

Sammanfattning

På spelmarknaden har spelbolagen fördelarna på sin sida, men denna teoretiska fördel är betydligt mindre på marknaden för amerikansk basket jämfört med exempelvis europeisk fotboll. Eftersom oddsen på spelmarknaden bygger på sannolikheter och dessa är subjektiva vill jag med en logistisk modell skatta vinstsannolikheter för utfallen i basketmatcher och undersöka om man med dessa kan erhålla vinster spelmarknaden för amerikansk basket. Detta testas över säsongen 2003-2004. I analysen är kontroll av att lagen spelar med olika hastigheter i spelet en viktig aspekt vilket ligger till grund för valet av variabler. Samband mellan variablerna och utfallen i matcherna (hemmavinst/bortavinst) skattas utifrån ett antal matcher med kända utfall. Detta används sedan för att skatta sannolikheterna för utfallen i basketmatcher där utfallen inte är kända. Om sannolikheten för något utfall i en match överstiger den sannolikhet som behövs för att göra vinster på spelmarknaden antas det att det spelas på detta utfall. Resultatet av studien visar att det är möjligt att göra vinster och att dessa kan öka betydligt om man selektivt väljer typ av spel eller vid vilken tidpunkt spelen läggs. Den teoretiska fördel som spelbolagen har visar sig också få stora konsekvenser på resultatet. Även om studien visar ett positivt resultat anses spelbolagens fördel vara för stor. Risken för framtida förluster anses överstiga möjligheten till konsekventa positiva vinster.

Innehållsförteckning

1. Inledning.....	1
1.1 PROBLEMFÖRMULERING.....	2
1.2 SYFTE.....	2
1.3 METOD.....	2
1.4 AVGRÄNSNINGAR.....	2
1.5 DISPOSITION.....	3
2. Bakgrund	4
2.1 IDROTTE BASKETBOLL	4
2.2 AMERIKANSKA BASKETLIGAN.....	5
2.3 ODDS PÅ SPELMARKNADEN.....	5
3. The possession scoring system.....	9
4. Logistisk regression.....	10
4.1 TEST AV VARIABLER OCH MODELLENS ANPASSNING.....	11
4.1.1 Wald statistika	11
4.1.2 Likelihood ratio test.....	11
4.1.3 Cox & Snell och Nagelkerkes R^2	12
4.1.4 Omnibus test för modell koefficienter.....	13
5. Empirisk undersökning.....	13
5.1 DATAINSAMLING.....	13
5.2 VARIABLER I MODELLEN	14
5.3 MODELLEN.....	18
5.3.1 Korrelation mellan oberoende variabler.....	19
5.3.2 Modellspecifikation	19
5.4 BERÄKNING AV SANNOLIKHETER.....	23
6. Resultat av studien.....	25
6.1 SPEL TILL SLUTODDSEN.....	25
6.2 SPEL TILL STARTODDSEN.....	29
7. Slutsats.....	31
Litteraturförteckning:.....	33

1. Inledning

Spelmarknaden är idag en miljardindustri och antalet aktörer på marknaden växte explosionsartat i takt med att internet nådde var mans hem. Spel om pengar har fascinerat människor i årtionden och många drömmer om att göra storvinsten. I stort sätt varje vecka hör och ser vi reklam om att man kan vinna miljonbelopp även om sannolikheten för detta är oerhört liten. Men visst är det trevligt att få drömma sig bort en stund. Att sätta några tior på exempelvis Lotto eller en fotbollsmatch för att höja spänningen i vardagen kan i ju för sig vara en rolig grej som inte är så kostsamt. Tyvärr finns det också en baksida av spel som medför att vissa människor drabbas av ”speldjävulen” och spelar bort både hus och hem. Hur man än vrider och vänder på det så är det alltid spelbolagen som har fördelarna. I kasinospel som exempelvis roulette har spelbolagen eller kasinona alltid en matematiskt säkerhetsställd fördel. Visst man vinna på kort sikt men för eller senare kommer kasinots fördel att ta spelarens pengar när antalet spel blir tillräckligt stort. Vid spel på exempelvis sportevenemang sätts oddsen¹ av människor. Detta betyder således att oddsen är subjektiva. Det finns ingen möjlighet att bestämma sannolikheterna för utfallen i en match utifrån ett experiment. Man kan inte upprepa en basketmatch under identiska förhållanden ett oändligt antal gånger. En match spelas en gång och inget annat. Detta borde medföra att man kan vinna pengar på detta givet att man är bättre på att bestämma sannolikheterna för utfallen än vad spelbolagen gör. Men givetvis har spelbolagen ändå fördelen på sin sida då de inte ger odds som är i paritet med sannolikheterna för utfallen. På ett matchspel i exempelvis fotboll håller de internetbaserade bolagen en återbetalningsprocent² på ungefär 90% vilket medför att de kan ha relativt säkra marginaler. Dessa marginaler eller riskpremium är betydligt lägre på den amerikanska basketligan där vissa bolag kan ha en så hög återbetalningsprocent som 98% vilket gör att chansen för vinst är betydligt högre. Om det är möjligt att erhålla vinster på en sådan marknad vill jag undersöka med hjälp av en statistisk modell.

¹ Proportionen mellan vinst och insats. Förklaras utförligare i kapitel 2.3.

² Anger hur stor andel av den totala insatsen från spelarna som spelbolaget betalar ut i form av vinster. Med 98% kommer spelbolaget betala ut 98 kr för varje 100 kr som spelas (teoretiskt sätt).

1.1 Problemformulering

Är det möjligt att med hjälp av en statistisk modell erhålla vinster på spelmarknaden för basket eller har spelbolagen en allt för stor fördel?

1.2 Syfte

Syftet med uppsatsen är att med en logistisk modell skatta sannolikheter för utfallen i basketmatcher och undersöka om dessa kan användas på spelmarknaden för att erhålla vinster.

1.3 Metod

Utifrån 1189 matchprotokoll från säsongen 2003-2004 hämtas statistik om lagets prestationer och dessa transformeras om till nya tempofria³ variabler. Med hjälp av en logistisk regression skattas samband mellan variablerna och utfallen i basketmatcher (hemmavinst/bortavinst) där utfallen redan är kända. De skattade koefficienterna och värdena på variablerna som baseras på lagens tidigare prestationer används sedan för att beräkna sannolikheter för hemmavinst respektive bortavinst i basketmatcher där utfallen inte är kända. Om dessa sannolikheter för något utfall överstiger spelbolagets sannolikhet i sådan grad att det är (hypotetiskt sett) vinstgivande att spela utfallet antas det att 100 kr spelas på det resultatet. För analysen används programmen SPSS och Excel.

