

A Statistical Analysis of NFL Quarterback Rating Variables

Seminar: Statistik im Sport



von

Norbert Krautenbacher

betreut durch

Sebastian Kaiser

bei Prof. Dr. Friedrich Leisch, Manuel Eugster, Sebastian Kaiser

am Institut für Statistik

Ludwig-Maximilians-Universität München

Inhaltsverzeichnis

1 Einführung.....	3
2 Spielregeln generell und die Rolle des Quarterbacks.....	4
2.1 Spielkonzept.....	4
2.2 Spielfeld.....	5
2.3 Wichtigste Methoden um zu punkten.....	6
2.4 Passspielzug und Position des Quarterbacks.....	6
3 Das Quarterback-Rating.....	8
4 Hinführung zu den Problemstellungen.....	8
4.1 Veränderungen des Quarterback-Ratings und seiner Variablen im Laufe der Zeit.....	8
4.2 Beziehungen zwischen den Variablen des Quarterback Ratings.....	9
5 Die Daten.....	10
6 Strukturbrüche in Zeitreihen.....	13
6.1 Strukturbruchtest.....	13
6.2 Ergebnisse der Strukturbruchtests.....	14
7 Graphentheorie.....	20
7.1 Analyse von Zusammenhängen durch den PC-Algorithmus.....	20
7.2 Kausale Beziehungen zwischen den vier Quarterback-Rating-Variablen.....	22
8 Fazit und Diskussion der Methoden.....	29
9 Weitere Statistiken.....	30
Quellen- und Literaturverzeichnis.....	32

1 Einführung

Durchsucht man das Internet nach Online-PC-Spielen zum American Football, findet man unter anderem auf <http://www.shockwave.com/> das Browser-Spiel „Two Minute Football 3D 2010“¹. Man spielt dabei die angreifende Mannschaft in einem Football-Spiel und hat dabei zwei Minuten Zeit, um eine bestimmte Anzahl an Punkten gegen die gegnerische Verteidigung zu erzielen. Am Spielende erhält man schließlich sein Ergebnis: „You lost“ oder „You won“, die erreichte Anzahl an Punkten und einen Wert für ein sogenanntes Quarterback-Rating. Dafür, dass es sich bei Letzterem lediglich um einen statistischen Wert zur Beschreibung der Leistung eines einzigen Spielers aus einer elfköpfigen Mannschaft handelt – eben dem Quarterback – scheint diesem Wert doch eine hohe Bedeutung zugeschrieben zu werden, erst recht, wenn er selbst in einem wohl ohne zu großen Aufwand programmierten Browser-Spiel noch berechnet und präsentiert wird. Auch während eines realen Football-Spiels der National Football League, der Profiligen des American Footballs in den USA, kommt es bei dessen Fernsehübertragung oder im Anschluss daran in diversen amerikanischen Sportsendungen häufig unter den Kommentatoren zu Diskussionen über diesen Wert und es stellt sich heraus, dass er oft nicht verstanden oder als nicht sehr sinnvoll angesehen wird. Diese Arbeit hat das Ziel, das Quarterback-Rating zu verstehen, indem mithilfe zweier Methoden, der Strukturbruchanalyse und der Graphentheorie, untersucht wird, in welcher Beziehung die einzelnen Variablen, die dieses Rating ausmachen, zueinander stehen und wie sie sich über die Zeit verändert haben. Diese Arbeit ist dabei auf das Paper „A Statistical Analysis of NFL Quarterback Rating Variables“ aus dem „Journal of Quantitative Analysis in Sports“ von Derek Stimmel gestützt². Sämtliche Methoden und Ergebnisse gehen – sofern nicht anders gekennzeichnet – aus diesem Paper hervor.

1 Shockwave: „Two Minute Football 3D 2010“

2 D. Stimmel: "A Statistical Analysis of NFL Quarterback Rating Variables"

2 Spielregeln generell und die Rolle des Quarterbacks^{3 4 5 6 7}

2.1 Spielkonzept

Beim American Football ist es Ziel zweier Mannschaften aus je elf Spielern, den Spielball in Form eines Rotationsellipsoids entweder in die gegnerische Endzone zu bringen oder ein „Field Goal“ zu erzielen, um dabei möglichst viele Punkte zu erreichen. Die angreifende Mannschaft – die „Offense“ – hat hierbei die Möglichkeit durch Werfen oder Laufen Raum zu gewinnen, um der Endzone näher zu kommen, während die verteidigende Mannschaft – die „Defense“ – versucht, diese hiervon abzubringen und selbst in Ballbesitz zu kommen. Schafft es die Offense in vier Versuchen einen Raumgewinn von zehn sogenannten „Yards“ zu erreichen, bekommt sie erneut vier Versuche um weiteren Raumgewinn zu erlangen. Gelingt ihr dies nicht, so wechselt das Angriffsrecht. Gewinner ist schließlich die Mannschaft, die am Ende die meisten Punkte erzielt hat.

3 Wikipedia: „Quarterback“

4 Wikipedia: „American Football“

5 Wikipedia: „Offensive Line“

6 Osnaball.de: „American Football – Offense“

7 talkamericanfootball.co.uk: „The Rules of American Football“

2.2 Spielfeld

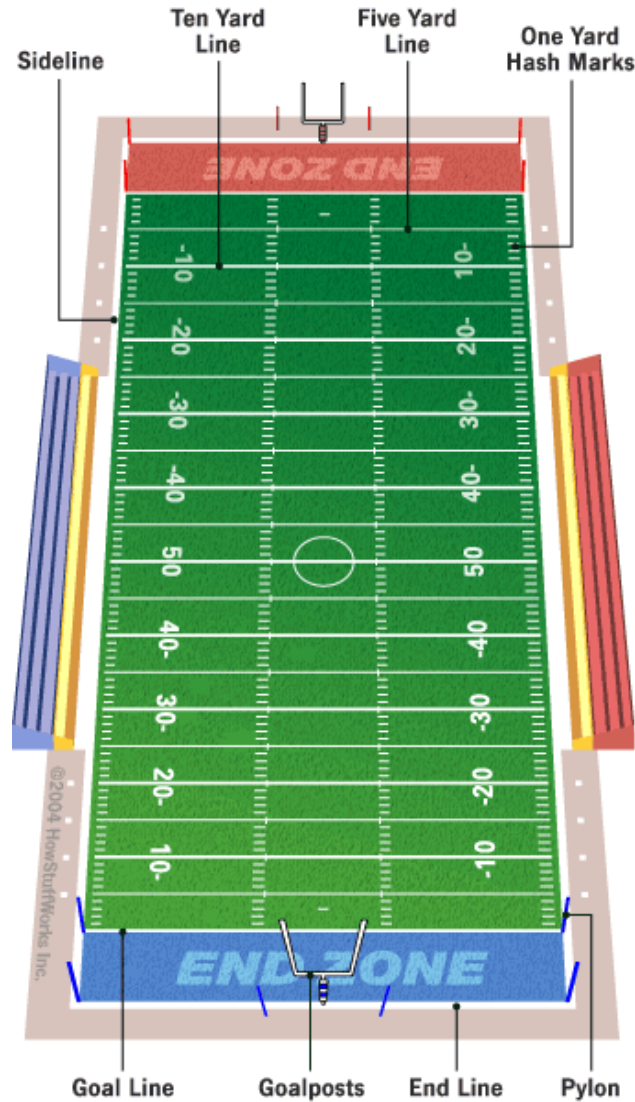


Abbildung 1: American Football Feld (von: K. Bonsor: „How American Football Works“)

Das Spielfeld (Abbildung 1) ist, ähnlich wie ein Fußballfeld, circa 110 Meter lang und circa 50 Meter breit und ist in 120 Yards bzw. 12 gleichgroße Abschnitte zu je 10 Yards eingeteilt, wobei der jeweils letzte Abschnitt am Ende des Spielfeldes die sogenannte „Endzone“ darstellt. In dieser befinden sich stimmgabelförmige sogenannte „Goal Posts“, welche die Tore darstellen. In Längsrichtung besitzt das Spielfeld zwei parallele Markierungen, die „Hash Marks“. Sollte ein Versuch außerhalb dieser Linien enden, so beginnt der nächste Versuch an der nächst gelegenen Hash Mark.

2.3 Wichtigste Methoden um zu punkten

Für die Bewertung der Qualität des Quarterbacks ist die Methode, die am meisten Punkte einbringt, die entscheidende, der „Touchdown“. Er wird mit sechs Punkten gewertet und wird erlangt, indem der Ball entweder durch einen Laufspielzug in die gegnerische Endzone getragen oder durch einen Passspielzug in dieser gefangen wird.

Drei Punkte können dadurch erzielt werden, dass der Ball durch die gegnerischen Goal Posts gekickt wird. Dies wird in der Regel dann versucht, wenn es als aussichtslos erscheint, noch einen Touchdown im aktuellen Angriff zu erzielen.

2.4 Passspielzug und die Position des Quarterbacks

Zu Beginn eines Versuchs wird der Spielball in der Offense, üblicherweise vom „Center“ hinter sich an den Quarterback übergeben. Dieser hat nun neben den Möglichkeiten, selbst einen Laufspielzug zu beginnen („Scrambling“) oder einem weiteren Spieler den Ball für einen Laufspielzug zu überreichen, die eines Passspielzugs. Er versucht hierbei den sogenannten „Receivern“, welche eine vorher festgelegte Route laufen, einen Pass zuzuwerfen.

Der versuchte Pass kann nun vom Receiver gefangen werden („Catch“ bzw. „Completion“, also vollständiger Pass), den Boden berühren oder ins Aus („Out of Bounds“) geworfen werden – man spricht in beiden Fällen von „Incompletion“ – oder von der Defense abgefangen werden („Interception“). Bei einer Incompletion ist der Versuch beendet, wobei der nächste Versuch von der Startposition des vorherigen beginnt. Bei einem Catch kann der Spieler so weit laufen, wie es ihm möglich ist, bis er von einem gegnerischen Spieler zu Boden gebracht („getackelt“) wird. Der Versuch ist wieder beendet. Der nächste Versuch beginnt jedoch an der Position, wo der Spieler getackelt wurde. Die imaginäre Querlinie in dieser Stelle wird als „Line of Scrimmage“ bezeichnet. Es kommt auch vor, dass der Spieler in Ballbesitz den Ball verliert, bevor er getackelt wird. In diesem Fall gibt es keine Spielunterbrechung und der Ball kann von jedem Spieler – auch der Defense – aufgenommen werden. Man spricht bei dieser Art des Ballverlusts von einem „Fumble“. Diese Schilderung des Spielverlaufs deutet an, welche wichtige Rolle der Quarterback in dieser Sportart spielt. Er ist die Schlüsselfigur für die Offense und muss die vom Trainer geplanten Spielzüge umsetzen und seinen Mitspielern vermitteln. Zusätzlich muss er versuchen, schnell auf die Handlungen der Defense zu reagieren und sich daran anpassen, um sich für die bestmögliche

Spieloption zu entscheiden.

Abbildung 2 zeigt die Aufstellungen der Offense (unten) und der Defense (oben).



