



Zijn de wedkantoren te verslaan?

De efficiëntie van quoteringen in de Eredivisie

Bachelorscriptie Algemene Economie

Onderwijsinstelling: Erasmus Universiteit Rotterdam

Erasmus School of Economics

Auteur: Martijn van Gijlswijk

Studentnummer: 427334

Begeleider: Prof. Dr. A.J. Dur

Tweede beoordelaar: Dr. T.L.P.R. Peeters

Datum: 11 juli 2018

Samenvatting

In dit onderzoek staat de volgende vraag centraal: *Hoe efficiënt zijn de quoteringen van wedkantoren in het voorspellen van de uitkomsten van wedstrijden in de Eredivisie?* Om deze vraag te onderzoeken is er informatie verzameld over het eredivisieseizoen 2016/2017. Zo zijn de marktwaardes van de teams in dit model meegenomen, het onderling resultaat, het wel of niet spelen op kunstgras, het aantal afwezige spelers, eventuele tussentijdse wedstrijden, verschillende *moving average* variabelen die slaan op het aantal voor-en tegendoelpunten en het aantal punten van zowel het thuisteam als het uitteam en de quoteringen van de wedkantoren. Allereerst werd er gekeken in hoeverre de quoteringen verklaard konden worden aan de hand van de eerder genoemde variabelen. De variatie in de quoteringen van de wedkantoren kon voor ruim 91% verklaard worden. Vervolgens is er aan de hand van een binair probit model, en de BIC-waardes hiervan, onderzocht welk forecastmodel het beste in staat is om te voorspellen of het thuisteam daadwerkelijk wint of niet. Alleen de marktwaardes van het thuis- en het uitteam bleken verklarende kracht te hebben. Vervolgens is dit *out-of-sample* getest in het daaropvolgende eredivisieseizoen. Het model is *out-of-sample* getest met behulp van twee wedstrategieën, namelijk de unit bet strategie en de unit win strategie. Uiteindelijk bleken beide methodes niet winstgevend. Aangezien de quoteringen geen verklarende kracht hebben in het forecastmodel, kan er geconcludeerd worden dat ze niet efficiënt zijn in het voorspellen van uitkomsten van eredivisiewedstrijden. Dit gaat tegen de bestaande literatuur in, aangezien die aangeeft dat forecast modellen die gebaseerd zijn op de quoteringen van de wedkantoren beter blijken te zijn, of tenminste even goed, als forecast modellen die enkel gebaseerd zijn op sport gerelateerde verklarende variabelen (Forrest, Goddard, & Simmons, 2005; Song, Boulier, & Stekler, 2007).

Inhoudsopgave

Introductie	4
Theoretisch kader	5
Data	7
Methodologie	14
Resultaten	18
Conclusie	23
Discussie	24
Bijlage	27
Referenties	33

Introductie

Het wedden op sportwedstrijden is al jaren populair onder reguliere gokkers en professionals. De laatste jaren heeft het echter ook de aandacht getrokken van onderzoekers. Wedkantoren berekenen de quoteringen van sportwedstrijden op basis van statistieken. De gokker heeft dan de keuze om deze quotering te accepteren en dus in te zetten, of om niet in te zetten. Het doel van de wedkantoren is om op lange termijn winst te behalen. Dit realiseren zij door een bepaalde premium te rekenen, de winstmarge. Deze marge moet hoog genoeg zijn om een lange-termijn winst te behalen, maar niet zo hoog zijn dat de gokker bij een concurrent gaat gokken. De quotering van een wedkantoor kan worden gezien als een voorspelling door experts van de uitslag van een wedstrijd (Pope en Peel, 1989). De centrale vraagstelling van dit paper is:

Hoe efficiënt zijn de quoteringen van wedkantoren in het voorspellen van de uitkomsten van wedstrijden in de Eredivisie?

Aangezien het niet bekend is welk model wedkantoren gebruiken om deze quoteringen te berekenen, wordt er eerst onderzocht hoe deze quoteringen tot stand komen. In dit paper wordt er gepoogd om dit model zo goed mogelijk na te bootsen door gebruik te maken van verschillende statistieken die voor de aanvang van de wedstrijd bekend zijn. Vervolgens wordt er voor het seizoen 2016/2017 een binair probit model gemaakt, die de uitkomsten van de Eredivisie zo goed mogelijk voorspellen. Dit model wordt vervolgens out-of-sample getest voor het seizoen 2017/2018. Uiteindelijk wordt er gekeken of in het jaar 2017/2018 het wedkantoor verslagen kan worden, aan de hand van twee wedstrategieën, door het efficiënte model van het seizoen ervoor te gebruiken.

In 2011 is er al eens onderzoek gedaan naar de efficiëntie van de quoteringen van wedkantoren in de Premier League voor het seizoen 2006/2007. Uit dit onderzoek bleek dat de quotering efficiënt is als voorspeller van de uitkomsten van de Premier League (Xu, 2011). Echter bleek ook het aantal gele kaarten dat het thuisteam de vorige wedstrijd kreeg een significante variabele. Het model dat deze variabele en de quoteringen van het wedkantoor bevatte, bleek bij toepassing op het jaar 2007/2008 geen winstgevende strategie (Xu, 2011). Doordat in dit paper de

hoeveelheid variabelen is uitgebreid en gebruik wordt gemaakt van *moving average* variabelen, in plaats van variabelen die enkel verwijzen naar de vorige wedstrijd, denk ik een model te kunnen ontwikkelen dat wel winstgevend is wanneer het out-of-sample wordt gebruikt.

In het vervolg van dit paper ga ik verder in op de al bestaande literatuur omtrent dit onderwerp. Vervolgens bespreek ik de methodiek en data die bij dit empirisch onderzoek zijn gebruikt. Dan wordt er verder ingegaan op de resultaten van dit onderzoek. Ten slotte wordt het paper afgesloten met een conclusie en een discussie.

Theoretisch kader

Voorspellingen

Voetbal is een van de populairste sporten in de wereld. Wereldwijd wedden miljoenen mensen op deze sport. Wat aangenomen wordt door de meeste gokkers is dat kennis over voetbal leidt tot betere voorspellingen. Dit is in de realiteit vaak echter niet het geval. Het gevoel dat gokkers door hun voetbalkennis betere voorspellingen denken te kunnen maken is te verklaren door de "*illusion of control*" (Khazaal et al., 2012). De "*illusion of control*" slaat op het feit dat mensen geloven dat ze controle kunnen uitoefenen, of tenminste de uitkomst kunnen beïnvloeden, van een bepaalde gebeurtenis (Langer, 1975).

Er kunnen echter wel betere voorspellingen gemaakt worden met behulp van data-analyse. Zo zijn er *trading-companies* die met behulp van kansberekening toch een lange-termijnwinst proberen te behalen (Chellel & Massa, 2017). Brown en Sauer (1993) hebben bewezen dat er observationele variabelen zijn die gebruikt kunnen worden bij het voorspellen van professionele basketbalwedstrijden, die niet in de quoteringen verwerkt zijn.

Quoteringen

Wedkantoren houden de precieze berekeningen van de quoteringen geheim. Wel is er bekend dat de quoteringen gebaseerd zijn op statistieken en de publieke opinie. Naar verwachting zullen de variabelen in mijn model enkel een deel van de quoteringen verklaren, aangezien er geen maat voor publieke opinie in mijn model zit verwerkt en het model een gelimiteerd aantal variabelen bevat.

Het is maar de vraag of de quoteringen van wedkantoren efficiënt zijn. Zo heeft een onderzoek naar de *Association Football betting market* in het Verenigd Koninkrijk aangetoond dat er bewijs is voor inefficiëntie van de quoteringen, al werd er geen winstgevende wedstrategie gevonden gedurende deze sample periode (Pope & Peel, 1989). De wedmarkt met vaste quoteringen lijkt daarom te voldoen aan het belangrijkste criteria voor markt efficiëntie, namelijk dat er geen abnormale winsten behaald kunnen worden met behulp van één wedmodel. Wel werd er informatie gevonden dat de quoteringen niet alle beschikbare informatie reflecteren, waardoor gokkers bepaalde strategieën kunnen toepassen om hun verliezen te beperken (Pope & Peel, 1989).

Ook werd er bewijs gevonden dat de quoteringen voor de Engelse voetbal competitie in de jaren 1999 en 2000 niet geheel efficiënt zijn. Onderzoekers waren hier in staat een winst te genereren van +8% in mei en april in de jaren 1999 en 2000 (Goddard & Asimakopoulos, 2004). Echter waren zij in sommige andere maanden niet in staat winst te genereren, waardoor er in het seizoen 1999 een totale winst van 0.3% werd behaald. In het seizoen 2000 zorgde de implementatie van het model voor een verlies van 5.5%. Wel bewezen zij dat er naast variabelen die gebaseerd zijn op historische wedstrijdstatistieken, ook andere relevante verklarende variabelen (zoals afstand tussen de stadions van het thuisteam en uitteam, promotie en degradatie zorgen en het wel of niet actief zijn in de bekercompetitie) zijn met verklarende kracht.