1.4 Avgränsningar

Studien begränsas till den amerikanska basketligan⁴ över säsongen 2003-2004. Variablerna i modellen baseras på statistik om lagets prestationer, d.v.s. statistik om de enskilda spelarna bortses från. Studien begränsas till grundspelet⁵ och sträcker sig således inte över slutspelet⁶. Oddsen som används är rörliga vilket innebär att de kan vara olika från en tidpunkt till en annan. I uppsatsen används de odds som släpps i samband med att marknaden för matchen öppnar samt de odds som är strax innan matchstart. Dessa odds kommer från ett spelbolag, Pinnaclesports. Då detta bolag har den högsta

³ Tempofria variabler förklaras utförligare i kapitel 3.

⁴ National Basketball Association (NBA)

⁵ Ordinarie säsongen där alla lag spelar lika många matcher.

⁶ De lag som vunnit mest i grundspelet går till slutspelet. Slutspelet spelas i utslagsform tills enbart ett lag kvarstår, vilket kan titulera sig mästare.

återbetalningsprocenten på marknaden antas den representera hela spelmarknaden för den amerikanska basketligan.

1.5 Disposition

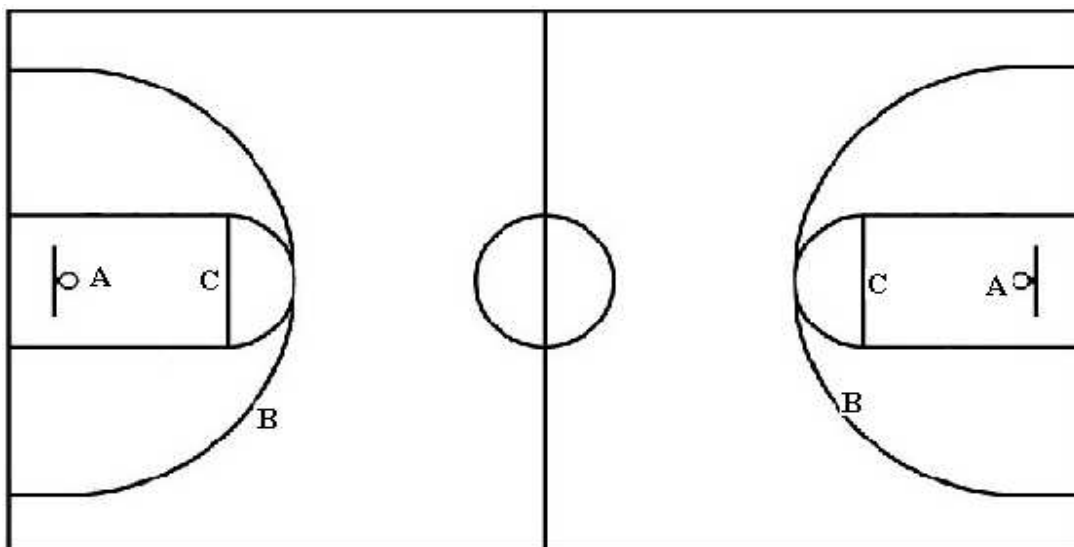
I kapitel två ges en förenklad förklaring om hur idrotten basket spelas, hur den amerikanska basketligan är uppbyggd samt en förklaring om oddsen. Vidare i kapitel tre förklaras varför tempofria variabler är viktigt för analysen. I det fjärde kapitlet ges teori om logistik regression och för testen att bestämma lämplig modell. Den empiriska analysen kommer i kapitel fem där datainsamling, variabelval, modellspecifikation, och beräkning av vinstsannolikheterna visas. Resultatet på spelmarknaden ges i kapitel 6. Slutligen redovisas slutsatsen av uppsatsen i kapitel 7.

2. Bakgrund

2.1 Idrotten basketboll

Det finns många termer och regler inom basket vilket gör att det inte alltid är helt enkelt att förstå sporten fullt ut. Här förklaras endast grundprinciperna på ett förenklat sätt. I en basketmatch möts två lag med fem spelare i varje lag på en rektangulär bana med korgar (eller mål) på var sin kortsida, vilket symboliseras med (A) i figur 1. Poäng erhålls genom att kasta bollen i motståndarens korg. Det lag som gjort mest poäng vid matchens slut har vunnit. Runt korgen finns en trepoängslinje (B) och om man skjuter utanför denna linje och sätter bollen i korgen får man tre poäng medan man får två poäng om skottet görs innanför linjen. Ett lag kan även göra poäng genom frikast. Ett lag får generellt sätt göra frikast om en spelare i motståndarlaget gör en foul, d.v.s. en otillåten kroppskontakt etc. Frikastet görs från frikastlinjen (C) och vid ett sådant kan inte motståndarlaget försvara sig utan det kan ses som ”straff mot öppen korg”. För ett lyckat frikast erhåller laget en poäng och oftast tilldelas laget två frikast vid en foul. En basketmatch kan endast sluta med att ett av lagen har vunnit matchen, det kan således inte sluta oavgjort. Har båda lagen lika många poäng vid matchens slut går det till övertid tills det att det ena laget vunnit. För fullständigt regelverk se (Jackson 2008).

Figur 1: Basketbollplan



2.2 Amerikanska basketligan

National Basketball Association (NBA) som den amerikanska basketligan heter består av 29 lag som är indelade i den västra och den östra konferensen. Varje konferens är i sin tur indelad i tre divisioner. Basketsäsongen börjar i slutet av oktober och varje lag spelar 82 matcher (hälften på hemmaplan och hälften på bortaplan) fram till mitten av april då de lag som presterat bäst går till slutspel. Eftersom avstånden, geografiskt sätt, är stora i USA spelar lagen inte lika många matcher mot varje motståndare. Ett lag möter de andra lagen som spelar i samma konferens tre till fyra gånger medan de endast möter de lagen som ligger i den andra konferensen två gånger. Till skillnad från exempelvis fotboll där ett lag generellt sett spelar varannan match på hemmaplan och varannan på bortaplan är det vanligt att i NBA spela flera matcher i rad på hemmaplan eller bortaplan. Ett lag kan exempelvis spela fyra matcher i rad på hemmaplan för att därefter åka på en "roadtrip" och spela upp till sju matcher i rad på bortaplan.

