger (2015) verwendeten Modelle besser abschneiden als die EWMA-Prozesse, wurde bereits in Kapitel 6 erläutert, es sei aber nochmals erwähnt. Der Grund hierfür liegt an der wesentlich geringeren Parameterzahl der EWMA-Prozess-Modelle. Der größte Wert wird beim allgemeinen Ansatz mit einem Quantilswert von 0.395 erreicht. Beim Median ist die Reihenfolge der besten Werte gleich den 25%-Quantilswerten. Der Median des empirischen Modells liegt eindeutig über dem der anderen Modelle. Beim 75%-Quantil gestaltet sich die Reihenfolge anders. Hier schneidet das gewichtete dynamische Modell am schwächsten ab und der allgemeine Ansatz fällt nicht mehr von den anderen ab. Dies liegt daran, dass es durch die klaren Aussagen des gewichteten dynamischen Modells prinzipiell auch zu mehr höheren Werten kommen kann. Beim allgemeinen Ansatz liegen die Werte, wie bereits erwähnt, alle relativ ähnlich. Der Median und das 75%-Quantil unterscheiden sich bei diesem Ansatz daher nur um 0.125 Scorepunkte. Die Werte für den Mittelwert sind alle sehr ähnlich. Die Reihenfolge der besten Werte spielt aufgrund der geringen Aussagekraft dieses Wertes keine Rolle. Da die Differenzen der einzelnen Quantile, mit Ausnahme derer des allgemeinen Ansatzes, aber ebenfalls relativ ähnlich sind, kann angezweifelt werden, ob aufgrund der ermittelten Scores Aussagen getroffen werden können, welche Modelle am geeignetsten für die Vorhersage zukünftiger Ereignisse sind.

7.2 Ergebnisvorhersage

Nachdem die prädiktiven rank probability scores betrachtet wurden, folgt eine Anwendung der Modelle, welche für die Praxis relevant ist. Es soll getestet werden, wie viele Spiele einer Saison durch die vorgestellten Modelle richtig getippt werden. Dafür werden zwei unterschiedliche Ansätze betrachtet.

7.2.1 Simulationsstudie

Als erstes wird getestet, wie viele Spiele mit Hilfe einer Simulationsstudie richtig getippt werden können. Hierfür interessiert im Folgenden nur, ob ein Spiel mit einem Sieg, einer Niederlage oder Unentschieden endete. Die Unterscheidung, mit wie vielen Toren Unterschied das Spiel endete, fällt weg. Zunächst werden dafür die Wahrscheinlichkeiten der Spielausgänge für jedes einzelne Spiel und jedes Modell berechnet. Anschließend werden alle 252 Partien ab dem 7. Spieltag zehntausendmal mit den ermittelten Wahrscheinlichkeiten simuliert. Die Ergebnisse dieser Simulation finden sich in Tabelle 7.2 wieder.

Es sind jeweils die durchschnittlichen Anzahlen richtig getippter Spiele in Prozent ange-

	stat	dyn	clust	mw	EWMA	EWMA1	emp
Minimum	0.313	0.298	0.286	0.290	0.270	0.290	0.254
0.05-Quantil	0.365	0.381	0.349	0.357	0.333	0.349	0.313
Durchschnitt	0.410	0.423	0.396	0.405	0.380	0.395	0.360
0.95-Quantil	0.456	0.468	0.444	0.452	0.429	0.444	0.409
Maximum	0.512	0.540	0.496	0.520	0.488	0.508	0.476

Tabelle 7.2: Ergebnisse der Simulationsstudie bei 10000 Simulationen pro Modell: Einfaches BTL (stat), gewichtetes BTL (dyn), Modell mit Penalisierung und globalem Heimeffekt (clust), Modell mit Penalisierung, globalem Heimeffekt und Berücksichtigung der Startaufstellung (mw), exponential-gewichteter moving-average Prozess mit gemeinsamen Startwerten (EWMA), exponential-gewichteter moving-average Prozess mit individuellen Startwerten (EWMA1) und ein Modell mit gemittelten Ergebniswahrscheinlichkeiten (emp)

geben. Bei allgemeiner Wahrscheinlichkeitsverteilung werden, wie erwartet, im Schnitt am wenigsten Spiele richtig getippt. Im Fall der schlechtesten Simulation werden nur 25.4% aller Spiele richtig getippt, das entspricht nur knapp jedem vierten Spiel. Beim einfachen Bradley-Terry Modell werden durchschnittlich 41% der Spiele richtig vorhergesagt, was

7 Prädiktiver Modellvergleich

einer Anzahl von über 103 richtig getippten Spielen entspricht. Bei der besten Simulation werden sogar mehr als die Hälfte der Spiele richtig vorhergesagt. Beim gewichteten dynamischen Modell liegen die Zahlen der richtig getippten Spiele etwas höher als beim Modell ohne Gewichtung. Es weist im Schnitt 42.3% der Spiele das richtige Ergebnis zu. Es werden also im Schnitt fast 107 von 252 Spielen richtig getippt. Bei der besten Simulation dieses Modells wurden 54% der Spiele richtig getippt. Dies entspricht einer Zahl von gut 136 Spielen und liefert das beste Ergebnis aller durchgeführten Simulationen. Alle Modelle bis auf den allgemeinen Ansatz erhalten für das 5%-Quantil mindestens Werte in Höhe von 0.33. Daraus folgt, die Modelle liefern normalerweise bessere Ergebnisse als die Annahme, die Spielausgänge seien zufällig und jedem Ausgang wird die gleiche Wahrscheinlichkeit zugerechnet. Daher erscheinen die Modelle auf den ersten Blick als gute Möglichkeit, Spielausgänge aufgrund ihrer prognostizierten Ergebnisse zu tippen. Jedoch wurde bereits in Kapitel 2 erwähnt, dass in der Saison 2014/15 47.4% aller Heimspiele gewonnen wurden. In den vorherigen Saisons war die Zahl der Heimsiege ähnlich hoch. In den letzten acht Spielzeiten wurden immer über 40% aller Heimspiele gewonnen und niemals mehr als ein Drittel aller Auswärtsspiele. In der Saison 2008/2009 lag der Anteil der Heimsiege sogar bei 48% [vgl. fussball.wettpoint.com (2016)].

7.2.2 Tipp der höchsten Wahrscheinlichkeit

Wie gerade beschrieben, liegen bei einer Simulationsstudie große Schwankungen bei der Anzahl der richtig vorhergesagten Ergebnisse vor und die Anzahl der richtig getippten Spiele ist ebenso nicht zufriedenstellend. In der Realität wird normalerweise auch nur ein einziges Mal auf den Ausgang eines Spieles gesetzt und nicht zehntausendmal. Daher soll nun noch ein Ansatz betrachtet werden, bei welchem tatsächlich jedes Spiel nur ein einziges Mal getippt wird. Nachfolgend wird immer auf das Ergebnis gesetzt, welches vom Modell als am wahrscheinlichsten vorhergesagt wird. Dabei ergeben sich die Ergebnisse aus Tabelle 7.3.

Bei diesem Ansatz liegen die Anzahlen der richtig getippten Spiele bei allen Modellen höher als das 95%-Quantil der Ergebnisse der entsprechenden Modelle aus der Simulationsstudie. Das geclusterte Modell mit globalem Heimeffekt, der exponential-gewichtete moving-average Prozess mit individuellen Startwerten und das geclusterte Modell mit Berücksichtigung der Startaufstellungen sagen mehr als die Hälfte aller Ergebnisse richtig vorher. Auch die anderen Modelle liefern ähnlich gute Ergebnisse. Daher erscheint die in

7 Prädiktiver Modellvergleich

Modell	richtige Tipps	Anteil richtiger Tipps
stat	124	0.492
dyn	124	0.492
clust	128	0.508
mw	130	0.516
EWMA	119	0.472
EWMA1	127	0.504
emp	123	0.488

Tabelle 7.3: Anzahl richtiger Ergebnisse bei 252 Spielen und Tipp auf das wahrscheinlichste durch das Modell ermittelte Ergebnis: Einfaches BTL (stat), gewichtetes BTL (dyn), Modell mit Penalisierung und globalem Heimeffekt (clust), Modell mit Penalisierung, globalem Heimeffekt und Berücksichtigung der Startaufstellung (mw), exponential-gewichteter moving-average Prozess mit gemeinsamen Startwerten (EWMA), exponential-gewichteter moving-average Prozess mit individuellen Startwerten (EWMA1) und ein Modell mit gemittelten Ergebniswahrscheinlichkeiten (emp)

diesem Abschnitt vorgestellte Vorgehensweise die effektivere zu sein. Es sei noch angemerkt, dass beim Ansatz mit allgemeinen Ergebniswahrscheinlichkeiten jedes Spiel auf einen Heimsieg getippt wird.