Abbildung 2: Aufstellungen (von: K. Bonsor: „How American Football Works“); bearbeitet

Vor gegnerischen Tacklern wird der Quarterback von der „Offensive Line“ geschützt, bestehend aus „Left Tackle“, „Left Guard“, Center, „Right Guard“ und „Right Tackle“. Der „Tight End“ steht neben der Offensive Line und ist bei einem Passspielzug neben den „Wide Receivern“ ebenfalls ein möglicher Fänger für den vom Quarterback geworfenen Ball, welcher ebenfalls seine festgelegte Passroute läuft. Im Falle des Laufspiels dient er dem Quarterback ebenfalls als Blocker. „Halfback“ und „Fullback“ fungieren bei einem Passspiel ebenfalls als Blocker, andernfalls sind sie als mögliche Läufer gedacht, wenn ihnen der Quarterback den Ball für einen Lauf überlässt.

Es lässt sich erahnen, dass es durch so viele unterschiedliche Positionen eine Vielzahl an Spielzugmöglichkeiten und Spielsituationen gibt, weshalb es eine Vielzahl von Theorien zur Taktik gibt und man häufig beim Football vom „Rasenschach“ spricht.

3 Das Quarterback-Rating

Beim Quarterback-Rating handelt es sich um einen Wert, der eine Bewertung der Passfähigkeiten des Quarterbacks darstellen soll und somit diese mit den Passfähigkeiten anderer Quarterbacks vergleichbar machen soll. Die Quarterback-Rating-Formel besteht aus einer Linearkombination von vier Variablen:

durchschnittliche Anzahl der erreichten Yards pro Passversuch (DY), Prozentsatz der vollständigen Pässe pro Passversuch (VP), Prozentsatz der erzielten Touchdowns pro Passversuch (TD) und Prozentsatz der geworfenen Interceptions pro Passversuch (IC).

Damit wird das Quarterback-Rating (QBR) folgendermaßen berechnet:

$$QBR = 100 \cdot \left(\underbrace{\frac{0.25(DY - 3)}{6}}_{\text{Yards-Komponente}} + \underbrace{\frac{0.05 \cdot (VP \cdot 100 - 30)}{6}}_{\text{VP-Komponente}} + \underbrace{\frac{20 \cdot TD}{6}}_{\text{TD-Komponente}} + \underbrace{\frac{(2.375 - 25 \cdot \text{Int})}{6}}_{\text{Int-Komponente}} \right)$$

Zusätzlich müssen einige Bedingungen erfüllt sein:

Keine der vier Komponenten darf negativ werden, womit beispielsweise der Prozentsatz der vollständigen Pässe nicht weniger als 30% betragen darf. Ist dies trotzdem der Fall, so wird der Wert auf 30% gesetzt (so dass sich als VP-Komponente 0 ergibt ($0.05 \cdot (0.30 \cdot 100 - 30) = 0$)).

Weiterhin darf keiner der Zähler der vier Brüche größer als 2.375 sein, wodurch einem Spieler, der ein größeres VP als 77.5% hat, trotzdem nur exakt 77.5% angerechnet werden ($0.05 \cdot (0.775 \cdot 100 - 30) = 2.375$).

Somit nimmt das Quarterback Rating einen Wert zwischen 0 und 158.3 an

$$\left(100 \cdot 4 \cdot \frac{2.375}{6} = 158.\bar{3} \right) .$$

4 Hinführung zu den Problemstellungen

4.1 Veränderungen des Quarterback-Ratings und seiner Variablen im Laufe der Zeit

Um eine erste Problemstellung zu ermitteln, wendet man sich vorerst der grundsätzlichen Kritik am Quarterback-Rating zu, wobei diese in drei Bereiche eingeteilt wird:

Einen ersten Kritikpunkt stellt die schwere Verständlichkeit der Formel dar, was unmittelbar zur

Folge hat, dass auch ihre Ergebnisse nur schwer zu deuten sind. Betrachtet man die Formel, ist es schwierig zu sehen, wie sich die Kombination der Variablen auf diese Weise begründet. Es besitzen beispielsweise die durchschnittlich erreichten Yards pro Passversuch und die vollständigen Pässe pro Versuch eine unterschiedliche Skalierung ($DY \in \mathbb{R}_0^+$; $VP \in [0;1]$), weshalb beide Variablen transformiert werden müssen, um in der Formel kombiniert werden zu können.

Ein weiteres Problem zieht die Skalierung mit sich. Anstatt eine natürliche Skala von beispielsweise 0 bis 100 heranzuziehen, bewegt sie sich zwischen 0 und 158.3..

Als zweiter Kritikpunkt lässt sich heranziehen, dass das Quarterback-Rating ein unangemessenes Maß für die Leistung eines Quarterbacks liefert, da einige Faktoren vernachlässigt werden. Erzielt der Quarterback ein erfolgreiches Ergebnis durch Scrambling, wird dies nicht berücksichtigt, erreichen die Receiver hingegen nach einem Catch noch einige Yards, so verbessert dies das Rating, ohne, dass dies zwangsläufig des Quarterbacks Verdienst war.

Man gibt sich hier mit diesen jeweiligen beiden kurzen Begründungen zufrieden und richtet nun das Augenmerk auf folgenden bedeutenden Kritikpunkt.

Das Quarterback Rating macht es schwer möglich, die Quarterbacks aus verschiedenen Zeiträumen miteinander zu vergleichen. Ein Nachteil der Formel ist dabei, dass sie auf die durchschnittliche Leistung der Quarterbacks im Jahr 1970 gestützt ist, was aber nicht allein dafür verantwortlich ist, dass die Ratings in den verschiedenen Zeiträumen schlecht vergleichbar sind. Jedoch ist das Rating über die Zeit hinweg angestiegen und es wurden immer wieder einige Regeländerungen vorgenommen, welche das Passspiel begünstigen sollten. Dies und manch weitere Gründe führen in den Zeitreihen zu sogenannten Strukturbrüchen. Vergleiche sind hier schließlich nicht völlig unmöglich, aber sie müssen für die verschiedenen Abschnitte angepasst werden. In Kapitel 6 werden durch Strukturbruchtests die einzelnen Strukturbrüche aufdeckt und im Detail nach den Gründen dafür gesucht.

4.2 Beziehungen zwischen den Variablen des Quarterback Ratings

Eine weitere interessante Fragestellung beim Quarterback Rating ist, welche Beziehung zwischen den vier Variablen in der Formel zueinander herrscht. Die Graphentheorie ist dabei sowohl ein Weg, um den Sachverhalt genauer zu erklären, als auch ein Mittel um diesen zu untersuchen. Somit kann ein einzelner Passversuch folgendermaßen dargestellt werden (Abbildung 3).

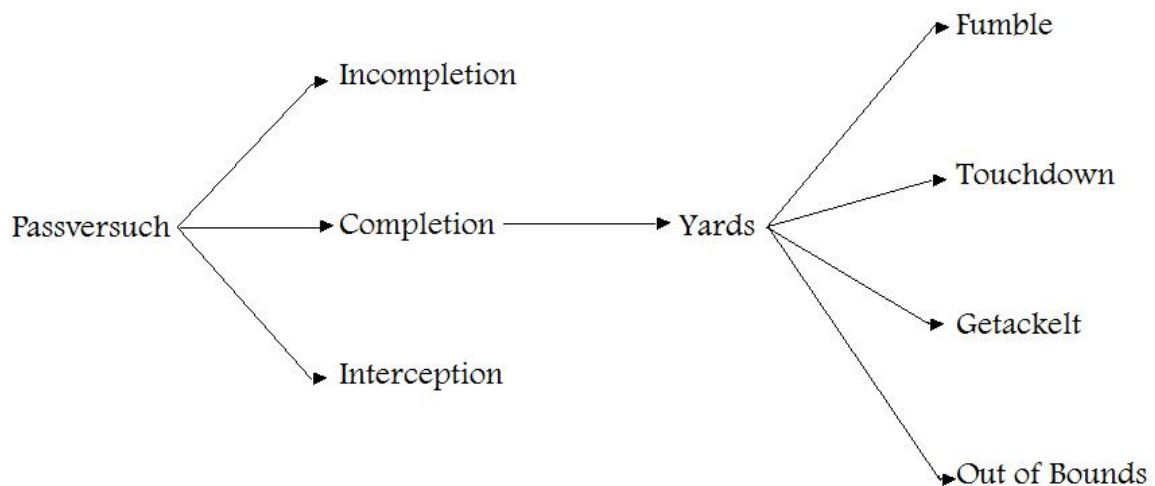


Abbildung 3: kausale Beziehungen für einen Passversuch

„ $A \rightarrow B$ “ habe hierbei die Bedeutung „A 'verursacht' B“, womit aber hier nicht gemeint sei, dass A hinreichend ist, um daraus B zu folgern – man würde hierbei von kausaler Suffizienz sprechen – sondern dass im statistischen Sinne das Ereignis B vom Ereignis A abhängt, während der bedingte Zusammenhang der Variablen in diesem System gegeben ist. Abbildung 3 zeigt, dass das Quarterback Rating beinahe jede Möglichkeit erfasst, wie ein Passversuch ausgehen könnte. Wie bereits bekannt, ist jede Variable relativ zum Passversuch definiert, was sich nun aber auf die kausalen Beziehungen unter den Variablen auswirkt. Das folgende Beispiel soll dies klarlegen: Bei einem einzelnen Versuch ist es offensichtlich, dass ein vollständiger Pass einen Touchdown durch Pass verursacht, denn der vollständige Pass muss vor dem Touchdown durch Pass erfolgen. Es stellt sich jedoch die Frage, ob in mehreren Versuchen, das heißt während eines Spieles oder gar einer ganzen Saison, der Prozentsatz der vollständigen Pässe den Prozentsatz der Touchdowns verursacht. Dies ist der zweite Aspekt, der unter die Lupe genommen wird, er wird in Kapitel 7 behandelt.

5 Die Daten ⁸

Für die Analyse werden Daten aus den Saisons der NFL (National Football League) von der offiziellen NFL-Website benutzt. Dabei wird die Anzahl der Quarterbacks so eingeschränkt, dass nur die, die ein Minimum von 14 versuchten Pässen pro Spiel erreicht haben, untersucht werden. Damit wird vermieden, dass Spieler anderer Positionen miteinbezogen werden, da es beispielsweise auch vorkommt, dass diese einen Pass werfen.

⁸ Wikipedia: „National Football League“

Verfügbar sind Daten aus der NFL schon ab 1937, im Paper von Derek Stimel werden jedoch Stichproben erst ab dem Jahr 1960 verwendet, da vor diesem Jahr nie mehr als zwölf Quarterbacks die untere Grenze von 14 Passversuchen pro Spiel übertroffen hatten. Nach 1960 schafften dies zumindest jeweils 20 Quarterbacks – mit Ausnahme des Jahres 1973, dort waren es lediglich 17 – ab 1980 sogar rund 30. Ein Grund dafür mag sein, dass 1960 auch dem Startjahr der AFL (American Football League) entspricht, welche zur NFL hinzugefügt wurde, nachdem aus AFL die AFC und aus der ursprünglichen NFL die NFC entstanden war, welche schließlich kombiniert wurden. Dadurch kamen zur Stichprobe einige weitere Teams und damit Quarterbacks hinzu. Für die in den nächsten Kapiteln beschriebenen Analysen wurde für die vier Variablen und das Quarterback Rating jeweils der jährliche Mittelwert (arithmetisches Mittel), sowie die dazugehörige Standardabweichung berechnet.