Lagen i de olika konferenserna har historiskt sett spelat lite olika. Lagen från den västra kännetecknas generellt sett av starkt spel från spelarna närmast korgen och spelar ett snabbare mer öppet spel över hela planen än vad lagen i den östra gör. Lagen i den östra konferensen spelar däremot oftast en mer disciplinerad fysisk basket med fokus på försvarsspel och starkt spel från de yttre spelarna. Lagen i denna konferens spelar oftast i ett långsammare tempo än lagen i den västra (NBA Media Ventures). Skillnader i spelhastigheter är en viktig aspekt som ligger till grund för valet av variabler i modellen som senare används.

2.3 Odds på spelmarknaden

Spelbolagen på den globala spelmarknaden erbjuder möjligheten att spela på det mesta inom sport men även på politik, finansmarknaderna, prisutdelningar eller hur vädret kommer bli. Vid spel på exempelvis vilket lag som kommer att vinna i en basketmatch är det sällan så att sannolikheten att lag A kommer vinna är lika stor som den att lag B kommer vinna. Det ena laget är oftast bättre, eller har någon form av fördel som hemmaplan eller liknande. Ett lag kan exempelvis ha ungefär 80% chans att vinna medan det andra laget har 20% att vinna. Detta reflekteras i oddsen som ger en proportion

mellan vinst och insats. En oddskvot definieras som kvoten mellan sannolikheten att händelsen inträffar och att den inte gör det. (Sharma 1996, 317ff)

$$\text{Oddskvot} = \frac{P}{1-P} \quad (1)$$

Där P är sannolikheten att händelsen inträffar. Detta innebär att om vinstsannolikheten för lag A är 80% och lag B är 20% så kommer lag A att vinna i 80% = 0,8 av fallen. Oddskvoten blir då $0,8/(1-0,8)=4$. Detta innebär att om man satsar 4 kronor på lag A och laget vinner så får man tillbaka 4 kr plus 1 kr i vinst. Formel 1 ovan kan skrivas om som

$$\begin{aligned} \text{Oddskvot} &= \frac{P}{1-P} \Rightarrow \text{Oddskvot}[1-P] = P \\ \Rightarrow \text{Oddskvot} - [\text{Oddskvot} \times P] &= P \Rightarrow \text{Oddskvot} = P + (\text{Oddskvot} \times P) \\ \Rightarrow \text{Oddskvot} &= P(1 + \text{Oddskvot}) \Rightarrow \frac{\text{Oddskvot}}{1 + \text{Oddskvot}} = P \end{aligned} \quad (2)$$

Sannolikheten för att lag A ska vinna utifrån oddskvoten är alltså $4/(1+4)=0,8$. Man kan beteckna odds på spelmarknaden på olika sätt och ett av de är i decimalform. Ett odds på exempelvis 2,40 betyder att man vinner 2,4 gånger insatsen. En insats på 100 kr ger alltså 240 kr tillbaka om utfallet man satsar på vinner, vilket motsvarar 140 kr i vinst om man räknar bort insatsen. Oddset för kan beräknas genom $1/P$, där P är sannolikheten för att händelsen ska inträffa. Med de sannolikheter som gavs ovan skulle alltså oddsen på vinst för lag A vara lika med $1/0,8 = 1,25$. Vilket är det inverterade värdet från formel (1) plus ett. Addering av värdet ett för det motsvarar vinsten (insats 4 kr, vinst 1 kr ger $5/4=1,25$). Odds i decimalform är det som används i denna uppsats. Ju större sannolikhet ett lag har att vinna desto lägre blir alltså oddset för detta. Generellt sätt förekommer rörlig oddssättning vilket medför att oddsen justeras efter hur spelarna spelar. Det främsta målet är att göra riskfria vinster. Spelbolagen ger inte odds som är i paritet med sannolikheten för att på så sätt kunna balansera insatserna. Antag ett spelobjekt med två möjliga utfall. Spelbolagen har värderat sannolikheterna för utfallen till 50-50 och ger då oddsen 1,91-1,91. Dessa sannolikheter grundar sig främst på två faktorer. Främst den estimerade sannolikheten för att respektive lag vinner men detta justeras av den andra faktorn, hur

man tror att spelarna kommer spela (Vaughan-Williams 2002, 242). Om vi antar vidare att endast två personer spelar för 100 kr vardera och att de spelar på olika utfall i matchen ovan. Detta betyder att spelbolaget får in 200 kr och när matchen är avslutad endast behöver betala ut 191 kr, det vill säga en riskfri vinst på 9 kr. Det spelar således ingen roll för spelbolaget vilket lag som vinner. Skulle mer pengar än förväntat komma in på det ena utfallet sänker man oddset på detta för att locka spelare till att spela på det andra utfallet. Justeringar för spelbeteende är en viktig faktor för att hålla nere risken. Vissa lag har en stor supporterskara (Barcelona, Manchester United etc. i fotboll) och spelbolagen får alltid mycket pengar på dessa lag. Även om de skulle kunna ge ett odds på exempelvis 1,30 på vinst väljer man kanske att ge ett odds på 1,20 istället då man vet att pengarna kommer strömma in ändå och det finns ingen anledning att bära risk i onödan.

De odds som används i denna uppsats kommer från spelbolagen Pinnaclesports. Pinnaclesports används dels för att det är ett av de bolag som har högst återbetalningsprocent på marknaden (98%) för basket och dels för att det har fått högsta betyg (A+) från den oberoende undersökaren Sportbookreviw. Detta medför att det är ett tryggt och välskött företag som betalar ut en spelares eventuella vinster. Bolaget har licens inom EU-området vilket gör att eventuella vinster som kan göras med hjälp av modellen är skattefria enligt EU-lag. För mer information se sportbookreviw.com. Oddsen som kommer att användas är både de som bolaget startat med och de som är strax innan matchstart. Anledningen till att odds vid två olika tidpunkter används är för att se om tidpunkten för vid vilket spelet läggs kan vara avgörande för resultatet.

För matchen Utah Jazz mot Boston Celtics den 12 december 2003 ger Pinnaclesports startoddsen 1,49 för hemmalaget och 2,85 på bortalaget. Fram till matchstart har oddsen rört sig till 1,37 - 3,40 (Tip-Ex International Limited). Oddsjusteringar behöver inte alltid bero på insatsstrukturen, det kan även vara ett resultat av ny information, som skada på en nyckelspelare etc. (Vaughan-Williams 2002, 136). Spelbolagets teoretiska fördel för startoddsen kan ses i tabell 1.

