

Samenhangen tussen eindstanden eredivisie

E.P. Martens

Onderzoeksvragen

1. In hoeverre hangen de eindstanden van eredivisieclubs tussen twee opeenvolgende jaren met elkaar samen?
2. Hoe neemt deze samenhang af bij verder uit elkaar liggende jaren?
3. Is deze samenhang tussen eindstanden in de loop der tijd sterker dan wel zwakker geworden?

Conclusies

1. de samenhang tussen de eindstand van het ene jaar en die uit het voorgaande jaar lijkt tamelijk hoog, gemiddeld 0,76; de laatste 7 jaren ligt deze zelfs rond 0,85. Toch betekent een dergelijk getal dat er nog de nodige verschuivingen kunnen hebben plaatsgevonden.
2. de samenhang neemt geleidelijk af als verder uit elkaar liggende jaren met elkaar worden vergeleken: van 0,76 (1 jaar verschil) tot 0,44 (6 jaren verschil).
3. de samenhangen tussen de eindstanden van verschillende jaren zijn in de loop der tijd duidelijk groter geworden, zeker als het om verder uit elkaar liggende jaren gaat: bij een verschil van drie seizoenen is deze samenhang gestegen van rond de 0,3 halverwege jaren tachtig tot boven de 0,6 begin deze eeuw.

Artikel

Samenhangsmaten

De samenhangen tussen de eindstanden van voetbalclubs in de eredivisie kunnen worden gekwantificeerd met een maat voor correlatie. Omdat de eindstand een *rangordening* is van voetbalclubs (die bepalend is voor Champions League, Europees voetbal, degradatie e.d.), is de beste statistische maat de Spearman correlatiecoëfficiënt, hoewel Pearson's correlatiecoëfficiënt hier niet veel van afwijkt [1]. Dit is een rangorde-correlatiecoëfficiënt die kan variëren van -1 tot $+1$ en weergeeft in hoeverre de stand van het ene jaar overeenkomt met dat van het andere jaar. Een correlatie van 1 betekent dat de stand van het ene jaar exact gelijk is aan die van het andere jaar. Hoewel deze maat wat minder gevoelig is voor uitbijters en de verdelingen, blijft het een samenvatting voor een groot aantal waarnemingspunten. Een correlatie van bijv. 0,87 kan dan ook op heel verschillende manieren zijn ontstaan, zoals blijkt uit de laatste twee plaatjes van figuur 1.

Methodologisch probleem

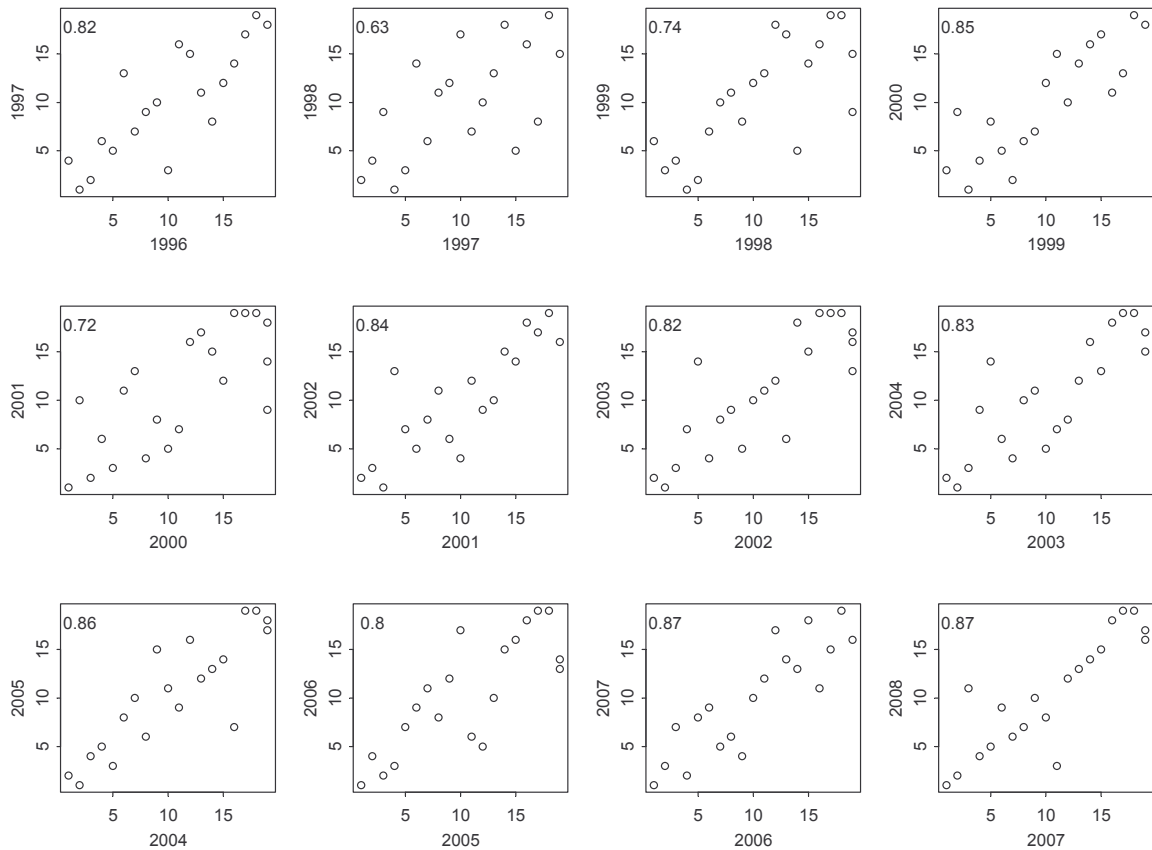
Een methodologisch probleem doet zich voor bij de berekening van correlaties tussen jaren. Per definitie degraderen namelijk één of meer clubs waardoor er geen waarneming is in het daaropvolgende jaar. Het weglaten van deze waarneming(en) zou de inhoudelijke conclusie verkeerd beïnvloeden, zeker als correlaties tussen verderliggende jaren worden berekend [2]. Het meest te verdedigen is om de gedegradeerde clubs wel mee te nemen in de berekening met een score van bijv. 19.