7.3 Wettquoten und Gewinnmöglichkeiten

Beim Tipp auf die höchste Gewinnwahrscheinlichkeit fällt auf, dass die Ergebnisse für alle Modelle ähnlich ausfallen und in etwa genauso viele Spiele wie bei durchgehendem Tipp auf die Heimmannschaft richtig vorhergesagt werden. Ist nur die Anzahl der richtigen Tipps von Bedeutung, ist kein Modell zu favorisieren. Wie sieht es aber aus, wenn nicht die Anzahl der richtigen Tipps das wichtigste Merkmal darstellt, sondern ein möglichst hoher Gewinn erzielt werden soll? Wettanbieter bieten beim Tipp der einzelnen Spiele unterschiedliche Quoten an, um Anreize zu setzen riskante Tipps abzugeben. Daher kann es aus profitorientierten Gründen oft sinnvoll sein, sein Geld nicht auf das favorisierte Team zu setzen. Dies sei an den Quoten für den FC Bayern München verdeutlicht. Der Verein dominierte in den letzten paar Jahren die Bundesliga und ist seit Jahrzehnten immer im vorderen Tabellenbereich zu finden. Soll ein sicherer Tipp abgegeben werden, so sollte jedes Spiel auf einen Sieg des FC Bayern getippt werden. Jedoch ließe sich damit nicht viel Geld verdienen, da die Quoten sehr niedrig sind. Im Schnitt waren die Quoten für

7 Prädiktiver Modellvergleich

einen Sieg der Münchner in der vergangenen Saison bei 1.33. Bei einem Einsatz von einem Euro wäre also ein Gewinn von 33 Cent erzielt worden. Folglich würden bereits Verluste eintreten, wenn Bayern schon jedes vierte Spiel nicht gewinnt. Riskante Tipps können daher theoretisch eine bessere Rendite einbringen. Beim Spiel zwischen Bayern München und dem Hamburger SV am 21. Spieltag lag die Quote für einen HSV-Sieg bei 28.5, was der höchsten Quote der ganzen Saison entsprach. Bei einem Euro Einsatz wäre ein Gewinn von 27.50 Euro bei einem Sieg des Hamburger SV erzielt worden.

Auch bei den Wettquoten lässt sich der Heimvorteil der Mannschaften erkennen. Die Quoten für einen Heimsieg lagen in der Saison 2014/2015 im Schnitt bei 2.87, die Quoten für ein Unentschieden bei 4.09 und die Quoten für einen Auswärtssieg bei 4.54. Bei ständigen Tipps auf einen Heimsieg lässt sich also wesentlich weniger Geld verdienen als bei Tipps auf Unentschieden oder Auswärtssiege. Aus diesen Gründen werden nun verschiedene Ansätze zur Platzierung gezielter Wetten auf die Daten der Saison 2014/15 angewendet und so überprüft, ob sich mit einem der Modelle Gewinne hätten erzielen lassen.

7.3.1 Simulationsstudie

Zunächst sollen die Ergebnisse aus Unterabschnitt 7.2.1 mit den Wettquoten verknüpft werden. Das Vorgehen ist dabei folgendermaßen: Es wird auf jedes Spiel ein Euro gesetzt. War der Tipp richtig, so wird dem Konto die Quote minus dem Einsatz hinzugefügt. War der Tipp falsch, beträgt der Verlust einen Euro. Tabelle 7.4 zeigt die Ergebnisse dieser Simulation.

Bei dieser Betrachtung fällt nun auf, dass der allgemeine Ansatz hier nicht die schlechtesten Ergebnisse liefert. Im Durchschnitt beträgt der Verlust bei diesem Modell nach 252 getippten Spielen 1.83 Euro. Am meisten Geld verliert man im Durchschnitt überraschenderweise bei den Prognosen des geclusterten Modells mit Berücksichtigung der Startaufstellungen. Hier liegt der durchschnittliche Verlust bei 4.70 Euro. Im bestmöglichen Fall liegt der Gewinn bei diesem Modell bei 103.77 Euro, bei der schlechtesten Simulation liegt der Verlust jedoch bei 86.66 Euro. Beim einfachen Bradley-Terry Modell wird im Schnitt ein Gewinn von 3.00 Euro erzielt. Jedoch ist die Spanne der Simulationsergebnisse sehr hoch. Bei Betrachtung der Spanne, in welcher 90% der Simulationsergebnisse liegen, fällt auf, dass diese zwischen einem Verlust von 34.07 Euro und einem Gewinn von 41.18 Euro schwankt. Beim gewichteten dynamischen Modell ist der durchschnittliche Gewinn

7 Prädiktiver Modellvergleich

	stat	dyn	clust	mw	EWMA	EWMA1	emp
Minimum	-77.34	-100.15	-76.75	-86.66	-89.28	-82.41	-83.57
0.05-Quantil	-34.07	-32.95	-38.83	-42.10	-42.45	-43.49	-42.89
Durchschnitt	3.00	3.34	-1.11	-4.70	-3.34	-5.60	-1.83
0.95-Quantil	41.18	40.51	38.28	33.98	36.55	33.30	39.89
Maximum	90.61	88.42	79.05	103.77	86.10	83.98	101.40

Tabelle 7.4: Ergebnisse der Simulationsstudie zur Gewinnberechnung bei 10000 Simulationen pro Modell: Ergebnisse der Simulationsstudie bei 10000 Simulationen pro Modell: Einfaches BTL (stat), gewichtetes BTL (dyn), Modell mit Penalisierung und globalem Heimeffekt (clust), Modell mit Penalisierung, globalem Heimeffekt und Beachtung der Startaufstellung (mw), exponential-gewichteter moving-average Prozess mit gemeinsamen Startwerten (EWMA), exponential-gewichteter moving-average Prozess mit individuellen Startwerten (EWMA1) und ein Modell mit gemittelten Ergebniswahrscheinlichkeiten (emp)

sogar noch etwas höher und liegt bei 3.34 Euro. Doch auch hier sind die Schwankungen sehr hoch. Aufgrund der hohen Schwankungen ist das Risiko sehr hoch und dieser Ansatz zur Gewinnoptimierung nicht zu empfehlen.

7.3.2 Tipp der höchsten Wahrscheinlichkeit

Eine zweite Betrachtungsmöglichkeit besteht darin, die Spiele wie in Unterabschnitt 7.2.2 nach den höchsten Gewinnwahrscheinlichkeiten des Modells zu tippen. In Verknüpfung mit den entsprechenden Wettquoten ergeben sich die Ergebnisse aus Tabelle 7.5.

Im Vergleich zum vorherigen Ansatz lassen sich jetzt bessere Ergebnisse erzielen. Das klassische Bradley-Terry Modell, das gewichtete Modell, das geclusterte Modell mit globalem Heimeffekt und der exponential-gewichtete moving-average Prozess mit gemeinsamen Startwerten liefern nahezu identische Gewinne. Die Rendite bei diesen Modellen liegt bei etwa 7%. Beim EWMA-Prozess mit individuellen Startwerten liegt der Verlust bei 3.49 Euro. Beim geclusterten Modell mit globalem Heimeffekt und Berücksichtigung der Startaufstellungen lässt sich kaum ein Gewinn im Vergleich zu anderen Modellen erzielen. Dies ist etwas überraschend, da durch die zusätzliche Information der Startaufstellungen eine Verbesserung erwartet wurde. Beim Ansatz mit allgemeinen Ergebniswahrscheinlichkeiten werden, wie erwähnt, alle Spiele auf einen Heimsieg getippt. Mit dieser Methode ließe sich ein Gewinn von 31.41 Euro erzielen, was einer Rendite von 12.5% entspricht. Für die vorliegenden Daten scheint ein durchgehender Tipp auf Heimsiege scheinbar die beste

7 Prädiktiver Modellvergleich

Modell	richtige Tipps	Anteil	Gewinn	Gewinn pro Spiel
stat	124	0.492	17.97	0.071
dyn	124	0.492	17.56	0.070
clust	128	0.508	17.79	0.071
mw	130	0.516	1.08	0.004
EWMA	119	0.472	17.59	0.070
EWMA1	127	0.504	-3.49	-0.014
emp	123	0.488	31.41	0.125

Tabelle 7.5: Anzahl richtiger Ergebnisse und Höhe des Gewinns bei 252 Spielen und Tipp auf das wahrscheinlichste durch das Modell ermittelte Ergebnis: Einfaches BTL (stat), gewichtetes BTL (dyn), Modell mit Penalisierung und globalem Heimeffekt (clust), Modell mit Penalisierung, globalem Heimeffekt und Berücksichtigung der Startaufstellung (mw), exponential-gewichteter moving-average Prozess mit gemeinsamen Startwerten (EWMA), exponential-gewichteter moving-average Prozess mit individuellen Startwerten (EWMA1) und ein Modell mit gemittelten Ergebniswahrscheinlichkeiten (emp)

Tippmethode darzustellen.