Tabell 1: Bolagets teoretiska fördel och vilka odds som vore rättvisa för spelarna

Spel	Spelbolagets odds	Motsvarar sannolikhet	"Fair" odds	Motsvarar sannolikhet
Hemmlaget	1,49	$100/1,49=67,11$	1,52	65,67
Bortalaget	2,85	$100/2,85=35,09$	2,91	34,33
Totalt:		102,2		100

Utifrån tabell 1, kan ses att spelbolagen har en teoretisk fördel på 2,2 %. Detta beräknas genom att dividera sannolikheten i utfallsrummet (100%) med oddset för respektive utfall och summera detta. Värdet överstiger 100 vilket betyder att det är spelbolaget som har fördelen. Hade värdet varit under 100 hade spelaren haft fördelen, vilket betyder att insatserna kan viktas på sådant sätt att spelaren kan erhålla riskfria vinster.

Om oddsen var satta så att sannolikheterna tillsammans är lika med 100% , vilket kallas rättvisa odds (varken spelare eller spelbolag har någon fördel) skulle oddsen vara 1,52 för hemmalaget och 2,91 för bortalaget. Detta beräknas genom:

$$\text{Justerad sannolikhet för hemmalaget} = \frac{\text{Sannolikhet som spelbolaget ger}}{\text{Den teoretiska fördel som spelbolaget har}} \quad (3)$$

$$\text{Rättvist odds för hemmalaget} = \frac{\text{Högsta möjliga sannolikhet i utfallsrummet}}{\text{Justerade sannolikheten för hemmalaget}} \quad (4)$$

Den justerade sannolikheten för hemmalaget blir alltså med den sannolikhet som spelbolaget ger (0,6711) och den teoretiska fördelen (1,022) lika med $0,6711/1,022 = 0,6567$. Det rättvisa oddsen för hemmalaget är lika med $1/0,6567 = 1,52$. Beräkningarna görs på motsvarande sätt för bortalaget. Skillnaden på att få 1,49 eller 1,52 i odds kan kanske anses försumbar, men denna lilla skillnad kan få väldigt stora konsekvenser på resultatet när antalet spel blir tillräckligt stort, vilket kommer visas senare.

3. The possession scoring system

The possession scoring system är ett system som utvecklats för att samla in så mycket information som möjligt om en basketmatch. Systemet fokuserar på spelaren med bollen och följer bollen från spelare till spelare tills bollen övergår till motståndarna, antingen genom skottförsök (missat eller i korgen) eller genom en turnover (misstag som leder till att bollen övergår till motståndarlaget). Det viktiga med metoden är att visa hur lagen byter innehav av bollen under matchen och att båda lagen under en match innehar bollen lika många gånger⁷. Båda lagen kommer således ha lika många chanser att göra poäng.

Olika lag spelar olika, vissa lag fokuserar på starkt försvarsspel och vårdar bollen inom laget längre tid innan de kommer till avslut, andra lag spelar ett snabbare mer öppet spel. När två lag ska jämföras måste det tas hänsyn till att de i de tidigare matcherna spelat med olika tempo. Det snabbspelande laget (som mött andra snabbspelande lag) kommer att ha haft bollen fler gången och därmed haft fler möjligheter att göra poäng än vad laget som spelat mot andra långsamma lag. Men när dessa lag möts kommer de att ha bollen lika många gånger. Det innebär att man inte kan använda exempelvis antal gjorda poäng som variabel. De variabler som används i modellen måste vara så kallat tempofria. För mer utförlig information se (Oliver 2003). Antalet bollinnehav som lagen haft under en match kan approximativt beräknas som:

$$\text{Antal bollinnehav} = \text{FGA} - \text{OR} + \text{TO} + 0,45 \times \text{FTA} \quad (5)$$

I (5) är FGA antalet skottförsök från banan. Dessa kan resultera i att bollen går i korgen och då övergår givetvis bollen direkt till motståndaren eller att man missar. En miss som återtas av laget, kallas offensiv rebound (OR) och innebär att bollinnehavet inte förändras och därför subtraherar man för detta. Turnovers (TO) är antalet misstag som laget gör som leder till förändring av bollinnehav. Slutligen är FTA antalet frikastförsök. Man

⁷ Under en match som inte går till övertid är det möjligt att ett lag kan ha som mest två bollinnehav mer än dess motståndare men över en säsong är det inte troligt att ett lag kommer ha konsekvent mer eller mindre bollinnehav än motståndarna.

multiplikerar FTA med 0,45⁸ för att få bästa skattningen på antalet frikastförsök som avslutar ett bollinnehav. Eftersom inte alla frikast avslutar ett bollinnehav då ett lag kan tilldelas flera frikast. (Oliver 2003, 8ff).

4. Logistisk regression

Valet av att använda en logistisk regression beror på att man skatta sannolikheten för att en händelse ska inträffa/inte inträffa. I en binär logistisk regression är den beroende variabeln binär, d.v.s. kan endast anta två värden (noll och ett). Detta motsvarar i uppsatsen ett för hemmavinst och noll för bortavinst. De oberoende variablerna kan vara diskreta, kontinuerliga, binära eller en blandning av dessa. Den logistiska funktionen är begränsad inom intervallet 0-1 vilket gör den lämplig att använda då det är de enda möjliga utfallen i studien för basket. I många sammanhang uttrycks sannolikheter som odds, till exempel på spelmarknaden. Odds och sannolikheter uttrycker egentligen samma sak men i olika form och det är enkelt att konvertera mellan dessa. (Sharma 1996, 317ff).

$$\text{Odds}(Y = 1 | X_1, X_2, \dots, X_p) = \frac{P(Y = 1 | X_1, X_2, \dots, X_p)}{1 - P(Y = 1 | X_1, X_2, \dots, X_p)} \quad (6)$$

Där Y är den beroende variabeln och P är sannolikheten att den beroende variabeln antar värdet ett, givet p oberoende variabler (X_1, X_2, \dots, X_p). Oddset är oddskvoten som gavs i (1)

Som visades i (2) kan detta skrivas om som

$$P(Y = 1 | X_1, X_2, \dots, X_p) = \frac{\text{Odds}(Y = 1 | X_1, X_2, \dots, X_p)}{1 + \text{Odds}(Y = 1 | X_1, X_2, \dots, X_p)} \quad (7)$$

Den logistiska regressionsmodellen kan skrivas som:

$$P(Y = 1 | X_1, X_2, \dots, X_p) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}} \quad (8)$$

⁸ Från början användes värdet 0,40 men det har senare justerats till 0,45 då detta värde stöds bättre av empiriska studier.