Im folgenden sind hierfür die Zeitreihen zu abgebildet (Abbildungen 4 – 6)

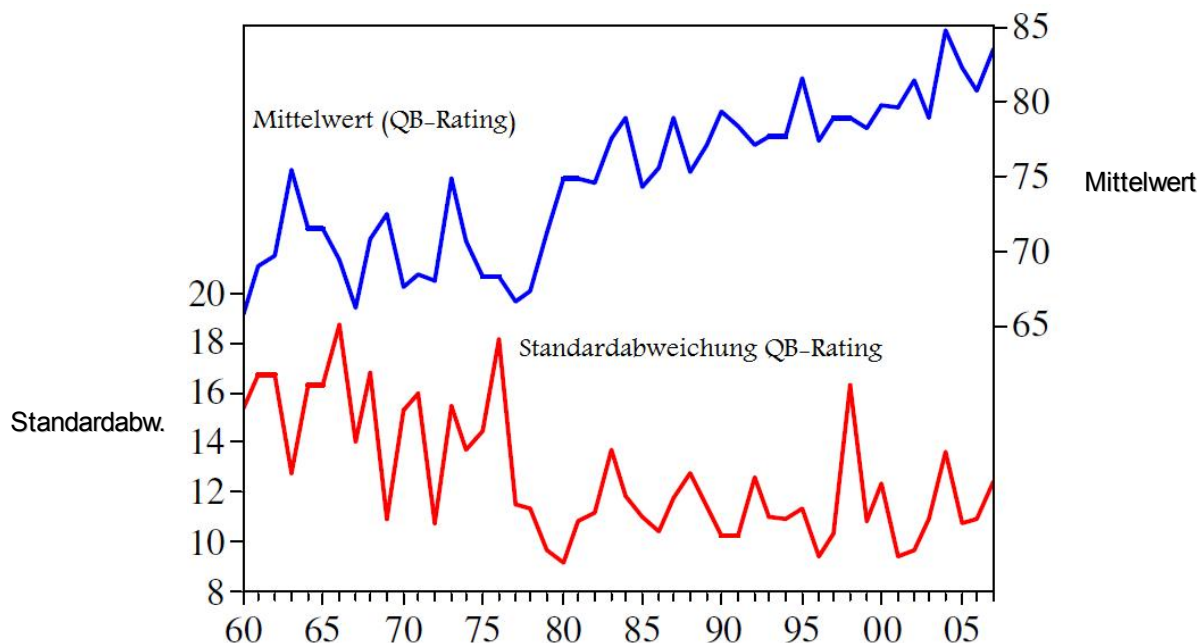


Abbildung 4: **Mittelwert** und Standardabweichung des Quarterback-Ratings von 1960 bis 2007; von: D. Stimel: "A Statistical Analysis of NFL Quarterback Rating Variables"); bearbeitet

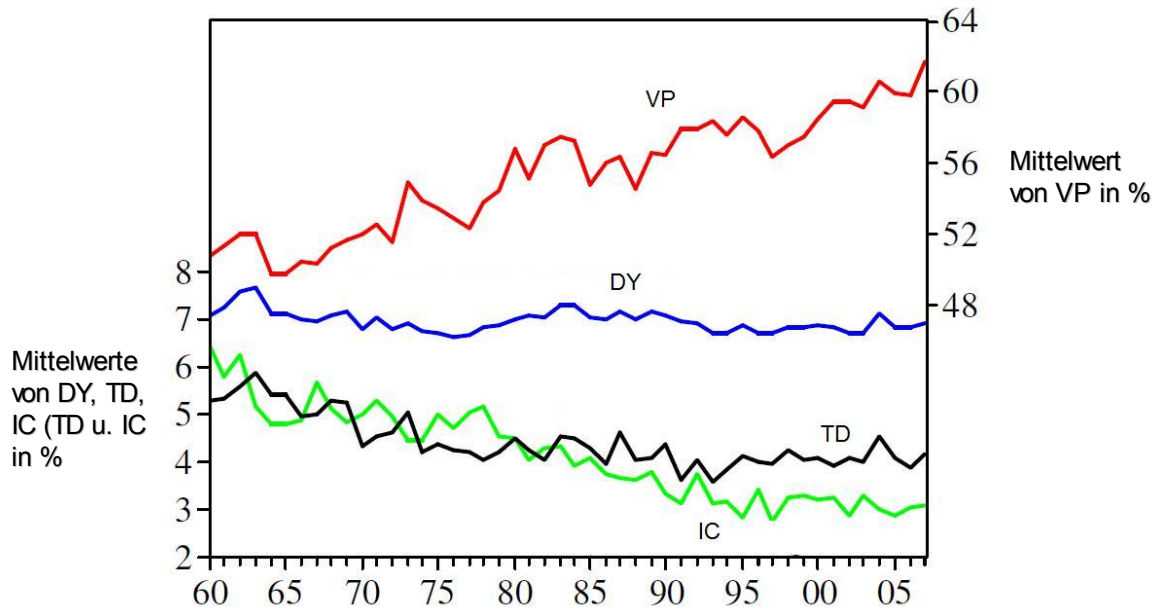


Abbildung 5: **Mittelwerte** der Quarterback-Rating-Variablen von 1960 bis 2007;
 von: D. Stimel: "A Statistical Analysis of NFL Quarterback Rating Variables"); bearbeitet

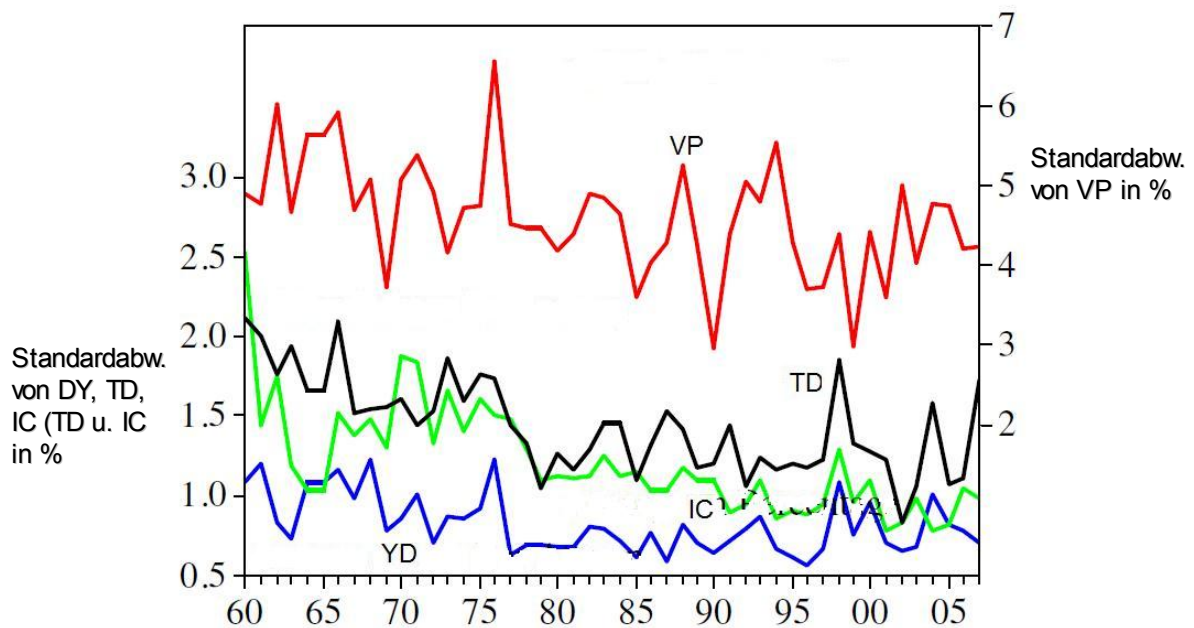


Abbildung 6: **Standardabweichungen** der Quarterback-Rating-Variablen von 1960 bis 2007;
 von: D. Stimel: "A Statistical Analysis of NFL Quarterback Rating Variables"); bearbeitet

In Abbildung 4 fällt auf, dass der Mittelwert des Quarterback-Ratings gestiegen und die Standardabweichung gesunken ist. Ein bedeutender Sprung scheint dabei in den späten 70ern in beiden Zeitreihen zu erkennen zu sein. Wenn man die Zeitreihen für die einzelnen Variablen ins Auge fasst, ist ein Fall der Standardabweichungen ebenfalls in den späten 70ern festzustellen (Abbildung 6). Betrachtet man Abbildung 5, ist kaum ein Bruch im Mittelwert des Anteils der vollständigen Pässe zu entdecken, er scheint konstant seit 1960 gestiegen zu sein. Die durchschnittlich erreichten Yards haben sich nur wenig über die Zeit hinweg verändert. Der Touchdownprozentsatz ist gesunken, wie auch der Anteil der Interceptions, jedoch sank dieser in den späten 70ern noch tiefer ab.

Im folgenden Kapitel wird nun nach Erklärungen für diese Zeitreihenverläufe gesucht.

6 Strukturbrüche in Zeitreihen

6.1 Strukturbruchtest⁹

Um die Veränderungen des Quarterback-Ratings über den Zeitraum von 1960 bis 2007 zu untersuchen, wird nun ein einfaches Regressionsmodell aufgestellt. Dabei werden Strukturbrüche mithilfe von Tests aufgedeckt werden, wobei man von Strukturbruch spricht, wenn sich die Regressionsparameter während einer Zeitreihe plötzlich ändern, das heißt nicht durchgehend konstant bleiben. Folgendes Modell diene hierbei als Regressionsmodell:

$$y_t = \alpha_j + \varepsilon_t$$

$$t = T_{j-1} + 1, \dots, T_j; \text{ mit } j = 1, \dots, m; \text{ wobei } T_0 = 0 \text{ und } T_{m+1} = T$$

y ist die zu untersuchende Zeitreihe, α die Intercept, ε die Störgröße, m die Anzahl der Brüche – damit gibt es $m+1$ Abschnitte, also $m+1$ verschiedene α – und T die Anzahl der Beobachtungen. Das Modell besteht also mit der Intercept nur aus einem Regressionsparameter, welcher sich über die Zeitreihe hinweg verändern kann. Es wird damit auf Verschiebungen im Mittelwert der Zeitreihen getestet, das heißt im Mittelwert aus den Mittelwerten und aus den Standardabweichungen.

Es besteht nun das Problem, dass sowohl die Parameter als auch die Daten der Strukturbrüche

⁹ Wikipedia: „Strukturbruch“

(T_1, \dots, T_m) zu schätzen sind. Als Lösung wird die Minimierung der Residuenquadratsummen angeboten. Derek Stimel stützt sich hierbei auf einen von Bai und Perron^{10 11} entwickelten Algorithmus, der auf dynamischer Programmierung beruht.

Damit erhält man die geschätzten Strukturbruchdaten $(\hat{T}_1, \dots, \hat{T}_m)$ als folgende Minimierung:

$$(\hat{T}_1, \dots, \hat{T}_m) = \arg \min_{T_1, \dots, T_m} S_T(T_1, \dots, T_m) ,$$

mit $S_T(T_1, \dots, T_m)$ als die Residuenquadratsummen.

Damit minimieren die Strukturbruchdaten die Residuenquadratsummen.