7.3.3 Tipp nach Rentabilität

Wie zu Beginn von Abschnitt 7.3 beschrieben, bieten Wettanbieter unterschiedliche Quoten an, um einen Anreiz für riskante Tipps zu schaffen. Im Folgenden wird die Rentabilität solcher Tipps betrachtet. Hierfür werden die von den Modellen errechneten Ergebniswahrscheinlichkeiten mit den Quoten verknüpft und anschließend wird überprüft, ob sich ein Tipp auf dieses Spiel lohnt oder nicht. Die Rentabilität eines Tipps ergibt sich dabei folgendermaßen: Angenommen, ein Modell weist der Heimmannschaft in einer Partie eine Siegeswahrscheinlichkeit von 45% zu und die Quote des Wettanbieters für einen Heimsieg beträgt 2.1, dann ergibt sich die erwartete Rückzahlung bei einem solchen Tipp aus dem Produkt der beiden Zahlen und liegt bei 0.945. Da diese kleiner ist als der Einsatz, wäre ein solcher Tipp nicht rentabel und es sollte nicht auf dieses Ergebnis gesetzt werden. Abschließend werden nun noch zwei unterschiedliche Wettstrategien, basierend auf der Rentabilität der Tipps, betrachtet.

Rentabilität bei höchster Wahrscheinlichkeit

Bei dieser Wettstrategie wird wie in Unterabschnitt 7.3.2 jedes Spiel auf die Mannschaft mit der höchsten Gewinnwahrscheinlichkeit gesetzt, jedoch mit der Einschränkung, dass nur Spiele getippt werden, welche als rentabel klassifiziert werden. Tabelle 7.6 zeigt die Ergebnisse einer solchen Vorgehensweise.

Es ergeben sich größere Unterschiede bei der Anzahl der getippten Spiele. Während

Modell	getippte Spiele	richtige Tipps	Anteil	Gewinn	Gewinn pro Spiel
stat	159	69	0.434	19.43	0.122
dyn	192	89	0.464	21.12	0.110
clust	116	46	0.390	18.86	0.163
mw	104	47	0.452	10.79	0.104
EWMA	152	52	0.342	7.08	0.047
EWMA1	127	51	0.402	-2.459	-0.019
emp	126	49	0.389	36.38	0.289

Tabelle 7.6: Anzahl richtiger Ergebnisse und Höhe des Gewinns bei Tipp auf das wahrscheinlichste durch das Modell ermittelte Ergebnis und Berücksichtigung der Rentabilität: Einfaches BTL (stat), gewichtetes BTL (dyn), Modell mit Penalisierung und globalem Heimeffekt (clust), Modell mit Penalisierung, globalem Heimeffekt und Berücksichtigung der Startaufstellung (mw), exponential-gewichteter moving-average Prozess mit gemeinsamen Startwerten (EWMA), exponential-gewichteter moving-average Prozess mit individuellen Startwerten (EWMA1) und ein Modell mit gemittelten Ergebniswahrscheinlichkeiten (emp)

beim gewichteten Modell nur 60 Spiele als nicht rentabel klassifiziert werden und deshalb noch 192 Spiele getippt werden können, liegt die Zahl der getippten Spiele beim Modell unter Berücksichtigung der Startaufstellungen bei nur knapp über 100 Spielen. Bei allen Modellen hat sich der Anteil der richtig getippten Spiele verringert. Beim EWMA-Prozess mit gemeinsamen Startwerten werden nur noch gut ein Drittel aller Spiele richtig prognostiziert. Aufgrund der unterschiedlichen Anzahl getippter Spiele ist es nun besser, die Modelle nicht anhand ihrer Gewinne, sondern anhand der zu erwartenden Renditen miteinander zu vergleichen. Dabei fällt auf, dass sich beim einfachen Bradley-Terry Modell, beim gewichteten Modell und beim geclusterten Modell mit globalem Heimeffekt die Renditen zwischen 5% und 9% erhöhen. Bei Berücksichtigung der Startwerte lässt sich die Rendite auf über 10% steigern und das Modell, welches nur Heimsiege vorhersagt, steigert seine Rendite auf 28.9%. Bei den restlichen zwei Modellen sinkt die Rendite im Vergleich

7 Prädiktiver Modellvergleich

zum Tipp auf alle Spiele. Es scheint, als sei auch bei dieser Methode ein grundsätzlicher Tipp auf Heimerfolge zu bevorzugen.

Tipp auf höchste Rentabilität

Ein letzter Wettansatz testet die zu erwartenden Ergebnisse bei ausschließlicher Tipp auf die höchste Rentabilität eines Spiels. Es werden nun für alle Ereigniswahrscheinlichkeiten und Quoten die jeweiligen erwarteten Rückzahlungen berechnet. Anschließend wird auf das Ergebnis gesetzt, welches die höchste Rentabilität aufweist ungeachtet der zugrundeliegenden Wahrscheinlichkeit des Ergebnisses. Liegt die Rentabilität bei einem Spiel für alle drei möglichen Ergebnisse niedriger als das Risiko, so wird auf dieses Spiel nicht gesetzt. Tabelle 7.7 zeigt die entsprechenden Ergebnisse.

Bei einer solchen Vorgehensweise sinken die Anteile der richtig getippten Ergebnisse

Modell	getippte Spiele	richtige Tipps	Anteil	Gewinn	Gewinn pro Spiel
stat	241	81	0.336	19.02	0.079
dyn	245	95	0.386	32.17	0.131
clust	241	63	0.261	-2.99	-0.012
mw	234	69	0.295	-5.47	-0.103
EWMA	239	65	0.272	3.88	0.016
EWMA1	250	72	0.288	0.820	0.003
emp	232	65	0.280	29.48	0.127

Tabelle 7.7: Anzahl richtiger Ergebnisse und Höhe des Gewinns bei Tipp auf das Ergebnis mit der höchsten Rentabilität: Einfaches BTL (stat), gewichtetes BTL (dyn), Modell mit Penalisierung und globalem Heimeffekt (clust), Modell mit Penalisierung, globalem Heimeffekt und Berücksichtigung der Startaufstellung (mw), exponential-gewichteter moving-average Prozess mit gemeinsamen Startwerten (EWMA), exponential-gewichteter moving-average Prozess mit individuellen Startwerten (EWMA1) und ein Modell mit gemittelten Ergebniswahrscheinlichkeiten (emp)

erwartungsgemäß deutlich ab. Teilweise werden nur noch knapp ein Viertel aller Spiele richtig vorhergesagt. Auch die Renditen fallen bei fast allen Modellen deutlich. Einzige Ausnahme ist das gewichtete Modell. Bei diesem werden mit Abstand die meisten Spiele richtig getippt und die Rendite liegt bei über 13%. Obwohl beim Ansatz mit allgemeinen Ergebniswahrscheinlichkeiten nun nicht mehr alle Spiele auf einen Heimsieg getippt werden, schneidet dieses Modell auch bei der vorgestellten Betrachtungsweise sehr gut ab. Es lässt sich eine Rendite von 12.7% erzielen.

8 Zusammenfassung und Diskussion

Ziel dieser Arbeit war es, verschiedene Ansätze zu betrachten, um die Spielstärke der Fußballbundesligisten in der Saison 2014/2015 zu modellieren. Zunächst wurde hierfür das statische Bradley-Terry Modell behandelt. Nach einer kurzen Vorstellung dieses Modells wurde es zu einem ordinalen Modell erweitert. Zusätzlich wurde noch ein globaler und ein teamspezifischer Heimeffekt eingeführt, welche getrennt voneinander zu betrachten sind. Durch seinen Zusammenhang mit den generalisierten linearen Modellen konnte auf die vorgestellten Modelle die bekannte goodness-of-fit Statistik sowie die Likelihood-ratio-Statistik angewandt werden. Die Anwendung dieser Modelle zeigte, dass eindeutig ein Heimeffekt für die verwendeten Daten existiert. Ein teamspezifischer Heimeffekt brachte im Vergleich zu einem globalen Heimeffekt jedoch keine Verbesserung. Die geschätzten Stärken der Mannschaften spiegelten die tatsächlichen Platzierungen der Mannschaften zwar relativ gut wider, jedoch zeigten die quasi-Standardabweichungen auf, dass sich die meisten Spielstärken nicht signifikant voneinander unterscheiden lassen. Dies warf die Vermutung auf, dass die Spielstärken der einzelnen Mannschaften geclustert sein könnten. Kapitel 4 beschäftigte sich daher mit dem Clustering der Spielstärken durch Regularisierungsmethoden. Zunächst wurde hier das Ranking LASSO vorgestellt. Diese Methodik belegt die Bradley-Terry Likelihood mit einem Strafterm für alle paarweisen Spielstärkekoeffizienten. Dieser Strafterm ist eine Verallgemeinerung des fusionierten LASSO-Strafterms von Tibshirani et al. (2005). Um keine inkonsistenten Schätzer zu erhalten, werden die Inhalte des Strafterms stärker gewichtet als die Abnahme der Effekte. Ein Nachteil dieses Vorgehens besteht darin, dass die Methodik erst angewandt werden kann, sobald jedes Team nicht mehr jedes Spiel gewonnen oder verloren hat, da sonst die Spielstärken dieser Teams als $+\infty$ oder $-\infty$ geschätzt werden. Bei einer Unterscheidung der Ergebnisse in fünf Kategorien schwächt sich diese Bedingung etwas ab. Hier ist eine Modellierung möglich, sobald jedes Team nicht mehr durchgehend dasselbe Ergebnis erzielt hat. Als Auswahlkriterien für das Clustering werden das Akaike und das Bayessche Informationskriterium verwendet. Bei Betrachtung des BIC als Auswahlkriterium und globalem Heimeffekt lassen sich die Mannschaften in sechs verschiedene Cluster einteilen, wobei diese teilweise sehr ähnlich sind, so dass eine Einteilung in vier Cluster vertretbar scheint. Die ersten drei Cluster sind dabei von den vier stärksten Teams der Liga besetzt. Diese Teams hatten sich in der abgelaufenen Saison deutlich von den restlichen Teams abgesetzt, so dass diese Einteilung sinnvoll erscheint.

