

A Statistical Analysis of NFL Quarterback Rating Variables

Seminar: Statistik im Sport

Norbert Krautenbacher

Institut für Statistik
Ludwig-Maximilians-Universität München

19. Dezember 2009



Gliederung

- 1 Spielregeln und die Rolle des Quarterbacks (QB)
- 2 Das Quarterback-Rating
- 3 Problemstellungen
- 4 Die Daten
- 5 Strukturbrüche in Zeitreihen
- 6 Graphentheorie
- 7 Fazit und Diskussion
- 8 Weitere Statistiken

19.12.2009

Norbert Krautenbacher

2

1 Spielregeln und die Rolle des QBs



a) Spielkonzept

- zwei Mannschaften zu je 11 Spielern
- Ziel: Befördern des Spielballs in die gegnerische Endzone oder Erzielen eines „Field Goals“, um dabei möglichst viele Punkte zu erreichen
- „Offense“ (angreifende Mannschaft) hat Möglichkeit durch Werfen oder Laufen Raum zu gewinnen
- „Defense“ (verteidigende Mannschaft) versucht diese davon abzubringen und selbst in Ballbesitz zu gelangen
- Offense hat 4 Versuche um Raumgewinn von 10 sog. Yards zu erreichen
→ neue 4 Versuche; bei Nichtgelingen → Ballwechsel



19.12.2009

Norbert Krautenbacher

3

b) Spielfeld

- ca. 110 m lang, ca. 50 m breit
- eingeteilt in 120 Yards bzw. 12 Abschnitte zu je 10 Yards mit
- jeweils letzter Abschnitt entspricht Endzone und besitzt sog. „Goal Posts“ als Tore
- Zwei parallele Linien in Längsrichtung, sog. „Hash Marks“
- endet Versuch außerhalb der Hash Marks
→ nächster Versuch an nächstgelegener Hash Mark



19.12.2009

Norbert Krautenbacher

4



c) Wichtigste Methoden um zu punkten

- Am meisten Punkte bringende Methode: der „Touchdown“ (6 Punkte)
- Touchdown wird erlangt, indem Ball entweder durch Laufspielzug in gegnerische Endzone getragen oder durch Passspielzug darin gefangen wird
- Weitere Methode: Ball wird durch gegnerische Goal Posts gekickt, sog. Field Goal (3 Punkte)
- Field Goal wird in der Regel versucht, wenn Touchdown in aktuellen Versuche als aussichtslos erscheint

d) Passspielzug und die Position des Quarterbacks

- QB hat nach Erhalt des Balls Möglichkeit des eigenen Laufspiels („Scrambling“), der Übergabe an weiteren Spieler oder die des Passspielzugs.
- Pass kann
 - vom Receiver gefangen werden (vollständiger Pass)
 - Boden berühren oder ins Aus geworfen werden (unvollständiger Pass)
 - von Spieler der Defense abgefangen werden („Interception“)
- Bei vollständigem Pass kann Spieler soweit laufen bis er von Defense-Spieler zu Boden gebracht („getackelt“) wird
→ Nächster Versuch beginnt an dortiger Position („Line of Scrimmage“)
- Bei unvollständigem Pass: Ende des Versuchs
- Bei Ballverlust ohne vorherigen Tackle: Spiel läuft weiter („Fumble“)



1 Spielregeln und die Rolle des QBs d) Passspielzug und die Position des Quarterbacks



- QB ist Schlüsselfigur der Offense: muss geplante Spielzüge umsetzen und sich an Handlungen der Defense anpassen
- QB wird durch „Offensive Line“ (LT, LG, C, RG, RT) vor gegn. Defense beschützt
- TE und WR sind für QB mögliche Ballfänger; HB und FB sind mögliche Ballträger für Laufzug, sonst Blocker
- Vielzahl möglicher Taktiken (→ „Rasenschach“)



2 Das Quarterback-Rating

- stat. Wert; soll Passfähigkeiten des QBs bewerten und Vergleich zu weiteren QBs gewährleisten
- Linearkombination von vier Variablen:
 - Durchschnittl. Anzahl der err. Yards pro Passversuch (**DY**)
 - Anteil der vollst. Pässe pro Passversuch (**VP**)
 - Anteil der erzielten Touchdowns pro Passversuch (**TD**)
 - Anteil der geworfenen Interceptions pro Passversuch (**IC**)
- Damit wird QB-Rating (QBR) folgendermaßen berechnet:

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



3 Problemstellungen

a) Veränderungen des QB-Ratings über die Zeit

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

Bedingungen für diese Formel:

- keine der vier Komponenten darf negativ werden
→ VP mindestens 30% $(0.05 \cdot (0.30 \cdot 100 - 30) = 0)$
- keiner der Zähler der vier Brüche darf größer als 2.375 sein
→ VP höchstens 77.5% $(0.05 \cdot (0.775 \cdot 100 - 30) = 2.375)$
- QB-Rating ist Wert zwischen 0 und 158.3 $(100 \cdot 4 \cdot \frac{2.375}{6} = 158.\bar{3})$

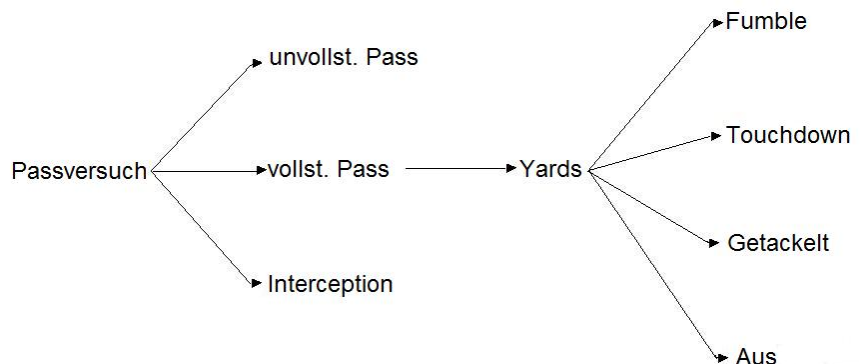
- erste Problemstellung ergibt sich aus Kritik an QB-Rating
- allgemeine Kritik an QB-Rating:
 - schwere Verständlichkeit → Ergebnisse schwierig zu deuten
 - QB-Rating als Maß für Leistung eines QBs unangemessen
 - Vergleich von QBs aus verschiedenen Zeiträumen nur schwer möglich; Gründe wie Regeländerungen führten im Laufe der Zeit zur Erhöhung des QB-Ratings
- Untersuchung der Veränderungen des QB-Ratings im Laufe der Zeit durch Strukturbruchanalyse:
Suchen nach sog. Strukturbrüchen in den Zeitreihen und den dafür vorliegenden Gründen (→ Gliederungspunkt 5)



3 Problemstellungen

b) Beziehungen zwischen den Variablen des QB-Ratings

- Anwendung der Graphentheorie um Beziehungen zueinander zu untersuchen
- kausale Beziehung zwischen Variablen für einen Passversuch leicht darstellbar:



3 Problemstellungen

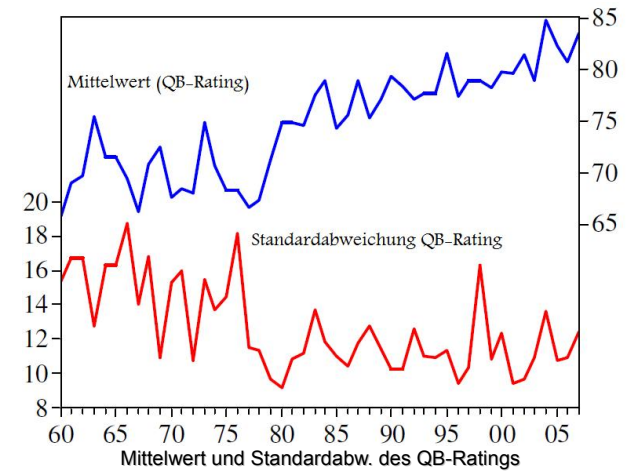
b) Beziehungen zwischen den Variablen des QB-Ratings

- QB-Rating erfasst fast jede Möglichkeit für Ausgang eines Passversuches
- „A → B“ :<=> „A 'verursacht' B“, ohne Annahme von kausaler Suffizienz, d. h. A nicht hinreichend, um B zu folgern
- für Untersuchung auf ganze Saison ist zu beachten: jede Variable ist relativ zum Passversuch definiert
→ es bestehen nicht unbedingt die selben Kausalitäten.
- Bsp:
für einen Passversuch: vollst. Pass → Touchdown durch Pass
für ganze Saison: VP → TD ??? (→ Gliederungspunkt 6)

