LUNDS UNIVERSITET
Ekonomihögskolan

Effektiv simulering av fotbollsmatcher

Författare: Johan Hansson & Jesper Wallman

Handledare: Björn Holmquist

STAM 01 Examensarbete HT 2013

Sammanfattning

Denna studie sammanställer en beprövad statistisk modell, vilken skattar förväntat antal gjorda mål i fotbollsmatcher i den svenska fotbollens division 1. Modellen är upprättad genom en Poissonregression vilken gör det möjligt att beräkna två lags relativa styrka mot varandra i form av hur många mål de förväntas göra. Studiens syfte är att med 1092 matcher som grund se om modellen kan prediktera utfall i fotbollsmatcher med högre träffsäkerhet än spelmarknaden. Detta mäts genom spelbolaget Bet365 där spel placeras på marknadsformerna ”Asiatiskt handikapp” och ”Antal mål”. Slutligen undersöks om det med hjälp av implementerade spelstrategier går att utvinna/ öka en potentiell positiv avkastning. Studien kan inte påvisa att skattningarna är bättre än Bet365, då de använder en tillräckligt stor vinstmarginal för att täcka upp eventuell ineffektivitet i prissättningen. Däremot påvisas att det med strategier går att utvinna en positiv avkastning för referenssäsongen.

Nyckelord: Fotboll, Odds, Sannolikhet, Poissonregression, Strategiskt spelande, Asiatiskt Handikapp och Antal mål

Abstract

This study put together a well established statistical model which estimates the expected number of goals being scored in football games in the Swedish 1 division (the third tier). The model is based on a Poisson regression and makes it possible to calculate the relative strength between two teams in terms of how many goals they are expected to score. The study is based on data from 1092 games and investigates whether it is possible to predict with higher accuracy than the betting market. Higher accuracy is measured through bets being places with Bet365 on the “Asian Handicap” as well as “Total goals” markets. Finally the study examines if there is a possibility to generate or increase a potential positive return. The study shows no evidence that the model can predict with better accuracy than Bet365, mainly because their winning margins are large enough to cover up for possible inefficiencies in pricing. However it is shown that there is a possibility to generate a positive return for the reference season when strategic betting is applied.

Keywords: Football, Odds, Probability, Poisson, Strategic Gambling, Asian Handicap and Total Goals

Innehållsförteckning

1. INLEDNING 4

1.1 Bakgrund 4

1.2 Frågeställning 5

1.3 Syfte 5

1.4 Avgränsningar 5

2. FOTBOLL OCH ODDS 6

2.1 Division 1 i Sverige 6

2.2 Odds och spelmarknader 6

3 STATISTISK TEORI 8

3.1 Fotbollsresultat som en Poissonfördelning 9

3.2 Poissonregression 10

4 METOD – POISSONREGRESSION MÖTER FOTBOLL 12

4.1 Data och Parametrar 12

4.2 Poissonregression för fotbollsdata 15

4.3 Spelstrategier 20

5. ANALYS 23

6. RESULTAT 25

7. VIDARE FORSKNING 26

REFERENSLISTA 27

1. Inledning

Under det senaste decenniet har marknaden för vadslagning vid sportevenemang expanderat kraftigt. Denna expansion tillskrivs en ökad tillgänglighet som gjorts möjlig genom att bolag idag kan erbjuda spel på ett globalt plan via internetplattformer (Lotteriinspektionen 2011). Utvecklingen har medfört att traditionella vadhållningsagenter (traditionella spelbolag) idag konkurrerar med spelbörser, vilka påminner om en derivatmarknad. På dessa spelbörser kan spelare i realtid både köpa och sälja egna spel direkt till andra konsumenter (Betfair 2012a).

1.1 Bakgrund

Det större intresset har även hittat in i de akademiska korridorerna, där flertalet studier undersökt huruvida spelmarknaden är effektiv eller inte. Detta har i huvudsak gjorts genom att testa om spelmarknaden drivs av arbitrageintentioner (Sinkey & Logan 2011). Gemensamt konstaterar dessa studier att prissättningen på spelmarknaden mellan olika spelbolag inte fluktuerar nämnvärt. Detta stärker hypotesen om att prissättning på spelmarknaden är en effektiv aktivitet (Dixon & Coles 1997, Copton & Sigler 2012, Golec & Tamarkin 1991, Wever & Aadland 2012). Gemensamt för studierna är att de mäter effektiviteten på spelmarknaden med utgångspunkt i populära sporter i den absoluta eliten.

Dixon och Coles (1997) presenterade en statistisk modell som på ett effektivt sätt kan prediktera utfall i fotbollsmatcher. Denna studie var starten för flertalet kvantanalytiker som idag med statistiska verktyg försöker prediktera resultat i fotbollsmatcher. Dessa kvantanalytiker handlar på ovan nämnda odds-handelsplatser med stora likvida medel och är en av anledningarna till att spelmarknaden för fotbollens största ligor idag kan anses vara effektiv. Med stora likvida medel i rullning är de beroende av marknader med en hög omsättning, vilket fört dem in till de absoluta toppligorna (Quant Sports 2012).

I sin artikel från 2005 presenterar Wikberg skillnader mellan professionell och amatörmässig idrott. Wikberg (2005) understryker att ett mindre allmänintresse och ekonomiska faktorer är de huvudsakliga skiljaktigheterna. Trots detta erbjuder spelbolag omfattande spel på fotboll i lägre divisioner. Med ett trögrörligt informationsflöde (Betting Advice 2012), mindre allmänhetsintresse och svagare finansiell styrka skiljer sig prissättningen på dessa marknader väsentligt från toppligorna. Därför tvingas spelbolagen följa sina konkurrenter och reagera på prisförändringar sett till den risk de är benägna att acceptera (Spelaspel.se 2012).

Dixon och Coles (1997) diskuterar vidare i sin studie att deras modell är utformad att fungera på vilken liga och fotbollsmatch som helst. Eftersom Dixon och Coles (1997) med framgång använt statistisk applicering för prissättning av fotbollsmatcher i den engelska toppfotbollen, är det intressant att se hur en liknande matematisk modell presterar för fotbollsmatcher i lägre divisioner. Om modellen effektivt kan skatta utfall öppnar det en möjlighet att undersöka huruvida prissättning av fotbollsmatcher i lägre divisioner är en effektiv aktivitet eller ej. Om det skulle föreligga ineffektivitet öppnar det en möjlighet att genom strategiskt spelande kunna generera vinst och potentiellt även maximera denna.

1.2 Frågeställning

Den presenterade bakgrunden har lett fram till följande frågeställning:

- Går det med en matematisk modell prediktera utfall i fotbollsmatcher och med detta som grund, genom spelande generera en positiv avkastning?

- Går det utvinna strategier för hur spel ska utföras som kan förbättra avkastningen?

1.3 Syfte

Den här studien undersöker om det med statistiska verktyg (beprövad statistisk Poissonmodell Dixon och Coles 1997) går att modellera fotbollsmatcher i lägre divisioner. Skattningarna från modellen används sedan för jämförelse med spelbolag i syfte att se om denna kan prediktera med större styrka, det vill säga vara mer träffsäker. Träffsäkerheten mäts genom att undersöka ifall modellen kan generera en positiv avkastning. Vidare testas om det går att utvinna strategier som kan förbättra resultatet.

1.4 Avgränsningar

Studien avgränsar sig till att innehålla matcher i Svenska division 1 södra och division 1 norra. Totalt har det samlats in resultat för säsongerna 2010, 2011 och 2012 vilka omfattar 1092 matcher. Studien behandlar spelmarknader av formen asiatiskt handikapp och antal mål (se avsnitt 2.2).

I studien tas ingen hänsyn till spel utanför ligan, det vill säga exempelvis cupmatcher. Anledningen till detta är att spel mot lag i andra ligor inte anses vara relevant för den statistiska modell som används. Det tas vidare ingen hänsyn till matchens relevans, exempelvis om det gäller upp-/nedflyttning, ej heller andra tänkbara förklarande variabler såsom kortsiktig form, skador och nyförvärv. Detta för att det är ytterst svårt att hitta rimliga vikter till dessa parametrar samt att detta skulle medföra svårlösta datainsamlingsproblem. Odds har samlats in för den sista halvan av säsong 2012, det vill säga de sista 13 omgångarna. Dessa odds används som referenssäsong och undersöker hur modellen presterar.