Correlaties tussen twee opeenvolgende jaren

Voor de gehele periode 1981-2008 lag de gemiddelde (rangorde-)correlatie op 0,76. Dit lijkt hoog, maar gegeven de kennis die voetbalkenners hebben over het bestaan van topclubs, subtop, middenmoters en degradatiekandidaten, is het nog niet eens zo hoog.

De eindstand voor een willekeurige club in het ene jaar is dus slechts ten dele een graadmeter voor de eindnotering in het daaropvolgende jaar. Hieronder zijn de scatterplots en correlaties weergegeven tussen twee opeenvolgende jaren voor de periode 1996 t/m 2008.

FIGUUR 1: Rangorde-correlaties tussen de eindstanden eredivisie van twee opeenvolgende jaren, 1996-2008



Voor de periode 1996-2008 variëren deze correlatiecoëfficiënten van 0,63 tot 0,87. Vergelijk het patroon in de laatste twee plaatjes: in beide gevallen is er een correlatie van 0,87 berekend, terwijl het beeld erg verschillend is. Tussen 2008 en 2007 waren slechts twee clubs (AZ en NAC) die heel anders scoorden, terwijl er tussen 2006 en 2007 meer kleine wisselingen geweest zijn.

Correlaties tussen twee verder uit elkaar liggende jaren

Het zal duidelijk zijn, dat naarmate de periode langer is tussen twee eindstanden (2, 3, 4, 5 of 6 jaar), de correlatie zal afnemen. Interessanter is de mate waarin dat gebeurt. Het blijkt dat deze slechts zeer geleidelijk afneemt (tabel 1), hetgeen inhoudelijk te begrijpen is: er is namelijk een behoorlijk aantal clubs dat over een vrij lange periode vergelijkbaar scoort (bovenin, middenmoot, onderin) [3].

TABEL 1: Rangorde-correlaties tussen twee eindstanden eredivisie, van 1-6 jaar verschil