Kuypers (2000) was wel in staat een totale winst in het seizoen 1994/1995 te behalen. Hij vond dat de quoteringen van Engelse voetbalwedstrijden inefficiënt waren. Hij stelde een forecast model op voor het seizoen 1993/1994. Vervolgens

testte hij dit model een seizoen later. Met dat model was hij in het 1994/1995 in staat om een totale winst van 7% te behalen na het betalen van belastingen.

Ondanks dat quoteringen niet altijd volledig efficiënt zijn, kunnen zij wel voorspellende waarde hebben voor de uitkomst van een wedstrijd. Zo vonden Štrumbelj en Šikonja (2010) dat de quoteringen van voetbalwedstrijden van sommige wedkantoren meer voorspellende waarde hebben dan de quoteringen van andere wedkantoren. Ook kwamen zij erachter dat effectiviteit van de quoteringen toegenomen is over tijd. Forecast modellen die gebaseerd zijn op de quoteringen van de wedkantoren blijken beter te zijn, of tenminste even goed, als forecast modellen die enkel gebaseerd zijn op sport gerelateerde verklarende variabelen (Forrest, Goddard, & Simmons, 2005; Song, Boulier, & Stekler, 2007).

De verwachting is daarom dat de quoteringen van voetbalwedstrijden in de eredivisie voor het seizoen 2016/2017 inefficiënt zijn, maar wel een voorspellende waarde hebben voor de uitkomst van de eredivisiewedstrijden. Door alle variabelen met een voorspellende waarde in het forecast model te verwerken wordt er vermoed dat er, met behulp van het forecast, model winst gemaakt kan worden in het jaar 2017/2018.

Data

De dataset die gebruikt is voor dit onderzoek bevat verschillende variabelen die informatie geven over de eredivisiewedstrijden die gespeeld zijn in het seizoen 2016/2017. De dataset bevat de volgende variabelen:

Winst thuisteam

De dummyvariabele winst thuisteam (Winst TT) heeft de waarde één wanneer het thuisteam de wedstrijd won. Wanneer het thuisteam gelijkspelde of verloor heeft deze variabele de waarde nul.

Quoteringen

De dataset bevat informatie over de quoteringen van alle eredivisiewedstrijden van verschillende online wedkantoren. De quoteringen in de dataset zijn de quoteringen die als eerste gepubliceerd zijn door de wedkantoren zelf. De quoteringen kunnen in die periode tot de aanvang van de wedstrijd nog iets wijzigen. Een voorbeeld wanneer de wedkantoren hun quoteringen in de loop van de tijd iets aan kunnen passen is wanneer er in de loop van de periode niet proportioneel veel op het thuisteam in wordt gezet. De wedkantoren kunnen dan de quotering, die slaat op de winst van het thuisteam, iets naar beneden aanpassen om gokkers te ontmoedigen om op het thuisteam in te zetten. De quoteringen voor een gelijkspel en die voor de winst van het uitteam kunnen dan iets omhoog aangepast worden. Een wedkantoor kan dit doen om zijn risico te spreiden. Aangezien er geen data beschikbaar is van de gemiddelde quotering tot de aanvang voor de wedstrijd, is er voor gekozen om toch de eerst bekende quoteringen te gebruiken. Wel moet men zich realiseren dat de quotering bij aanvang van de wedstrijd iets kunnen verschillen van de quoteringen in deze dataset. De dataset bevat de quoteringen van het wedkantoor Bet365. Er is voor de quotering van Bet365 gekozen, omdat deze de laagste winstmarge rekenen (en dus de hoogste quoteringen geven) van alle wedkantoren die hun quoteringen van de eredivisie gepubliceerd hebben. De verwachting is dat hoe lager de quotering met betrekking tot winst van het thuisteam is, hoe hoger de kans dat het thuisteam wint. Dit is de verwachting omdat bij een lage quotering men minder geld terug krijgt wanneer de voorspelling daadwerkelijk uitkomt. Dit betekent dat de wedkantoren ook verwachten, na het uitvoeren van een uitgebreide analyse, dat het thuisteam wint.

Moving average variabelen

Het model bevat ook verschillende *moving average* variabelen die gebaseerd zijn op de vorige één, drie, vijf of zeven gespeelde competitiewedstrijden van het betreffende seizoen. Er zullen vier verschillende modellen die allemaal *moving average* variabelen bevatten die gebaseerd zijn op verschillende aantal wedstrijden. Deze variabelen zijn: gemiddelde doelpunten voor in de laatste wedstrijden van het thuisteam (MA GVTT) en uitteam (MA GVUT), gemiddelde doelpunten tegen in de

laatste wedstrijden van het thuisteam (MA GTTT) en uitteam (MA GTUT) en gemiddelde aantal behaalde punten in de laatste wedstrijden van het thuisteam (MA PTT) en uitteam (MA PUT). De variabelen MA PTT en MA PUT kunnen enkel waardes aannemen die zich bevinden in het interval nul tot drie, aangezien er gemiddeld niet meer dan drie punten behaald kunnen worden per wedstrijd. De verwachting is dat de variabelen MA GVTT, MA GTUT, MA PTT positief gerelateerd zijn aan de kans dat het thuisteam wint en dat de variabelen MA GVUT, MA GTTT en MA PUT hier een negatief effect op hebben. Deze variabelen moeten in het model verwerkt worden, omdat deze variabelen de huidige vorm van de beide teams voor aanvang van de wedstrijd beschrijven.

Onderling resultaat

Ook het onderling resultaat (OR) in het verleden zal invloed hebben op hoe groot de kans is dat het thuisteam wint. Sommige speelstijlen zullen bepaalde teams beter liggen dan andere speelstijlen en zullen dus leiden tot een grotere winkans. Ook vangt deze variabele andere verschillen tussen de beide teams die constant zijn over tijd, zoals de afstand die het uitteam moet reizen naar het stadion van het thuisteam. De afstand tussen stadions kan voor thuisvoordeel zorgen, omdat de bezoekende ploeg vermoeid is door het reizen. Smith, Ciacciarelli, Serzan en Lambert (2000) hebben hier een klein effect voor gevonden. De variabele OR is tot stand gekomen door het gemiddelde aantal punten van het thuisteam in de laatste drie officiële thuiswedstrijden met elkaar te vergelijken. Oefenwedstrijden zijn niet meegenomen in de variabele OR, omdat in oefenwedstrijden vaak niet alle spelers die gewoonlijk tot het basiselftal behoren worden opgesteld. De variabele kan alleen waardes aannemen tussen het interval nul en drie. De verwachting is dat hoe meer punten het thuisteam in de laatste drie officiële onderlinge thuiswedstrijden tegen deze tegenstander heeft behaald, hoe groter de kans wordt geacht dat het thuisteam wint.

Marktwaarde

Voor dit model zijn de marktwaardes van de eredivisieteams gebruikt die gepubliceerd zijn op www.transfermarkt.nl. De marktwaarde van een speler is gebaseerd op kwaliteiten en potentie van een speler, niveau van de club, prestaties van de speler in het heden en verleden, de leeftijd van een speler en de positie van de speler. De marktwaardes van de speler worden bepaald door non-experts. Door de marktwaardes van alle spelers bij elkaar op te tellen, kunnen de marktwaardes van de selectie bepaald worden. De marktwaardes van de selectie kunnen door prestaties van de club in dat seizoen of door transfers wijzigen in de loop van het seizoen. Het model bevat de volgende twee marktwaarde variabelen: marktwaarde thuisteam (Markt TT) en marktwaarde uitteam (Markt UT). Thomas Peeters (2018) heeft in zijn onderzoek gevonden dat een model met de marktwaardes van www.transfermarkt.nl, het aantal spelers en het thuisvoordeel de prestaties van voetbalteams beter kan voorspellen dan dat de Fifa ranking en de ELO ranking dat kunnen. De verwachting is dat een hogere Markt TT leidt tot een verhoogde kans dat het thuisteam wint, aangezien de kwaliteit van de selectie dan hoger is. Een hogere Markt UT leidt waarschijnlijk tot een lagere kans dat het thuisteam wint, aangezien de tegenstander dan meer kwaliteit bezit. De marktwaardes worden twee keer per maand door de website gepubliceerd. De marktwaardes van de teams zijn gemeten in miljoenen euro's.