2. Fotboll och odds

Oavsett om du ser en fotbollsmatch hemma i TV-soffan eller på plats är du omgiven av marknadsföring från spelbolag och möjlighet till spel. Tillgängligheten till spel är konstant och det enda som krävs för att placera ett spel idag är en enhet med internetuppkoppling. Liksom tillgängligheten har utbudet ökat lavinartat. Spelbolag erbjuder idag spel på ett stort antal fotbollsligor runt om i världen.

2.1 Division 1 i Sverige

Inom den svenska herrfotbollen är Allsvenskan den högsta divisionen tätt följd av Superettan. Därefter skuggar division 1 (den tredje divisionen) vilken innehåller 28 lag. De 28 lagen är delade i två (södra och norra) med 14 lag i vardera division. Lagen i respektive division spelar 2 matcher mot varandra, dessa är fördelade på en hemmamatch samt en bortamatch och summerar till totalt 26 matcher per lag och säsong. Vid vinst erhålles 3 poäng, vid ett oavgjort resultat erhålles 1 poäng och vid en förlust tilldelas laget 0 poäng. I slutet av säsongen summeras antalet poäng och det lag med flest poäng inom respektive division blir uppflyttad till Superettan (andra divisionen). Lagen på andra plats inom respektive division får möjligheten att spela kval mot lag i Superettan med möjlighet till befordran. De lagen som slutar på de tre sista platserna i respektive division blir degraderade och ersatta av lag från Division 2 (den fjärde divisionen). Om två lag slutar på samma poäng så bestäms deras placering efter målskillnad (gjorda mål minus insläppta mål) och vid oavgjort även här avgör antalet gjorda mål (SvFF 2012).

2.2 Odds och spelmarknader

I vadslagningssammanhang förknippas odds med förhållandet mellan vinst och insats vid spel om pengar (Weisstein 2013), således kan odds förknippas med sannolikheten för ett visst utfall. Beroende på vilket land man befinner sig i kan odds presenteras på olika sätt. I denna studie presenteras odds i decimalform vilket också från svensk ståndpunkt är det mest vedertagna.

Som tidigare nämnts, har spelmarknaden på global front växt kraftigt, vilket lett till uppkomsten av nya spelmarknader (Lotteriinspektionen 2011). Två marknader som utvecklats i denna expansion är spel på asiatiskt handikapp och antal mål. Det asiatiska handikappet har som namnet antyder sitt ursprung i Asien. Spelets utformning innebär att ett lag får ett försprång, ett positivt handikapp eller missgynnas av ett negativt handikapp. Detta handikapp adderas eller subtraheras från slutresultatet (se tabell 2.1) innan vinnaren kan utses (Betfair 2012b).

Asiatisk Handikapp	Odds	Sannolikhet
Hemmalag (-0,5)	1,95	0,5128
Bortalag(+0,5)	1,95	0,5128
Summa		1,0256

Tabell 2.1 exempel på två Asiatiska handikappspel.

I exemplet ovan illustreras att bortalaget innan matchens start tilldelas +0,5 mål. Detta innebär att om en match slutar oavgjort eller om bortalaget vinner kommer spel på bortalaget stå som vinnare (hemmalagsspel står som förlorare) eftersom +0,5 mål ska adderas på bortalagets totala målskörd. Det motsatta gäller om hemmalaget vinner matchen med ett mål (eller mer). Då kommer följaktligen de som satsat på hemmalaget att vinna sitt spel eftersom endast -0,5 mål ska subtraheras från hemmalagets målskörd. Anledningen till att man adderar/ subtraherar "icke hela mål" är att spelet endast ska ha två möjliga utfall, det vill säga hemmavinst eller bortavinst (och inte oavgjort). Fallet 0,5 är bara ett exempel och handikappet anpassas beroende på lagens relativa styrka. Under extrema förhållanden kan det uppkomma handikapp så stora som 5,5, vilket gör att handikapp kan anpassas till en match mellan i princip vilka lag som helst. Det bör påpekas att sannolikheterna inte summerar till ett, utan 1,0256 vilket kan framstå som aningen besynnerligt. Utan någon utförligare förklaring konstateras att detta representerar oddspremien som spelbolag tar ut på sina spel, det vill säga förväntad vinstmarginal på långsikt (Bet365 2012). Sannolikheten för respektive utfall räknas ut på följande sätt: Sannolikhet $= \frac{1}{Odds}$.

Marknaden för ”antalet mål” bygger på hur många mål som produceras under en fotbollsmatch. Likt asiatiskt handikapp används även för ”antalet mål” halva utfall så marknaden kan delas in i endast två möjliga utfall. Efter en match summeras hemmalagets och bortalagets totala målskörd ihop innan en vinnare kan utses. Odds på marknaden över antalet mål kan se ut enligt följande:

Antal mål	Odds	Sannolikhet
Över 2,5	1,57	0,6369
Under 2,5	2,35	0,4255
Summa		1,0624

Tabell 2.2 exempel på två antal mål spel.

Om en fotbollsmatch slutar till exempel 2-1, det vill säga en total målskörd på 3 mål (eller fler) kommer spel på över 2,5 mål stå som vinnare. Om matchen slutar 1-1, totalt 2 mål (eller färre) vinner spel på under 2,5 mål. Även marknaden ”antalet mål” erbjuds med skiftande målskörd och sträcker sig från över/under 0,5 mål till över/under 6,5 mål (i extrema fall ännu högre).

3 Statistisk teori

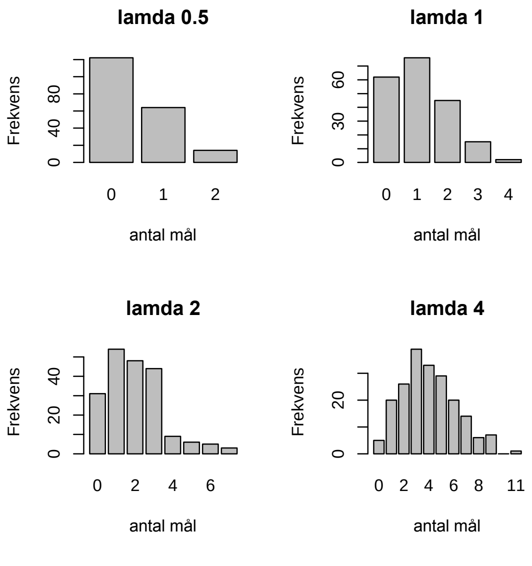
Enligt Groot (2008) kan antalet mål i en fotbollsmatch ses som utfallet av en Poissonfördelad variabel. Per definition kallas en diskret slumpvariabel X en Poissonvariabel när dess sannolikhetsfunktionen ser ut enligt följande:

$$p(x) = \exp(-\lambda)\frac{\lambda^x}{x!} \quad (3.1)$$

där $x = 0,1,2, \ldots$. Här representerar λ väntevärdet i form av ett positivt reellt tal. Sannolikhetsfunktionen (formel 3.1) anger sannolikheten att slumpvariabeln X antar ett specifikt värde x. Sannolikheterna för hela utfallsrummet summerar till 1:

$$\sum_{x=0}^{\infty} \exp(-\lambda)\frac{\lambda^x}{x!} = 1 \quad (3.2)$$

Den Poissonfördelade variabeln antalet mål i en fotbollsmatch, kan ses som en diskret variabel (Groot 2008). I figur 3.1 nedan illustreras att Poissonfördelningar med låga väntevärden (λ) är skeva från vänster till höger och att höga väntevärden ger en allt mer symmetrisk fördelning.

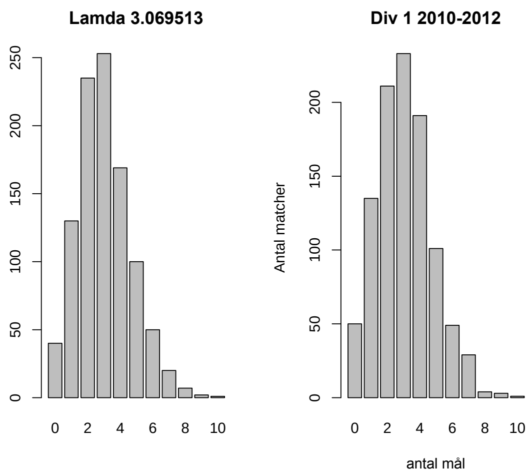


Figur 3.1 Poissonfördelningar med olika väntevärden (lamdavärden)

Genom sannolikhetsfunktionen går det att härleda väntevärde och varians för Poissonfördelningen. Om X är ungefär $Poiss(\lambda)$ är väntevärde lika med λ och varians lika med λ. Värt att notera är att väntevärde och varians alltså är identiska för Poissonfördelningen. För vidare matematiska bevis hänvisas läsaren till Blom m.fl. (2004).