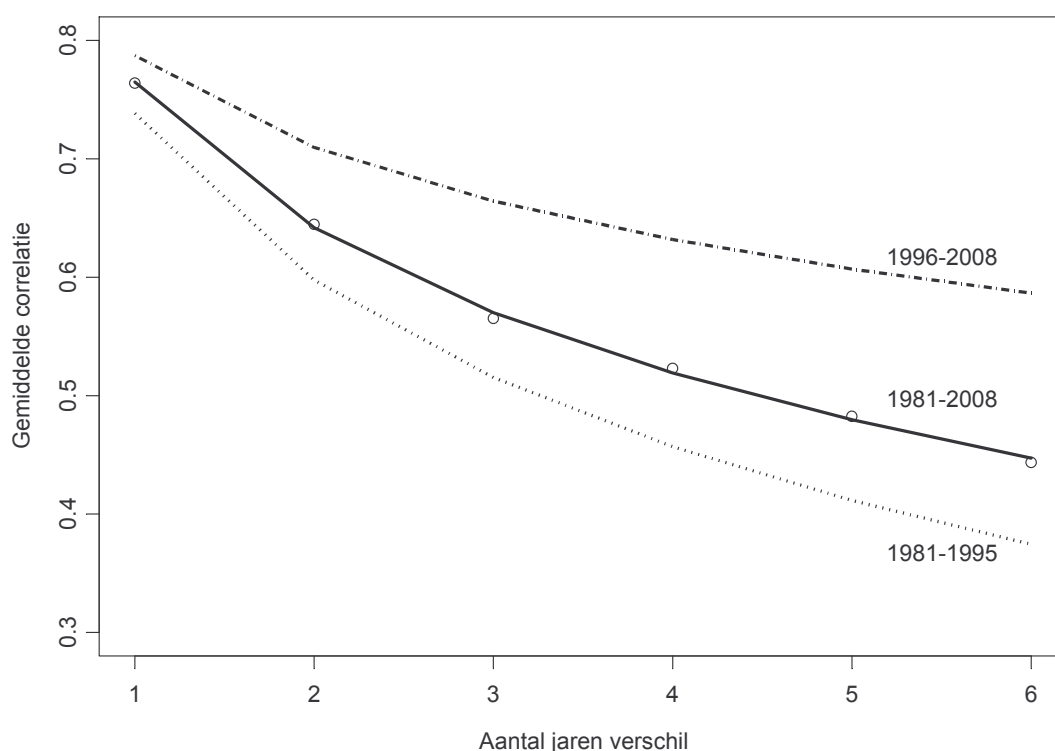
aantal jaren verschil	gemiddelde correlatie 1981-2008
1	0,76
2	0,64
3	0,57
4	0,52
5	0,48
6	0,44

Uit figuur 2 blijkt dat de afname van deze samenhangen bijzonder goed te beschrijven is als een logarithmische curve.

Verloop van de correlaties in de tijd

Als vervolgens de periode 1981-2008 in twee delen wordt gesplitst (voor en na de invoering van het driepuntensysteem), dan blijkt dat de samenhangen aanzienlijk hoger liggen in de afgelopen periode vergeleken met de periode voor 1996. Dat betekent dat er van jaar tot jaar minder verrassingen zijn dan in het verleden het geval was. Dat geldt met name als over een langere periode wordt gekeken. Zo ligt de gemiddelde correlatie tussen eindstanden met drie seizoenen verschil in de periode voor 1996 op 0,51, terwijl deze vanaf 1996 op 0,64 ligt.