Relationen mellan sannolikheten (P) och de oberoende variablerna är icke-linjär (8). Medan relationen mellan oddset och de oberoende variablerna är linjär (9). (Sharma 1996, 317ff).

$$\ln\left(\frac{P(Y=1|X_1, X_2, \dots, X_p)}{1 - (P(Y=1|X_1, X_2, \dots, X_p))}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (9)$$

Tolkning av koefficienterna för de oberoende variablerna skall göras med hänsyn till effekten på logaritmen av oddsen och inte för sannolikheten. Estimaterna av parametrarna i den logistiska regressionen erhålls genom en iterativ process (maximum likelihood)(Sharma 1996, 317ff). I uppsatsen ges dessa estimat av SPSS.

4.1 Test av variabler och modellens anpassning

4.1.1 Wald statistika

I en linjär regression kan ett vanligt t-test användas för att undersöka om en enskild koefficient i modellen är signifikant skild från noll. För motsvarande test i en logistisk regression används Walds statistika. Nollhypotesen i testet är definierad till att koefficienten inte är signifikant skild från noll, d.v.s. det finns ingen signifikant relation mellan koefficienten och utfallet Afifi, Clark och May 2004, 291-292).

$$Wald = \frac{b^2}{SE(b)^2} \sim \chi^2(1) \text{ under } H_0 \quad (10)$$

I (10) representerar b värdet på den estimerade koefficienten β och $SE(b)$ är dess standardfel. Statistikan följer en χ^2 -fördelning med en frihetsgrad. (Afifi, Clark och May 2004, 291-292).

4.1.2 Likelihood ratio test

Modellens anpassning till data kan testas med ett likelihood ($-2\log L$) test. Likelihooden (L) definieras som sannolikheten att den estimerade modellen representerar data och

nollhypotesen som testas är således att modellen passar data. För att testa nollhypotesen transformeras L till $-2\log L$. Statistikan $-2\log L$ följer en χ^2 -fördelning med $n-q$ frihetsgrader, där n är antalet observationer och q är antalet parametrar i modellen (Sharma 1996, 323-324).

Statistikan kan även användas för att testa om addering av ytterligare oberoende variabler signifikant förbättrar modellens anpassning. För detta kan ett χ^2 - differens test användas. Differensen mellan $-2\log L$ för "ursprungliga modellen" och modellen där nya oberoende variabel lagts till följer en χ^2 -fördelning med lika många frihetsgrader som skillnaden i frihetsgrader mellan de båda modellerna. (ibid, 323-324).

4.1.3 Cox & Snell och Nagelkerkes R^2

Både Cox & Snell och Nagelkerkes R^2 motsvarar R^2 -värdet i en linjär regression och anger hur stor andel av variationen i datat som förklaras av modellen. Måttet som Cox & Snell utvecklat baseras på log likelihooden av den nya modellen ($LL(ny)$) och log likelihooden av den ursprungliga modellen ($LL(ursprunglig)$) samt stickprovsstorleken n . (Aiken et. al 2002, 502-503).

$$R_{\text{Cox \& Snell}}^2 = 1 - \exp\left[-\frac{2}{n}(LL(ny) - LL(ursprunglig))\right] \quad (11)$$

Problemet med måttet är att det inte når det maximala värdet ett. (Field 2005, 223) . Måttet når ett maximalt värde av 0,75 när proportionen i stickprovet är lika med 0,5 (Aiken et.al, 502-503).

Nagelkerkes R^2 är en modifiering av Cox and Snells R^2 så att det maximala värdet ett kan nås. Detta görs genom att dividera Cox and Snell med det maximala värdet som det kan nå givet proportionen i stickprovet. Värdet på Nagelkerkes mått är således generellt sätt högre än Cox & Snells (ibid, 502-503).

$$R_{\text{Nägelkerke}}^2 = \frac{R_{\text{Cox \& Snell}}^2}{1 - \exp\left[\frac{2(LL(\text{Ursprunglig}))}{n}\right]} \quad (12^9)$$

4.1.4 Omnibus test för modell koefficienter

Omnibustest för modellens koefficienter är ett annat mått för validiteten av modellen. Testet påminner om F-test i en linjär regression och testar nollhypotesen, alla koefficienter signifikant lika med noll. I testet är χ^2 -värdet lika med skillnaden mellan χ^2 -värdet för modellen med bara konstant och χ^2 -värdet för den fullständiga modellen (konstant och oberoende variabler) med lika många frihetsgrader som skillnaden i antalet parametrar mellan modellerna. (Gamst, Guarino och Meyers 2005, 239).

5. Empirisk undersökning

5.1 Datainsamling

En intressant aspekt med den amerikanska basketligan (liksom med många andra amerikanska ligor) är att det finns mycket omfattande statistik om varje lag och spelare. För att erhålla statistiska mått på hur varje lag har presterat i varje enskild match har samtliga matchprotokoll gått igenom under säsongen 2003-2004. Detta medför att data har samlats in från 1189 matcher, vilket varit en mycket tidskrävande process. Data har samlats in från databasebasketball och hoopstats. Variablerna har sedan beräknats om till tempofria variabler för att kunna kontrollera för att de olika lagen spelar med olika hastigheter. Den slutliga delen som undersöks är om det är möjligt att göra överavkastning på spelmarknaden. De odds som används i undersökningen är från spelbolaget Pinnaclesports och har köpts in från Tip-Ex International Limited. Dessa kan av upphovsrättsliga skäl inte redovisas i uppsatsen.

⁹ Formel från (Field 2005, 223)

5.2 Variabler i modellen

Dean Oliver beskriver i sin bok "Basketball on Paper: Rules and Tools for Performance Analysis" fyra faktorer för att lyckas i basket. Dessa fyra faktorer är samtliga tempofria och den första faktorn anses ha störst betydelse för att vinna matcher, den andra näst störst betydelse och så vidare. Effektiviteten i skotten är den första faktorn och enkelt förklarar, ju fler lyckade skott per skottförsök desto större chans att vinna. Den andra faktorn är antalet misstag (turnovers) per bollinnehav som leder till förändring av bollinnehavet. Ett lag som gör många misstag kommer få mindre möjligheter till att göra poäng och därmed ha en lägre chans till att vinna. Offensiv rebounding, säger hur stor andel av de missade skottförsöken som laget återtar bollen innan motståndarlaget hinner ta bollen. Ett lag som tar många returer kommer få fler chanser att göra poäng och därav större vinstmöjlighet. Den slutliga faktorn är frikast. Denna faktor mäter skickligheten i att sätta sina frikast relativt till antalet skottförsök från banan. Dessa faktorer mäts för både anfallsspelet och försvarsspelet i laget. Samtliga faktorer kommer användas i modellen och förklaras mer ingående nedan.