3.1 Fotbollsresultat som en Poissonfördelning

För att se om studiens data över de hittills spelade 1007 matcherna följer en Poissonfördelning är det lämpligt att plotta detta. Genom att plotta ackumulerat antal mål bredvid en Poissonfördelning med samma medelvärde som studiens data, $\lambda = \frac{3091}{1007} \approx 3{,}067$, ges en övergripande bild om fördelningarnas överensstämmelse. Återigen bör understrykas att endast 1007 matcher har används vid skattningen vilket beror på att testet genomfördes mitt under pågående referenssäsongen. Nedan illustreras ett histogram över 1007 slumptal från en Poissonfördelning med lamdavärde (medelvärde) 3,07 samt ett histogram över det observerade antalet gjorda mål i matcherna.



Figur 3.2 Histogram över 1007 slumptal med lamdavärde 3,07 (t.v.) samt histogram över insamlad data (t.h.)

En snabbgranskning av stapeldiagrammen ovan ger en indikation på att antalet gjorda mål i Sveriges division 1 skulle kunna följa en Poissonfördelning, eftersom diagrammen någorlunda överensstämmer. Dixon och Coles (1997) belyser dock en problematik med denna jämförelse. Eftersom syftet med modellen i fråga är att anpassa en individuell Poissonfördelning för respektive lag och match är det kontrasterande att genom det ackumulerande antalet mål titta på hela datamaterialet som endast en Poissonfördelning (mer om detta i avsnitt 4.2). Däremot stärker detta hypotesen om att vi jobbar med data från en Poissonfördelning. Innan vi går vidare till nästa del vill vi undersöka huruvida denna hypotes är statistiskt signifikant eller ej. Detta görs genom ”Pearson’s chi-två goodness-of-fit-test” vilket är ett test för att undersöka kategorisk räknedata. Testet jämför om aktuell data överensstämmer med specifik fördelning enligt följande formel:

$$\chi^2 = 2\sum_{i=1}^{c} O_i \ln\left(\frac{O_i}{E_i}\right) \quad (3.3)$$

Här är O_i det observerade värdet, E_i är det skattade värdet sett till den Poissonfördelningen som observationerna antas följa och c motsvarar antalet klasser (totalt 11 stycken) över ackumulerat antal mål i de i matcherna (StatsDirect 2013a). Innan testet utförs upprättar vi våra hypoteser:

H_0: Fördelningen för ackumulerat antal gjorda mål följer en Poissonfördelning
H_1: Fördelningen för ackumulerat antal gjorda mål följer inte en Poissonfördelning

I tabellen nedan återfinns observerade och skattade värden:

Antal mål	**Observerade**	**Skattade**
0	50	46,77
1	135	143,56
2	211	220,33
3	233	225,43
4	191	172,99
5	101	106,20
6	49	54,33
7	29	23,82
8	4	9,14
9	3	3,12
10 eller fler mål	1	0,96

Tabell 3.1 Frekvens av antalet gjorda mål samt skattade värden för anpassad Poissonfördelning.

Dessa värden används för att räkna ut ”goodness-of-fit” enligt formel (3.3). Testet ger ett χ^2 lika med 9,44 med 9 frihetsgrader, vilket ger ett p-värde på 0,40. Därför kan nollhypotesen inte förkastas och studiens data över antalet gjorda mål i division 1 mellan säsongerna 2010-2012 ser ut att följa en Poissonfördelning.

3.2 Poissonregression

Poissonregressionen är en modell baserad på kategorisk räknedata. I denna regression antas den beroende variabeln y_i följa en poissonfördelning, där logaritmen av väntevärdet beskrivs med en linjärkombination av kända förklarande variabler. Den beroende variabeln y_i motsvarar antalet gånger en händelse inträffat. Variabeln x_i skildrar den linjära vektorn av förklarande variabler, det vill säga de variabler som har en effekt på värdena för y_i (Cameron & Pravin 1998). Eftersom y_i beror på x_i och följer Poissonfördelningen får regressionen följande sannolikhetsfunktion:

$$p(y_i|x_i) = \frac{\exp(-\mu_i)\,\mu_i^{y_i}}{y_i!} \qquad (3.4)$$

för $i = 1, \dots, k$. Här är $\mu_i = \mu(x_i, \beta) = E[y_i|x_i] = \exp(x_i^T\beta)$ väntevärdet. Vidare antas att Y_i är ungefär $Poiss(\mu_i)$, vilket gör att Y_i får följande utseende:

$$Y_i \approx Poiss(\exp(x_i^T\beta)) \qquad (3.5)$$

Storleken på produkterna i väntevärdesvektorn måste överensstämma för att göra matrismultiplikationen och skattningen möjlig. Detta är möjligt då y_i är en skalär, x_i^T en ”1 x d” vektor, β en ”d x 1” vektor och alltså är $x_i^T\beta$ av samma dimension som y_i. Parametrarna β skattas genom ”maximum likelihoodestimation” (MLE).

3.3 Maximum Likelihood

Den skattade parametern i MLE räknas fram som en funktion av parametervektorn (Rao & Toutenberg 1999). På så vis är denna skattning framtagen ur likelihoodfunktionen (formel 3.6) vilken tar produkten av alla tätheter (formel 3.5) från $i = 1, \dots, k$, logaritmerar dessa och skapar underlag för att lösa uttrycket. Formeln ser ut enligt följande:

$$\mathcal{L}(\beta;\ y_i, x_i) = log\left(\prod_{i=1}^{k} p(y_i|x_i)\right) \qquad (3.6)$$

Om vi använder formel 3.4 där $\mu_i = \exp(x_i^T\beta)$, fås följande ML-funktion:

$$\mathcal{L}(\beta;\ y_i, x_i) = log\left(\prod_{i=1}^{k} \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}\right)$$
$$= log\left(\prod_{i=1}^{k} \frac{\exp(x_i^T\beta)^{y_i}\ \exp(-\exp(x_i^T\beta))}{y_i!}\right)$$
$$= \sum_{i=1}^{k} log\left(\frac{\exp(y_i\,x_i^T\beta)\exp\left((-\exp(x_i^T\beta)\right)}{y_i!}\right)$$
$$= \sum_{i=1}^{k} ((y_i\,x_i^T\beta - \exp(x_i^T\beta) - \log(y_i!))$$

Parametern β skattas sedan genom att hitta maxvärdet för funktionen. Detta kan effektivt göras genom differentiering av likelihood-funktionen med avseende på parameter β_j, där $j = 1, \dots, d$.

$$\frac{\delta}{\delta\beta_j}\mathcal{L}(\beta) = \sum_{i=1}^{k}(y_i\, x_{ji} - x_{ji}\exp(x_i^T\beta))$$
$$= \sum_{i=1}^{k}(y_i - \exp(x_i^T\beta))x_{ji}$$

Uttrycket sätts sedan till noll för att kunna lösa maxvärdet för skattningen på β. Det är ur dessa ekvationer vi erhåller våra estimat (Cameron & Trivedi 1998).

$$\sum_{i=1}^{k}\left(y_i - \exp\left(x_i^T\hat{\beta}\right)\right)x_{ji} = 0 \quad (3.7)$$

för $j = 1,2, \dots, d$. Cameron och Trivedi (1998) diskuterar vidare att det inte finns en analytisk lösning på $\hat{\beta}$, utan att denna behöver lösas numeriskt. För att komma runt denna problemdynamik finns dataverktyg som effektivt kan estimera skattningarna i formel 3.7 . Den statistiska mjukvaran R är en bra miljö för detta då den redan är utrustad med en med inbyggd funktion för att utföra Poissonregressionen.

4 Metod – Poissonregression möter fotboll

Första steget för att kunna utföra en Poissonregression på fotboll är att samla in data. Det finns flertalet organisationer som håller sig med databaser över resultat och statistik för den svenska fotbollens division 1 serier. En väl strukturerad sådan plattform är Sportbladets (Aftonbladet 2013) statistiska databas, Betradar. Denna databas innehåller resultat och tabeller för division 1 från 2006 fram till idag. Nedan presenteras vilken data som inhämtats och hur denna implementerats.