8 Zusammenfassung und Diskussion

In Kapitel 5 wurde nun versucht herauszufinden, ob sich die Spielstärke der Mannschaften durch einzelne Kovariablen erklären lässt. Dazu wurde das Modell so erweitert, dass sich Kovariablen, welche eindeutig einem Verein zugeschrieben werden können ins Modell mit aufnehmen lassen. Zunächst beschränkte sich diese Erweiterung auf Kovariablen, welche sich für jede Mannschaft über die Saison nicht verändern. Dies waren der Lizenzspielertat und der Marktwert der einzelnen Mannschaften. Später wurde das Modell noch so angepasst, dass auch spieltagsabhängige Kovariablen modelliert werden können. Diese Erweiterung fand Anwendung bei der Betrachtung der durchschnittlichen Marktwerte der Startaufstellungen der Mannschaften. Bei der Analyse der Ergebnisse für die erstgenannte Erweiterung fiel auf, dass trotz Berücksichtigung der Marktwerte oder auch der Vereinsbudgets vier der ersten fünf Mannschaften der Abschlusstabelle als stärkste Mannschaften klassifiziert wurden. Nur der FC Bayern München fällt hier ins Mittelfeld ab. Werden die Marktwerte als unabhängige Variablen und die ermittelten Spielstärken als abhängige Variablen betrachtet und führt mit diesen eine lineare Regression durch, so ist festzustellen, dass sich über 55% der Streuung durch ein lineares Modell und fast 63% der Streuung durch ein additives Modell erklären lassen. Daraus lässt sich schließen, dass die gewählte Kovariable einen erheblichen Teil der Spielstärken der Mannschaften erklären kann. Bei Verwendung der durchschnittlichen Marktwerte der Startaufstellungen als Kovariablen gehört der FC Bayern München überraschenderweise trotzdem zu den stärksten Teams bei den ermittelten Spielstärken. Dies lässt vermuten, dass diese Wahl der Kovariablen eventuell doch nicht so geeignet ist. Bei Betrachtung eines Regressionsmodelles, wie gerade beschrieben, lassen sich jedoch fast 57% bei einem linearen Modell bzw. über 69% bei einem additiven Modell der Streuung des Modells so erklären.

Ein anderer Ansatz legte einen exponential-gewichteten moving-average Prozess zur Modellierung der Spielstärke der einzelnen Mannschaften zugrunde und wurde in Kapitel 6 beschrieben. Bei einem solchen Ansatz werden wesentlich weniger Parameter geschätzt als bei der zuvor vorgestellten Methodik. Daher ist das Ziel eines solchen Modells, Ergebnisse zu erhalten, welche möglichst nahe an denen des klassischen Bradley-Terry Modells liegen. Der Vorteil dieses Ansatzes liegt darin, dass dynamische Spielstärken für jede Mannschaft erhalten werden. Für jede Mannschaft werden getrennte Spielstärken für Heim- und Auswärtsspiele modelliert. Die Heimspielstärke lässt sich mit Hilfe eines Glättungsparameters sowie der vergangenen Ergebnisse der Heimspiele einer Mannschaft iterativ ermitteln. Die Auswärtsspielstärke ergibt sich äquivalent. Bei der Wahl von gemeinsamen Startwerten für jede Mannschaft lassen sich bei den ermittelten Stärkekurven die Veränderungen während der Saison gut erkennen. Bei individuellen Startwerten für jede Mannschaft sind

8 Zusammenfassung und Diskussion

diese Verläufe nicht zu erkennen. Die Spielstärke ergibt sich vor allem aus den Endplatzierungen der Vorsaison. Zum Vergleich der Modelle mit dem klassischen Bradley-Terry Ansatz wurde der rank probability score eingeführt. Dieser gibt an, wie gut die geschätzten Ergebnisse auf die tatsächlichen Daten passen. Für beide EWMA-Modelle finden sich sowohl Argumente für als auch Argumente gegen ihre Anwendung. Ein Nachteil dieser Betrachtungsweise ist die strikte Trennung zwischen Heim- und Auswärtsspielen. Es wäre wünschenswert, wenn sich die getrennten Betrachtungen zu einem gemeinsamen Modell zusammenfügen ließen. Ein weiterer Ansatz, um ein dynamisches Modell zu erhalten, ist dem vorgestellten klassischen Ansatz aus Kapitel 3 sehr ähnlich. Durch Modellierung der Spielstärken für jeden Spieltag einzeln ergibt sich bereits ein dynamisches Modell. Zusätzlich wurden aber noch die Likelihoods mit Hilfe einer Kernfunktion gewichtet. Somit sollen weiter zurückliegende Spiele weniger Einfluss auf die Spielstärke erhalten. Um zu entscheiden, welche Bandbreite für die Kernfunktion gewählt werden soll, wurde ein prädiktives Modell auf die Daten der Saison 2013/2014 angewandt. Durch den nun gewählten Ansatz lässt sich der rank probability score deutlich verbessern. Auch die Spielstärkenentwicklung der einzelnen Vereine über die Saison ist bei diesem Modell gut erkennbar. Hier ist jedoch wieder zu beachten, dass sich vor allem zu Beginn der Saison aufgrund der geringen Anzahl an Daten sehr hohe quasi-Standardabweichungen ergeben. Eine signifikante Unterscheidung der Spielstärke der Mannschaften ist daher meist nicht möglich.

Für einen Anwender ist von Interesse, wie gut die Modelle prognostizieren. Kapitel 7 beschäftigte sich daher mit der Prognose von Ergebnissen. Zunächst wurde der jeweilige rank probability score ausgewählter Modelle miteinander verglichen. Hierbei konnten nur kleine Unterschiede zwischen den Modellen festgestellt werden. Daher wurde anschließend für diese Modelle überprüft, wie gut sie Spielausgänge vorhersagen. Bei einer Simulationsstudie schnitt dabei das Modell mit den gewichteten Likelihoods am besten ab. Bei ausschließlichen Tipps auf die Heimmannschaft liegt die Zahl der durchschnittlich getippten Spiele jedoch deutlich höher als bei diesem Modell. Aufgrund dessen wurde getestet, wie hoch die Erfolgsquote bei Tipps auf das wahrscheinlichste vom Modell prognostizierte Ergebnis liegt. Dadurch wurde erreicht, dass sich die Trefferquote auf über 50% erhöht. Als Letztes wurden zusätzlich noch die Wettquoten für alle Spiele betrachtet. Da die Quoten für Heimsiege geringer sind als für Auswärtsspiele, wird vermutet, dass sich trotz einer geringeren Anzahl richtig getippter Spiele durch gezielt platzierte Wetten höhere Gewinne erzielen lassen. Bei einer Simulation von 10000 Durchgängen konnte man feststellen, dass sich nur bei drei Modellen im Schnitt ein geringer Gewinn

8 Zusammenfassung und Diskussion

erzielen lässt. Die Schwankungen der Ergebnisse sind zusätzlich sehr hoch. Tipps auf den wahrscheinlichsten Spielausgang lieferten bessere Ergebnisse. Bei vier Modellen beträgt die Rendite etwa 7%. Ausschließliches Tippen auf Heimsiege führt zu einer Rendite von 12.5%. Wird zusätzlich die Rentabilität eines Tipps berücksichtigt, lassen sich die Renditen deutlich steigern. Das beste in dieser Arbeit vorgestellte Modell weist eine Rendite von über 16% auf. Bei ausschließlichen Tipps auf einen Heimsieg unter Berücksichtigung der Rentabilität liegt die Rendite bei 28.9%. Eine Wettstrategie, bei der die Wetten nur nach der größten Rentabilität platziert werden, verschlechtert für die meisten Modelle den erwarteten Gewinn. Nur beim gewichteten dynamischen Modell kann so die Rendite auf über 13% gesteigert werden. Abschließend sollen die Ergebnisse nochmals hinterfragt und diskutiert werden.