4 Die Daten

- Daten aus den Saisonen der NFL (National Football League)
- nur QBs, mit Minimum von 14 Pässen pro Spiel;
- Daten ab 1960, da vorher wenige QBs 14-Pässe-Grenze überschritten, möglicherweise, da 1960 AFL (American Football League) mit NFL verschmolz → mehr Mannschaften
- für Analyse: Berechnung der Mittelwerte und Standardabweichungen pro Saison für QB-Rating und die vier QB-Variablen

Quarterback-Rating



19.12.2009

Norbert Krautenbacher

13

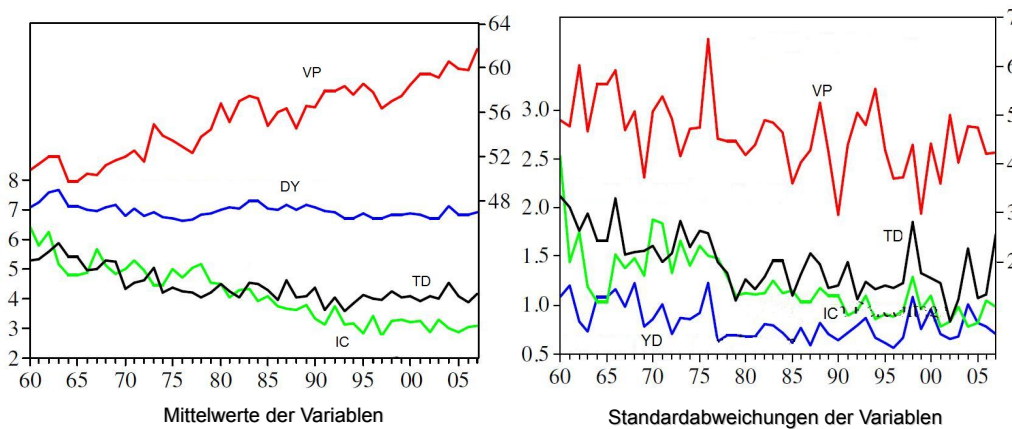
19.12.2009

Norbert Krautenbacher

14

4 Die Daten

Quarterback-Rating-Variablen



19.12.2009

Norbert Krautenbacher

15

19.12.2009

Norbert Krautenbacher

16

5 Strukturbrüche in Zeitreihen

a) Strukturbruchtest

- Ziel: Untersuchung der Veränderung des QB-Ratings über Zeit:
 - Aufstellen eines Regressionsmodells
 - Ermitteln von Strukturbrüchen durch stat. Tests
- Strukturbruch: plötzliche Änderung der Regressionsparameter in einer Zeitreihe



- > einfaches Regressionsmodell:

$$y_t = \alpha_j + \varepsilon_t$$

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

m : Anzahl der Brüche → m+1 Abschnitte bzw. Intercepts

T : Anzahl der Beobachtungen

- > Test auf Verschiebungen im Mittelwert der Zeitreihen
- > Schätzung der Strukturbruchdaten (T_1, \dots, T_m) :

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

mit $S_T(T_1, \dots, T_m)$ als Residuenquadratsumme



Anzahl der Strukturbrüche durch Sup-F-Test:

H_0 : "Es gibt l Strukturbrüche" gegen H_1 : "Es gibt l+1 Strukturbrüche"

Teststatistik: $\text{Sup } F(l+1|l)$

(kleinste obere Schranke der F-Statistiken für die Hypothesen)

Vorgehensweise:

→ Durchführung eines F-Tests mit $l = 0$ (H_0 : "kein Bruch" gegen H_1 : "1 Bruch")

→ Weiter Testen mit $l = 1, \dots, m \in \mathbb{N}$, mit m so, dass H_0 zum ersten mal nicht verworfen wird

also letzter Test: H_0 : "m Brüche" gegen H_1 : "m+1 Brüche"



Annahmen für die durchzuführenden Tests:

- > Durch Bai und Perron vorgegeben: trimmender Anteil und größtmögliche Anzahl von Brüchen
- > trimmender Anteil * T = Mindestlänge eines Abschnitts
- > trimmender Anteil = 0.15; T = 48 (von 1960 bis 2007)
 - Mindestlänge: 7 Jahre (exakt 7.2)
- > größtmögliche Anzahl von Brüchen: 5, also höchstens 6 Abschnitte
- > 1%-Signifikanzniveau



b) Ergebnisse der Strukturbruchtests

QB-Rating

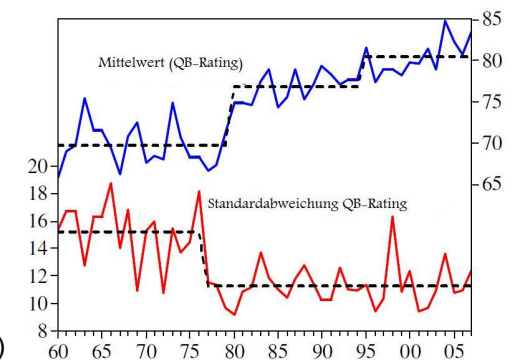
Mittelwert:

- > Bruch 1979: Regeländerungen (1978, 1979)
 - Receiver konnten freier manövrieren
 - Vorteil für QB

- > Bruch 1990: Verstärkung dieser Regeln (90er)
 - 1994: Einführung einer Gehaltsobergrenze („Salary Cap“)

Standardabweichung:

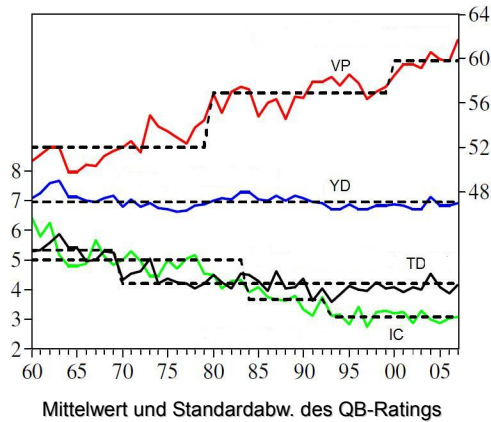
Bruch: 1976 (→ wie Variablen S.21)



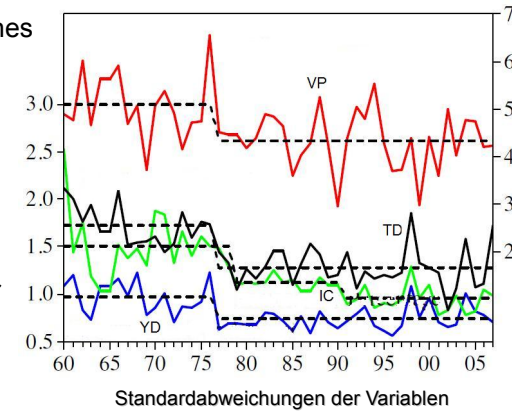
Mittelwert und Standardabw. des QB-Ratings

Mittelwerte

- > DY: keine Strukturbrüche
- > VP und IC: (weitgehend übereinstimmende Strukturbrüche)
 - > beide Brüche wie QB-Rating begründbar
- > TD:
 - > Bruch 1969 ab 1970 bemerkbarer Abfall der Touchdowns allgemein

Standardabweichungen:

- > Alle Variablen außer IC haben gleiches Bruchdatum:
 - Bruch 1976 genannte Regeländerungen
- > IC:
 - > Bruch 1978 ebenfalls auf diese Regeländerungen zurückführbar
 - > Bruch 1990 möglicherweise ebenfalls durch Verstärkung der Regeln

Schlussfolgerungen:

- > Vergleiche zwischen QBs aus verschiedenen Zeitabschnitten kaum möglich, da Spiele unter unterschiedlichen Bedingungen stattfanden (Regeln)
- > TD und DY scheinen Einflussfaktoren für steigendes QB-Rating zu sein, da ihre Brüche sich angleichen
- > Verbesserung der Leistung der QBs bzw. des Passspiels, aber Schwankungsbreite zwischen sehr guter und durchschnittlicher Leistung verringert

6 Graphentheorie

a) Analyse von Zusammenhängen

- > Graphentheorie: Beschäftigung mit Eigenschaften von Graphen
- > Abhängigkeiten, Nicht-Abhängigkeiten und Richtungen der Abhängigkeiten können aus Information über Zusammenhänge extrahiert werden → Grafisch darstellbar
- > Zugrunde liegende Idee: „kausale Markovbedingung“:
Zwei Variablen A und B sind **unkorreliert unter der Menge C**, wenn gilt:
 - > $A \leftarrow C \rightarrow B$, oder
 - > $A \rightarrow C \rightarrow B$
- > Zusätzlich zur kausalen Markovbedingung gilt:
A und B sind **korreliert unter C** wenn gilt:
 - > A und B sind bedingt unabhängig, gegeben einer Menge von Variablen, welche weder C noch ihre Nachfahren beinhaltet und $A \rightarrow C \leftarrow B$ (C ist „Unshielded Collider“)