Blessures, schorsingen en andere afwezigen

Niet alle selectiespelers zijn elke wedstrijd beschikbaar voor een wedstrijd. Dit kan komen omdat zij geschorst zijn doordat zij gedurende het seizoen vijf gele kaarten hebben gekregen (met een schorsing van één wedstrijd tot gevolg), of een rode kaart gekregen hebben. De duur van de schorsing is dan afhankelijk van de ernst van de overtreding. Ook (langdurige) blessures kunnen ervoor zorgen dat spelers competitiewedstrijden missen. Er zijn ook andere redenen waarom selectiespelers niet bij een wedstrijd aanwezig kunnen zijn. Dit kan komen door privéredenen, zoals de geboorte van een kind of het overlijden van een familielid, of interlandverplichtingen, zoals de Africa Cup, die gehouden werd tijdens de reguliere competitie. De data over het aantal blessures, schorsingen en andere afwezigen is

afkomstig van Voetbal International. Dit voetbalblad publiceert voor aanvang van de eerste wedstrijd van de speelronde een overzicht met het aantal afwezig van elke eredivisieclub. De variabelen aantal blessures, schorsingen en andere afwezig van de thuisploeg (BS TT) en van de uitploeg (BS UT) bevatten informatie over het aantal afwezig van elke club in elke speelronde. Ook de spelers waarvan, op het moment van publiceren, het niet zeker is of zij mee kunnen zijn meegenomen. Deze keuze is gemaakt, omdat bij de publicatie van de quoteringen het ook nog niet bekend was of deze spelers daadwerkelijk mee konden spelen. Het is dus aannemelijk dat die de quoteringen ook beïnvloed heeft. De verwachting is dat wanneer de variabele BS TT toeneemt, de kans dat het thuisteam wint afneemt. Dit komt omdat het aantal spelers dat de trainer voor die wedstrijd tot zijn beschikking heeft, dan af is genomen, wat mogelijk leidt tot een afname in kwaliteit van het team. Het is te verwachten dat het tegenovergestelde het geval zal zijn wanneer de variabele BS UT toeneemt.

Tussentijdse wedstrijden: KNVB Beker, Europa League en Champions League

De tussentijdse wedstrijden zullen naar verwachting ook invloed hebben op de competitiewedstrijden. Zo zorgen tussentijdse wedstrijden ervoor dat teams minder tijd hebben om zich voor te bereiden op de competitiewedstrijden. Ook kan vermoeidheid een rol gaan spelen wanneer wedstrijden zich snel achter elkaar opvolgen. Een tussentijdse wedstrijd voor de thuisploeg, beïnvloedt naar verwachting de kans dat het thuisteam wint dus negatief. Dit is in overeenstemming met de bevindingen van Goddard & Asimakopoulos (2004), die hebben gevonden dat een gespeelde tussentijdse bekerwedstrijd van het thuisteam, de kans op winst van het thuisteam verkleint. Ook was een van hun bevindingen dat wanneer het uitteam een tussentijdse wedstrijd heeft gespeeld, de kans dat het thuisteam wint hoger is. De volgende tussentijdse wedstrijden kunnen er gespeeld worden: KNVB Beker, Europa League en Champions League. Alleen de officiële tussentijdse wedstrijden zijn meegenomen in het model. Tussentijdse wedstrijden zijn enkel meegenomen in het model, wanneer zij maximaal vijf dagen voor de aanvang van de competitie gespeeld werden. Deze keuze is gemaakt, omdat teams na vijf dagen genoeg tijd hebben om zich volledig voor te bereiden en zich weer volledig op te

laden voor de competitiewedstrijd. De variabelen voor een tussentijdse wedstrijd staan in het model als *TTBeker_CL_EL* voor het thuisteam en als *UTBeker_CL_EL* voor het uitteam. Dit is een dummyvariabele die de waarde één aanneemt wanneer zij inderdaad een tussentijdse wedstrijd hebben gespeeld. Indien dit niet het geval is neemt de variabele de waarde nul aan.

Ook is het mogelijk dat teams rekening houden met de tussentijdse wedstrijden voordat deze daadwerkelijk is gespeeld. Denk hierbij aan het sparen van de sterspelers tijdens de competitiewedstrijd, zodat zij onvermoeid aan bijvoorbeeld een Champions League wedstrijd kunnen beginnen. Wanneer het thuisteam daarom binnenkort een tussentijds wedstrijd heeft, is de kans dat het thuisteam wint mogelijk verkleint. Daarom zijn er nog twee dummyvariabelen aan het model toegevoegd. Wanneer binnen 5 dagen na het competitieduel een tussentijdse wedstrijd plaatsvindt, neemt de dummyvariabelen de waarde één aan. De variabelen voor deze tussentijdse wedstrijden die kort na het competitieduel gespeeld worden, staan in het model als *TTBeker_CL_EL_n* voor het thuisteam en als *UTBeker_CL_EL_n* voor het uitteam.

Kunstgras

Vertrouwdheid en bekendheid met het voetbalveld (leereffecten), kunnen voor thuisvoordeel zorgen. Vooral het hebben van een uitzonderlijk speelveld zorgt voor dit effect (Nevill & Holder, 1999). In de Eredivisie zijn er geen uitzonderlijkere voetbalvelden dan kunstgrasvelden. Zelfs tussen de kunstgrasvelden in de eredivisie zitten grote verschillen in structuur. Daarom zit ook de dummyvariabele *Kunstgras* in het model. Deze variabele neemt de waarde één aan wanneer de thuisclub op kunstgras speelt en de waarde nul aan wanneer de thuisclub op natuurgras speelt. Dat kunstgras zorgt voor een thuisvoordeel, wordt ook ondersteund door het onderzoek dat is uitgevoerd door van Ours en van Tuijl (2017). Zij vonden dat kunstgrasclubs in de Eredivisie een extra doelsaldo van 0,53 doelpunten hebben in een wedstrijd met natuurgrasclubs. Daarom is de verwachting dat het hebben van kunstgras de kans op winst van de thuisploeg verhoogt. Zes van de achttien eredivisieclubs spelen op kunstgras.

Doordat het model *moving average* variabelen bevat die terugkijken op de vorige zeven, vijf, drie of één wedstrijden bevatten de vier modellen niet evenveel variabelen. Het eredivisieseizoen bestaat uit 34 speelrondes, met elk 9 wedstrijden. Aangezien de *moving average* variabelen terugkijken op een bepaald aantal wedstrijden, bestaan de modellen niet uit 306 wedstrijden. Zo bevat het model met de *moving average* variabelen alleen de wedstrijden van speelronde acht tot en met 34, oftewel 243 wedstrijden. De andere modellen bestaan daarom uit 261 wedstrijden, 279 wedstrijden en 297 wedstrijden.

De gemiddelden van de samples blijven redelijk gelijk over tijd (zie tabel 1). Wat opvalt is dat het model dat de *moving average* variabelen minder tussentijdse wedstrijden bevat. Dit komt omdat dit model niet meer de eerste KNVB-bekerronde bevat. Aangezien elk eredivisieteam aan deze bekerronde deelnam, bevat het model 18 tussentijdse wedstrijden minder dan de vorige modellen. Wat ook opvalt is dat het aantal blessures, schorsingen en andere afwezigingen toeneemt naarmate het model minder observaties bevat. Dit komt omdat aan het begin van het seizoen er nog weinig schorsingen en blessures zullen zijn ten opzichte van later in het seizoen. De verschillen in de marktwaardes tussen de teams zijn groot. Zo heeft Go Ahead Eagles aan het begin van de herfst een marktwaarde van 6.2 miljoen, terwijl Ajax rond de winterstop een marktwaarde had van 128.2 miljoen.

Tabel 1: Beschrijvende statistieken met de gemiddeldes van de variabelen in de verschillende modellen

Model	Observaties	Winst TT	ORthuis	Markt..TT	Markt..UT	BSTT	BSUT	TTBeker..n	UTBeker..n	TTBeker/CL/E	UTBeker/CL/I	Kunstgras
1 MA	297	46.80%	1.65	34.83	34.03	2.91	3.01	14%	15%	14%	15%	33%
3 MA	279	47.67%	1.63	34.73	34.85	2.93	3.1	14%	15%	15%	15%	33%
5 MA	261	47.12%	1.66	34.55	34.81	2.98	3.15	14%	16%	16%	16%	34%
7 MA	243	47.44%	1.65	34.63	35.02	3	3.15	11%	13%	13%	12%	33%

Methodologie

Genormaliseerde kansen

Een wedstrijd heeft drie mogelijke uitkomsten, namelijk winst voor het thuisteam, een gelijkspel of winst voor het uitteam. Wedkantoren geven voor elk van deze uitkomst een andere quotering. Deze quoteringen kunnen worden gezien als bepaalde kansen die wedkantoren aan een gebeurtenis toeschrijven. Zo vermenigvuldigde Bet365 tijdens de wedstrijd Fc Utrecht- Fc Groningen op 16 september 2016 de gokker zijn inzet maal 1.57 bij winst van het thuisteam, maal 4 bij een gelijkspel en maal 5.5 bij winst van het uitteam.