Um schließlich die Anzahl der Brüche herauszufinden, bieten Bai und Perron mehrere Methoden an.^{12 13} Stimel entscheidet sich hier für die sequentielle Methode, welche bei Bai und Perron¹⁴ zu finden ist. Dabei wird ein sup-F-Test verwendet, ein Test, der sich für unbekannte Bruchpunkte eignet¹⁵, bei welchem

H_0 : " es gibt l Strukturbrüche " gegen H_1 : " es gibt $l+1$ Strukturbrüche " getestet wird. Die Teststatistik $\sup F(l+1|l)$ ist die kleinste obere Schranke der F-Statistiken für die Hypothese H_1 gegen die Hypothese H_0 , während der zusätzliche Bruch über alle möglichen Brüche variiert wird.¹⁶ Bai und Perron geben kritische Werte für verschiedene trimmende Anteile und eine größtmögliche Anzahl von Strukturbrüchen vor.^{17 18} Die trimmenden Anteile ergeben multipliziert mit der Anzahl der gesamten Beobachtungen die Mindestlänge eines Abschnitts. Nun können also nacheinander – beginnend bei $l = 0$ – jeweils l Brüche gegen $l+1$ Brüche getestet werden, solange, bis die kleinere Anzahl der Brüche beibehalten wird.

6.2 Ergebnisse der Strukturbruchtests¹⁹

Bei der Umsetzung der Tests wurde ein Maximum von fünf Strukturbrüchen, d. h. sechs mögliche Abschnitte und ein trimmender Anteil von 0.15 angenommen. Damit liegt mit 48 Beobachtungen

10 Bai, J. and P. Perron (1998) "Estimating and Testing for Multiple Structural Changes in a Linear Model,"

11 Bai, J. and P. Perron (2003a) "Computation and Analysis of Multiple Structural *Change Models*"

12 Bai, J. and P. Perron (1998) "Estimating and Testing for Multiple Structural Changes in a Linear Model,"

13 Bai, J. and P. Perron (2004) "Multiple Structural Change Models: A Simulation Analysis"

14 Bai, J. and P. Perron (1998) "Estimating and Testing for Multiple Structural Changes in a Linear Model,"

15 A. Zeileis, F. Leisch, K. Hornik, C. Kleiber, „Testen auf Strukturbruch – Implementierung in R und Anwendung“

16 Prodan, R. (2008) "Potential Pitfalls in Determining Multiple Structural Change Models with an Application to Purchasing Power Parity"

17 Bai, J. and P. Perron (1998) "Estimating and Testing for Multiple Structural Changes in a Linear Model,"

18 Bai, J. and P. Perron (2003b) "Critical Values for Multiple Structural Change Tests"

19 Wikipedia: „Salary Cap“

(1960 – 2007) die minimale Abschnittslänge bei (auf ganze Jahre gerundeten) 7 Jahren. Für die Standardfehler in diesem Regressionsmodell wird eine Methode nach Andrews²⁰ verwendet. Für den sup-F-Test wird das Signifikanzniveau auf 1% gesetzt. Im folgenden werden in Abbildung 7 und 8 die durch diesen Test geschätzten Strukturbruchdaten dargestellt, wie auch die dazu berechneten 90%-Konfidenzintervalle und die geschätzten Mittelwerte für die Zeitreihen der Mittelwerte sowie für die der Standardabweichungen für die jeweiligen Abschnitte (welche in diesem Modell den Intercepts α entsprechen).

Quarterback Rating						
Strukturbruchdaten	1979	1994	Abschnitte	1960-1979	1980-1994	1995-2007
90%-KI	[1978,1981]	[1989,1996]	α pro Abschnitt	69.64	76.79	80.45
			Standardabw.	0.51	0.59	0.63
DY						
Strukturbruchdaten	-		Abschnitte	1960-2007		
90%-KI			α pro Abschnitt	6.98		
			Standardabw.	0.05		
VP						
Strukturbruchdaten	1979	1999	Abschnitte	1960-1979	1980-1999	2000-2007
90%-KI	[1978,1984]	[1996,2002]	α pro Abschnitt	52.06	56.89	59.84
			Standardabw.	0.28	0.28	0.45
IC						
Strukturbruchdaten	1983	1992	Abschnitte	1960-1983	1984-1992	1993-2007
90%-KI	[1982,1989]	[1991,1997]	α pro Abschnitt	4.99	3.68	3.10
			Standardabw.	0.09	0.15	0.12
TD						
Strukturbruchdaten	1969		Abschnitte	1960-1969	1970-2007	
90%-KI	[1968,1971]		α pro Abschnitt	5.35	4.19	
			Standardabw.	0.09	0.05	

Abbildung 7: Strukturbruchtests für die Mittelwerte (sup-F-Test mit 1% Signifikanzniveau)

Als erstes beschäftigt man sich mit den Ergebnissen für das Quarterback Rating. Bei den Mittelwerten (Abbildung 7) wurde 1979 und 1994 jeweils ein Strukturbruch gefunden, welche nach Stimel beide auf Regeländerungen beruhen. 1977 wurde eine Regel eingeführt, die es den Verteidigern verbot, mehr als einen Körperkontakt mit den Receivern einzugehen. Ein Jahr später wurde es den Verteidigern erlaubt, Kontakt mit den Receivern innerhalb von fünf Yards vor der

²⁰ Andrews, D.W.K. (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation"

Line of Scrimmage zu haben, jedoch war der Kontakt nach den fünf Yards eingeschränkt. Diese beiden Regeländerungen öffneten den Receivern mehr Möglichkeiten, um ihre Manöver durchzuführen und sollte somit auch dazu geführt haben, die Möglichkeiten der Quarterbacks für einen Passversuch zu steigern.

Mitte der 90er Jahre wurden diese Regeln schließlich noch einmal verstärkt. Beispielsweise wurde 1996 die Fünf-Yards-Kontakt-Regel noch mehr verstärkt (Ein Stoß der Verteidiger mit Helm bzw. Kopf gegen einen Receiver wurde als Faul gewertet). Um direkt auf das Jahr des Strukturbruchs einzugehen, war es möglicherweise der 1994 eingeführte Salary Cap, der einige Auswirkungen auf die Leistung der Spieler mit sich brachte. Es handelt sich beim Salary Cap um eine Gehaltsobergrenze, die einerseits zur Kontrolle der Gehälter, andererseits zur Gewährleistung finanzieller Ausgeglichenheit der Klubs eingeführt wurde. Er könnte dazu geführt haben, dass wirtschaftsschwache Mannschaften ihr Talent aufzustocken hatten, was zu einer höheren Chancengleichheit geführt haben dürfte und somit auch zur Erhöhung des Quarterback-Ratings.

Wirft man nun einen Blick auf die vier Quarterback-Rating-Variablen, fällt auf, dass der Prozentsatz der vollständigen Pässe und der der Interceptions weitgehend übereinstimmende Strukturbrüche aufweisen. Für den Bruch aus dem Jahr 1979 und dem der Interceptions, vier Jahre später, mag sich sicherlich dieselbe Begründung, wie für den Bruch des Quarterback-Ratings im Jahr 1979 heranziehen lassen (Begünstigung der Receiver). Auch die Ursache für den zweiten Bruch des Quarterback-Ratings im Jahr 1994 dürfte ebenfalls die Ursache für den Bruch in den Interceptions zwei Jahre vorher und den der vollständigen Pässe fünf Jahre nachher sein (weitere Begünstigung der Receiver und Salary Cap).

Für die durchschnittlich erreichten Yards zeigt der Test keinerlei Strukturbrüche.

Der Strukturbruch im Prozentsatz der Touchdowns kann nur schwer durch eine Regeländerung gerechtfertigt werden. Die offizielle Verschmelzung von NFL und AFL zu der Zeit mag sicher nahe legen, darin eine Ursache zu finden, jedoch gibt es keinen ersichtlichen Hinweis, dass dies den Touchdownprozentsatz in einer Weise beeinflusst. Jedoch war 1970 ein bemerkbarer Abfall im Mittelwert der grundsätzlich geworfenen Touchdowns in einer Saison festzustellen, was möglicherweise zu diesem Strukturbruch 1969 geführt haben mag.

Quarterback Rating						
Strukturbruchdaten	1976	Abschnitte	1960-1976	1977-2007		
90%-KI	[1974,1978]	α pro Abschnitt	15.20	11.29		
		Standardabw.	0.43	0.32		
DY						
Strukturbruchdaten	1976	Abschnitte	1960-1976	1977-2007		
90%-KI	[1974,1982]	α pro Abschnitt	0.97	0.74		
		Standardabw.	0.03	0.03		
VP						
Strukturbruchdaten	1976	Abschnitte	1960-1976	1977-2007		
90%-KI	[1970,1983]	α pro Abschnitt	5.09	4.34		
		Standardabw.	0.15	0.11		
IC						
Strukturbruchdaten	1978	1990	Abschnitte	1960-1978	1979-1990	1991-2007
90%-KI	[1978,1983]	[1984,1991]	α pro Abschnitt	1.51	1.12	0.95
			Standardabw.	0.05	0.07	0.06
TD						
Strukturbruchdaten	1976	Abschnitte	1960-1976	1977-2007		
90%-KI	[1976,1978]	α pro Abschnitt	1.72	1.28		
		Standardabw.	0.05	0.04		

Abbildung 8: Strukturbruchtests für die Standardabweichungen (*sup-F-Test* mit 1% Signifikanzniveau)

Geht man nun auf die Ergebnisse der Strukturbruchtests für die Standardabweichungen ein (Abbildung 8), so fällt ein Ergebnis sofort ins Auge. Der Test liefert für die geworfenen Interceptions zwei Strukturbrüche (1978 und 1990), während er für alle anderen Variablen – das Quarterback-Rating miteinbezogen – nur einen Strukturbruch ausgibt und zwar jedes Mal für das Jahr 1976. Somit zeigen alle fünf Zeitreihen je einen Strukturbruch Mitte bis Ende der 70er Jahre, welche wohl wieder auf die Regeländerungen zurückzuführen führen sind, welche das Passspiel begünstigen sollten.

Aus den eben dargelegten Ergebnissen kann man nun einige Schlussfolgerungen ziehen. Die erste ist – wie in Kapitel 3 angesprochen – der dritte Kritikpunkt zur Formel des Quarterback-Ratings. Es ist nicht vertretbar Quarterbacks durch das Quarterback-Rating über verschiedene Zeiträume hinweg zu vergleichen ohne dabei Unterschiede in den verschiedenen Abschnitten der Zeitreihen zu berücksichtigen, da sie nicht unter den selben Bedingungen bzw. Spielregeln spielten.

Des Weiteren scheinen der durchschnittliche Prozentsatz der vollständigen Pässe und der

Prozentsatz der durchschnittlichen Interceptions Einflussfaktoren für das steigende Quarterback-Rating zu sein, da ihre Brüche sich nah an die Brüche des Quarterback Ratings angleichen.

Auch mögen die Testergebnisse den Rückschluss zulassen, dass die durchschnittliche Leistung der Quarterbacks bzw. das Passspiel sich verbessert hat, wohingegen sich die Schwankungsbreite zwischen sehr guter und durchschnittlicher Leistung verringert hat, wie die Ergebnisse für die Standardabweichungen andeuten. Das könnte wiederum zur Folge haben, dass der Wert eines sehr guten Quarterbacks im Vergleich zu einem durchschnittlichen Quarterback sich verringert haben könnte (bei alleiniger Betrachtung der Passqualitäten).

An letzter Stelle dieses Kapitels können die Zeitreihen erneut als Grafik, diesmal mit den entsprechenden Strukturbrüchen, betrachtet werden (Abbildungen 9 – 11).

Es ist zu sehen, dass bei den Variablen nicht so sehr Trends vorkommen wie Mittelwertverschiebungen. Das könnte bedeuten, dass es von außen kommende „Erschütterungen“ gibt, welche diese Umbrüche eher ausmachen, wie zum Beispiel Regeländerungen, als irgendeine innerhalb entstandene Resonanz der Zwischenbeziehungen dieser Variablen, wie zum Beispiel das Spielen von kurzen Pässen.