Zusammenhänge dieser Art findet der sogenannte PC-Algorithmus mit folgenden Anweisungsschritten:

- (1) jede Variable sei mit jeder anderen Variablen ungerichtet verbunden. ($A - B$)
- (2) Teste unbedingte Korrelation für jedes Paar von Variablen mit Fisher-z-Test mit Nullhypothese: "die zwei Variablen sind unabhängig"
Bei Beibehalten: Entferne Verbindung
- (3) Für restl. Variablen: teste bedingte Korrelation für jedes Paar bedingt unter einer dritten Variablen
Bei Beibehaltung der Nullhypothese: Entferne Verbindung
Wiederhole Test für übrige verbundenen Paare bedingt unter zwei Variablen, drei Variablen usw.
- (4) Teste übrige verbundene Variablen, um Unshielded Colliders zu finden. Falls fündig, richte die Verbindungen folgendermaßen: $A \rightarrow C \leftarrow B$
- (5) Falls es Beziehung $A \rightarrow B - C$ gibt, so richte die unger. Verbindung so aus: $A \rightarrow B \rightarrow C$
- (6) Falls es Beziehung $A - B$ gibt und gerichtete Verbindungen von A zu B durch weitere Variablen existieren, so richte die Verbindung so aus: $A \rightarrow B$

19.12.2009

Norbert Krautenbacher

25



b) kausale Beziehungen zwischen den vier Variablen

- > Anwendung des implementierten PC-Algorithmus Tetrad III (v. Spirtes et a.)
- > Verwendung der Daten erst ab 2000, da von da an keine Strukturbrüche
- > Vor Anwendung treten 2 Schwierigkeiten auf:
 - > keine kausale Suffizienz
Lösung: Algorithmus nimmt dies nicht an:
„ $A \circ \rightarrow B$ “ : \Leftrightarrow „A verursacht B“ v „A und B besitzen gem. Ursachen die nicht in Modell mit einbezogen sind“
 - > Algorithmus nicht direkt auf Daten mit Trend anwendbar
Lösung: ab 2000 scheint in keiner Zeitreihe ein Trend zu bestehen
- > Damit ist Algorithmus anwendbar

19.12.2009

Norbert Krautenbacher

26



Durchführung des Fisher-z-Tests für die einfachen Korrelationen
(Schritt 2 des PC-Algorithmus)

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

Korrelationen zwischen den QB-Variablen

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

p-Werte der Unabhängigkeitstests (Fisher-z-Tests)

→ Nullhypothese kann wird jedes Mal verworfen

→ Verbindungen bleiben bestehen

19.12.2009

Norbert Krautenbacher

27



Schritt 3 (Test für bedingte Korrelation):

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

p-Werte der bedingten Unabhängigkeitstests (Fisher-z-Tests)

Annahme eines 10%-Signifikanzniveaus

→ Verbindung VP o-o IC bleibt bestehen

→ IC o-o TD, DY o-o IC und VP o-o TD werden getrennt

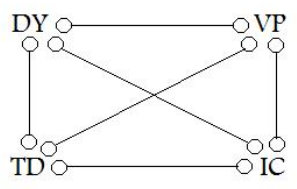
19.12.2009

Norbert Krautenbacher

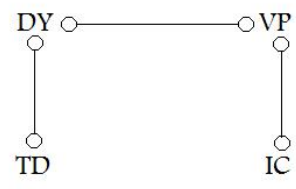
28



Sog. Skelett des kausalen Graphen ist damit bestimmt:



Verbindungen nach Schritt 1 und 2



Verbindungen nach Schritt 3

Ermitteln der Richtungen der Graphen:

- > Schritt 4: Algorithmus sucht nach Unshielded Colliders
- > VP wird als Unshielded Collider zwischen DY und IC identifiziert (DY → VP ← IC)
- > Schritt 5 und 6 führen zu keinen weiteren Ergebnissen
- > Problem: für DY und TD konnte kann Richtung durch Algorithmus gefunden werden
- > Lösung: Benutzung der Ergebnisse der durchgeführten Strukturbruchtests
- > Zwischen 1960 und 1976 ein Strukturbruch bei Mittelwerten und Standardabw. für die beiden Variablen (im Mittelwert der Touchdowns 1969)
- > Annahme dass Zeitabschnitt 1960 - 1976 dieselben grundlegenden kausalen Beziehungen zw. den DY und TD herrschen wie 2000-2007