Stel dat ik van mijn inzet 100 euro wil maken. Dan moet ik de volgende bedragen inzetten op de 3 uitkomsten om van mijn inzet 100 euro te maken.

$$\text{Winst thuisteam: } \frac{100}{1.57} = 63.69$$

$$\text{Gelijkspel: } \frac{100}{4} = 25$$

$$\text{Winst uitteam: } \frac{100}{5.5} = 18.18$$

Wanneer je deze inzetten bij elkaar optelt ($63.69 + 25 + 18.18 = 106.87$) kom je niet op 100 euro uit. Dit komt omdat de wedkantoren een zekere winstmarge doorberekenen. De winstmarge die de wedkantoren doorberekenen, kan op de volgende manier berekend worden:

$$1 + \text{winstmarge} = \frac{1}{\text{quotering winst thuisploeg}} + \frac{1}{\text{quotering gelijkspel}} + \frac{1}{\text{quotering winst uitploeg}}$$

In dit geval is de doorberekende winstmarge dus gelijk aan: $\left(\frac{1}{1.57} + \frac{1}{4} + \frac{1}{5.5}\right) - 1 = 6.876\%$. Nu de winstmarge voor deze wedstrijd bekend is, kan de echte kans berekend worden die de wedkantoren toeschrijven aan de winst van de thuisploeg. Dit kan gedaan worden door de quoteringen te corrigeren voor de winstmarge die de wedkantoren doorberekenen.

$$\text{Genormaliseerde kans dat het thuisteam wint} = \frac{\frac{1}{\text{quotering winst thuisploeg}}}{1 + \text{winstmarge}}$$

De kans dat het thuisteam wint tijdens de wedstrijd Fc Utrecht- Fc Groningen is volgens Bet365 gelijk aan $\frac{1}{1.06876} = 59.60\%$. De genormaliseerde kans dat het thuisteam wint (Kans TT) wordt gebruikt om de oorsprong van de quoteringen van wedkantoren te verklaren.

Verklaren van de quoteringen

Met behulp van een lineaire regressie, met de genormaliseerde kans dat het thuisteam wint (Kans TT) als afhankelijke variabele, wordt er onderzocht in hoeverre het model met de eerder beschreven variabelen de quoteringen kan verklaren. Er wordt gekeken naar de R^2 . De R^2 wordt ook wel de determinatiecoëfficiënt genoemd en geeft aan welk gedeelte van de variatie in de afhankelijke variabele verklaart kan worden door de verschillende onafhankelijke variabelen. Hoe hoger de R^2 is, hoe meer verklarende waarde het model heeft.

Forecast model seizoen 2016/2017

Vervolgens wordt er een forecast model gemaakt, die zo goed mogelijk kan voorspellen of het thuisteam gaat winnen. Dit model wordt gemaakt door een binaire probit regressie uit te voeren. Als afhankelijke variabele wordt de variabele winst thuisteam (Winst TT) gebruikt. Zoals eerder aangegeven kan deze dummyvariabele enkel de waarde één of nul aannemen. De in het data-gedeelte beschreven variabelen en de genormaliseerde kansen dat het thuisteam wint volgens het wedkantoor (Kans TT) kunnen elk als mogelijke parameter in het uiteindelijke model dienen. Er is gekozen voor een binair probit model, omdat de kans dat het thuisteam wint enkel tussen de 0% en 100% kan liggen. Wanneer er een lineair probit model gebruikt zou worden, zou de kans op winst voor het thuisteam volgens het model ook negatief en boven de 100% kunnen zijn. Dit is in de praktijk echter onmogelijk.

Om het beste forecast model te kiezen, moet er een afweging gemaakt worden tussen het aantal parameters (die de verklarende kracht van het model versterken) en de ruis die de niet-significante parameters met zich meebrengen (overfitting). Het uiteindelijke model wordt gekozen door te kijken naar de *Bayesian information criterion* (BIC) (Schwarz, 1978). Het model met de laagste *Bayesian information*

criterion (BIC) waarde wordt als beste model beschouwd. Er moet wel rekening gehouden worden met het feit dat de aantal observaties in de te vergelijken modellen gelijk zijn. Een groter aantal observaties zal automatisch leiden tot een hogere BIC waarde, ook al hoeft dat model niet meer voorspellende kracht te hebben. Dit blijkt ook uit de formule van de BIC waarde: $BIC = -2\ln(L) + k\ln(n)$. De L staat hier voor de gemaximaliseerde waarde van de *likelihood* functie. De k is gelijk aan het aantal parameters en de n staat voor het aantal observaties. Een toename in observaties zal dus automatisch leiden tot een hogere BIC waarde.

Wanneer het uiteindelijke model enkel de variabele genormaliseerde kans op winst van het thuisteam volgens het wedkantoor bevat (Kans TT), en het toevoegen van andere parameters het model niet verbeteren, kan er gesteld worden dat deze efficiënt is. Indien het uiteindelijke model helemaal niet de genormaliseerde kans op winst van het thuisteam volgens het wedkantoor bevat, kan er gesteld worden dat naast dat de quoteringen niet efficiënt zijn, ze ook geen voorspellende waarde hebben.

Model testen voor seizoen 2017/2018: Unit bet strategie

Nadat er een forecast model is gecreëerd voor het seizoen 2016/2017, wordt dit *out-of-sample* getest voor het daaropvolgende seizoen. Er wordt onderzocht of er na implementatie van het model winst gemaakt wordt.

De voorspelde waardes die uit het model voor het seizoen 2017/2018 komen, zijn gelijk aan Z-scores. Deze Z-scores worden omgezet tot p-scores, zodat ze gebruikt kunnen worden om de verwachte winst te berekenen. De verwachte winst kan als volgt berekend worden.

$$E [R_{y_{ijt}}] = (\text{quotering thuisploeg } (y_{ijt}) \times f_{ijt-1} (y_{ijt}) - 1$$

De variabele y_{ijt} beschrijft het uiteindelijke resultaat tussen thuisploeg i en uitploeg j ten tijde van t en de variabele f_{ijt-1} is gelijk aan de voorspelde p-score die berekend is met de informatie die beschikbaar was voor de wedstrijd. Alleen wanneer de verwachte winst groter dan nul is zal er één euro ingezet worden. Er zal gekeken worden naar de quotering voor winst van de thuisploeg. Deze strategie wordt de unit bet strategie genoemd. Indien de verwachte winst negatief is, zal er

niet ingezet worden. Wanneer het thuisteam daadwerkelijk wint en er €1,- is ingezet, zal dit bedrag vermenigvuldigd worden met de quoterings van het wedkantoor voor de winst van het thuisteam. Wint het thuisteam uiteindelijk niet, dan zal het ingezette bedrag verloren gaan. Vervolgens worden al deze winstbedragen bij elkaar opgeteld en alle verliesbedragen van dit bedrag afgehaald. Uiteindelijk kan er gekeken worden of de wedkantoren daadwerkelijk verslagen zijn. Wanneer het bedrag positief uitvalt, is het mogelijk om met het forecast model de wedkantoren te verslaan.

Unit win strategie

De unit win strategie zorgt mogelijk voor een grotere winst als bij de unit bet strategie. Zo vond Thomas Peeters (2018), dat de unit win strategie voor hogere winsten (of kleinere verliezen) zorgt dan de unit bet strategie. Aangezien er bij de unit bet strategie er standaard €1,- in wordt gezet, zijn de opbrengsten bij zeer zekere gebeurtenissen klein. De unit win strategie corrigeert hiervoor, door zo in te zetten dat de winst altijd gelijk is aan €1,- euro. Om dit te bereiken wordt er ingezet op de volgende manier, bij elke wedstrijd waar de verwachte winst positief is.

$$\frac{1}{\text{Quoterings thuisteam } (\gamma_{ijt}) - 1}$$

Een nadeel van deze methode is, dat er bij wedstrijden waarvan het vrij zeker is dat het thuisteam wint, een groot geldbedrag ingezet moet worden om een winst van €1,- behalen. Daarom zorgt deze methode er in het algemeen voor, dat de gokker gemiddeld gezien een groter risico neemt dan bij de unit bet strategie.

Passieve wedstrategie

Zelfs wanneer het uiteindelijk gewonnen bedrag negatief uitvalt, kan de wedstrategie toch nog enige effectiviteit hebben. Aangezien de wedkantoren winstmarges rekenen, is de verwachte winst van een niet-onderbouwde wedstrategie negatief. Wanneer de wedstrategie er simpelweg voor zorgt dat de winst minder negatief is, heeft deze dus enige effectiviteit. Daarom worden beide wedstrategieën vergeleken met de passieve wedstrategie. Bij de passieve wedstrategie wordt er op elke thuiswedstrijd €1,- op de thuisploeg gezet. Op deze

manier kan er gekeken worden of een eventuele winst door een wedstrategie niet komt, omdat er in dat seizoen toevallig veel vaker een thuisploeg won, zonder dat dit met het wedmodel te maken heeft.