4.1 Data och Parametrar

Då spelbolag ogärna exponerar odds för redan spelade matcher har insamlingen av odds behövts göras successivt när matcherna har spelats. Dessa odds har hämtats från spelföretaget Bet365 (2012) då de har ett bra marknadsutbud av odds för fotbollens division 1 i Sverige. Dessutom genomfördes en jämförelse med andra spelbolag vilken endast visade på marginella skillnader i odds, varpå vi valde att fokusera på odds från endast ett spelbolag. I denna oddsinsamling har det hänt att vissa matcher saknat odds (huvudsakligen för matcher i sista omgången) och därmed är spel inte möjliga att genomföra (vilket förklarar bortfallet). Oddsen har samlats in från omgång 14 för referenssäsongen, detta för att alla lag ska ha spelat mot varandra minst en gång när parametrarna skattas.

Totalt omfattar en säsong 364 matcher fördelade på 182 matcher $(14 * 13 = 182)$ per division. Vi har valt att använda data från tre säsonger och kommer därför ha ett datamaterial totalt omfattande $182 * 3\ (säsonger) * 2\ (divisioner) = 1092$ matcher. Valet av tre säsonger gjordes med hänsyn till att tävlingsbestämmelserna för division 1 ändrades efter 2009 års säsong. Vidare belyser Dixon och Coles (1997) att estimaten endast blir marginellt bättre när data från mer än 2 säsonger adderas. Efter 10 säsonger är effekten av adderad data meningslös (Dixon & Coles 1997).

För analys av våra data används Excel för att konvertera datafilen till csv-format, ett format som är kompatibelt med R. Våra matcher matas in i kronologisk ordning och visas i R enligt följande:

Match	Hemmalag	Bortalag	Mål Hemmalag	Mål Bortalag
1	Ytterby	Rosengård	1	2
2	Torslanda	Frölunda	0	2
3	Qviding	Norrby	0	2
4	Kristianstad	Värnamo	2	1
5	Husqvarna	Skövde	2	2
. . .	. . .	. . .	. . .	. . .
180	Sylvia	LB07	4	1
181	Frölunda	Torslanda	2	0
182	Rosengård	Ytterby	3	1

Tabell 4.1 – Resultat från Division 1 Södra säsong 2010

Inom fotboll (och sport generellt) finns det många studier som fastställer att det råder en fördel av att spela på hemmaplan. Dixon och Coles (1997) diskuterar detta faktum och inkluderar i sin Poissonmodell en parameter för fördel av att spela på hemmaplan. Hur mycket påverkan spel på hemmaplan innebär har inte konstaterats. Däremot stöttar flertalet studier de facto att hemmaplansfördelen är närvarande i all fotboll oavsett liga eller land (Pollard 2008). Om denna fördel antas stämma även för Sveriges division 1 bör en hemmaparameter implementeras i Poissonmodellen. För att statistiskt testa detta kan man genom ett traditionellt parat t-test avgöra skillnaden mellan medelvärdena för gjorda hemma- och bortamål (Körner & Wahlgren

2006). Om antalet hemmamål är större än antalet gjorda bortamål innebär det att det i division 1 görs fler mål på hemmaplan. Detta är synonymt med att det råder en hemmaplansfördel, då lagen generellt sett gör fler mål (presterar bättre) på hemmaplan. Den uppmärksamma läsaren kan påstå att effekten av hemmaplansfördel skiljer sig från lag till lag. Det ligger säkerligen en sanning i detta argument, men för att göra modelleringen mindre komplex har vi valt att fokusera på hemmaplansfördel som en global parameter. Dixon och Coles (1997) samt Pollard (2008) diskuterar detta faktum och använde med framgång i tidiga Poissonmodeller samma generella antagande.

Om X_1 representerar det totala antalet hemmamål i ett stort antal matcher och X_2 det totala antalet bortamål i samma antal matcher, så kan dessa anses vara approximativt normalfördelade. Om Y_{H_i} representerar antal mål som ett hemmalag gör i match i, så är $X_1 = Y_{H1} + Y_{H2} + \cdots + Y_{Hg}$. Om det förväntade antalet mål ett hemmalag gör är $\mu_1 = E\left(Y_g\right)$ och variansen är $\sigma_1^2 = Var\left(Y_g\right)$ så är X_1 ungefär normalfördelad $N(g\mu_1, \mathrm{g}\sigma_1^2)$. Ur detta följer att $\bar{Y}_H = \frac{X_1}{g}$ är ungefär normalfördelad $N(\mu_1, \frac{\sigma_1^2}{g})$. Samma gäller för antalet gjorda mål av ett bortalag som är ungefär normalfördelad $N(g\mu_2, g\sigma_2^2)$ och $\bar{Y}_B = \frac{(Y_{B1}+Y_{B2}+\cdots+Y_{Bg})}{g} = \frac{X_2}{g}$ blir då approximativt normalfördelad $N(\mu_2, \frac{\sigma_2^2}{g})$. Här är $\mu_2 = E(Y_B)$ det förväntade antalet gjorda mål ett bortalag gör i match nummer i, och $\sigma_2^2 = Var(Y_B)$ (Körner & Wahlgren 2006). Med följande bakgrund kan man testa hypotesen $H_0\colon \mu_1 = \mu_2$ mot alternativhypotesen $H_1\colon \mu_1 > \mu_2$ med följande teststorhet:

$$Z = \frac{(\bar{Y}_H - \bar{Y}_B)}{\sqrt{\frac{S_H^2}{g} + \frac{S_B^2}{g}}} \qquad (4.1)$$

där

$$s_H^2 = \frac{1}{g-1}\sum_{i=1}^{g}(Y_{Hi} - \bar{Y}_H)^2 \qquad (4.2)$$

Och motsvarande för s_B^2 där H, H_i och $\bar{Y}_H$ ersätts med B, B_i och $\bar{Y}_B$.

Dessa funktioner finns inbyggda i R:

$$Z = \frac{(1{,}76 - 1{,}33)}{0{,}08} = 5{,}38$$

I uttrycket ovan är 1,76 och 1,33 medelvärdena över antalet gjorda hemma- och bortamål. Här är den skattade standardavvikelsen för skillnaden i de två medelvärdena 0,08. I R fås ett P-värde fram som ligger långt under 1 %. Det bör påpekas att p-värdet måste multipliceras med 2 eftersom att vi använder ett ensidigt test. Trots detta förkastas nollhypotesen och det kan konstateras att det råder en fördel att spela på hemmaplan.

4.2 Poissonregression för fotbollsdata

Nästa steg är att utföra Poissonregressionen. Vi låter n (14) representera antalet lag i respektive division (uppdelad i södra och norra) och g (182) är lika med antalet matcher. Då kommer modellen att använda $i = 1, \dots, 2g$ (det vill säga $k = 2g$) eftersom vi undersöker antalet mål gjorda av hemmalag och bortalag i g antal matcher.

Regressionen av Y på X med utgångspunkt från formel (3.5) kommer att modellera resultaten (antalet mål) för våra fotbollsmatcher. Vektor Y_i skapas genom att alla resultat för respektive lag utgör dess element. Då varje match innehåller två bitar information (hemmalagets och bortalagets målskörd) kommer $Y_i \in \mathbb{N}_0$ representera ett heltal från 0 och uppåt där stora tal är osannolika. Detta gör att vektorn Y^T får följande utseende:

$$Y^T = \left(y^1_{i_1,j_1}, y^1_{j_1,i_1}, y^2_{i_2,j_2}, y^2_{j_2,i_2}, \dots, y^g_{i_g,j_g}, y^g_{j_g,i_g}\right) \qquad (4.3)$$

Där $y^c_{a,b}$ representerar antalet gjorda mål av lag a mot lag b i match c. I slutet av detta avsnitt kommer vi visa att det inför varje match finns behov att upprätta ett nytt väntevärde λ. Därför måste matcherna matas in i kronologisk ordning då det lägger grunden för att x_i^T kan tilldela ett lag ”rätt” parametervärden i nästkommande match.