Beroende variabel:

Den beroende variabeln (*Resultat*) är binär och antar värdet ett om hemmalaget vinner och värdet noll för bortavinst.

Oberoende variabler:

De oberoende variablerna i modellen fokuserar främst på de fyra faktorerna ovan för att mäta lagets prestationer i anfallsspel samt försvarsspel. Variablerna är enligt följande för anfallet:

Offensiv skotteffektivitet: Mäter hur effektivt laget är på att sätta bollen i korgen per skott utifrån banan. Måttet tar hänsyn till att ett lyckat trepoängsskott är värt 0,5 mer än en lyckad tvåpoängare.

$$\text{Offensiv skotteffektivitet} = \frac{FGM + (0,5 \times 3PM)}{FGA} \quad (13)$$

I (10) är FGM antalet gånger som bollen går i korgen vid skott utifrån banan. Måttet 3PM är antalet av dessa gånger som skottet görs utanför trepoängslinjen och slutligen är FGA det totala antalet skottförsök. Att sätta en tredjedel av trepoängsförsöken är lika bra som att sätta hälften av tvåpoängsförsöken. Ju högre värde variabeln har desto högre effektivitet har laget i det offensiva spelet.

Offensiva turnovers: Mäter hur många misstag per bollinnehav som laget gör sådant att bollen övergår till motståndarna. Kommer man inte till avslut kan man inte heller göra poäng.

$$\text{Andel Turnovers} = \frac{\text{Turnovers}}{\text{Antal bollinnehav}} \quad (14)$$

Ett högt värde på variabeln innebär att laget gör många misstag (mindre möjligheter att göra poäng) och således minskar chansen att laget vinner. Värdet är i procent och variablerna i modellen heter *Home_Oturn* och *Away_Oturn* för hemma- respektive bortalag.

Offensiv rebounding: Mäter hur ofta det anfallande laget tar kontroll av bollen igen efter ett missat skott.

$$\text{Andel Offensiv rebounding} = \frac{\text{Lagets antala offensiva rebounds}}{\text{Lagets antala offensiva rebounds} + \text{motståndarens antala defensiva rebounds}} \quad (15)$$

Lagets andel offensiva rebounds är antalet gånger som laget återtar bollen vid missat skottförsök och motståndarens defensiva rebounds är antalet gånger som motståndarlaget tar bollen vid ett sådant tillfälle (10). Värdet är i procent och variabeln *Home_Oreb* mäter värdet för hemmalaget och *Away_Oreb* är motsvarande värde för bortalaget. Högt värde på variabeln innebär att laget är bra på återta bollen efter ett missat skottförsök (vilket leder till en ny möjlighet att göra poäng) och ökar chansen att laget vinner.

Offensiva frikast: Hur många poäng laget gör genom frikast relativt till antalet skott som görs utifrån banan. Frikast är en viktig del i basket som ofta underskattas. Speciellt i jämna matcher kan det vara just skickligheten i att sätta sina frikast som är den avgörande faktorn mellan vinst och förlust.

$$\text{Frikast skicklighet} = \frac{\text{FTM}}{\text{FGA}} \quad (16)$$

I formel (13) är FTM (Free Throws made) antal satta frikast relativt till totala antalet skottförsök under en match (FGA). I modellen utgörs detta mått av variablerna *Home_Ofree* och *Away_Ofree*.

Av dessa fyra faktorer säger turnovers och offensiv rebounding hur många möjligheter till att göra poäng som laget själv skapar medans skotteffektivitet och frikast säger hur effektivt laget är på att omvandla dessa poängmöjligheter till poäng. Måtten för försvaret är motsvarande som för anfallet men omvänt.

Defensiv skotteffektivitet: Mäter hur effektivt motståndarlaget är på att göra poäng på laget. Ett högt värde på variabeln är en indikation på att ett svagt defensivt spel i laget och ska således minska chansen för att laget vinner.

$$\text{Defensiv skotteffektivitet} = \frac{\text{Motståndarens FTM} + 0,5 \times \text{Motståndarens 3PM}}{\text{Motståndarens FGA}} \quad (17)$$

För att mäta effektiv skotteffektivitet används i modellen variablerna *Home_efgdiff* och *Away_efgdiff*. Dessa variabler är beräknade som Offensiv skotteffektivitet – Defensiv skotteffektivitet. Variablerna mäter alltså hur mycket mer effektivt anfallet är jämfört med försvaret om variabeln är positiv och vice versa.

Defensiva turnovers. Mäter motståndarnas misstagsandel och ju högre värde desto mindre möjligheter för motståndarna att göra poäng och därmed ökad vinstchans för

laget. Ett högt värde kan vara en indikation på att laget är bra på att stressa motståndarlaget till misstag etc.

$$\text{Andel Defensiva Turnovers} = \frac{\text{Motståndarens antal turnovers}}{\text{Antal bollinnehav}} \quad (18)$$

Defensiv rebounding: Hur bra försvaret är på att ta bollen från motståndarna då de missat ett skottförsök. Ju högre värde på variabeln desto bättre är laget på att ta bollen vilket minskar antalet möjligheter för motståndaren att göra poäng och ökar därmed lagets vinstchanser. I modellen utgörs detta av variablerna Home_Drebs och Away_Drebs för hemma- respektive bortalag.

$$\text{Andel Defensiva rebounds} = \frac{\text{Lagets antala defensiva rebounds}}{\text{Lagets antala defensiva rebounds} + \text{Måttståndarnas antala offensiva rebounds}} \quad (19)$$

Defensiva frikast: Variabeln mäter på samma sätt som för offensiva frikast men med den skillnaden att det nu är hur effektivt motståndarlaget är i sina frikast. Hur effektivt motståndarlaget är kan i princip laget inte påverka, eftersom de inte kan försvara sig vid frikast. Ett lågt värde på variabeln skulle dock kunna vara ett resultat av att laget har en fanatisk publik som lyckas störa motståndarlaget så pass mycket att koncentrationen minskar och mer missar uppstår.