Resultaten

Quoteringen verklaren

Allereerst gaat er gekeken worden in welke mate de verschillende modellen in staat zijn om de quoteringen te verklaren aan de hand van de R^2 . Elk model bevat de variabelen Onderling resultaat in thuiswedstrijden, Marktwaaarde thuisteam, Marktwaaarde uitteam, Blessures en schorsingen thuisteam, Blessures en schorsingen uitteam, Tussentijdse wedstrijd thuisteam (voor en na het competitieduel), Tussentijdse wedstrijd uitteam (voor en na het competitieduel) en de variabele Kunstgras. Het toevoegen van deze variabelen zal de R^2 altijd doen stijgen, behalve als de coëfficiënt van de variabele precies gelijk is aan nul (dan zal het onveranderd blijven). Alle modellen bevatten naast de eerder genoemde variabelen de *moving average* variabelen voor het aantal behaalde punten, aantal doelpunten voor en het aantal doelpunten tegen van het thuisteam en het uitteam. Model 1 bevat de *moving average* variabelen die gebaseerd zijn op de afgelopen 7 wedstrijden, model 2 kijkt terug op de afgelopen 5 wedstrijden, de *moving average* variabelen van model 3 zijn gebaseerd op de afgelopen 3 wedstrijden en model 4 kijkt enkel terug op de laatste wedstrijd.

Tabel 2: R-kwadraat van de verschillende modellen

	R-kwadraat	Aantal observaties
Model 1	91.20%	243
Model 2	89.39%	261
Model 3	87.98%	279
Model 4	86.81%	297

Uit de resultaten blijkt dat het model dat de *moving average* variabelen bevat die terug kijkt op de afgelopen 7 wedstrijden, het grootste deel van de variatie in de kans dat het thuisteam wint volgens de wedkantoren kan verklaren, namelijk 91.20% (Zie tabel 2 en tabel 7). Dit betekent dat het model redelijk goed in staat is om de quoteringen van het wedkantoor (Kans TT) te kunnen verklaren. Het aantal observaties bij gebruik van de verschillende modellen is niet gelijk. Zo kan model 1 pas gebruikt worden vanaf speelronde acht, omdat er dan pas data beschikbaar is over de eerste zeven speelrondes. Daarentegen kan model 4 al gebruikt worden vanaf speelronde twee, omdat dit enkel speelronde één nodig heeft om informatie te verschaffen over de *moving average* variabelen. Het feit dat de R-kwadraat van model 1 het hoogst is, heeft niks te maken met het aantal observaties. Zo heeft model 1 nog steeds de meest verklarende kracht wanneer bij alle modellen het aantal observaties wordt teruggebracht tot 243.

Forecast model 2016/2017

Om het beste forecast model te kiezen wordt er een binaire probit regressie uitgevoerd en wordt er gekeken naar het model met de laagste BIC waarde. Om de verschillende modellen met elkaar te kunnen vergelijken, is het van belang dat enkel modellen met dezelfde aantal observaties met elkaar vergeleken worden. Een lager aantal observaties zal er namelijk automatisch voor zorgen dat de BIC-waarde omlaag gaat, zonder dat dit de voorspellende waarde van het model verbeterd.

Allereerst zijn er vier regressies uitgevoerd met als afhankelijke variabele of de thuisploeg uiteindelijk won of niet (Winst TT) en als onafhankelijke variabelen de kans dat het thuisteam wint volgens de wedkantoren (Kans TT) en de modellen 1, 2, 3 en 4. De *moving average* variabelen bleken in alle modellen zwaar insignificant wat ervoor zorgde dat het de voorspellende kracht van het model schaadde (*overfitting*) (zie tabel 11). Wanneer de BIC-waarde van deze modellen vergeleken werd met de BIC-waarde van de regressie zonder de *moving average* variabelen (met hetzelfde aantal observaties), bleek de BIC-waarde lager zonder de *moving average* variabelen (zie tabel 12, 13 & 14).

Aangezien het model nu geen *moving average* variabelen bevat, kunnen de BIC-waardes met 306 observaties weer met elkaar vergeleken worden. De vorige

regressie wordt nu opnieuw uitgevoerd, maar dan worden ook speelronde 1 tot en met 7 nu in de regressie meegenomen (zie tabel 15). Zoals in de regressie te zien zijn de variabelen OR thuis, UTBeker_CL_EIn, BSTT en BSUT zwaar insignificant. Wederom wordt er hierdoor voor ruis gezorgd, dat de voorspellende kracht van het model verlaagt. Het verwijderen van deze variabelen uit het model, zorgt voor een verlaagde BIC waarde (zie tabel 16, 17 & 18).

Andermaal kan de voorspellende kracht van het model verbeterd worden door de ruis uit het model te filteren, oftewel insignificante variabelen te verwijderen. Dit maal worden de variabelen UTBeker_EL_CL, TTBeker_CL_EIn en Kunstgras uit het model verwijderd. De BIC waarde zal hierdoor opnieuw verder afnemen (zie tabel 18, 19 & 20).

Vervolgens worden de variabelen Kans TT en TTBeker_EL_CL uit het model verwijderd, wat opnieuw de BIC waarde zal laten afnemen (zie tabel 20, 21 & 22). De laatste stap in het proces is het verwijderen van de insignificante constante term. Dit laat wederom de BIC waarde dalen (zie tabel 22, 23 & 24).

De regressie die over blijft is het beste in staat om te voorspellen of het thuisteam wel of niet gaat winnen in het seizoen 2016/2017.

$$\gamma_{ijt}^* = 0.012718 \times \text{Marktwaarde TT}_{it} + -0.0157231 \times \text{Marktwaarde UT}_{jt}$$

De afhankelijke variabele γ_{ijt} is een Z-score die het resultaat beschrijft het van de wedstrijd ten tijde van t tussen de thuisploeg i en de uitploeg j. Het resultaat van de wedstrijd kan een winst voor de thuisploeg zijn of geen winst voor de thuisploeg zijn. Uit de resultaten blijkt dat de quoteringen van de wedkantoren niet efficiënt zijn en geen voorspellende waarde hebben, aangezien de regressie niet de variabele kans op winst van het thuisteam volgens de wedkantoren (Kans TT) bevat. Dit impliceert dat er een betere manier is om te voorspellen of het thuisteam daadwerkelijk wint of niet, dan door te kijken naar de quoteringen. Hierdoor zouden de wedkantoren verslagen kunnen worden.

Out-of-sample test voor het seizoen 2017/2018: Unit bet strategie

Wanneer het bovenstaande model gebruikt zou worden voor het seizoen 2016/2017, zou er ondanks de winstmarge die de wedkantoren rekenen een winst behaald worden van €14.02 (zie tabel 3) met de unit bet strategie.

Tabel 3: Unit bet strategie seizoen 2016/2017

Totaal aantal inzetten	137
Totale opbrengst	€151.02
Winst	€14.02
Winstpercentage	9.28%

Gokkers kunnen van dit model echter geen gebruik maken, aangezien dit model pas na het seizoen gemaakt kan worden. Wat interessanter is voor gokkers, is of het model ook in andere seizoenen winst oplevert. Daarom wordt het model out-of-sample getest voor het seizoen 2017/2018.

Wanneer de gokker precies geld inzet zal uitgelegd worden aan de hand van het volgende voorbeeld. De eerste wedstrijd van het seizoen 2017/2018 is Ado den Haag tegen Fc Utrecht. Ado den Haag heeft ten tijde van het eerste wedstrijd een marktwaarde van 9.3 miljoen euro en Fc Utrecht van 26.7 miljoen euro.

$$\gamma_{ijt}^* = 0.012718 \times 9.3_{it} + -0.0157231 \times 26.7_{jt} = -0.30152937$$

Deze Z-score worden vervolgens omgezet in een p-score. De Z-score die uit de formule voortkomt is gelijk aan een p-score van 0.38. Wanneer de P-score x de quotering groter is dan één euro, zal er ingezet worden. Aangezien de quotering voor de thuisploeg hier gelijk is aan 3.6, zal er ingezet worden ($0.38 \times 3.6 = 1.37$). In dit geval bleek het thuisteam uiteindelijk niet te winnen, waardoor de gokker uiteindelijk zijn euro verloor. Mocht het thuisteam wel hebben gewonnen, dan wordt de inzet van één euro vermenigvuldigt met de quotering van het thuisteam.