Parametervektorn beskrivs av två parametrar vardera för de n lagen. Första delen i vektorn består av anfallsparametern α_k och den andra delen av försvarsparametern γ_k. Vektorn kommer se ut enligt följande:

$$\beta^T = (\alpha_{1,}, \dots, \alpha_n, \gamma_1, \dots, \gamma_n) \qquad (4.4)$$

$x_i^T \in \mathbb{R}^d$ bildar flera vektorer som tillsammans formar matrisen där $i = 1, \dots, 2g$ likt fallet för Y_i. x_i^T bildas så att rätt resultat för respektive lag plockas ut och används för att skatta respektive lags parametrar. Transponatet av vektor x_i (det vill säga x_i^T) multipliceras enligt regler för matrismultiplikation med parametervektorn (vilken är en kolumnvektor). Om detta ska utföras för en match mellan lag_a och lag_b kommer vi placera en etta (1) på $lag_a's$ plats och minus ett (-1) på $lag_b's$ plats. Resterande element i x_i^Tvektorn kommer vara fyllda med nollor för att ignorera alla andra lags parametrar. Vektorn ser ut enligt följande:

$$x^T = (0, \dots, 1, \dots, 0\ , 0, \dots, -1, \dots, 0) \quad (4.5)$$

Om detta stämmer kommer dimensionerna i multiplikationen att addera upp korrekt. Det bör återigen poängteras att vi använder väntevärdesvektorn från avsnitt 3.2 där $Y = (Y_1, Y_2, \dots, Y_k)^T$ motsvarar antalet gjorda mål, $X = (X_1, X_2, \dots, X_{2g})^T$plockar ut rätt matcher för parameterskattning och β är parametervektorn. Storleken på X är därmed $2g\ x\ 2n$ samtidigt som β bildar en $2n\ x\ 1$ vektor. När $X\beta$ multipliceras skapas en vektor av storleken $2g\ x\ 1$ vars dimension överensstämmer med Y. Väntevärdena i m (4.6) beräknas därmed på följande sätt när lag_a möter lag_b:

$$m = (\exp(\alpha_a - \gamma_b), \exp(\alpha_b - \gamma_a), \dots) \quad (4.6)$$

Som tidigare nämnts är $\exp(\alpha_a - \gamma_b)$ väntevärdet över antal gjorda mål av lag_a vid möte med lag_b. För att räkna ut hur många mål ett lag kommer göra används skillnaden i skattningarna mellan de respektive lagen.

Låt oss nu återvända till diskussionen om att införa en hemmaparameter. Om denna parameter inte implementeras spelar det ingen roll i vilken ordning våra resultat matas in, det vill säga om ett hemmalag ses som bortalag och vice versa. Resultatet i avsnitt 4.1 styrker argumentet att införa en hemmaparameter i modellen, vilket innebär att parametervektorn måste förändras. Det är sedan tidigare känt att $n = 14$ representerar antalet lag i respektive division. Eftersom modellen nu ska skatta anfallsparameter, försvarsparameter och ta hänsyn till hemmafördel måste parametermatrisen ”förlängas” ett steg, vilket gör att längden då blir $d = 2n + 1$.

Vi väljer att kalla hemmaparametern δ. Detta gör att parametervektorn som följer formel 4.4 får följande utseende:

$$\beta^T = (\alpha_{1,} \dots, \alpha_n, \gamma_1, \dots, \gamma_n, \delta) \quad (4.7)$$

Formel 4.5 kommer därmed förändras. Här är det viktigt att kunna skilja på hemma- och bortalaget. Detta kan implementeras genom att förlänga matrisen med en etta i slutet vilket gör det möjligt för ett hemmalags β-värde att plocka ut och skatta denna parameter.

$$x_i^T = (0, \dots, 1 \dots, 0, 0, \dots, -1, \dots, 0, 1) \quad (4.8)$$

För bortalaget kommer denna matris se ut enligt följande (där ettan i slutet av 4.8 ersatts med en nolla):

$$x_j^T = (0, \dots, 1 \dots, 0, 0, \dots, -1, \dots, 0, 0) \qquad (4.9)$$

Efter denna implementering är dimensionen på $x_i^T = 1\ x\ (2n + 1)$ vilket innebär att X får dimension $2g\ x\ (2n + 1)$. Denna multiplikation är fortfarande möjlig eftersom $X\beta$ innanför exponenten i väntevärdet, $\mu_i = \exp(x_i^T \beta)$, har samma dimension ($2g\ x\ 1$) som Y. Se exemplet nedan:

Match	Hemmalag	Bortalag	Hemmamål	Bortamål
1	Skövde	Lund	3	3
2	LB07	Skövde	1	3
3	Lund	LB07	4	1

Tabell 4.2 exempel på fotbollsresultat

Låt säga att matcherna ovan representerar en hypotetisk "miniserie", då skulle vektorerna Y och β samt matris X vilka beskrevs ovan att se ut enligt följande:

$$Y^T = (3,3,1,3,4,1) \qquad (4.10)$$

$$\beta^T = (\alpha_{Skövde}, \alpha_{Lund}, \alpha_{LB07}, \gamma_{Skövde}, \gamma_{Lund}, \gamma_{LB07}, \delta) \qquad (4.11)$$

$$\boldsymbol{X} = \begin{array}{c} \\ match1hemma \\ match1borta \\ match2hemma \\ match2borta \\ match3hemma \\ match3borta \end{array} \begin{array}{c} \begin{array}{ccccccc} SKÖ & LU & LB07 & SKÖ & LU & LB07 & HEMMA \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{pmatrix} \end{array} \qquad (4.12)$$

När hemmaparametern finns implementerad får multiplikationen av $X\beta$ följande utseende:

$$(X\beta)^T = (\alpha_S - \gamma_{LU} + \delta, \alpha_{LU} - \gamma_S, \alpha_{LB} - \gamma_S + \delta, \alpha_S - \gamma_{LB}, \alpha_{LU} - \gamma_{LB} + \delta, \alpha_{LB} - \gamma_{LU}). \quad (4.12)$$

Y representerar därför väntevärdet μ över hur många mål ett specifikt lag kommer göra i en specifik match (mot ett specifikt motstånd).

Med aktuella fotbollsdata i R används kod från bilaga 1 för att kunna utföra parameterskattningen. Bilaga 2 plockar ut de lag som är av intresse (eliminerar de lag som inte längre spelar i division 1). Nedan presenteras exempel på skattningar för division 1 södra efter omgång 14:

Lag	Anfall	Försvar
Örgryte	0,77	0,64
Lund	0,38	0,66
Oddevold	0,54	0,32
Kristianstad	0,34	0,22
Trollhättan	0,59	0,07
Skövde	0,30	0,13
Sylvia	0,60	0,12
Karlstad	0,38	0,03
Qviding	0,31	0,38
Utsikten	0,22	0,21
LB07	0,35	0,13
Sleipner	0,42	-0,02
Norrby	0,45	0,01
Gauthiod	0,34	-0,11
Hemmaparameter	0,30	

Tabell 4.3- Exempel på skattade parametrar

Dessa parametrar kommer ändras från omgång till omgång eftersom parametervärdet skattas med utgångspunkt i hur många mål ett lag gjort och släpt in i sina föregående matcher. Därmed kommer anfalls- och försvarsparametern ändras för varje lag och omgång. Därmed kan man konstatera att anfalls- och försvarsparametern för respektive lag också anpassas/ ändras för varje lag och omgång. Dessa parametrar används sedan som underlag för att skatta förväntad målskörd för respektive lag i nästkommande match. I punktlistan nedan ges ett exempel hur detta utförs i hemmamatchen Trollhättan mot Gauthiod:

- Estimerat antal mål för Trollhättan, $exp(0{,}591 - (-0{,}107) + 0{,}2952) \approx 2{,}70$
- Estimerat antal mål för Gauthiod, $exp(0{,}342 - 0{,}074) \approx 1{,}31$

I avsnitt 3.2 bevisades att studiens data för ackumulerat antal mål följer en Poissonfördelning. Vidare beskrivs antagandet att Poissonfördelningarnas väntevärden ändras för varje lag från match till match. Så länge likelihood funktionen är specificerad kan detta testas med ett likelihood kvottest (Cameron & Pravin 1998). Likelihood kvottestet undersöker om väntevärdet skiljer sig från det verkliga utfallet. I fotbollssammanshang innebär det att man jämför logkvoten mellan det faktiska resultatet och väntevärdet för att se om Poissonfördelningarnas väntevärden skiljer sig för varje lag och match (Cameron & Pravin 1998). Testet ser ut enligt följande

$$-2 * \sum_{i=1}^{k} y_i * \log\left(\frac{y_i}{\bar{y}}\right) \qquad (4.13)$$

där y_i motsvarar antalet gjorda mål i en specifik match och $\bar{y}$ representerar medelvärdet över antalet gjorda mål i de skattade matcherna. Cameron & Pravin 1998 bevisar att detta test asymptotiskt följer en χ^2 med k-1 frihetsgrader, där k representarar antalet skattade parametarar. Eftersom vi skattar en anfallsparameter och en försvarsparameter för varje match kommer $k = 2g = 364$. Innan vi går vidare och utför testet upprättar vi följande hypoteser:

H_0: De olika Poissonfördelningarnas väntevärden är desamma
H_1: De olika Poissonfördelningarnas väntevärden är olika

	χ^2-värde	**Frihetsgrader**	**P(>χ^2)**
LikelihoodRatio	154.67	363	0,000

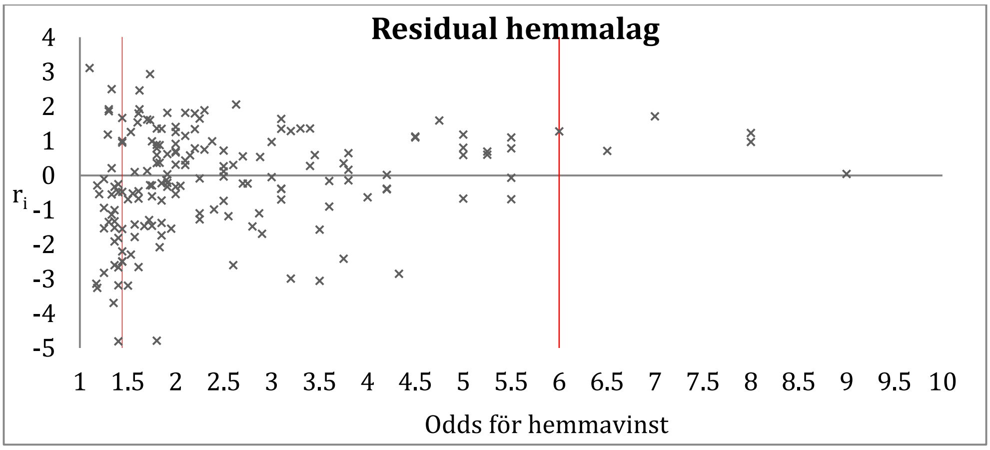
Tabell 4.4Likelihood kvot test.

P-värdet ligger väldigt nära noll och vi förkastar därmed nollhypotesen. Därmed konstaterar vi att det finns incitament att upprätta individuella Poissonfördelningar med skiftande väntevärde för respektive lag och match.

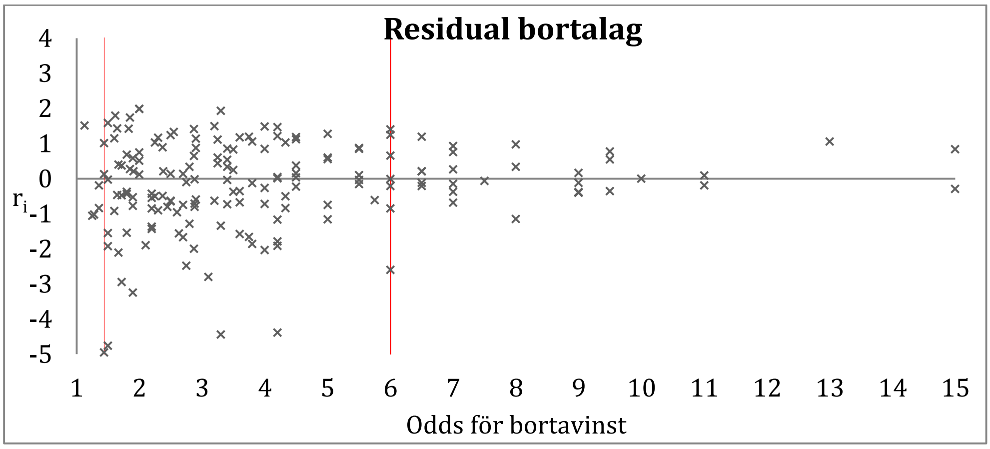
För att kontrollera precisionen i modellen är det lämpligt att genomföra en residualanalys. Trots att Poissonregressionen som tidigare nämnts inte har en felterm i dess naturliga form är det viktigt att se om modellen systematiskt estimerar fel eller ej. Om det råder en skillnad mellan storheterna observerat och estimerat antal mål kan det innebära att modellen inte plockar upp all relevant information, alternativt är bättre på att skatta en viss typ av matcher. Då vi arbetar med räknedata uppbyggd ur likelihoodfunktionen finns ett antal sätt att kontrollera modellens residualer. Vi har valt att titta på skillnaden mellan observerat och skattat antal mål vilket ger en god bild om modellen över- eller underestimerar en viss typ av matcher (StatsDirect 2013b, Winkelmann 1997, Cameron & Trived 1998). Residualerna räknas ut genom följande:

$$r_i = y_i - \hat{\mu}_i \qquad (4.13)$$

där y_i är det observerade värdet och $\hat{\mu}_i$ modellens skattade värde. Följande resultat erhålles:



Figur 4.1 Residualer hemmalag



Figur 4.2 Residualer bortalag

I residualplottarna går det utläsa att felskattningarna är större när hemmalaget ses som favorit (har ett lågt odds). Därav finns det incitament att tro att om dessa skattningar utesluts skulle vi få ett bättre resultat. Vidare går det utläsa att det endast återfinns ett fåtal observationer med låga odds för bortalaget, detta är ytterligare ett tecken på att ett lag ett generellt är starkare på hemmaplan, det vill säga medeloddset för bortalaget är större.

4.3 Spelstrategier

Prediktionen av antalet mål i respektive match är framtagen ur Poissonmodellen och lägger grunden för den fortsatta analysen. Med utgångspunkt i skattade värden för hemmalagets- samt bortalagets målskörd väljs de positioner med högst potentiell avkastning (högsta oddset). Därför kommer ett spel för varje match placeras på det asiatiska handikappet och ett spel på marknaden för antalet mål. Ur detta har två strategier utvecklats. Den första strategin är rent matematiskt framtagen och följer logiskt av de gränser som ges för aktuellt spel och den andra strategin är en empiriskt framarbetad strategi. För att göra det överskådligt fortsätter vi på exempelmatchen mellan Trollhättan och Gauthiod från ovan. Estimerat antal mål i denna match är 2,70 för Trollhättan och 1,31 för Gauthiod vilket ger en differens i favör till Trollhättan på 1,39 (2,7-1,31). Med den matematiska strategin skulle det för asiatiskt handikapp innebära att aktuella spel är -0,5 samt >+0,5 på Trollhättan och >+1,5 på Gauthiod (se tabell 4.4). Detta eftersom lagen med dessa handikapp adderade/ subtraherade uppvisar en positiv målskörd. Uträkning görs enligt följande:

- Trollhättan: $1{,}39 - 0{,}5\ (\text{handikapp}) = 0{,}89\ \text{mål} > 0\ \mathit{mål}$
- Gauthiod: $-1{,}39 + 1{,}5\ (\text{handikapp}) = 0{,}11\ \text{mål} > 0\ \mathit{mål}$

Vi fortsätter med testmatchen, adderar ihop antalet mål och får en förväntad total målskörd på 4,01 (2,70+1,31). Detta öppnar en möjlighet för spel på över 3,5 mål samt under 4,5 mål där vi återigen för varje match och spel väljer positionen med det högsta oddset. Denna procedur görs för samtliga matcher där odds finns tillgängliga. Strategin sammanfattas nedan som strategi 1:

Strategi 1	
Asiatisk hcp	Spel aktuellt när estimerat antal mål hemma minus borta
Hemmalag +0,5	>-0,5
Hemmalag -0,5	>0,5
Hemmalag +1,5	>-1,5
Hemmalag -1,5	>1,5
Hemmalag +2,5	>-2,5
Hemmalag -2,5	>2,5
	Spel aktuellt när estimerat antal mål borta minus hemma
Bortalag +0,5	>-0,5
Bortalag -0,5	>0,5
Bortalag +1,5	>-1,5
Bortalag -1,5	>1,5
Bortalag +2,5	>-2,5
Bortalag -2,5	>2,5
Antal mål	Spel aktuellt när estimerat antal mål hemma plus borta
Över 1,5	>1,5
Under 1,5	<1,5
Över 2,5	>2,5
Under 2,5	<2,5