$$\text{Defensiv frikast} = \frac{\text{Motståndarens FTM}}{\text{Motståndarens FGA}} \quad (20)$$

Samtliga variabler ovan är beräknade utifrån hur lagen har presterat i de tre senaste matcherna och ett medelvärde av detta har beräknats. Anledningen till att det inte separeras på hur laget spelar på hemmaplan respektive bortaplan är att det ofta kan vara många matcher däremellan. I NBA spelar lagen ofta ett antal hemmamatcher i rad och därefter kan de exempelvis åka på en "roadtrip" och spela upp till sju matcher på bortaplan. Att använda hemmastatistik igen när de kommer tillbaka från sådan roadtrip antas inte vara lämpligt då förändringar i spel och form kan ha hunnit förändrats under

denna tid. Dessa variabler fokuserar alltså på lagens form. Eftersom uppsatsen fokuserar på lagstatistik och inte statistik för de enskilda spelarna antas det att genom att endast fokusera på de tre senaste matcherna ska variablerna snabbt kunna fånga upp om någon nyckelspelare blivit skadad etc.

Hemmalagets sannolikhet: Som ett generellt mått används hemmalagets vinstsannolikhet som spelbolaget ger då marknaden för matchen öppnar. Detta är känt vid speltillfället. Eftersom de andra variablerna ovan fokuserar på kortsiktighet (form) antas variabeln *Sannolikhet* fånga upp ett långsiktigt styrkeförhållande mellan lagen samt hemmalagets grad av hemmafördel¹⁰.

5.3 Modellen

För att skatta sannolikheterna för utfallen måste först ett samband mellan utfallen och de oberoende variablerna identifieras. För detta måste utfallen givetvis vara kända. Basketsäsongen sträcker sig från slutet av oktober till mitten av april och alla matcher i oktober och november används till att ta fram en initial modell. Denna modell och värdena på variablerna i första omgången i december används sedan för att beräkna sannolikheterna för hemma- respektive bortavinst i varje match. Om dessa sannolikheter för något utfall överstiger spelbolagets sannolikhet i sådan grad att det är (hypotetiskt sett) vinstgivande att spela utfallet antas det att 100 kr spelas på det resultatet. När omgången är spelad är dessa utfall kända och vi har mer information vilket gör att dessa observationer läggs till de övriga kända utfallen. Modellen skattas igen och de nya parametervärdena används för att räkna sannolikheterna i nästkommande omgång. Därefter läggs dessa till och modellen skattas igen. Detta förfarande fortsätter tills vi når sista omgången i mitten av april.

¹⁰ Lag spelar generellt sätt bättre på hemmaplan än bortaplan och för vissa lag är den skillnaden större än för andra. Spel på hemmaplan innebär att man spelar på den arena som man är van vid och att majoriteten av publiken stödjer laget. Faktorer som kan påverka eventuellt domaren att ta obekväma beslut mot hemmalaget.

5.3.1 Korrelation mellan oberoende variabler

För att välja en lämplig modell undersöks korrelationen mellan de oberoende variablerna. Korrelationen mellan varje par av variabler baseras på de kända observationerna från oktober-november. Utfallen i december och framåt antas vi inte ännu känna till.

Tabell 2: Korrelationen mellan varje par av oberoende variabler

	<i>Home_ EfgDiff</i>	<i>Away_ EfgDiff</i>	<i>Slh</i>	<i>Home_ Dturn</i>	<i>Home_ Orebs</i>	<i>Home_ Drebs</i>	<i>Home_ Oturm</i>	<i>Home_ Ofree</i>	<i>Home_ Dfree</i>	<i>Away_ Orebs</i>	<i>Away_ Drebs</i>	<i>Away_O turn</i>	<i>Away_ Dturn</i>	<i>Away_ Ofree</i>	<i>Away_ Dfree</i>
H_EfgDiff	1,00														
Aefgdifff	0,05	1,00													
Slh	0,18	-0,26	1,00												
H_Dturn	-0,01	-0,03	0,07	1,00											
H_Orebs	-0,16	0,06	-0,06	-0,05	1,00										
H_Drebs	-0,10	0,00	0,08	-0,24	0,01	1,00									
H_Oturm	0,18	-0,01	-0,22	0,16	-0,05	0,01	1,00								
H_Ofree	-0,05	-0,02	0,00	0,09	-0,08	0,01	-0,13	1,00							
H_Dfree	-0,13	0,14	-0,26	-0,07	0,04	0,07	0,14	0,07	1,00						
A_Orebs	-0,05	-0,10	0,05	0,14	0,01	0,01	-0,02	0,01	-0,01	1,00					
A_Drebs	0,23	-0,07	0,02	-0,05	-0,02	-0,08	0,08	0,06	-0,10	-0,08	1,00				
A_Oturm	-0,02	-0,05	0,29	-0,01	-0,06	0,03	-0,04	-0,06	-0,11	0,17	0,02	1,00			
A_Dturn	-0,16	-0,13	-0,05	0,13	-0,06	0,01	0,04	-0,01	-0,01	0,13	-0,18	0,15	1,00		
A_Ofree	-0,13	-0,07	-0,09	0,17	-0,03	0,03	0,02	0,00	0,03	0,01	-0,14	-0,01	0,17	1,00	
A_Dfree	0,00	0,03	0,21	-0,05	-0,06	-0,05	-0,06	0,01	-0,05	-0,02	0,02	0,15	-0,06	-0,02	1,00

Utifrån tabell 2 kan ses att korrelationen inte överstiger 0,3 mellan något par av variabler. Korrelationen ligger nära noll för samtliga variabelpar och antagandet om oberoende variabler kan antas uppfyllt.

5.3.2 Modellspecifikation

Skotteffektiviteten ses som den viktigaste faktorn i basket och variablerna för detta inkluderas först i modellen (modell 1). I modell 2 inkluderas även variablerna för turnovers som antas vara den näst viktigaste faktorn och därefter variablerna för rebounds (modell 3), frikast (modell 4) och slutligen sannolikhetsvariabeln (modell 5).