Für die vorliegenden Daten lassen sich mit Hilfe der vorgestellten Modelle leichte Gewinne durch gezielte Wetten erzielen. Da auch Wettanbieter ihre Quoten über statistische Modelle modellieren, welche wesentlich komplexer sind als die in dieser Arbeit vorgestellten Modelle mit maximal einer Kovariablen, kann angezweifelt werden, dass die Modelle allgemeingültig sind und auch für zukünftige Partien gute Ergebnisse liefern. Es ist jedoch trotzdem überraschend, dass das Modell mit Berücksichtigung der Startaufstellungen keine besseren Ergebnisse liefert als die anderen Modelle, da hier eigentlich für jedes Spiel zusätzliche Information zur Verfügung steht. Dies könnte wohl vor allem an der Wahl des BIC-Kriteriums bei der Clusterbildung liegen. Dieses ist sehr restriktiv und wird daher nur wenige Cluster erzeugen. Durch die Einteilung in Cluster weist das Modell sehr ausgeglichene Ergebniswahrscheinlichkeiten aus. Daher beraubt sich das Modell klarer Aussagen. Ebenso kann die Wahl der Kovariablen hinterfragt werden. Die in dieser Arbeit benutzten Kovariablen beruhten hauptsächlich auf Schätzungen, welche hohe Fehler aufweisen können. Interessant wäre ebenfalls, eine Verbindung zwischen den einzelnen Modellen herzustellen, wie zum Beispiel zwischen den gewichteten Likelihoods und verschiedener zusätzlicher Kovariablen.

Die Modellierung von Sportereignissen ist ein sehr weites Feld, in welchem sich ständig neue Möglichkeiten ergeben und welches sich in den nächsten Jahren wohl stetig weiterentwickeln wird. Besitzt man noch zusätzliches Insiderwissen wie Matthew Benham, so kann mit statistischer Modellierung im Sportbereich auch sehr viel Geld verdient werden.

A Verwendete Software

Alle für diese Arbeit durchgeführten Analysen wurden mit der statistischen Software *R* (RCoreTeam (2016)) durchgeführt. Zur Modellierung der Koeffizientenschätzer in Kapitel 3 wurde das Softwarepaket *ordBTL* (Casalicchio (2013)) verwendet. Die quasi-Standardabweichungen konnten mit Hilfe des Paketes *qvcalc* (Firth (2015)) ermittelt werden. Zur Identifizierung von Clustern sowie der Erweiterung des Modells für erklärende Variablen diente ein von Gunther Schauburger zur Verfügung gestellter Code als Grundlage. Zur Ermittlung des Zusammenhangs zwischen den ermittelten Spielstärken und den erklärenden Variablen wurde das Programmpaket *mgcv* (Wood (2015)) verwendet. Der Programmiercode für die von Cattelan et al. (2013) durchgeführten Analysen ist im Internet frei zugänglich und diente als Ansatz für die Anwendungen der in dieser Arbeit betrachteten Daten. Sie stehen auf der Internetseite

<http://www.blackwellpublishing.com/rss>

zur Verfügung. Um Kernfunktionen zur Gewichtung der Likelihoods verwenden zu können wurde mit dem Paket *gplm* (Mueller (2014)) gearbeitet. Zur Aufbereitung der Daten aus der Saison 2013/2014 war das Paket *plyr* (Wickham (2015)) nötig.

B Anhang

Cluster	Mannschaften	Spielstärke
1	FC Bayern München	0.64
2	VfL Wolfsburg	0
3	Borussia Mönchengladbach	-0.05
4	Bayer 04 Leverkusen	-0.27
5	Borussia Dortmund	-1.04
6	FC Augsburg, FC Schalke 04	-1.06
7	1.FC Köln, 1.FSV Mainz, Eintracht Frankfurt	-1.15
8	1899 Hoffenheim, SC Freiburg, Werder Bremen	-1.16
9	Hamburger SV, Hannover 96, Hertha BSC Berlin	-1.17
	VfB Stuttgart	
6	SC Paderborn	-1.43

Tabelle B.1: Cluster der Spielstärken bei globalem Heimeffekt unter Verwendung des AIC-Kriteriums

B Anhang

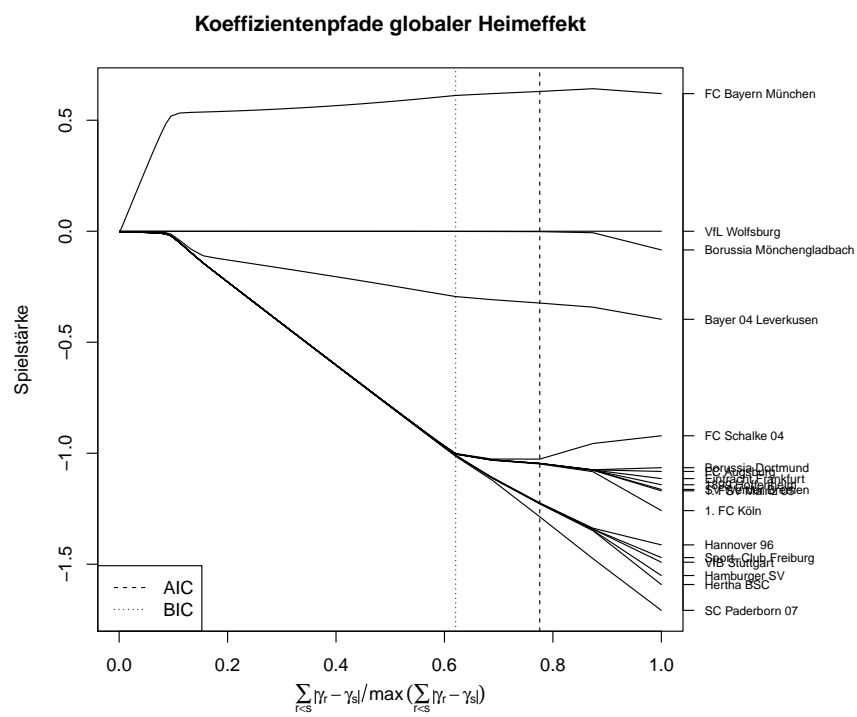


Abbildung B.1: Koeffizientenpfade bei globalem Heimeffekt unter Verwendung eines adaptiven L_1 -Strafterms und drei Kategorien für die Ergebnisse

B Anhang

Cluster	Mannschaften	Spielstärke
Heimstärke		
1	FC Bayern München	1.06
2	VfL Wolfsburg	0.52
3	Borussia Mönchengladbach	0.45
4	Bayer 04 Leverkusen	0.18
5	Borussia Dortmund, Schalke 04	-0.33
6	Eintracht Frankfurt, FC Augsburg	-0.34
7	1.FSV Mainz 05	-0.54
8	TSG Hoffenheim	-0.55
9	SV Werder Bremen	-0.80
10	1. FC Köln, Hamburger SV, Hertha BSC, SC Freiburg	-0.86
11	Hannover 96, SC Paderborn 07	-1.12
12	VfB Stuttgart	-1.20
Auswärtsstärke		
1	FC Bayern München	0.72
2	VfL Wolfsburg	0.00
3	Borussia Mönchengladbach	-0.08
4	Bayer Leverkusen	-0.35
5	1.FC Köln	-0.93
6	Borussia Dortmund	-1.18
7	FC Augsburg, Hannover 96, VfB Stuttgart	-1.19
8	1.FSV Mainz 05, Hertha BSC Berlin, SC Freiburg, SV Werder Bremen	-1.20
9	FC Schalke 04	-1.22
10	TSG 1899 Hoffenheim	-1.40
11	Eintracht Frankfurt, Hamburger SV, SC Paderborn 07	-1.52
Heimvorteil		
1	Eintracht Frankfurt	1.18
2	FC Schalke 04	0.89
3	TSG 1899 Hoffenheim, Borussia Dortmund, FC Augsburg	0.85
4	1.FSV Mainz 05, Hamburger SV	0.66
5	Bayer 04 Leverkusen, Borussia Mönchengladbach	0.53
6	VfL Wolfsburg	0.52
7	SC Paderborn 07, SV Werder Bremen	0.40
8	FC Bayern München, Hertha BSC Berlin, SC Freiburg	0.34
9	1.FC Köln, Hannover 96	0.07
10	VfB Stuttgart	-0.01

Tabelle B.2: Cluster der Spielstärken und des Heimvorteils bei teamspezifischem Heimeffekt unter Verwendung des AIC-Kriteriums