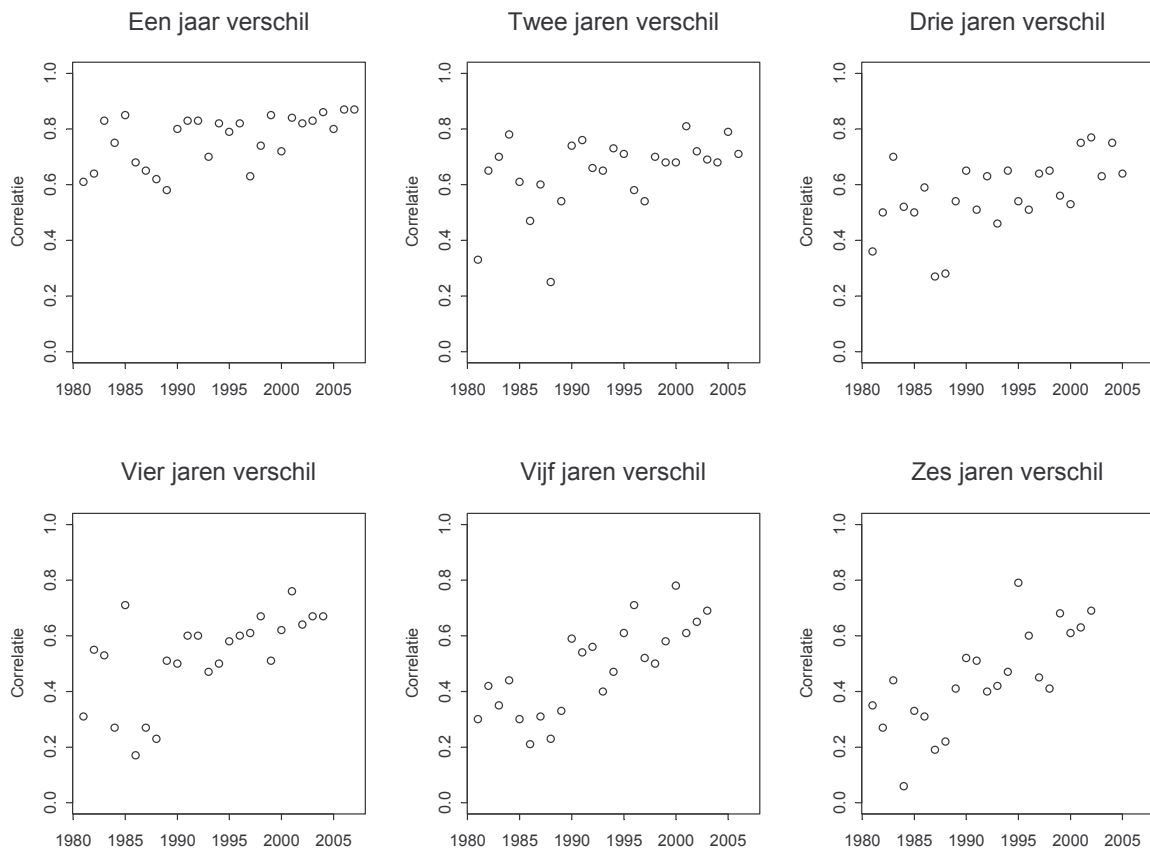
FIGUUR 2: Verloop gemiddelde correlatie tussen de eindstanden eredivisie naarmate er meer jaren tussen zitten, voor 1981-2008, 1981-1995 en 1996-2008



Dit kan nog beter worden geïllustreerd door niet twee groepen jaren te nemen, maar de tijd als glijdende schaal te beschouwen. Dat is gebeurd in figuur 3, waar de samenhangen tussen 1 t/m 6 jaren verschil in de tijd zijn uitgezet. Vooral de samenhangen over een langere periode (4, 5 of 6 jaren) zijn in de tijd fors toegenomen, van rond de 0,3 halverwege jaren tachtig tot boven de 0,6 na het jaar 2000.

Dit bevestigt het beeld dat uit figuur 2 naar voren kwam van eindstanden die door de jaren steeds meer op elkaar lijken. Dit lijkt in contrast te zijn met hetgeen er dit seizoen aan de hand is: de top drie presteert relatief slecht in vergelijking met andere jaren. In de correlatie over de gehele eindstand (na 29 speelronden) komt dat niet tot uitdrukking, want deze is nog steeds 0,78 met het voorgaande jaar. Dat komt omdat er toch ook veel clubs zijn die wel in de lijn van de verwachtingen presteren. Als we alleen de verschuivingen in de bovenste 5 zouden meten, dan zien we de verschuivingen in de cijfers wel direct terug.

FIGUUR 3: Verloop in de tijd van de correlaties tussen eindstanden eredivisie in de periode 1981-2008



Noten

[1] In "De wetten van Koning Voetbal" van Arnout Japsers, verschenen in *Natuur, Wetenschap en Techniek*, nr. 2 2009 (zie ook http://www.natutech.nl/00/nt/nl/47/artikel/24983/De_wetten_van_koning_Voetbal.html), wordt Pearson's correlatiecoëfficiënt gebruikt. Deze is bedoeld voor kwantitatieve gegevens en minder geschikt voor rangorde-gegevens zoals de eindstand van de eredivisie.

[2] In dezelfde publicatie wordt geselecteerd op niet-gedegreerde clubs. Dit is geen gelukkige keuze, omdat de uitspraken, zeker als over meer jaren wordt gekeken, sterk vertekend worden. Bovendien wordt later in de conclusie deze selectie genegeerd.

[3] De conclusie van de eerste wet uit de genoemde publicatie luidt: "De eindstand van een club voorspelt voor een groot deel het resultaat in het volgende seizoen. Maar het zegt nauwelijks iets over de verwachte prestaties over drie jaar. Clubs verschuiven blijkbaar langzaam in prestaties." Deze conclusie is onjuist: de samenhangen tussen verderliggende jaren is veel hoger dan in het artikel wordt gesuggereerd. De samenhang neemt slechts geleidelijk af, en is bovendien de laatste jaren fors toegenomen.