- > Bei weiterer Annahme, dass Beziehungen linear, kann folgendes einfaches Modell aufgestellt werden:

(I) $TD = \alpha_{TD} + \epsilon_{TD}$
 (II) $DY = \alpha_{DY} + \beta_{DY} \cdot TD + \epsilon_{DY}$
 (III) $DY = \mu_{DY} + v_{DY}$
 (IV) $TD = \mu_{TD} + \lambda_{TD} \cdot DY + v_{TD}$

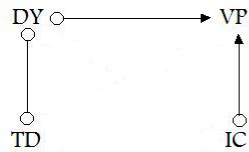
- > Wenn (I) und (II) präzise dann gilt TD → DY, für (III) und (IV) DY → TD
- > Falls (I) und (II) sind wahr. → Bruch 1969 müsste Gleichung (I) beeinflussen, nicht jedoch Gleichung (II)
- > (III) und (IV) müssten beide vom Bruch beeinflusst werden

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

- > Test (F-Test) bringt Ergebnisse, die die Vermutung bestätigen
- > (II) ist einzige stabile Gleichung
- > Damit ist TD → YD nahe liegend und letzte gesuchte Verbindung

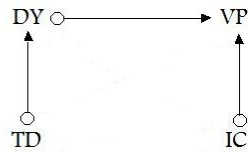


kausaler Graph nach Ausführung des Algorithmus:



Verbindungen nach Ausführung des PC-Algorithmus (Schritt 1 - 6)

kausaler Graph nach zusätzlicher Benutzung der durchgeführten Strukturbruchtests:



Graph für die kausalen Beziehungen der QB-Rating-Variablen



Erklärungen:

- IC → VP: unbedingter Korrr.-Koeffizient (-0.22) → neg. Zusammenhang
 - weniger geworfene Interceptions → mehr vollständige Pässe
- DY → VP: pos. Zusammenhang (Korr.-Koeffizient: 0.69);
 - möglicherweise Maßproblem: Resultat evtl. verzerrt, da Yards nach vollst. Pass nicht einbezogen
 - alternativ: QBs, die längere vollst. Pässe werfen, öffnen Spielfeld mehr, Defense verteilt sich mehr → höhere W-keit eines vollst. Passes
- TD → DY: pos. Korrelation (0.73)
 - macht Sinn: W-keit ist höher, dass QB Touchdowns wirft, wenn er im Durchschnitt längere Pässe wirft (damit mehr Yards)
 - In Nähe der Endzone: W-keit eines Touchdowns durch Pass gering



Schlussfolgerung:

- Beziehung nicht notwendigerweise unabh. voneinander (Gefahr eines „Double-Counting Effects“)
- Anwendung von Regression:
 - Für Schätzung der Einflüsse einzelner Spieler: Wirkung auf Ursache und nicht: Ursache für Wirkung



7 Fazit und Diskussion

Zeitreihen zu Mittelwerten und Standardabw.:

- Strukturbruchtests lieferten sinnvolle Ergebnisse: Regeländerungen um 1970 und 1990 im Einklang mit Brüchen → Vergleich verschiedener Abschnitte schwer möglich
- QB-Rating gestiegen durch Regeländerungen
 - Standardabw. ab 1970 geschrumpft, möglicher Grund: mehr geworfene Pässe → mehr Spiel vom QB → Verringerung d. Streuung (→ Verringerung d. Unterschieds v. guten u. durchschn. Qbs)
- Verwendung von sup-F-Test:
 - angewandt bei unbekanntnen Brüchen → hier der Fall
 - Möglichkeit der „Verfeinerung“ der Strukturbruchtests



8 Weitere Statistiken

Kausalanalyse:

- > wesentl. Unterschied zw. Beziehungen der QB-Rating-Variablen über Saison gegenüber einen Spielzug, da:
 - > Variablen relativ zum Passversuch definiert
 - > Variablen verändern sich über Saison
 - > Koppelung von Graphentheorie und Strukturbruchtests
 - Finden von drei Verbindungen
 - > Kritik: großzügige Annahme, dass Beziehungen zwischen Variablen in 2000 - 2007 gleich denen in 1960-1976
- > weitere Faktoren als nur Passspiel können einbezogen werden
 - > Betrachtung der Defense
 - > Analyse einzelner Spiele
 - > Anwendung auf weitere Sportarten