B Anhang

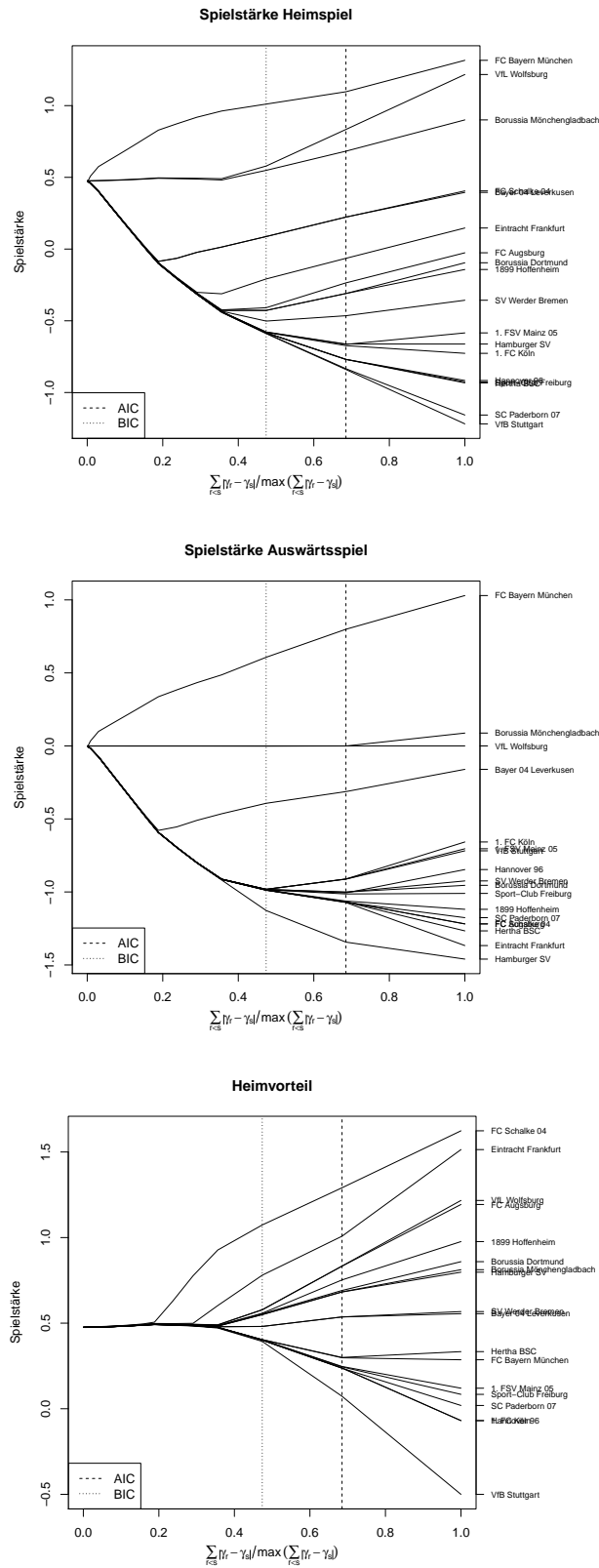


Abbildung B.2: Koeffizientenpfade bei teamspezifischem Heimeffekt unter Verwendung eines adaptiven L_1 -Strafterms und drei Kategorien für die Ergebnisse

B Anhang

Mannschaft	Etat	MW 23.10.2014	MW 01.02.2015
FC Bayern München	160	564.35(27)	544.05(25)
VfL Wolfsburg	75	174.45(28)	163.30(26)
Borussia Mönchengladbach	38	119.40(26)	128.65(26)
Bayer 04 Leverkusen	55	176.85(26)	179.30(24)
FC Augsburg	19	42.85(31)	47.55(31)
FC Schalke 04	78	208.25(31)	224.80(29)
Borussia Dortmund	73	344.45(30)	349.55(30)
TSG 1899 Hoffenheim	35	122.15(30)	116.65(27)
Eintracht Frankfurt	34	76.15(29)	73.98 (29)
SV Werder Bremen	30	55.00(27)	50.03(29)
1.FSV Mainz 05	29	79.05(30)	80.40(31)
1.FC Köln	25	41.25(25)	37.45(23)
Hannover 96	39	69.00(32)	73.93(33)
VfB Stuttgart	42	98.80(28)	95.25(28)
Hertha BSC Berlin	31	74.60(29)	70.85(27)
Hamburger SV	46	80.10(29)	78.50(30)
SC Freiburg	19	52.60(26)	54.85(26)
SC Paderborn 07	17	22.35(28)	24.95(28)

Tabelle B.3: Lizenzspieleretats und Marktwerte der einzelnen Vereine in Millionen Euro. In Klammern ist die Anzahl der gemeldeten Spieler des Vereins angegeben

B Anhang

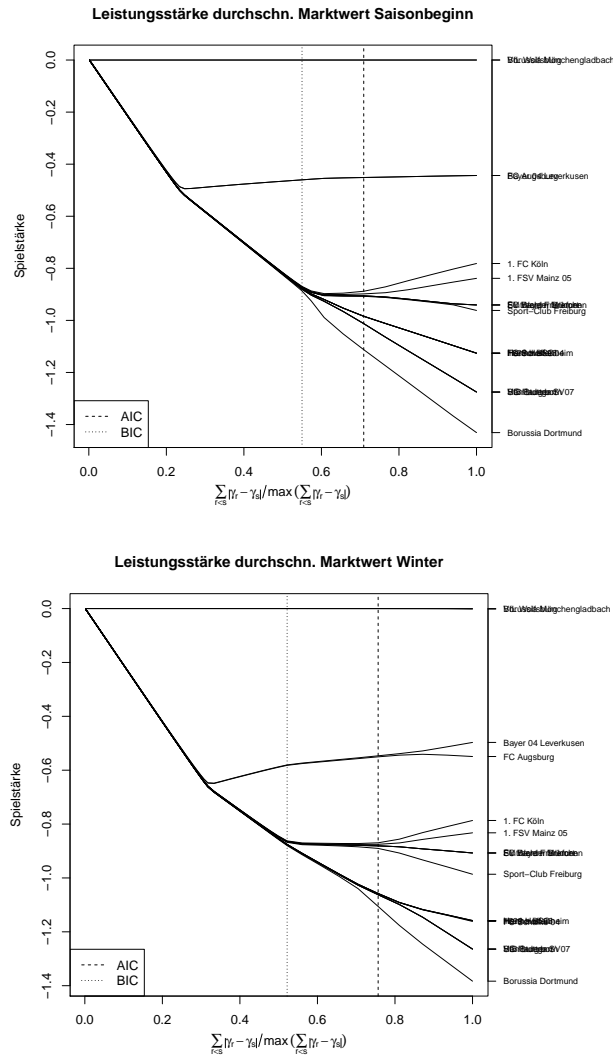


Abbildung B.3: Koeffizientenpfade der Spielstärke bei globalem Heimeffekt und dem durchschnittlichen Marktwert zu Saisonbeginn und zu Beginn der Winterpause als erklärende Variable