Dit proces wordt voor alle 306 wedstrijden herhaald. Implementatie van deze strategie zorgt voor een verlies van 10.99% (zie tabel 4).

Tabel 4: Unit bet strategie seizoen 2017/2018

Totaal aantal inzetten	139
Totale opbrengst	€123.73
Winst	-€15.27
Winstpercentage	-10.99%

Unit win strategie

Wanneer de unit win strategie zou worden toegepast in het seizoen 2016/2017 had dit voor een winst van 5.79 gezorgd (zie tabel 5). Toch kan ook dit model pas na dit seizoen gemaakt worden, waardoor gokkers het model niet op dit seizoen kunnen toepassen. Wat opvalt is dat de totale winst en het winstpercentage aanzienlijk lager zijn bij gebruik van deze wedmethode als die van de unit bet strategie.

Tabel 5: Unit win strategie seizoen 2016/2017

Totale inzet	€202.85
Gemiddelde inzet	€1.48
Totale opbrengst	€208.64
Totale winst	€5.79
Winstpercentage	2.85%

Het volgende seizoen kan de unit win strategie wel in de praktijk worden toegepast. Wanneer de verwachte return positief is, zal er afhankelijk van de quotering, zo ingezet worden dat de totale winst wanneer het thuisteam inderdaad wint gelijk is aan €1,-. Met deze strategie werd er in het seizoen 2017/2018 een verlies gehaald van €81,35 (zie tabel 6). Dit verlies is aanzienlijk groter als het verlies dat met de unit bet strategie werd gehaald.

Tabel 6: Unit win strategie seizoen 2017/2018

Totale inzet	€231.95
Gemiddelde inzet	€1.67
Totale opbrengst	€150.60
Totale winst	-€81.35
Winstpercentage	-35.07%

Passieve wedstrategie

Om de uitkomsten van de beide wedstrategieën te kunnen beoordelen, zullen deze vergeleken moeten worden met de passieve wedstrategie. Een passieve wedstrategie, waarbij er elke wedstrijd één euro op het thuisteam ingezet wordt, zou in het seizoen 2017/2018 resulteren in een totale opbrengst van 230 euro en 30 cent. Aangezien er in totaal 306 wedstrijden gespeeld worden, had deze strategie voor een verlies van 24,7% gezorgd. Als dezelfde passieve wedstrategie in het seizoen 2016/2017 toegepast zou worden, dan zou dit maar voor een verlies van 19 euro en 90 cent hebben gezorgd, oftewel een verliespercentage van 6.5%.

Conclusie

De hoofdvraag van dit paper is: *Hoe efficiënt zijn de quoteringen van wedkantoren in het voorspellen van de uitkomsten van wedstrijden in de Eredivisie?* Om de hoofdvraag te beantwoorden, werd er eerst gekeken naar in hoeverre de quoteringen verklaard kunnen worden aan de hand van mijn modellen. Het model dat de *moving average* variabelen bevat die terugkijkt op de laatste 7 wedstrijden, bleek het best in staat te zijn om de quoteringen te verklaren. Dit model is in staat om 91.20% van de variatie in de quoteringen te verklaren.

Vervolgens werd er met behulp van een binaire probit regressie het beste forecastmodel gekozen in het seizoen 2016/2017, aan de hand van het *Bayesian Information Criterion*. Een forecastmodel dat de marktwaardes van het thuis- en het

uitteam bevatte, bleek het best in staat om te voorspellen of het thuisteam de wedstrijd won of niet. Met dit forecastmodel was het mogelijk om de wedkantoren te verslaan in datzelfde seizoen. Aangezien de quoteringen geen voorspellende kracht hebben, kan er geconstateerd worden dat de quoteringen niet efficiënt zijn in het voorspellen van de uitkomsten van de wedstrijden van de eredivisie.

Het forecastmodel werd, aan de hand van twee verschillende wedstrategieën, ook nog getest in het seizoen 2017/2018. Met beide strategieën was het niet mogelijk om de wedkantoren te verslaan. Toch deed de unit bet strategie het aanzienlijk beter dan de passieve wedstrategie, wat duidt op enige effectiviteit van het wedmodel.

Discussie

Uit het onderzoek is naar voren gekomen dat de quoteringen niet efficiënt zijn in het voorspellen van eredivisiewedstrijden. Dit ging tegen mijn verwachtingen en tegen eerder onderzoek in, dat stelde dat forecast modellen die gebaseerd zijn op de quoteringen van de wedkantoren beter blijken te zijn, of tenminste even goed, als forecast modellen die enkel gebaseerd zijn op sport gerelateerde verklarende variabelen (Forrest, Goddard, & Simmons, 2005; Song, Boulier, & Stekler, 2007). Een reden hiervoor zou kunnen zijn dat de wedkantoren minder geavanceerd zijn in het voorspellen van uitslagen van Eredivisiewedstrijden, aangezien dit een relatief kleine competitie is, dan in het voorspellen van bijvoorbeeld Premier League wedstrijden, waar veel meer gokkers actief zijn. Echter zou dit er voor zorgen dat er kansen bestaan voor data-analisten om systematische winsten te behalen tijdens Eredivisiewedstrijden.

De verwachting was dat er met dit forecastmodel een winst gemaakt zou kunnen worden in het seizoen 2016/2017. Ondanks het feit dat dit onderzoek er met behulp van het forecastmodel niet in is geslaagd om een winst te behalen, was het verlies aanzienlijk lager bij gebruik van de unit bet strategie dan wanneer de passieve wedstrategie toegepast zou worden. In het seizoen 2017/2018 zou de passieve wedstrategie gezorgd hebben voor een verlies van 24.7%. Dit verlies is veel groter dan de verwachting. Verwacht zou worden dat dit verliespercentage ongeveer gelijk

zou zijn aan de winstmarge van de wedkantoren, wat in het seizoen 2017/2018 gelijk is aan 6.58%. Er kan dus gesteld worden dat het in het seizoen 2017/2018 sowieso een lastige opgave zou worden om een winst te behalen door op de winst voor het thuisteam in te zetten. In dit seizoen won het thuisteam namelijk veel minder vaak dan verwacht. Het zou daarom interessant zijn om het forecastmodel wederom te testen voor het komende eredivisieseizoen (2018/2019), om te zien of er dan wel een winst gemaakt zou kunnen worden.

De unit winst strategie bleek minder goed te presteren dan de unit bet strategie. Dit ging tegen de bevindingen van het onderzoek van Thomas Peeters (2018) in. Echter is er wel een substantieel verschil tussen beide onderzoeken. Zo werd er bij gebruik van mijn model enkel ingezet op winst van de thuisploeg, bij het onderzoek van Thomas Peeters werd er ook ingezet op gelijke spelen en overwinningen van de uitploeg.

Dit brengt ons gelijk bij de grootste limitatie van dit onderzoek. Door het gebruik van een binair probit model, wordt er enkel gekeken naar of het thuisteam wint of dat het thuisteam niet wint. Echter zijn er bij een voetbalwedstrijd niet twee uitkomsten, maar drie uitkomsten. Zo kan een voetbalwedstrijd ook eindigen in een gelijkspel. In dit onderzoek wordt een gelijkspel hetzelfde gewaardeerd als een nipte of als een ruime nederlaag van het thuisteam, namelijk als geen winst voor het thuisteam. Dit model is daarom niet in staat om gelijke spelen of winsten voor het uitteam te voorspellen. Aangezien er geen onderscheid wordt gemaakt tussen gelijke spelen en nederlagen van het thuisteam, zal naar alle waarschijnlijkheid ook de voorspellende kracht van overwinningen voor het thuisteam afnemen.

Andere limitaties van dit onderzoek zijn dat er enkel gekeken is naar de quoteringen van Bet365. Andere wedkantoren zijn misschien beter in staat om te voorspellen of het thuisteam wel of niet wint. Aangezien er geen andere data beschikbaar was, is er gebruik gemaakt van de eerst gepubliceerde quoteringen. Ondanks het feit dat de quoteringen vaak maar kleine veranderingen ondergaan, wijken deze quoteringen vaak toch iets af van de quoteringen vlak voor de aanvang van de wedstrijd. Een andere limitatie is dat de marktwaarde van alle clubs maar 2x per maand worden gepubliceerd. Ondanks dat het verschil tussen elke publicatie over

het algemeen niet heel groot is, wordt er geprefereerd dat deze data elke week beschikbaar komt. Dit zou de voorspellende kracht van het model naar alle waarschijnlijkheid vergroten. Ook is het gecreëerde forecast model maar gebaseerd op één seizoen. De voorspellende kracht van het model zou vergroot kunnen worden door gebruik te maken van data van meerdere seizoenen.