Strategi 2	
Asiatisk hcp	Spel aktuellt när estimerat antal mål hemma minus borta
Hemmalag +0,5	>-0,3
Hemmalag -0,5	>1
Hemmalag +1,5	>-0,5
Hemmalag -1,5	>2
Hemmalag +2,5	>-1,6
Hemmalag -2,5	>3
	Spel aktuellt när estimerat antal mål borta minus hemma
Bortalag +0,5	>-0,3
Bortalag -0,5	>1
Bortalag +1,5	>-0,5
Bortalag -1,5	>2
Bortalag +2,5	>-1,6
Bortalag -2,5	>3
Antal mål	Spel aktuellt när estimerat antal mål hemma plus borta
Över 1,5	>2
Under 1,5	<1
Över 2,5	>3
Under 2,5	<2

Över 3,5	>3,5	Över 3,5	>4
Under 3,5	<3,5	Under 3,5	<3
Över4,5	>4,5	Över4,5	>5
Under 4,5	<4,5	Under 4,5	<4

Tabell 4.4 Strategier

Att förlora ett spel är väldigt kostsamt, varför lejonparten av spel hos spelbolag läggs på favoriten, som har en lägre risk och då logiskt ger ett lägre odds. Detta faktum använder vi för att få fram en alternativ strategi där fler spel går in på bekostnad av ett lägre medelodds. Detta åstadkoms genom att ändra målskillnadsgränserna för när ett spel skall placeras, se strategi 2 i tabell 4.4. Denna strategi kommer per automatik vara mer benägen att välja spel där handikappet är positivt. Alternativkostnaden för detta är ett sjunkande medelodds. Anledningen till sänkt medelodds beror på att ett spel på ett positivt handikapp ger vinst även vid oavgjort resultat. För att kunna testa flertalet likartade strategier med nya prediktionsgränser valdes Microsoft Excel som hjälpmedel. Fördelen med detta är att nya gränser enkelt implementeras vilka genom IF satser plockar ut rätt odds sett till prediktionsgränserna (valda strategier).

Som tidigare beskrivits har modellen visat tecken på dålig precision när hemmalaget är stor favorit i matchen (se residualplottarna, figur 4.1 och figur 4.2). Detta är grunden för upprättande av strategier vilka eliminerar vissa matcher. Strategierna kan undersöka hur resultatet påverkas och vara grund för test av effektiv prissättning på fotbollsmatcher i lägre divisioner. Utgångspunkten för denna eliminering är matchresultatet som har tre utfall: hemmavinst (1), oavgjort (X) och bortavinst (2). I studien görs en brytpunkt för odds större eller lika med 6, detta för att kunna utesluta matcher med en klar favorit och en klar underdog. De valda gränserna har gjorts genom omfattande test, för att hitta den bästa avvägningen mellan förlorade observationer och minskad bias. Således innebär detta att om en match har ett odds vilket överstiger 6 för 1:a (hemmavinst) eller 2:a (bortavinst) kommer denna match att uteslutas.

Innan vi går vidare till diskussionsdelen bör det påpekas att allt som presenteras nedan utförs i svenska kronor (SEK, där en krona satsas på varje spel), förutom vinst vilken presenteras i procent.

5. Analys

Genom att applicera Poissonmodellen på den matematiska strategin har följande resultat erhållits:

Resultat strategi 1 a

	Asiatisk hcp	Antal mål	Totalt
Antal Spel	166	172	338
Avkastning	142,33	159,58	301,91
Vinst	-14,26%	-7,22%	-10,68%

Tabell 5.1 – Resultat strategi 1.

Här konstateras att det asiatiska handikappet omfattas av 166 spel vilka ger en avkastning på 142,33 det vill säga en förlust på 23,76 (-14,26%). För antal mål marknaden placeras totalt 172 spel till en förlust på -7,22%. Om asiatiska handikapp och antal mål spel adderas ihop uppnås en förlust på totalt -10,68%. Tidigare konstateras att modellen har problem med matcher med en klar favorit (och en underdog). Därför utesluts matcher med ett odds lika med eller större än 6 (se kapitel 4.3) med följande resultat:

Resultat strategi 1 b (matcher uteslutna)

	Asiatisk hcp	Antal mål	Totalt
Antal Spel	132	136	268
Avkastning	122,14	128,62	250,76
Vinst	-7,47%	-5,43%	-6,43%

Tabell 5.2 – Strategi 1 där matchodds över 6 uteslutits.

I tabellen ovan kan vi se att antalet spel för asiatiska marknaden rört sig från 166 till 132, det vill säga 34 spel färre. För över/ under marknader har 36 spel uteslutits, med totalt 136 återstående spel. Med eliminering av vissa spel kan vi se att den asiatiska marknaden presterar bättre men fortfarande uppvisar en förlust på -7,47%. Ett något bättre resultat erhålles på marknaden för antal mål där det negativa resultatet uppgår till -5,43%. När båda marknadernas spel adderas uppnås en förlust på -6,43%.

Om samma Poissonmodell används för strategi 2 erhålls följande resultat:

Resultat strategi 2 a

	Asiatisk hcp	Antal mål	Båda
Antal Spel	166	172	338
Avkastning	166,25	172,5	338,75
Vinst	0,15%	0,29%	0,22%

Tabell 5.3 – Strategi 2 alla matcher.

Det bör understrykas att denna strategi var en av de som uppvisade bäst resultat, men att flertalet andra strategier också kontrollerades. För strategi 2, med alla matcher inkluderade uppnår asiatiskt handikapp marknaden ett knappt positivt resultat, 0,15%, vilket skall jämföras med minus 14,26% för strategi 1. Marknaden för antal mål uppvisar också den ett positivt resultat på 0,29% vilket är en klar förbättring jämfört med strategi 1. Tillsammans uppgår den totala vinsten till 0,22%.

Uteslutandet av matcher med en kraftig favorit används för strategi 2 med följande resultat:

Resultat strategi 2 b (matcher uteslutna)

	Asiatisk hcp	Antal mål	Båda
Antal Spel	132	136	268
Avkastning	132,49	142,11	274,6
Vinst	0,37%	4,49%	2,46%

Tabell 5.4 – Strategi 2 där matchodds över 6 uteslutits.

I tabell 4.4 kan vi se att både marknaden för asiatisk handikapp och över/ under uppvisar positiva resultat, 0,37% respektive 4,49%. Resultatet vid sammanslagning av dessa två strategier är en vinst på 2,46% , vilket överstiger tidigare testade strategier. Vi kan se ett mönster som tyder på att implementering av en lämplig strategi tillsammans med uteslutande av vissa matcher gör att vi erhåller en modell som ur avkastningssynpunkt presterar bättre. När strategi 2 implementeras övergår resultaten för spel på asiatiskt handikapp och antal mål från negativt till positivt. Effekten av att ta bort matcher med en klar favorit för strategi 2 påverkar markanden för antalet mål markant. Därför finns en antydan att om matcher av denna karaktär är mer volatila vad gäller antalet gjorda mål än vad som är fallet för den relativa styrkan, det vill säga det asiatiska handikappet. Däremot kan det i studien inte statistiskt säkerställas att så är fallet. För att sammanfatta tabellerna ovan ser vi en förbättrad avkastning när utvalda matcher utesluts och strategier implementeras.

6. Resultat

Resultatdiskussionen har som huvudsaklig målsättning att besvara studiens problemformuleringar. Dessutom syftar den till att belysa iakttagelser författarna gjort under studiens gång. Genom användning av en beprövad statistisk Poissonmodell (Dixon och Coles 1997) upprättas skattningar av antalet gjorda mål för 182 matcher spelade i den svenska division 1 (södra och norra) för säsongen 2012. Det finns tecken i residualplottarna som pekar på att modellen systematiskt predikterar fel i matcher med en klar (hemma)-favorit. Något som stärker detta antagande är tabell 5.1-5.4 där resultatet förbättras när strategier implementeras och matcher med en klar favorit (och en klar underdog) utesluts.

Studien syftade vidare att undersöka om en beprövad statistisk modell kan prediktera odds för fotbollsmatcher Sveriges division 1 mer effektivt än spelmarknaden. Det går inte bekräfta att modellen predikterar effektivare än spelmarknaden. Om vi går tillbaka till tabell 5.1 kan vi se att spel på asiatiska och antal mål marknaden uppvisar ett negativt resultat. I detta negativa resultat (samt efterföljande resultat) är det dock värt att belysa att spelbolag tar ut en marginal, vilken för Sveriges division 1 fluktuerar runt 6 % för de två marknadstyperna som studien undersöker (Bet365 2012). Därmed går det fastställa att om marginalen hade varit mindre (vilket är fallet för större marknader) hade modellen ur avkastningssynpunkt presterat bättre. Om man således bortser från spelbolagens marginal kan man konstatera att modellen med relativt god träffsäkerhet kan användas för prissättning och prediktion av resultat i fotbollsmatcher.