B Anhang

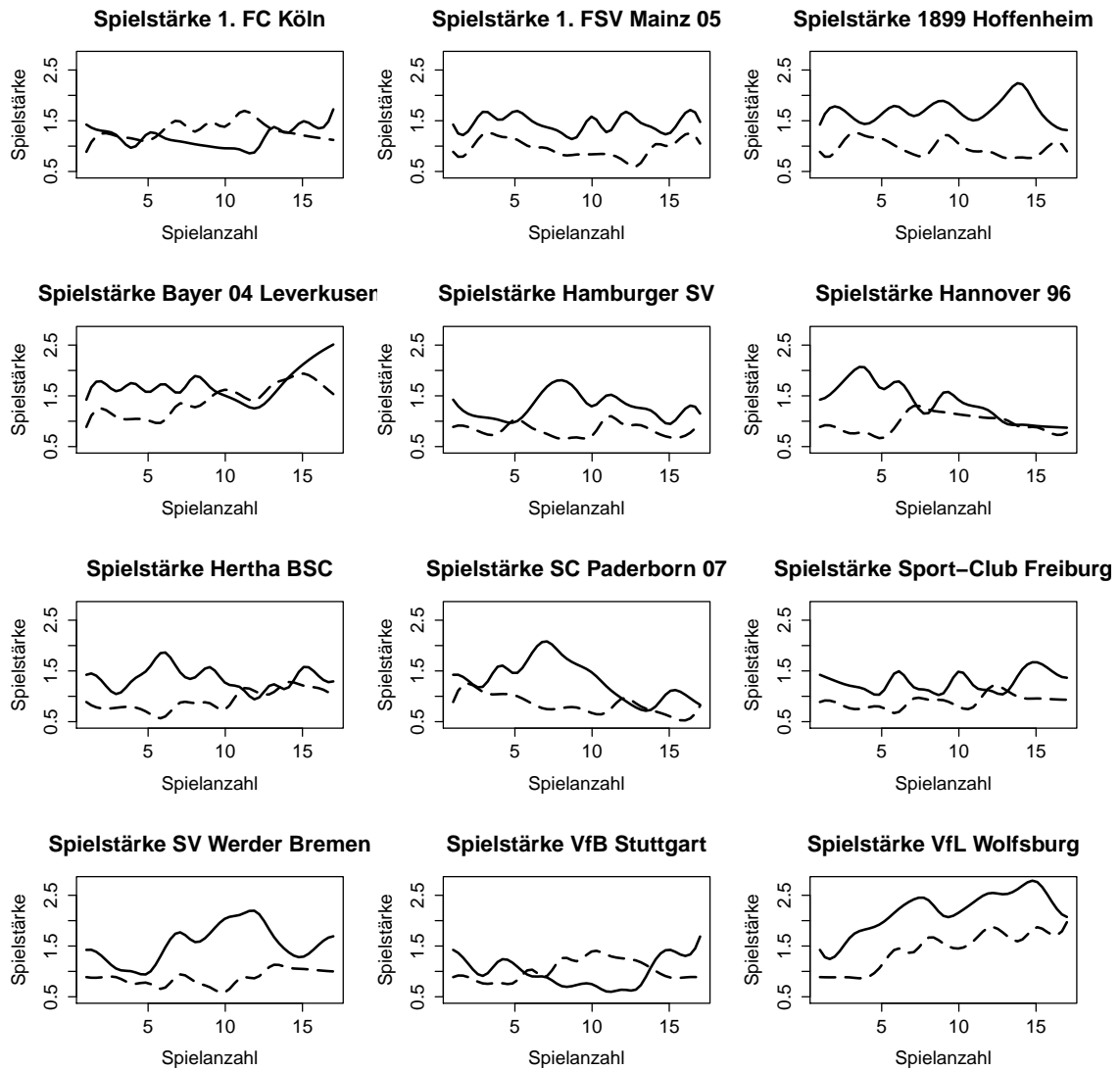


Abbildung B.4: Spielstärkeentwicklung der restlichen Mannschaften mit Unterscheidung zwischen Heim (durchgezogene Linie)- und Auswärtsspielstärke (gestrichelte Linie) bei gemeinsamen Startwerten

B Anhang

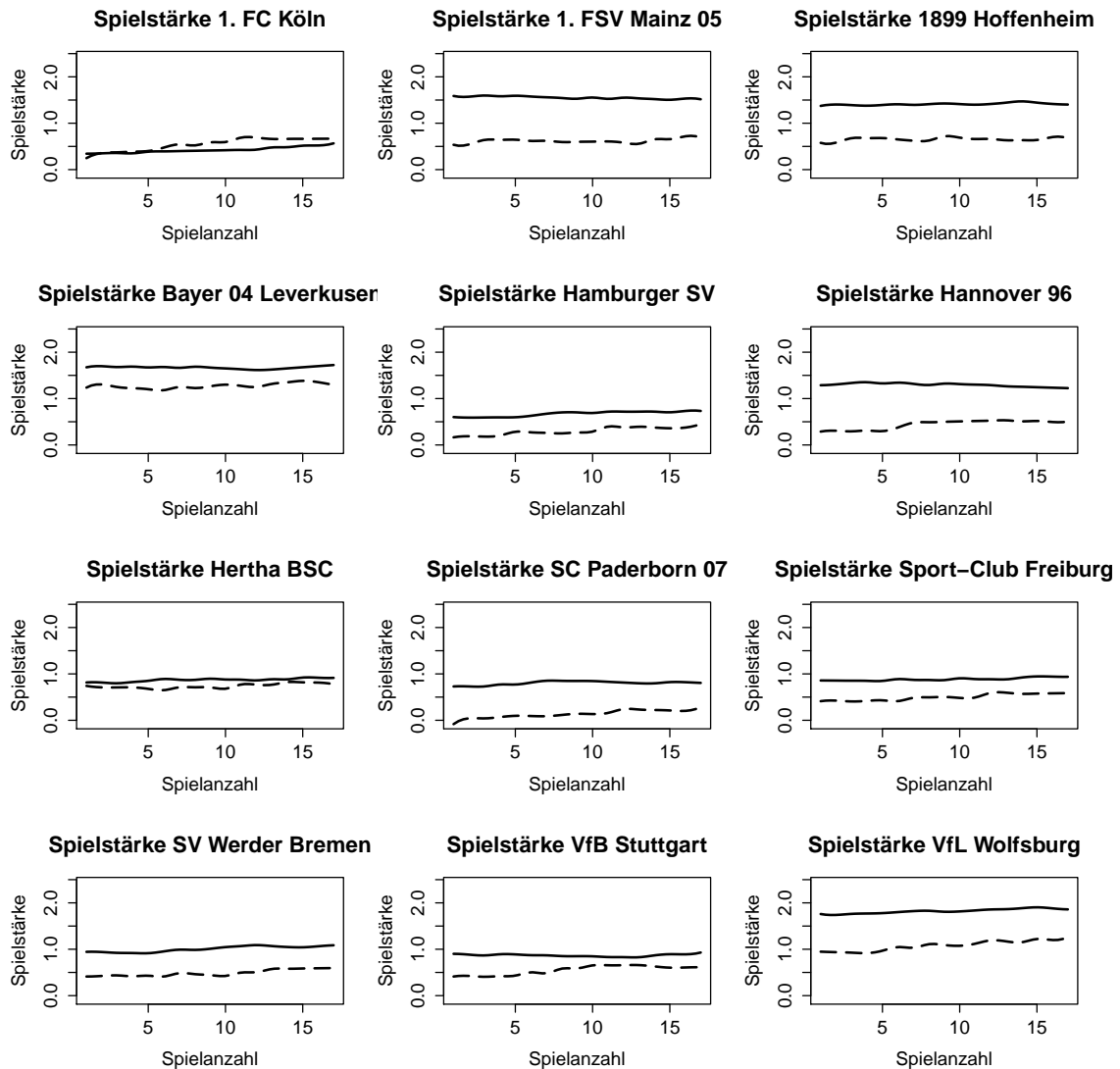


Abbildung B.5: Spielstärkeentwicklung der restlichen Mannschaften mit Unterscheidung zwischen Heim (durchgezogene Linie)- und Auswärtsspielstärke (gestrichelte Linie) bei individuellen Startwerten

B Anhang

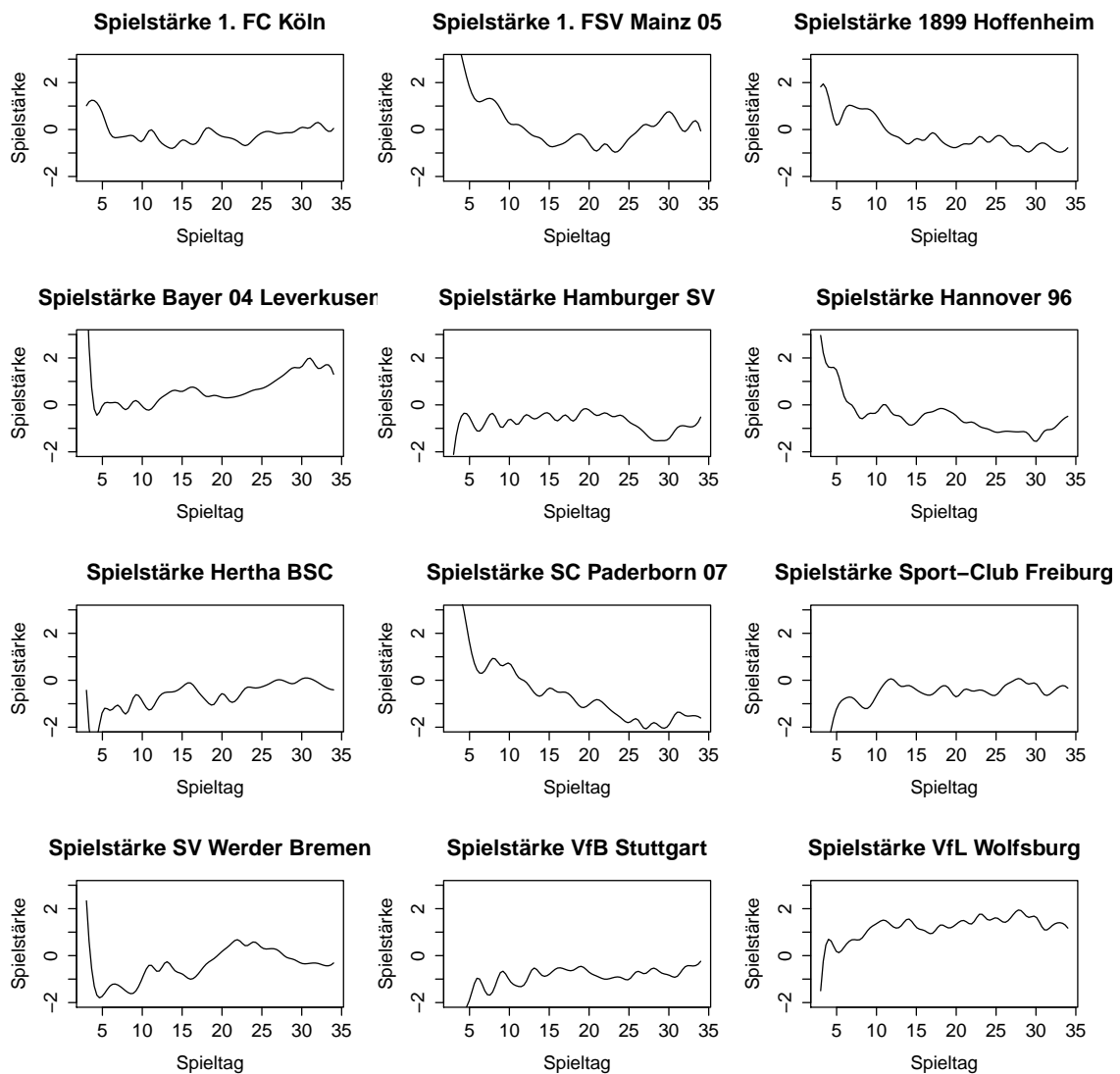


Abbildung B.6: Geschätzte Spielstärkekoeffizienten der restlichen Mannschaften bei gewichteter Likelihood vom 3. bis 34. Spieltag