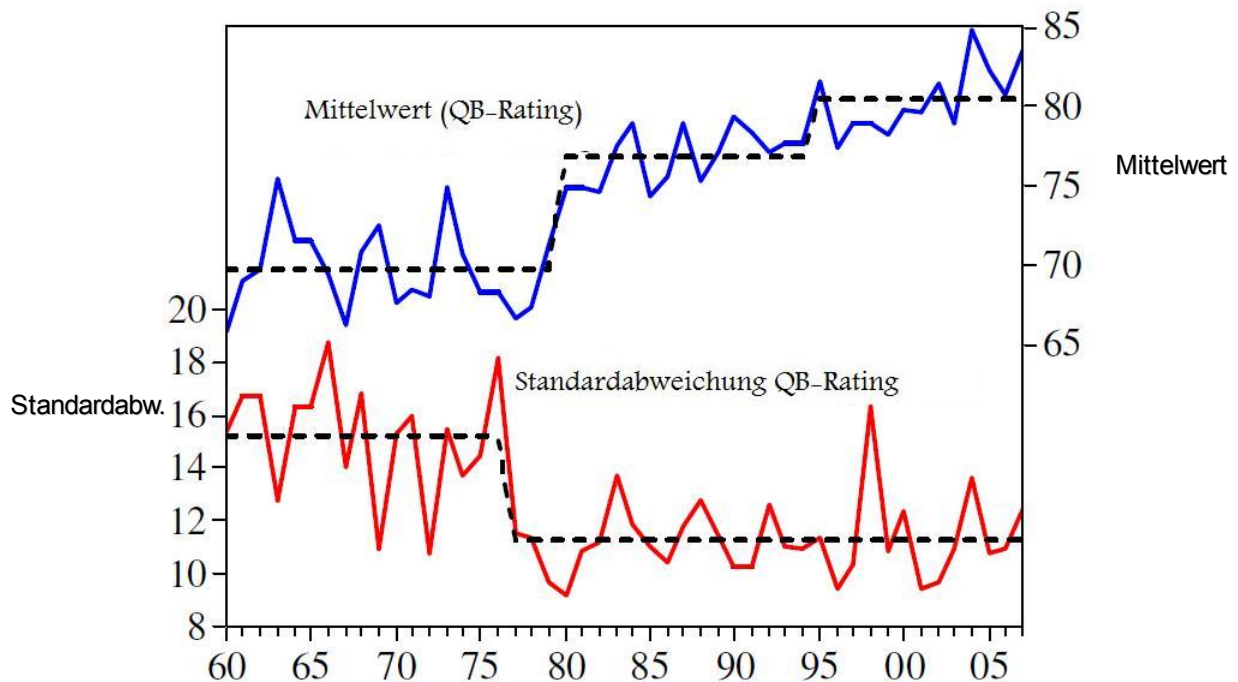


Abbildung 9: Mittelwert und Standardabweichung des Quarterback-Ratings (von: D. Stimel: "A Statistical Analysis of NFL Quarterback Rating Variables"); bearbeitet

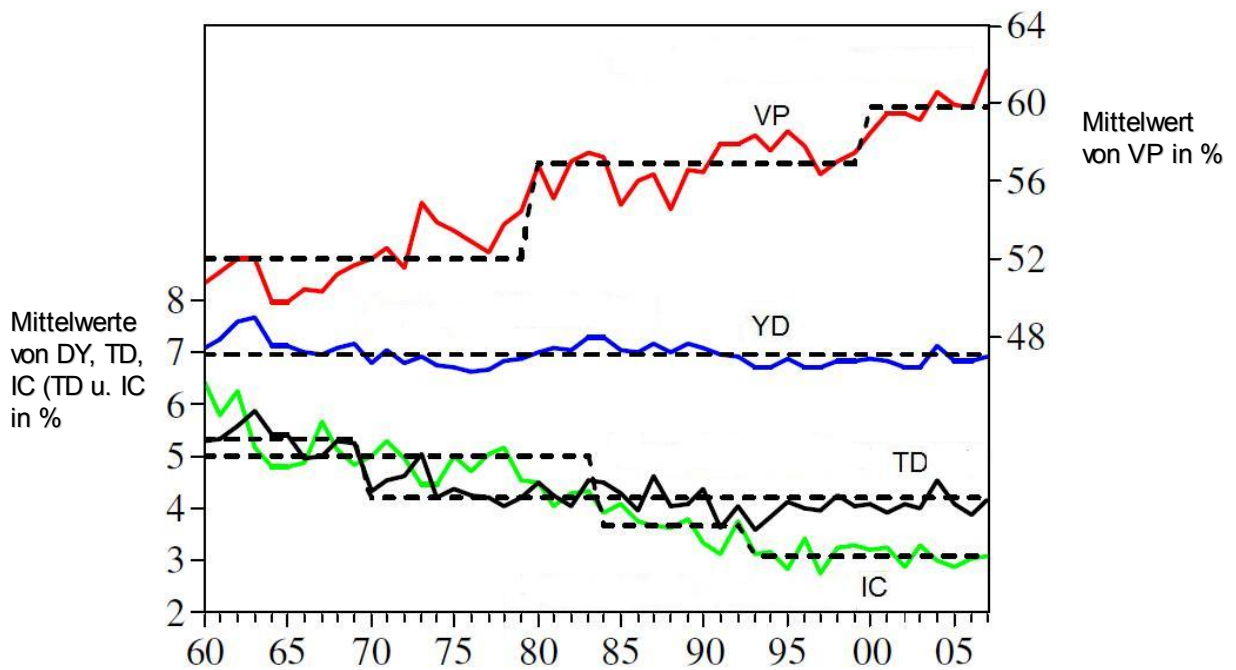


Abbildung 10: Mittelwerte der Quarterback-Rating-Variablen (von: D. Stimel: "A Statistical Analysis of NFL Quarterback Rating Variables"); bearbeitet

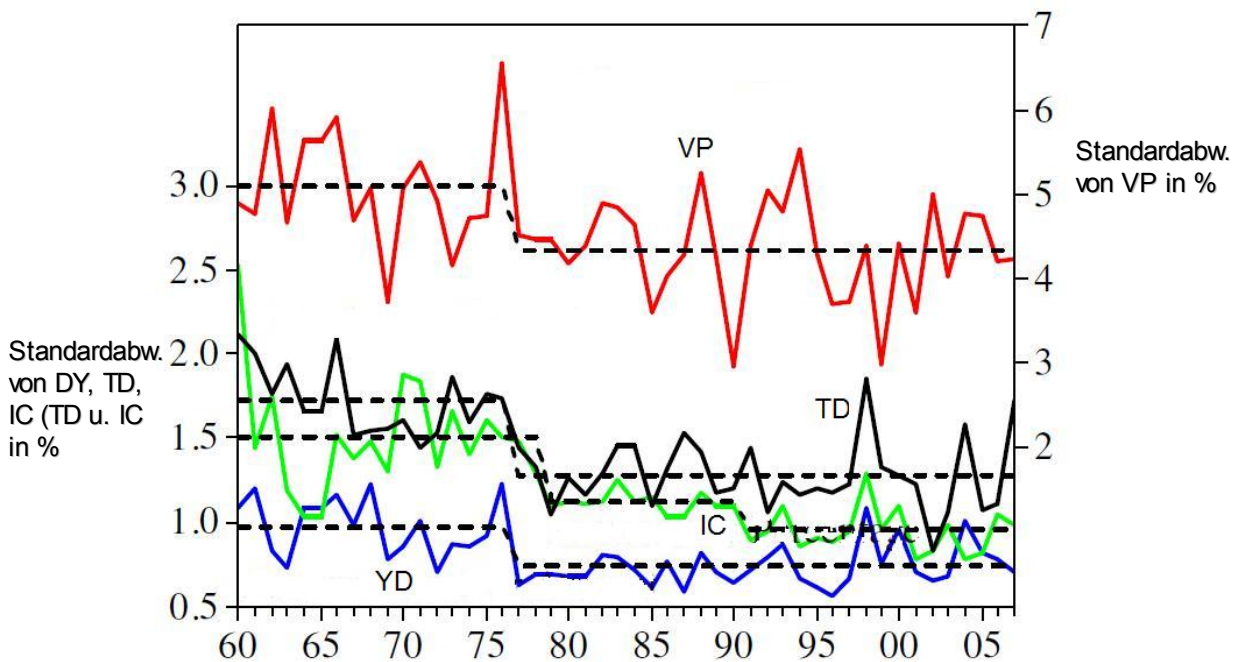


Abbildung 11: Standardabweichungen der Quarterback-Rating-Variablen (von: D. Stimel: "A Statistical Analysis of NFL Quarterback Rating Variables"); bearbeitet

7 Graphentheorie

7.1 Analyse von Zusammenhängen durch den PC-Algorithmus

Die Graphentheorie beschäftigt sich im Allgemeinen mit den Eigenschaften von Graphen, wobei ein Graph ganz allgemein definiert ist als eine Menge von Punkten zwischen denen Linien verlaufen.²¹ Damit bildet sie ein Konzept, mit welchem aus Informationen in Zusammenhängen zwischen Variablen eines Systems Abhängigkeiten, Nicht-Abhängigkeiten und die Richtungen der Abhängigkeiten extrahiert werden können. Solche Beziehungen können auch graphisch dargestellt werden.²² Zugrunde liegende Idee des Konzepts ist die „kausale Markovbedingung“, welche aussagt, dass zwei korrelierte Variablen, A und B, unkorreliert bedingt unter der Menge C sind, wenn gilt:

a) A und B haben eine gemeinsame Ursache bzw. Menge von Ursachen C, so dass gilt:

$$A \leftarrow C \rightarrow B,$$

oder

b) A und B haben eine zwischengeschaltete Ursache bzw. Menge von Ursachen C, so dass gilt:

$$A \rightarrow C \rightarrow B$$

Zusätzlich zur kausalen Markovbedingung gilt, dass A und B korreliert bedingt unter C sind, wenn gilt:

c) A und B sind bedingt unabhängig, gegeben einer Menge von Variablen, welche weder C noch ihre Nachfahren beinhaltet und es gilt:

$$A \rightarrow C \leftarrow B$$

Die Variable C wird dabei als „Unshielded Collider“ bezeichnet. Würde man A und B direkt in Verbindung setzen, so würde C zum „Shielded Collider“ und A und B wären nicht länger bedingt unabhängig.^{23 24}

Um in Daten solche Beziehungen zueinander zu finden, wurden spezielle Algorithmen entwickelt. Der am meisten allgemein gebräuchliche Algorithmus ist der sogenannte PC-Algorithmus. Spirtes et al.²⁵ haben hierzu ein Programm namens Tetrad III entwickelt, um diesen Algorithmus zu implementieren. Im Folgenden werden die einzelnen Anweisungsschritte an den Algorithmus beschrieben:

21 Wikipedia: „Graph (Graphentheorie)“

22 Hoover, K. D. (2003) „Some Causal Lessons From Macroeconomics,“

23 Hoover, K. D. (2003) „Some Causal Lessons From Macroeconomics,“

24 Demiralp, S. and K.D. Hoover (2003) „Searching for the Causal Structure of a Vector Autoregression“

25 Spirtes, P., R. Scheines, C. Meek, T. Richardson, C. Glymour, H. Hoijtink and A. Boomsma, (1996) *Tetrad 3: Tools for Causal Modeling*