I studien påvisas att implementering av strategiskt spelande leder till en positiv avkastning. Däremot bör det konstateras att materialet endast omfattas av 182 matcher under en säsong, vilket gör att resultaten inte kan anses uppfylla full validitet. Däremot går det att konstatera att modellen presterar väsentligt bättre när strategier implementeras. Det går i tabell 5.1-5.4 utläsa att uteslutandet av matcher bidrar till en högre avkastning. Vi fann att en gräns för uteslutande av matcher på ett odds lika med eller större än 6 passade aktuellt datamaterial bra. Dock finns det säkerligen ingen generell regel för vart den exakta brytpunkten går. För att sammanfatta det hela och svara på studiens frågeställningar står det klart att strategiskt spelande i princip är ett måste för att kunna uppnå en positiv avkastning, eftersom spelbolagens marginal är tillräckligt stor för att täcka eventuell ineffektivitet i sin prissättning.

7. Vidare forskning

Som slutsatsen antyder åstadkoms en positiv avkastning för Poissonmodellen när strategiskt spelande implementeras. Det bör understrykas precis som tidigare diskuterats att undersökningen endast utförs på 182 matcher vilket kan anses vara aningen snålt. Därför skulle en liknande undersökning kunna utföras på fler matcher (fler säsonger). Detta skulle kunna verka som en jämförelse för att se om det i långa loppet går att uppnå en positiv avkastning och samtidigt testa om den spelstrategi denna studie konstaterade som den bästa också är så på framtida matcher (säsonger).

Utöver detta skulle det vara intressant att undersöka hur tillförande av nya förklarande variabler påverkar prediktionen. En hypotes är till exempel att ett lag presterar signifikant sämre om de saknar en nyckelspelare i en match. Sådana tillägg av utförlig spelar- och lagstatistik skulle troligen göra det möjligt att fånga upp kortsiktiga mönster, vilket skulle kunna bidra med ytterligare en dimension i prediktionen som modellen i nuvarande utförande saknar. Vidare skulle man kunna undersöka huruvida hemmafördelen är större eller mindre för vissa lag. Om varje lag kan tilldelas en enskild hemmaparameter skulle prediktionen med stor sannolikhet kunna bli mer träffsäker. Troligen skulle detta göra att skattningar med en klar favorit (eller en klar underdog) skulle kunna göras med bättre styrka.

Referenslista

Aftonbladet (2013). *Samlad sportstatistik*. Tillgänglig:
http://stats.betradar.com/s4/?clientid=21#2_1,3_9,22_1,5_5802,9_overview,25_1 (2013-01-05)

Bet365 (2012). *Spelbolag*. Tillgänglig:
http://www.bet365.com/ (2012-10-01)

Betfair (2012a). *Fotboll*. Tillgänglig:
http://sports.betfair.com/football (2012-10-01)

Betfair (2012b). *Learning Betfair*. Tillgänglig:
http://learning.betfair.com/sv/videos/asiatiskt-handikapp.html (2012-11-16)

Betting Advice (2012). *New tipsters – specific leagues.* Tillgänglig:
http://www.bettingadvice.com/ (2012-09-21)

Blom, Gunnar, Enger, Jan, Englund, Gunnar, Grandell, Jan & Holst, Lars (2004). *Sannolikhetsteori och statistikteori med tillämpningar.* Lund: Studentlitteratur AB.

Cameron, Colin A. &Trived, Pravin K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Dixon, Mark J. & Coles, Stuart G. (1997). "Modelling Association Football Scores and Inefficiencies in the Football Betting Market". *Journal of the Royal Statistical Society. Series C (Applied Statistics*), *Vol. 46*, 265-280.

Compton, William & Sigler, Kevin (2012). "NBA Gambling Inefficiencies: A Second Look, The sport journal". *The Sports Journal, vol. 15*.

Golec, Joseph & Tamarkin, Maurry (1991). "The Degree of Inefficiency in the Football Betting Market". *Journal of Financial Economics*, *Vol 30*, 311-323.

Groot, Loek (2008). *Economics, Uncertainty and European Football: Trends in Competitive Balance (New Horizons in the Economics of Sport).* Cheltenham: Edward Elgar Pub.

Hogg, Robert V. & Tanis, Elliot A. (2005). *Probability & Statistical Inference*. Prentice Hall, 7th Edition.

Körner, Svante & Wahlgren, Lars (2006). *Statistisk Dataanalys*. Lund: Studentlitteratur AB.

Lotteriinspektionen (2011). *Spelmarknadens utveckling I Sverige och internationell 2010.* Strängnäs: Lotteriinspektionen.

Pollard, Richard (2008). "Home Advantage in Football: A Current Review of and Unsolved Puzzle." *Open Sports Sciences Journal*, Vol 1, 12 - 14.

Quant Sports (2012). *Powerful Sporting Analysis*. Tillgänglig: http://www.quantsports.co.uk/ (2012-10-11)

Rao, Radhakrishna C. & Toutenberg, Helge (1999) *Linear Models: Least Squares and Alternatives.* New York: Springer.

Sinkey, Michael & Logan, D. Trevor (2011). Does the Hot Hand Drive the Market? Evidence from Betting Markets. *Eastern Economic Journals*, Vol 1, 1-25.

Spelaspel.se (2012). *Så sätter spelbolagen oddsen, Nordenberg Andre.* Tillgänglig: http://www.spelaspel.se/odds/bettingskolan/sa-satter-spelbolagen-oddsen/ (2012)

StatsDirect (2013a). *Poisson Regression*. Tillgänglig: http://www.statsdirect.com/help/regression_and_correlation/poisson_regression.htm (2013-04-01)

StatsDirect (2013b). *Residualer*. Tillgänglig: http://www.statsdirect.com/help/default.htm#regression_and_correlation/poisson.htm (2013-04-24)

SvFF (2012). *SvFF:s Tävlingsbestämmelser 2012*, Tillgänglig: http://angermanland.svenskfotboll.se/ImageVault/Images/id_72324/scope_0/ImageVaultHandler.aspx (2013-01-31)

Wever, Sean & Aadland, David (2012). "Herd Behaviour and Underdog in the NFL". *Applied Economics Letters, Vol 19,* 93 - 97.

Weisstein, Eric W (2013). Odds mathworld, Tillgänglig: http://mathworld.wolfram.com/Odds.html (2013-10-09)

Wikberg, Karin (2005). *Amatör eller professionist? : studier rörande amatörfrågan i svensk tävlingsidrott 1903-1967*. Stockholm : SISU idrottsböcker.

Winkelmann, Rainer. (1997). *Econometric analysis of count data (Second, revised and enlarged ed.).* Berlin: Springer-Verlag.

Bilaga 1 Parameterskattning

```
Parameters<- function(games) {
teams<- sort(unique(c(games[,1], games[,2])), decreasing = FALSE)
n<- length(teams)
g<- nrow(games)
Y<- matrix(0,2*g,1)
for (iin1:g) {
Y[((2*i)-1)] <- games[i,3]
Y[(2*i)] <- games[i,4]
}
X<- matrix(0,2*g,((2*n)+1))
for (iin1:g) {
M<- which(teams == games[i,1])
N<- which(teams == games[i,2])
X[((2*i)-1),M] <- 1
X[((2*i)-1),N+n] <- -1
X[(2*i),N] <- 1
X[(2*i),M+n] <- -1
X[((2*i)-1),((2*n)+1)] <- 1
}
XX<- X[,-1]
parameters<- glm(formula = Y ~ 0 + XX, family = poisson)
Z<- c(0, coefficients(parameters))
P<- data.frame(row.names=teams, Attack=Z[1:n], Defence=Z[(n+1):(2*n)])
return(list(teams=P,home=as.numeric(Z[2*n+1])))
}
```

Bilaga 2 Välj ur rätt lag

```
interestedTeams<- c('Orgryte', 'Lund', 'Oddevold', 'Kristianstad',
'Trollhattan', 'Skovde', 'Sylvia', 'Karlstad BK',
'Qviding', 'Utsiktens BK', 'LB07', 'SLeipner', 'Norrby',
'IK Gauthiod')

teamIdx<- c()
for(teamininterestedTeams) {

 idx<- which(tolower(rownames(parametrar$teams)) == tolower(team))
 if(length(idx) >0) {
        teamIdx<- c(teamIdx, idx)
 }
}
parametrar$teams<- parametrar$teams[teamIdx, ]

parametrar
```