Zoals eerder genoemd zou het interessant zijn om het forecast model wederom te testen in het seizoen 2018/2019, om te zien of er dan wel een winst gemaakt kan worden. Ook zou het voor vervolgonderzoek interessant kunnen zijn om te kijken of de resultaten extern valide zijn, of dat er substantiële verschillen zijn tussen de verschillende Europese voetbalcompetities.

Bijlage

Source	SS	df	MS	Number of obs	=	243
Model	8.89791439	16	.556119649	F(16, 226)	=	146.40
Residual	.85850411	226	.003798691	Prob > F	=	0.0000
Total	9.7564185	242	.040315779	R-squared	=	0.9120
				Adj R-squared	=	0.9058
				Root MSE	=	.06163

Kansvanwinne-T	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
MAPTT7	.0077774	.0163163	0.48	0.634	-.0243742 .0399289
MAGVTT7	.044286	.0134863	3.28	0.001	.0177111 .070861
MAGTTT7	-.0302445	.0124951	-2.42	0.016	-.0548664 -.0056227
MAPUT7	.005227	.0156697	0.33	0.739	-.0256504 .0361044
MAGVUT7	-.0453658	.0130754	-3.47	0.001	-.071131 -.0196006
MAGTUT7	.0396493	.0127971	3.10	0.002	.0144325 .0648661
ORthuis	.0050035	.0056825	0.88	0.380	-.0061939 .0162009
MarktwaardeTT	.002928	.0002009	14.57	0.000	.0025321 .0033239
MarktwaardeUT	-.0029006	.0001893	-15.32	0.000	-.0032737 -.0025275
BSTT	-.004934	.0027415	-1.80	0.073	-.0103361 .0004681
BSUT	.0046348	.0024176	1.92	0.056	-.0001292 .0093988
TTBeker_CL_Eln	-.0136716	.0146133	-0.94	0.350	-.0424673 .0151242
UTBeker_CL_Eln	-.0231796	.0130753	-1.77	0.078	-.0489447 .0025854
TTBeker_CL_EL	.0063408	.0144516	0.44	0.661	-.0221363 .0348179
UTBeker_CL_EL	-.0037525	.0143751	-0.26	0.794	-.0320788 .0245738
Kunstgras	-.0186921	.0102058	-1.83	0.068	-.0388028 .0014186
_cons	.4241562	.0404881	10.48	0.000	.3443738 .5039386

Tabel 7: Lineaire regressie model 1

Source	SS	df	MS	Number of obs	=	261
Model	9.32061383	16	.582538364	F(16, 244)	=	128.53
Residual	1.1059002	244	.004532378	Prob > F	=	0.0000
Total	10.426514	260	.040101977	R-squared	=	0.8939
				Adj R-squared	=	0.8870
				Root MSE	=	.06732

Kansvanwinne-T	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
MAPTT5	.0037313	.0140113	0.27	0.790	-.0238673 .0313298
MAGVTT5	.0308967	.011861	2.60	0.010	.0075336 .0542598
MAGTTT5	-.0289434	.0107459	-2.69	0.008	-.0501099 -.0077768
MAPUT5	.0042396	.0147623	0.29	0.774	-.0248381 .0333173
MAGVUT5	-.0293549	.0117899	-2.49	0.013	-.052578 -.0061319
MAGTUT5	.029985	.011341	2.64	0.009	.0076462 .0523237
ORthuis	.0064345	.0060322	1.07	0.287	-.0054473 .0183164
MarktwaardeTT	.0030803	.0001966	15.67	0.000	.002693 .0034675
MarktwaardeUT	-.0030902	.0001852	-16.69	0.000	-.0034551 -.0027254
BSTT	-.004082	.0029264	-1.39	0.164	-.0098463 .0016823
BSUT	.0052616	.0025357	2.08	0.039	.0002669 .0102562
TTBeker_CL_Eln	.0118299	.0141628	0.84	0.404	-.016067 .0397269
UTBeker_CL_Eln	-.0101945	.013391	-0.76	0.447	-.0365712 .0161822
TTBeker_CL_EL	-.0020555	.0147791	-0.14	0.890	-.0311664 .0270555
UTBeker_CL_EL	-.009803	.0144368	-0.68	0.498	-.0382397 .0186337
Kunstgras	-.0179008	.0105308	-1.70	0.090	-.0386438 .0028421
_cons	.4336332	.0375707	11.54	0.000	.3596288 .5076375

Tabel 8: Lineaire regressie model 2

Source	SS	df	MS	Number of obs	=	279
Model	9.790119	16	.611882438	F(16, 262)	=	119.85
Residual	1.33765035	262	.005105536	Prob > F	=	0.0000
Total	11.1277694	278	.040027947	R-squared	=	0.8798
				Adj R-squared	=	0.8725
				Root MSE	=	.07145

Kansvanwinne-T	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
MAPTT3	.0088419	.0104833	0.84	0.400	-.0118004 .0294842
MAGVTT3	.015213	.0086355	1.76	0.079	-.0017909 .0322168
MAGTTT3	-.0142766	.0084749	-1.68	0.093	-.0309641 .002411
MAPUT3	.0045834	.0114584	0.40	0.689	-.0179789 .0271456
MAGVUT3	-.0195056	.0093532	-2.09	0.038	-.0379227 -.0010885
MAGTUT3	.0170646	.0090033	1.90	0.059	-.0006635 .0347926
ORthuis	.0068174	.0060519	1.13	0.261	-.0050991 .018734
MarktwaardeTT	.0032579	.000189	17.23	0.000	.0028856 .0036301
MarktwaardeUT	-.0032428	.0001801	-18.00	0.000	-.0035976 -.0028881
BSTT	-.0043791	.0029686	-1.48	0.141	-.0102244 .0014663
BSUT	.0046735	.0026052	1.79	0.074	-.0004562 .0098032
TTBeker_CL_Eln	.0158686	.0143382	1.11	0.269	-.0123642 .0441015
UTBeker_CL_Eln	-.011756	.0137763	-0.85	0.394	-.0388824 .0153703
TTBeker_CL_EL	.0059484	.0151974	0.39	0.696	-.0239762 .0358731
UTBeker_CL_EL	-.0114364	.0150123	-0.76	0.447	-.0409965 .0181236
Kunstgras	-.0248321	.0108421	-2.29	0.023	-.0461808 -.0034834
_cons	.4338802	.032142	13.50	0.000	.3705907 .4971696

Tabel 9: Lineaire regressie model 3

Source	SS	df	MS	Number of obs	=	297
Model	10.1744143	16	.635900893	F(16, 280)	=	115.18
Residual	1.54584204	280	.005520864	Prob > F	=	0.0000
Total	11.7202563	296	.039595461	R-squared	=	0.8681
				Adj R-squared	=	0.8606
				Root MSE	=	.0743

Kansvanwinne-T	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
MAPTT1	.0112039	.0066299	1.69	0.092	-.0018469 .0242546
MAGVTT1	-.0009164	.0057857	-0.16	0.874	-.0123054 .0104725
MAGTTT1	.0018052	.0048111	0.38	0.708	-.0076653 .0112757
MAPUT1	.0071457	.0059333	1.20	0.229	-.0045338 .0188252
MAGVUT1	-.0113547	.0049172	-2.31	0.022	-.021034 -.0016754
MAGTUT1	.0083875	.0055186	1.52	0.130	-.0024758 .0192508
ORthuis	.0075388	.0061367	1.23	0.220	-.0045411 .0196187
MarktwaardeTT	.0035262	.0001759	20.05	0.000	.0031799 .0038724
MarktwaardeUT	-.003434	.0001663	-20.65	0.000	-.0037613 -.0031066
BSTT	-.0019033	.0029969	-0.64	0.526	-.0078026 .003996
BSUT	.0063329	.0026053	2.43	0.016	.0012045 .0114613
TTBeker_CL_Eln	.0106887	.014505	0.74	0.462	-.017864 .0392414
UTBeker_CL_Eln	-.0088568	.0137924	-0.64	0.521	-.0360068 .0182932
TTBeker_CL_EL	.0082384	.0153103	0.54	0.591	-.0218995 .0383762
UTBeker_CL_EL	-.0116099	.014995	-0.77	0.439	-.0411272 .0179074
Kunstgras	-.0344088	.0108576	-3.17	0.002	-.0557817 -.0130358
_cons	.4145824	.0241329	17.18	0.000	.3670774 .4620874