- 1) Nimm an, dass **jede Variable** im System zu **jeder anderen Variablen** im System **verbunden** ist. Die Verbindung sei dabei **ungerichtet: $A - B$**
- 2) **Teste die unbedingte Korrelation** für **jedes Paar** von Variablen. Dabei handle es sich um einen **Fisher-z-Test**, bei dem die **Nullhypothese** aussage, dass die **zwei Variablen unabhängig** seien.
Entferne die Verbindung zwischen Variablen dort, wo der Test die **Nullhypothese beibehält**.
- 3) Für die **restlichen** Variablen, **teste die bedingte Korrelation** für **jedes Paar bedingt** unter einer **dritten Variablen** unter Verwendung des Fisher-z-Tests.
Entferne die Verbindungen dort, wo der Fisher-z-Test die **Nullhypothese beibehält**.
Wiederhole den Test für die übrigen verbundenen Paare **bedingt** unter **zwei** Variablen, **drei** Variablen und **so weiter**.
- 4) **Teste** die übrigen verbundenen Variablen, um **Unshielded Colliders** zu finden. Falls fündig, **richte die Verbindungen** folgendermaßen: $A \rightarrow C \leftarrow B$, wobei C ein Beispiel für einen Unshielded Collider darstellt
- 5) **Falls** es nun Mengen von Variablen der Beziehung $A \rightarrow B - C$ gibt, so richte die ungerichtete Verbindung auf folgende Weise aus: $A \rightarrow B \rightarrow C$
- 6) Falls es **zwei Variablen** mit der **Beziehung $A - B$** gibt und **gerichtete Verbindungen** von **A zu B** durch weitere Variablen existieren so **richte die Verbindung** so aus: $A \rightarrow B$

Die Schritte 1) – 3) ermitteln das sogenannte „Skelett“ des kausalen Graphen, welches die kausalen – aber noch ungerichteten – Verbindungen zwischen den Variablen darstellt. Die Schritte 4) – 6) sorgen schließlich für gerichtete Verbindungen. Ermittelte kausale Graphen werden als „getreu“ bezeichnet, wenn sie eine eindeutige Abbildung zur Verteilung der Daten darstellen. Jedoch ist der PC-Algorithmus nicht zwangsläufig fähig, für alle Verbindungen eine Lösung zu finden. Pearl ²⁶ verweist dabei auf das „Observational Equivalence Theorem“, welches besagt, dass eine durch einen getreuen Graphen dargestellte Verteilung durch einen weiteren Graphen beschrieben werden kann, welcher dasselbe Skelett und dieselben Unshielded Colliders besitzt. Existiert so eine Beobachtungsäquivalenz, so kann es sein, dass gerichtete Verbindungen in beide Richtungen zeigen können. Sprite et al.'s Programm Tetrad III lässt diese Verbindungen jedoch ungerichtet. Ebenfalls verzichtet Tetrad III auf die Implementierung von Schritt 6, in dem die ungerichtete Verbindung zweier Variablen ausgerichtet wird, sofern eine Verbindung über weitere Variablen besteht. Dieser

²⁶Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*

letzte Schritt muss also von Hand ausgeführt werden, nachdem das Programm ausgeführt worden ist.

7.2 Kausale Beziehungen zwischen den vier Quarterback-Rating-Variablen

Der PC-Algorithmus in Form von Tetrad III wird nun auf die Quarterback-Rating-Variablen angesetzt. Hierbei werden nur die Daten aus den NFL Saisons von 2000 bis 2007 verwendet, da aus den Strukturbruchtests in Kapitel 6 ab dem Jahr 2000 in allen Zeitreihen weder bei den Standardabweichungen noch bei den Mittelwerten Strukturbrüche zu finden waren.

Um zusätzlich zu gewährleisten, dass es sich um eine angemessene Unterstichprobe handelt, wurden mithilfe der Strukturbruchtestmethodik aus Kapitel 6 einfache Korrelationen zwischen den vier Quarterback-Rating-Variablen untersucht. Dabei wurde in der gesamten Stichprobe genau ein Bruch gefunden, nämlich in der Korrelation zwischen dem Anteil der vollständigen Pässe und dem der Interceptions. Nimmt man diesen Aspekt nun noch hinzu, so kommt man trotzdem zum Schluss, dass die Stichprobe von 2000 bis 2007 als geeignet erscheint, da dort bei keiner Analyse Strukturbrüche aufgetreten sind.

Bevor der Algorithmus anwendbar ist, muss man sich noch zwei Schwierigkeiten zuwenden.

Erstens ist es von Nachteil, dass der vom Algorithmus ermittelte Graph wahrscheinlich nicht kausal suffizient ist, was bedeutet, dass es Variablen geben kann, die nicht in diesem System miteinbezogen sind, jedoch wichtig für die kausalen Beziehungen sind. Das mag auch zu erwarten sein, da man sich mit den Variablen des Quarterback-Ratings nur auf das Passspiel beschränkt und weniger auf die gesamten Aspekte des Spiels. Aus diesem Grund wird das Programm ausgeführt, ohne kausale Suffizienz anzunehmen. Ausfindig gemachte Verbindungen werden durch „ $A \circ \rightarrow B$ “ dargestellt, was bedeutet, dass A B verursacht oder dass A und B eine Menge von gemeinsamen Ursachen besitzt, die nicht im Modell miteinbezogen wurden oder beides.

Die zweite Schwierigkeit ist, dass der Algorithmus nicht direkt auf Daten mit Trend angewandt werden kann, was häufig ein Problem darstellt, wenn man Graphentheorie in der Zeitreihenanalyse verwenden will. Bei diesem Datensatz handelt es sich um ein Panel, womit das Problem der Zeitreihen besteht. Allerdings fällt bei Betrachtung der Abbildungen 9 - 11 auf, dass ab 2000 kein Trend zu bestehen scheint. Es wird daher angenommen, dass die Zeitreihen in dieser Zeit stationär sind.

Nun kann der PC-Algorithmus auf die vier Variablen angewandt werden (durchschnittliche Anzahl

der erreichten Yards pro Passversuch (DY), Prozentsatz der vollständigen Pässe pro Passversuch (VP), Prozentsatz der erzielten Touchdowns pro Passversuch (TD) und Prozentsatz der geworfenen Interceptions pro Passversuch (IC)). Zwischen 2000 und 2007 gibt es exakt 260 Beobachtungen (jedes Jahr mindestens 30 Quarterbacks). Im folgenden ist die Korrelationsmatrix für die vier Variablen gezeigt (Abbildung 12).

Variablen	DY	VP	IC	TD
DY	1.00			
VP	0.69	1.00		
IC	-0.20	-0.22	1.00	
TD	0.73	0.53	-0.21	1.00

Abbildung 12: Korrelationen zwischen den QB-Variablen (2000-2007)

Als nächstes wurde der Fisher-z-Test für die einfachen Korrelationen durchgeführt, die dazugehörigen p-Werte sind dabei in Abbildung 13 zu finden. (Schritt 2 des PC-Algorithmus)

Variablen	DY	VP	IC	TD
DY	0.00			
VP	0.00	0.00		
IC	0.00	0.00	0.00	
TD	0.00	0.00	0.00	0.00

Abbildung 13: p-Werte der Unabhängigkeitstests

Den Ergebnissen der p-Werte zufolge wird die Nullhypothese jedes Mal verworfen, somit werden keine Verbindungen entfernt.

Schritt 3 des PC-Algorithmus kann nun angewandt werden. Auch bei diesen Tests auf Unabhängigkeit wird diese abgelehnt, abgesehen von vier Fällen, welche Abbildung 14 zeigt.

Geschätzte Verbindung	getestete Beziehung	bedingte Korrelation	p-Wert
VP o-o IC	Korr(VP,IC YD)	-0.12	0.06
IC o-o TD	Korr(IC,TD YD)	-0.09	0.14
DY o-o IC	Korr(DY,IC TD)	-0.07	0.24
VP o-o TD	Korr(VP,TD YD)	0.05	0.38

Abbildung 14: p-Werte für die bedingten Unabhängigkeitstests

Bei Betrachtung der p-Werte ist zu sehen, dass eine Entscheidung getroffen werden muss, ob für die Beziehung von TD und IC Unabhängigkeit angenommen oder abgelehnt werden soll. (für $p \leq \alpha$ lehnen wir H_0 ab). Man nimmt aber ein Signifikanzniveau von 10% an, um möglichst wenige Verbindungen entfernen zu müssen, womit die Verbindung bestehen bleibt. Damit ist das Skelett bestimmt (Abbildung 15).

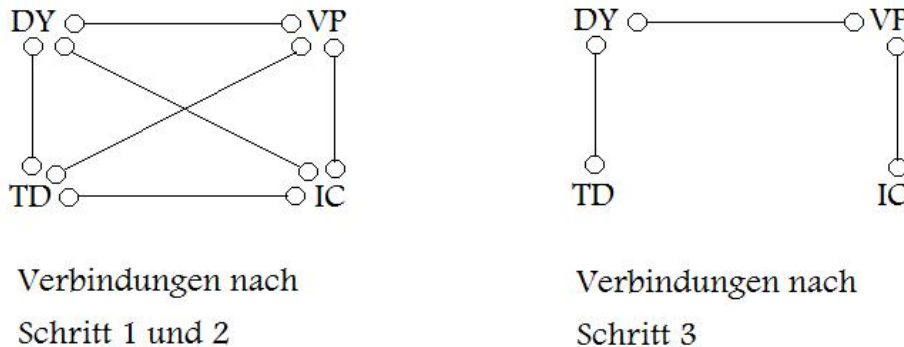


Abbildung 15: Verbindungen zu Beginn (links) und Skelett des kausalen Graphen (rechts)

Als Nächstes muss man sich für die Richtungen der Verbindungen entscheiden. In Schritt 4 sucht der Algorithmus nach Unshielded Colliders. Als Ergebnis wird dabei VP als Unshielded Collider zwischen DY und IC identifiziert, was bedeutet, dass IC und DY unbedingt unkorreliert, aber bedingt unter VP korreliert sind. Weiterhin wurden weder weitere Unshielded Colliders noch weitere Sachverhalte durch Schritt 5 und 6 gefunden, das heißt es konnten Richtungen für zwei der Verbindungen gefunden werden, nicht jedoch für die Verbindung zwischen DY und TD.

Da keine Entscheidung für die Richtung zwischen DY und TD durch den PC-Algorithmus getroffen werden kann, werden Zusatzinformationen benötigt. Zur Erinnerung: Aus Kapitel 4.2 ist zwar bereits bekannt, dass die Anzahl der Yards die Anzahl der Touchdowns verursacht, jedoch nur auf ein Spiel bezogen. In diesem Fall werden bekanntlich die Touchdowns und die durchschnittlichen Yards je pro Passversuch behandelt. Hinzu kommt, dass diese Variablen auf eine ganze Saison hinweg bezogen interessieren. Um den Sachverhalt zu klären wird eine Methode benutzt, die auch von Hoover, Sheffrin und Siegler^{27 28} verwendet wird: Man bedient sich der Ergebnisse aus den in Kapitel 6.2 durchgeführten Strukturbruchtests.