C Inhalt der CD

Auf beigelegter CD befinden sich folgende Ordner und Dateien:

- Ordner Masterarbeit:
Enthält die angefertigte Masterthesis im .pdf-Format
- Ordner R:
Enthält alle mit der Software R durchgeführten Analysen und ist in folgende Unterordner unterteilt:
 - Daten
 - Datenaufbereitung
 - Funktionen
 - Grafiken
 - Modelle
 - Prädiktion
 - workspaces
 - startcode.R
- Die Datei ReadMe.txt enthält zusätzlich die Beschreibungen aller R-Codes welche sich in den einzelnen Unterordnern befinden.

D Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die Masterthesis selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die aus fremden Quellen direkt oder indirekt übernommenen Gedanken als solche kenntlich gemacht habe.

Die Arbeit habe ich bisher keinem anderen Prüfungsamt in gleicher oder vergleichbarer Form vorgelegt. Sie wurde bisher nicht veröffentlicht.

München, den 29.02.2016

Literaturverzeichnis

- Agresti, A. (1992). Analysis of ordinal paired comparison data. *Appl. Stat.* 41(2), 287–297.
- Akaike, H. (1974). A new look at statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723.
- Biermann, C. (2013). Der beste Profiwetter der Welt. URL: <http://www.11freunde.de/interview/wie-setzt-man-die-perfekte-fussballwette-matthew-benham>. Zuletzt eingesehen: 15.02.2016.
- Bradley, R. A. und M. E. Terry (1952). Rank analysis of incomplete block designs, i. the method of paired comparison. *Biometrika* 39(3/4), 324–345.
- Buja, A., T. Hastie, und R. Tibshirani (1989). Linear smoothers and additive models. *Ann. Stat.* 17(2), 453–510.
- Casalicchio, G. (2013). ordBTL: Modelling comparison data with ordinales response. *R package version 7*.
- Cattelan, M., C. Varin, und D. Firth (2013). Dynamic bradley-terry modelling of sports tournaments. *Appl. Statist.* 62(1), 135–150.
- Clark, S. R. und J. M. Norman (1995). Home ground advantage of individual clubs in english soccer. *Statisticalian* 44, 509–521.
- Czado, C., T. Gneiting, und L. Held (2009). Predictive model assessment for count data. *Biometrics* 65, 1254–1261.
- Davidson, R. (1970). On extending the bradley-terry model to accomodate ties in paired comparison experiments. *Journal of the American Statistical Association* 65, 317–328.
- Fahrmeir, L. und G. Tutz (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association* 89(428), 1438–1449.
- Fahrmeir, L. und G. Tutz (2001). *Multivariate statistical modelling based on generalised linear models*. Springer-Verlag.
- Fan, J. und R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96, 1348–1360.

Literaturverzeichnis

- Firth, D. (2015). Quasi variances for factor effects in statistical models. *R package version 0.8-9*.
- Firth, D. und R. De Menezes (2004). Quasi-variates. *Biometrika* 91(1), 65–80.
- Friedman, J. H., T. Hastie, und R. Tibshirani (2010). Regularization path for generalized linear models via coordinate descent. *J. Stat. Softw.* 33(1), 1–22.
- fussball.wettpoint.com (2016). Ergebnis Verteilung 1.Bundesliga. URL: <http://fussball.wettpoint.com/liga/1-bundesliga-deutschland-vtg.html>. Zuletzt eingesehen: 14.02.2016.
- Gertheiss, J. und G. Tutz (2010). sparse modelling of categorial explanatory variables. *Ann. Appl. Stat* 4(4), 2150–2180.
- Glickman, M. E. und H. S. Stern (1998). A state-space model for national football league scores. *J. Am. Statist. Ass.* 93, 25–35.
- Harville, D. (1980). Predictions for national football league games via linear-model methodology. *J. Am. Statist. Ass* 75, 516–524.
- Knorr-Held, L. (2000). Dynamic rating of sports teams. *J. R. Stat. Soc. B* 49(2), 261–276.
- Luce, R. D. (1959). *Individual choice behaviour: A theoretical analysis*. Dover books on mathematics. Dover publications inc.
- Masarotto, G. und C. Varin (2012). The ranking lasso and its application to sport tournaments. *Ann. Appl. Stat.* 6(4), 1949–1970.
- McCullagh, P. (1980). Regression model for ordinal data (with discussion). *J. R. Stat. Soc. B* 42(2), 109–127.
- McCullagh, P. und J. Nelder (1989). *Generalized linear models*. Chapman & Hall.
- Mueller, M. (2014). Functions for estimating generalized partial linear models. *R package version 0.7-2*.
- Nedo, J. (2015). 20 Jahre Drei-Punkte Regel: Vom Schwachsinn zum Standard. URL: <http://www.tagesspiegel.de/sport/20-jahre-drei-punkte-regel-vom-schwachsinn-zum-standard/12190850.html>. Zuletzt eingesehen: 09.02.2016.

Literaturverzeichnis

- Oelker, M. und G. Tutz (2013). A general family of penalties for combining differing types of penalties in generalized structured models. *Technical Report 139, LMU, Department of Statistics*.
- ran.de (2015). 20 Jahre Drei-Punkte-Regel: Kein Vizekusen, kein Meister der Herzen. URL: <http://www.ran.de/fussball/bundesliga/news/20-jahre-drei-punkte-regel-kein-vizekusen-kein-meister-der-herzen-1035312>. Zuletzt eingesehen: 09.02.2016.
- Rao, P. und L. Kupper (1967). Ties in paired-comparison experiments: a generalization of the bradley-terry model. *Journal of the American Statistical Association* 62, 194–204.
- RCoreTeam (2016). R: A language and environment for statistical computing. URL: <http://www.r-project.org>.
- Reich, M. und V. Fritzen (2014). Fussballmillionäre im Vergleich- In der Bundesliga explodieren die Gehälter. URL: <http://www.handelsblatt.com/sport/fussball/fussballmillionaere-im-vergleich-in-der-bundesliga-explodieren-die-gehaelter/10715082.html>. Zuletzt eingesehen: 09.02.2016.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6(2), 461–464.
- Springall, A. (1973). Response surface fitting using a generalization of bradley-terry paired comparison model. *App. Stat.* 22(1), 59–68.
- Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, und K. Knight (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67, 91–108.
- transfermarkt.de (2016). URL: <http://www.transfermarkt.de/1-bundesliga/marktwerteverein/wettbewerb/L1>. Zuletzt eingesehen: 09.02.2016.
- Turner, H. und D. Firth (2012). Bradley-terry models in r: The bradleyterry2 package. *J.Stat. Softw.* 48(9), 1–21.
- Tutz, G. (1986). Bradley-terry-luce models with an ordered response. *J. Math. Psychol.* 30, 306–316.
- Tutz, G. (2012). *Regression for categorical data*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press.
- Tutz, G. und G. Schauburger (2015). Extended ordered paired comparison models with application to football data from german bundesliga. *AStA. Adv Stat Anal*, 209–227.

Literaturverzeichnis

- Wahl, G. (2016). Soccer analytics revolution underway at Benham's Brentford, Midtjylland. URL: <http://www.si.com/planet-futbol/2015/05/15/soccer-analytics-brentford-midtjylland-matthew-benham>. Zuletzt eingesehen: 15.02.2016.
- wettbuero.de (2016). Matthew Benham: vom Sportwetten-Millionär zum dänischen Meistermacher – sind Fußball und Sportwetten tatsächlich immer berechenbar? URL: <https://www.wettbuero.de/matthew-benham-vom-sportwetten-millionaer-zum-daenischen-meistermacher-sind-fussball-und-sportwetten-tatsaechlich-immer-berechenbar/>. Zuletzt eingesehen: 15.02.2016.
- Wickham, H. (2015). Tools for splitting, applying and combining data. *R package version. 1.8.1*.
- Wood, S. (2015). Mixed gam computation vehicle with gcv/aic/reml smoothness estimation. *R package version. 1.8-7*.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* 101, 1418–1429.