Tabel 10: Lineaire regressie model 4

WinstTT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
KansvanwinnenTT	2.044374	1.501827	1.36	0.173	-.8991525	4.987901
MAPTT7	-.0447772	.3718711	-0.12	0.904	-.7736311	.6840767
MAGVTT7	-.0728606	.3248316	-0.22	0.823	-.709519	.5637977
MAGTTT7	.0232645	.2918083	0.08	0.936	-.5486692	.5951983
MAPUT7	.1928444	.356033	0.54	0.588	-.5049674	.8906561
MAGVUT7	.1107877	.3133526	0.35	0.724	-.5033722	.7249475
MAGTUT7	.2140112	.2922806	0.73	0.464	-.3588483	.7868706
ORthuis	.1169578	.1254358	0.93	0.351	-.1288919	.3628074
MarktwaardeTT	.0151726	.0069643	2.18	0.029	.0015229	.0288223
MarktwaardeUT	-.008961	.0066444	-1.35	0.177	-.0219839	.0040619
BSTT	-.071937	.0602277	-1.19	0.232	-.1899812	.0461072
BSUT	-.0752026	.0555108	-1.35	0.176	-.1840017	.0335966
TTBeker_CL_Eln	-.191782	.364226	-0.53	0.599	-.9056519	.5220878
UTBeker_CL_Eln	-.3730889	.3329284	-1.12	0.262	-1.025617	.2794388
TTBeker_CL_EL	-.7825602	.3402232	-2.30	0.021	-1.449385	-.115735
UTBeker_CL_EL	.8851593	.3367145	2.63	0.009	.225211	1.545108
Kunstgras	.3129816	.2311232	1.35	0.176	-.1400115	.7659747
_cons	-1.574588	1.118368	-1.41	0.159	-3.766549	.6173733

Tabel 11: Binaire probit regressie model 1 met Kans TT

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	243	-168.1857	-131.2125	18	298.4251	361.3002

Tabel 12: BIC waarde model 1 met Kans TT

WinstTT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
KansvanwinnenTT	1.472158	1.27271	1.16	0.247	-1.022308	3.966624
ORthuis	.1142119	.1236184	0.92	0.356	-.1280756	.3564994
MarktwaardeTT	.0156955	.0066053	2.38	0.017	.0027495	.0286416
MarktwaardeUT	-.0093547	.0060903	-1.54	0.125	-.0212915	.0025821
BSTT	-.0702472	.0591005	-1.19	0.235	-.1860821	.0455876
BSUT	-.0747091	.0538938	-1.39	0.166	-.1803391	.0309208
TTBeker_CL_Eln	-.2191742	.3576751	-0.61	0.540	-.9202045	.481856
UTBeker_CL_Eln	-.3913785	.3292666	-1.19	0.235	-1.036729	.2539722
TTBeker_CL_EL	-.7585876	.33331	-2.28	0.023	-1.411863	-.105312
UTBeker_CL_EL	.8917761	.3271452	2.73	0.006	.2505833	1.532969
Kunstgras	.3045814	.2237462	1.36	0.173	-.133953	.7431159
_cons	-.7190063	.6972284	-1.03	0.302	-2.085549	.6475363

Tabel 13: Binaire probit regressie na verwijdering MA variabelen (gebaseerd op speelronde 8 t/m 34)

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	243	-168.1857	-132.2574	12	288.5148	330.4316

Tabel 14: BIC waarde na verwijdering MA variabelen

WinstTT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
KansvanwinnenTT	1.704401	1.053937	1.62	0.106	-.3612785 3.77008
ORthuis	.082336	.1072761	0.77	0.443	-.1279214 .2925933
MarktwaardeTT	.0108699	.005116	2.12	0.034	.0008428 .020897
MarktwaardeUT	-.008562	.0050307	-1.70	0.089	-.018422 .0012979
BSTT	-.0469536	.0520492	-0.90	0.367	-.1489682 .055061
BSUT	-.0473518	.0469328	-1.01	0.313	-.1393385 .0446348
TTBeker_CL_Eln	-.3168978	.283759	-1.12	0.264	-.8730553 .2392597
UTBeker_CL_Eln	-.2500254	.278685	-0.90	0.370	-.796238 .2961871
TTBeker_CL_EL	-.5772094	.2765136	-2.09	0.037	-1.119166 -.0352527
UTBeker_CL_EL	.5193909	.2693315	1.93	0.054	-.0084891 1.047271
Kunstgras	.2169012	.1951453	1.11	0.266	-.1655765 .5993789
_cons	-.8320888	.5566564	-1.49	0.135	-1.923115 .2589377

Tabel 15: Binaire probit regressie na verwijdering MA variabelen gebaseerd op alle speelrondes

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	306	-210.9971	-172.8021	12	369.6042	414.2873

Tabel 16: BIC waarde regressie na verwijdering MA variabelen gebaseerd op alle speelrondes

WinstTT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
KansvanwinnenTT	1.746749	1.029039	1.70	0.090	-.2701309 3.763628
MarktwaardeTT	.0115622	.0050617	2.28	0.022	.0016414 .0214829
MarktwaardeUT	-.0090546	.0049507	-1.83	0.067	-.0187578 .0006486
TTBeker_CL_Eln	-.3897286	.2499674	-1.56	0.119	-.8796558 .1001986
TTBeker_CL_EL	-.4851434	.2695593	-1.80	0.072	-1.01347 .0431831
UTBeker_CL_EL	.3958698	.2585413	1.53	0.126	-.1108619 .9026015
Kunstgras	.1799949	.1858477	0.97	0.333	-.18426 .5442498
_cons	-1.002729	.5119218	-1.96	0.050	-2.006077 .0006192

Tabel 17: Binaire probit regressie

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	306	-210.9971	-174.2995	8	364.599	394.3877

Tabel 18: BIC waarde regressie tabel 17

WinstTT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
KansvanwinnenTT	1.465697	1.003642	1.46	0.144	-.5014055 3.432799
MarktwaardeTT	.009194	.0048859	1.88	0.060	-.0003822 .0187703
MarktwaardeUT	-.0084621	.0048212	-1.76	0.079	-.0179115 .0009873
TTBeker_CL_EL	-.3182665	.2361877	-1.35	0.178	-.7811859 .144653
_cons	-.770374	.4686075	-1.64	0.100	-1.688828 .1480798

Tabel 19: Binaire probit regressie

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	306	-210.9971	-176.9529	5	363.9057	382.5236

Tabel 20: BIC waarde regressie tabel 19

WinstTT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
MarktwaardeTT	.0139848	.0025656	5.45	0.000	.0089563 .0190133
MarktwaardeUT	-.0142762	.0028344	-5.04	0.000	-.0198316 -.0087208
_cons	-.121735	.1362976	-0.89	0.372	-.3888735 .1454034

Tabel 21: Binaire probit regressie

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	306	-210.9971	-178.9067	3	363.8135	374.9842

Tabel 22: BIC waarde tabel 21

WinstTT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
MarktwaardeTT	.012718	.0021204	6.00	0.000	.008562	.016874
MarktwaardeUT	-.0157231	.0023713	-6.63	0.000	-.0203707	-.0110756

Tabel 23: Binaire probit regressie

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	306	.	-179.3056	2	362.6113	370.0584

Tabel 24: BIC waarde tabel 23

Referenties

Chellel, K., Massa, A. (2017, 30 Oktober). A Well-Known Quant Firm Is Looking for Traders Who Want to Bet on Sports. Geraadpleegd van

<https://www.bloomberg.com/news/articles/2017-10-30/sports-gambling-is-quant-trading-firm-s-way-to-beat-market-odds>

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International journal of forecasting*, 21(3), 551-564.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23(1), 51-66.

Khazaal, Y., Chatton, A., Billieux, J., Bizzini, L., Monney, G., Fresard, E., ... & Khan, R. (2012). Effects of expertise on football betting. *Substance abuse treatment, prevention, and policy*, 7(1), 18.

Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics*, 32(11), 1353-1363.

Langer, E. J. (1975). The illusion of control. *Journal of personality and social psychology*, 32(2), 311.

Nevill, A. M., & Holder, R. L. (1999). Home advantage in sport: An overview of studies on the advantage of playing at home. *Sports Medicine*, 28, 221-236

Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1), 17-29.

Pope, P. F., & Peel, D. A. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 323-341.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.

Smith, D. R., Ciacciarelli, A., Serzan, J., & Lambert, D. (2000). Travel and the home advantage in professional sports. *Sociology of Sport Journal*, 17, 364 – 385

Song, C., Boulier, B. L., & Stekler, H. O. (2007). The comparative accuracy of judgmental and model forecasts of American football games. *International Journal of Forecasting*, 23(3), 405-413.

Štrumbelj, E., & Šikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*, 26(3), 482-488.

van Ours, J., & van Tuijl, M. (2017). Thuisvoordeel kunstgras misschien toch geen mythe. *Me Judice*.

Xu, J. S. (2011). ONLINE SPORTS GAMBLING: A LOOK INTO THE EFFICIENCY OF BOOKMAKERS'ODDS AS FORECASTS IN THE CASE OF ENGLISH PREMIER LEAGUE. Unpublished undergraduate dissertation). University of California, Berkeley.