27 Hoover, K. D. and S. M. Sheffrin (1992) "Causation, Spending, and Taxes: Sand in the Sandbox or Tax Collector for the Welfare State"

28 Hoover, K. D. and M. Siegler (2000) "Taxing and Spending in the Long View: The Causal Structure of U.S. Fiscal Policy After 1791"

Bei den Mittelwerten in Abbildung 7 gab es nur einen Strukturbruch in den Zeitreihen der beiden Variablen - 1969 bei den Touchdowns. Die Zeitreihen der Standardabweichungen bleiben zwischen 1960 und 1976 für beide Variablen stabil. Für die weitere Untersuchung sei nun angenommen, dass die grundlegenden kausalen Beziehungen in diesem Abschnitt und in dem zu analysierenden dieselben seien, da man so in der Lage ist, weitere Informationen auszuschöpfen um die letzte Verbindung aufzuschlüsseln. Nimmt man nun weiterhin an, dass eine lineare Beziehung zwischen diesen beiden Variablen besteht, so sind folgende einfache Modelle möglich:

$$\begin{array}{ll}
 \text{(I)} & TD = \alpha_{TD} + \epsilon_{TD} \\
 \text{(II)} & DY = \alpha_{DY} + \beta_{DY} \cdot TD + \epsilon_{DY} \\
 \text{(III)} & DY = \mu_{DY} + \nu_{DY} \\
 \text{(IV)} & TD = \mu_{TD} + \lambda_{TD} \cdot DY + \nu_{TD}
 \end{array}$$

α und μ seien hierbei die Intercepts, β und λ die Steigungsparameter und ϵ und ν die Störgrößen.

Sollten sich Modelle (I) und (II) als präzise herauskristallisieren, so verursacht TD DY, sollten hingegen die Modelle (III) und (IV) als präzise gelten, so ist Umgekehrtes der Fall.

Als Erstes wird angenommen, dass Gleichung (I) und (II) wahr sind. Dann sollte der Strukturbruch von 1969 beim Touchdown-Anteil Gleichung (I) zu diesem Zeitpunkt beeinflussen, nicht jedoch Gleichung (II). Sollte man fälschlicherweise annehmen, Gleichung (III) und (IV) seien wahr, so sollte keine der beiden Gleichungen gegenüber diesem Strukturbruch invariant sein. Schließlich erzeugt der Strukturbruch aus 1969 auch genau diese Ergebnisse. Für die beiden Zeitspannen 1960 – 1969 und 1970 – 1977 ist die KQ-Schätzung nahe liegend, wessen Ergebnisse in Abbildung 16 dargestellt sind.

(I) $TD = \alpha_{TD} + \epsilon_{TD}$		
Stichprobe	Intercept	Steigung
1960-1969	5.34 (0.12)	-
1970-1976	4.46 (0.13)	-
Test auf Parameterunterschied	22.50 [0.00]	-
(II) $DY = \alpha_{DY} + \beta_{DY} \cdot TD + \epsilon_{DY}$		
Stichprobe	Intercept	Steigung
1960-1969	5.10 (0.17)	0.40 (0.03)
1970-1976	5.02 (0.15)	0.40 (0.03)
Test auf Parameterunterschied	0.48 [0.49]	0.35 [0.55]
(III) $DY = \mu_{DY} + v_{DY}$		
Stichprobe	Intercept	Steigung
1960-1969	7.21 (0.07)	-
1970-1976	6.81 (0.07)	-
Test auf Parameterunterschied	14.30 [0.00]	-
(IV) $TD = \mu_{TD} + \lambda_{TD} \cdot DY + v_{TD}$		
Stichprobe	Intercept	Steigung
1960-1969	-3.02 (0.63)	1.16 (0.09)
1970-1976	-4.01 (0.69)	1.24 (0.10)
Test auf Parameterunterschied	8.36 [0.00]	7.75 [0.01]

Abbildung 16: KQ-Schätzungen der Gleichungen (I)-(IV) mit F-Test oder Chow-Test als Test auf Parameterunterschied (es gab 213 Beobachtungen in der früheren und 156 in der älteren Stichprobe)

Dabei ist auffallend, dass Gleichung (II) die einzig stabile Gleichung für die beiden Stichproben ist. Damit ist die Lösung für die letzte Verbindung gefunden, denn dieses Resultat ist nur damit vereinbar, dass TD YD verursacht, mit Umgekehrtem jedoch nicht.

Demzufolge sind alle Richtungen der kausalen Verbindungen gefunden, skizziert in Abbildung 17.

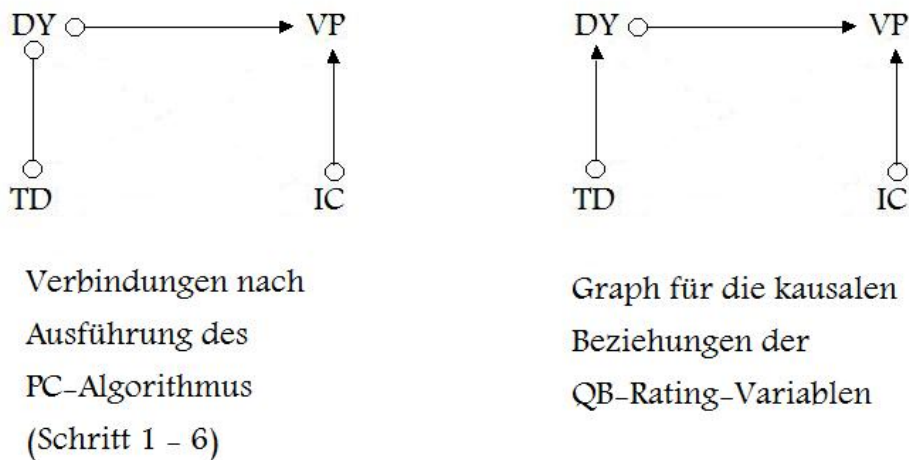


Abbildung 17: Kausaler Graph nach Ausführung des Algorithmus (links) und endgültiger kausaler Graph nach Anwendung der Strukturbruchtestergebnisse (rechts)

Für diese aufgeklärten Beziehungen wird jetzt nach Erklärungen gesucht.

Zuerst setze man sich mit der Beziehung zwischen dem Anteil der Interceptions und der vollständigen Pässe auseinander, also mit $IC \rightarrow VP$. Aus dem negativen Korrelationskoeffizienten (-0.22) aus der Korrelationsmatrix in Abbildung 12 sehen wir, dass die zwei Variablen negativ in Zusammenhang stehen, was auch schon intuitiv klar sein sollte, da ein hoher Anteil an geworfenen Interceptions im Verhältnis zu den Passversuchen dementsprechend aussagt, dass der Quarterback wahrscheinlich weniger fehlerfrei spielt, was einen geringeren Anteil an geworfenen vollständigen Pässen zur Folge hat. Dies dürfte im Übrigen auch ein Beispiel für das Problem der „ruhenden“ Variablen sein, welches im Allgemeinen besagt, dass es mindestens eine zusätzliche äußere Variable als gemeinsame Ursache für mindestens zwei Variablen im System gibt. Diese Variable – die ruhende Variable – sei in diesem Fall die „Genauigkeit“ des Quarterbacks, welches die gemeinsame Ursache für IC und VP sei. Falls dies der Wahrheit entspricht, mag das andeuten, dass der IC einen besseren Vertreter für die Genauigkeit des Quarterbacks als VP darstellt.

Als nächstes wird die Beziehung $DY \rightarrow VP$ in den Blick genommen. Die Variablen besitzen positiven Zusammenhang (Abbildung 12, Korrelationskoeffizient: 0.69), was nicht zwangsläufig nachvollziehbar sein mag, da man annehmen könnte, dass eine geringere Anzahl an durchschnittlichen Yards pro Passversuch Ursache für einen höheren Anteil an vollständigen Pässen sei, weil je kürzer ein Pass ist, desto höher sollte die Wahrscheinlichkeit für einen Catch und damit

einen vollständigen Pass sein. Möglicherweise besteht jedoch ein Maßproblem. Im zweiten Kritikpunkt von Kapitel 4.1 wurde bereits der Streitpunkt angesprochen, dass zu den durchschnittlich erreichten Yards pro Passversuch auch die erreichten Yards nach einem Catch zählen, was das Resultat unter Umständen verzerrt.

Alternativ bietet sich aber als Erklärung, dass Quarterbacks, die im Allgemeinen dazu neigen längere vollständige Pässe zu werfen, dadurch das Spielfeld mehr öffnen und so die Wahrscheinlichkeit eines vollständigen Passes erhöhen. Das hat nämlich zur Auswirkung, dass sich die gegnerische Defense dadurch mehr auf dem Feld verteilt. Ein Indiz für die Richtigkeit dieser Theorie gibt die NFL beispielsweise im Jahr 2007, in welchem die besten fünf Quarterbacks in DY auch fünf der wohl besten Quarterbacks dieser Liga überhaupt sind.

Zuletzt wird auch noch nach einer Erklärung für die durch die Strukturbruchtests ergründete Beziehung „TD \rightarrow DY“ gesucht. Nach Abbildung 12 herrscht zwischen beiden Variablen eine positive Korrelation (Korrelationskoeffizient: 0.73), was auch intuitiv wieder Sinn macht. Es ist wahrscheinlicher, dass ein Quarterback Touchdowns wirft, wenn er im Durchschnitt längere Pässe wirft (und damit mehr Yards gutmacht). Nähert sich ein Team der Endzone, so erhöht sich voraussichtlich die Wahrscheinlichkeit für einen Touchdown durch Laufspiel, wohingegen die Wahrscheinlichkeit für einen geworfenen Touchdown wahrscheinlich sinkt. Das bedeutet, dass Quarterbacks, die im Durchschnitt längere Pässe werfen, eher einen Touchdown werfen, als die, die im Durchschnitt kürzere vollständige Pässe werfen (da bei Versuchen kurz vor den Endzonen kaum noch geworfen wird, sondern Läufe vorgezogen werden). Auch hier kann eine ruhende Variable gefunden werden: Die Lauffähigkeit einer Mannschaft ist gemeinsame Ursache für TD und DY, was wiederum bedeuten mag, dass der Anteil der Läufe mehr den Anteil der gepassten Touchdowns als die durchschnittlichen Yards pro Versuch beeinflusst.

Am Ende dieser Kausalanalyse der Bewertungen über die Verhältnisse der einzelnen Variablen in der Quarterback-Rating-Formel (oder auch in anderen Formeln) kann eine allgemeine Schlussfolgerung gezogen werden. Wichtig ist die Tatsache, dass diese Verhältnisse nicht notwendigerweise unabhängig voneinander zu sein haben. Behandelt man beispielsweise IC und VP als unabhängig, kann man unabsichtlich den Einfluss von IC überschätzen, da Veränderungen in IC Veränderungen in VP verursachen, was letztendlich zu einem einseitigen „Double-Counting Effect“ führen würde. Ferner betont Stimel, dass es wichtig sei, hinsichtlich der Anwendung von Regression, um Einflüsse einzelner Spieler zu schätzen, dass es bei dieser Regression um die Wirkung auf die Ursachen geht und nicht um die Ursachen für die Wirkungen.

8 Fazit und Diskussion der Methoden

Es wurden in den letzten Kapiteln zwei Methoden vorgestellt, um statistische Beziehungen unter den Quarterback-Variablen zu analysieren: Strukturbruchtests und die Graphentheorie. Beide Methoden brachten interessante Ergebnisse und konnten sogar miteinander kombiniert werden. Abschließend sei ein Resümee über die gewonnenen Ergebnisse gehalten.

Als Erstes beschäftigte man sich mit den Zeitreihen zu den Mittelwerten der Standardabweichungen des Quarterback-Ratings, sowie der einzelnen Variablen. Die Berechnungen dieser im Paper konnten anhand der Daten nachvollzogen werden. Dabei sind keine Fehler aufgefallen.

Es wurden Strukturbruchtests durchgeführt, welche zu durchaus sinnvollen Ergebnissen geführt haben, da die Regeländerungen im Passspiel um 1970 und 1990 mit den gefundenen Strukturbrüchen in Einklang standen. Dadurch war eindeutig zu schlussfolgern, dass ein Vergleich von verschiedenen Zeitabschnitten somit kaum möglich ist bzw. es wenig Sinn macht, dies zu tun ohne dabei Regeländerungen zu berücksichtigen. Nimmt man jedoch Anpassungen an diesen Regeländerungen vor, so sind sinnvolle Vergleiche des Quarterbacks in verschiedenen Abschnitten denkbar.

Ein weiteres markantes Ergebnis des Strukturbruchtests war es, dass Regeländerungen im Laufe der Zeit zu einer Erhöhung des Quarterback-Ratings geführt haben, während nebenbei der Anteil der vollständigen Pässe gestiegen und der der geworfenen Interceptions dabei gesunken ist. Ab dem Bruch von 1970 schrumpften die Werte der Standardabweichungen, was möglicherweise damit zu tun hat, dass im Allgemeinen immer mehr Pässe pro Spiel geworfen wurden bzw. es immer mehr Spiel vom Quarterback gab, was bedeutet, dass sich die Streuung zwischen den Quarterbacks verringert hat. Aus diesem Grund sank auch der relative Unterschied zwischen der Leistung von sehr guten Quarterbacks und nur durchschnittlichen Quarterbacks.

Als Strukturbruchtest wurde ein sup-F-Test verwendet, welcher im Allgemeinen angewandt wird, wenn keine Bruchpunkte bekannt sind, was hier der Fall war. Es mag sicher noch möglich sein, noch feiner auf Strukturbrüche zu testen. Dadurch würden möglicherweise weitere Auswirkungen von Regeländerungen zum Vorschein kommen, da es in der Geschichte der NFL häufig (ab 1994 jährlich ²⁹) zu einer Regeländerung kam. Andererseits könnte eine zu feine Strukturbruchanalyse dazu führen, dass „bedeutungslose“ Brüche gefunden werden, d. h. Brüche, dessen Ursache nicht auf Regeländerungen sondern eher auf Zufall zurückzuführen sind.

Bei der durchgeführten Kausalanalyse konnte gezeigt werden, dass es einen wesentlichen

²⁹ Steelers Fever: „History of NFL Rules“

Unterschied zwischen den Beziehungen der Quarterback-Rating-Variablen über eine Saison und den Beziehungen dieser Variablen gibt, die nur auf einen Spielzug bezogen sind. Dies mag einerseits so sein, da sich die Eigenschaften der Variablen über die ganze Saison hinweg verändern, andererseits weil die Variablen immer relativ zum Passversuch definiert sind. Daher sind auch die Folgerungen und Interpretationen verschieden. Beispielsweise wurde gezeigt, dass der Anteil der vollständigen Pässe über eine ganze Saison hinweg andere Informationen beinhaltet, als der Anteil dieser innerhalb eines Spiels oder auch eines vollständigen Passes während eines Spielzugs.

Zu guter Letzt konnte man durch Koppelung der Graphentheorie mit den Strukturbruchtests drei Verbindungen bzw. Beziehungen zwischen den vier Variablen finden. Für diese wurden Erklärungen angeführt, für die jedoch laut Stimel weitere Analysen benötigt werden, um neben diesen weitere Möglichkeiten und Erklärungen wahrnehmen zu können.

Kritisch zu betrachten ist bei der Koppelung der beiden Methoden, dass man hierbei von der großzügigen Annahme ausgegangen ist, dass die kausalen Beziehungen zwischen den Variablen im Abschnitt 2000-2007 dieselbe ist wie 1960-1976. Man mag dies zwar vermuten, jedoch geht hier kein Nachweis dafür hervor.

Im Großen und Ganzen war die Anwendung dieser Methoden jedoch vertretbar und hat zu nachvollziehbaren und Sinn machenden Ergebnissen geführt.

9 Weitere Statistiken

Als Abschluss sei nun ein Ausblick auf weitere mögliche statistische Untersuchungen zu diesem Thema eingeblendet.

In erster Linie wäre es natürlich interessant weitere Faktoren mit in das Football-Spiel einzubeziehen als nur die des Passspiels. Hierzu sei auf jeden Fall der Aspekt des Laufspiels genannt, womit beispielsweise sowohl das Laufspiel des Quarterbacks (im Scrambling) als auch die der weiteren Spieler, beispielsweise der Receiver in Bezug auf erreichte Yards nach einem Catch, berücksichtigt werden könnten.

Ebenfalls wäre eine statistische Betrachtung der Defense sicher interessant und ebenso nicht unwichtig für Trainer einer Football-Mannschaft. Insbesondere werden in der NFL dafür Variablen für die Anzahl der Tackles, der Interceptions und der Fumbles, die beispielsweise zu einem Rückgewinn des Balls oder gar zu einem Touchdown führen, festgehalten.

Mit diesen Punkten würde man sich auch nicht nur auf das Quarterback-Rating beschränken,

sondern könnte sich auf eine Analyse des gesamten Spiels einlassen.

Hinzu kommt, dass hier nur Ergebnisse auf die gesamte Saison bezogen behandelt wurden. Analysen für einzelne Spiele könnten dabei sicherlich zu ebenfalls interessanten wie nützlichen Erkenntnissen führen.

Die Möglichkeit, diese Methoden auf den Sport anzuwenden, besteht jedoch sicherlich nicht nur beim American Football. Im Fußball wäre es zum Beispiel denkbar zu untersuchen, wie sich Variablen wie der Anteil der gewonnenen Zweikämpfe oder die Zahl der Ballkontakte pro Spieler im Laufe der Zeit verändern und ob es eventuell Strukturbrüche gibt oder gar ob es einen Zusammenhang zwischen diesen Variablen geben könnte und ist dies der Fall, welche Variable Ursache für die andere ist.

Damit steht fest, dass es noch eine Menge weiterer Aspekte des American Footballs gibt, die man statistisch beleuchten kann, aber auch, dass hier benutzte Methoden durchaus in weiteren Sportarten seine Anwendung finden können.

Quellen- und Literaturverzeichnis:

- D. Stimel: "A Statistical Analysis of NFL Quarterback Rating Variables", *Journal of Quantitative Analysis in Sports*, Vol. 5 : Iss. 2, Article 1. , 10.2202/1559-0410.1166, 2009
- L. Fahrmeir, Thomas Kneib, Stefan Lang: „Regression – Modelle, Methoden und Anwendungen“, Springer, Berlin Heidelberg, 2007
- Andrews, D.W.K. (1991) “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation”, *Econometrica*, 59, 817-858
- Bai, J. and P. Perron (1998) “Estimating and Testing for Multiple Structural Changes in a Linear Model,” *Econometrica*, 66, 47-78
- Bai, J. and P. Perron (2003a) “Computation and Analysis of Multiple Structural Change Models”, *Journal of Applied Econometrics*, 18, 1-22
- Bai, J. and P. Perron (2003b) “Critical Values for Multiple Structural Change Tests”, *Econometrics Journal*, 1, 1-7
- Bai, J. and P. Perron (2004) “Multiple Structural Change Models: A Simulation Analysis”, in Corbea, D., Durlauf, S., and B.E. Hansen (Eds.), *Econometric Essays*. Cambridge University Press
- Demiralp, S. and K.D. Hoover (2003) “Searching for the Causal Structure of a Vector Autoregression”, *Oxford Bulletin of Economics and Statistics*, 65, Supplement, 745-767
- Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge
- Prodan, R. (2008) “Potential Pitfalls in Determining Multiple Structural Change Models with an Application to Purchasing Power Parity”, *Journal of Business and Economics Statistics*, January, 26, 50-65
- Spirtes, P., R. Scheines, C. Meek, T. Richardson, C. Glymour, H. Hoijtink and A. Boomsma, (1996) *Tetrad 3: Tools for Causal Modeling*, program and user’s manual on the world wide web at <http://www.phil.cmu.edu/tetrad/tet3/master.htm>
- Hoover, K. D. (2003) “Some Causal Lessons From Macroeconomics,” *Journal of Econometrics*, 112, 121-125
- Hoover, K. D. and S. M. Sheffrin (1992) “Causation, Spending, and Taxes: Sand in the Sandbox or Tax Collector for the Welfare State”, *American Economic Review*, 82, 225-248
- Hoover, K. D. and M. Siegler (2000) “Taxing and Spending in the Long View: The Causal Structure of U.S. Fiscal Policy After 1791”, *Oxford Economic*

Internetseiten:

Wikipedia: „Quarterback“,
<http://de.wikipedia.org/wiki/Quarterback>
aufgerufen am 30.11.2009

Wikipedia: „American Football“,
http://de.wikipedia.org/wiki/American_football
aufgerufen am 30.11.2009

Wikipedia: „Offensive Line“,
http://de.wikipedia.org/wiki/Offensive_Line
aufgerufen am 01.12.2009

talkamericanfootball.co.uk: „The Rules of American Football“,
http://www.talkamericanfootball.co.uk/guides/rules_of_american_football.html
aufgerufen am 01.12.2009

Wikipedia: „National Football League“,
<http://de.wikipedia.org/wiki/NFL>
aufgerufen am 01.12.2009

Offizielle NFL-webseite,
<http://www.nfl.com/>
aufgerufen am 03.12.2009

Osnaball.de: „American Football – Offense“,
<http://www.osnaball.de/american-football-offense>
aufgerufen am 03.12.2009

A. Zeileis, F. Leisch, K. Hornik, C. Kleiber, „Testen auf Strukturbruch – Implementierung in R und Anwendung“,
<http://statmath.wu.ac.at/~zeileis/papers/OeSG-2001.pdf>
aufgerufen am 05.12.2009

Wikipedia: „Strukturbruch“,
<http://de.wikipedia.org/wiki/Strukturbruch>
aufgerufen am 05.12.2009

Wikipedia: „Structural break“,
http://en.wikipedia.org/wiki/Structural_break
aufgerufen am 05.12.2009

Steelers Fever: „History of NFL Rules“,
http://www.steelersfever.com/nfl_history_of_rules.html
aufgerufen am 07.12.2009

Wikipedia: „Salary Cap“,
http://de.wikipedia.org/wiki/Salary_Cap
aufgerufen am 07.12.2009

Wikipedia: „Graphentheorie“,
<http://de.wikipedia.org/wiki/Graphentheorie>
aufgerufen am 08.12.2009

Wikipedia: „Graph (Graphentheorie)“,
http://de.wikipedia.org/wiki/Graph_%28Graphentheorie%29
aufgerufen am 08.12.2009

Shockwave: „Two Minute Football 3D 2010“,
<http://www.shockwave.com/gamelanding/football3d.jsp>
aufgerufen am 09.12.2009

K. Bonsor: „How American Football Works“,
<http://entertainment.howstuffworks.com/football.htm/printable>
aufgerufen am 10.12.2009

Wikipedia: „Tight End“,
http://de.wikipedia.org/wiki/Tight_End
aufgerufen am 10.12.2009

Wikipedia: „Runningback“,
<http://de.wikipedia.org/wiki/Runningback>
aufgerufen am 10.12.2009