Tabell 3: Test av modellanpassning för fem olika modeller

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
-2Log L	243,166	236,228	222,975	213,381	198,166
Förändring -2log L	-	-6,938	-13,253	-9,594	-15,215
Skillnad i frihetsgrader mellan modellerna då variabler läggs till	-	4	4	4	1
Signifikans modelltest (-2 log L)	0,002	0,003	0,008	0,015	0,069
Signifikant förbättring av modellen då variabler läggs till (-2log L)	-	Nej	Ja	Ja	Ja
Andel korrekta klassificeringar	63,2%	62,7%	69,2%	70,8%	74,6%
Omnibustest signifikans	0,921	0,312	0,026	0,008	0,000
Cox & Snell R ²	0,001	0,038	0,104	0,149	0,217
Nagelkerkes R ²	0,001	0,051	0,142	0,204	0,296
Resultat Initiala odds	386	2018	4582	5206	5777
Resultat slutliga odds	1252	2818	5538	5941	6030

Från tabell 3 kan utläsas att $-2 \log L$ statistikan är sjunkande då man går från modell ett mot modell fem. Ju lägre värde på statistikan desto bättre är modellens anpassning. Lägst värde erhålls för den sista modellen. Enligt testet är det enbart för modell 5 som nollhypotesen inte kan förkastas. För de övriga modellerna förkastas nollhypotesen, modellen passar data, på 5%-nivån. Eftersom modellerna är nästade kan $-2 \log L$ statistikan även användas för att testa om modellen förbättras då de nya variablerna inkluderas. I den första modellen är värdet på $-2 \log L$ statistikan 243,166. När variablerna för turnovers inkluderas sjunker statistikan med 6,938 med en skillnad på 4 frihetsgrader. Detta är ingen signifikant förbättring av modellen då det inte överstiger det kritiska värdet på 5%-nivån ($\chi^2_{0,05}(4) = 9,49$). När de övriga variablerna inkluderas sjunker statistikan i sådan grad att förbättringen är signifikant i samtliga fall.

För omnibustestet kan nollhypotesen, alla koefficienter signifikant lika med noll, förkastas på signifikansnivån 5% för de tre sista modellerna. Enligt testet är modell 5 den bästa, lägst signifikansnivå.. Andelen korrekta klassificeringar ökar med modellerna och modell fem har högst andel korrekta klassificeringar med 74,6% vilket kan anses godkänt för modellen. Även värdena på Cox & Snell R² och Nagelkerkes R² ökar med modellerna och är högst för modell fem. Enligt Nagelkerke R² förklarar de oberoende variablerna 29,6% av variationen i matchresultatet. Detta värde är lågt, men dock högst för modell 5. Utifrån dessa statistiska test är modell 5 mest lämplig att använda i studien.

En sista faktor som används för att bestämma mest lämplig modell är att se till dess användningsområde. För varje modell testas vilket resultat den skulle ge på

spelmarknaden med de kända utfallen. Vinsten ökar för varje modell och har återigen bäst värde för den sista modellen. Den sista modellen antas vara mest lämplig då den har bäst värden på overalltesten, innefattar de viktigaste faktorerna och som kontroll för dess användningsområde så ger den högst vinst på spelmarknaden. Eftersom syftet med modellen är att beräkna sannolikheter i basketmatcher och samtliga faktorer anses vara viktiga ur en teoretisk synvinkel kommer samtliga variabler att återfinnas i modellen. Den valda modellen studeras nu mer i detalj.

Tabell 4: Estimat och signifikans av de oberoende variablerna i modell 5

Variabel	Estimat	Signifikans
Konstant	-11,108	0,032**
Home_EfgDiff	-0,029	0,415
Away_EfgDiff	0,037	0,341
Home_Oturn	0,020	0,781
Home_Dturn	-0,034	0,648
Away_Oturn	-0,285	0,001***
Away_Dturn	-0,084	0,261
Home_Orebs	0,092	0,011**
Home_Drebs	0,066	0,091*
Away_Orebs	0,068	0,106
Away_Drebs	0,068	0,074*
Home_Ofree	-0,009	0,785
Home_Dfree	-0,070	0,066*
Away_Ofree	0,059	0,103
Away_Dfree	0,034	0,337
Sannolikhet	0,053	0,000***

*** signifikans på 1%-nivån, ** 5%-nivån och * på 10%

I modellen är variabelerna *Sannolikhet* och bortalagets offensiva turnovers (*Away_Oturn*) signifikant på 1% nivån. Variabeln *Sannolikhet* har positivt tecken vilket betyder att ju högre hemmalagets sannolikhet är att vinna enligt spelbolagets initiala oddssättning desto högre är sannolikheten att hemmalaget vinner, vilket är rimligt. Variabeln *Away_Oturn* har ett negativt tecken vilket betyder att ju mer misstag som bortalaget gör desto lägre är sannolikheten att hemmalaget vinner, detta borde vara tvärtom. Variabeln *Home_Orebs* är signifikant på 5% nivån och har ett positivt tecken, vilket är korrekt enligt teorin. Ett högt värde på variabeln medför att laget är bra på att ta tillbaka bollen efter missat skottförsök vilket ger möjlighet till ett nytt försök att göra poäng. På 10%-nivån är

Home_Drebs signifikant med positivt tecken, återigen stödjer det vad variabeln ska mäta. Är laget bra på att ta bollen då motståndarna missat, minskar det chansen för bortalaget att återta bollen och vinstchansen för hemmalaget ökar således. Även *Home_Dfree* är signifikant på denna nivå och även tecknet på denna är som det bör vara. Ju sämre motståndarna är på att sätta sina frikast desto högre vinstchans för hemmalaget. Slutligen har *Away_Drebs* positivt tecken vilket inte stöds av teorin. Övriga variabler är inte signifikanta men antas ändå vara viktiga för sannolikhetsberäkningarna. Eftersom observationsmängden ökar så kan variabelernas effekt förändras med tiden. Det är således inte lämpligt att exkludera dessa variabler.

Tabell 5: Andel korrekta klassificeringar

Observerat utfall	Utfall enligt modellen		Procent korrekt
	Bortavinst (0)	Hemmavinst (1)	
Bortavinst (0)	39	29	57,4
Hemmavinst (1)	18	99	84,6
Total andel korrektklassificeringar			74,6

Modellen klassificerar korrekt i 74,6 procent av fallen, vilket kan ses som ett godkänt resultat (tabell 5). Men korrekt klassificering är inte det primära syftet i denna uppsats då det är sannolikheterna som är det viktiga. Antag exempelvis att observerat resultat är hemmavinst, men enligt modellen klassificeras det som bortavinst med 60% sannolikhet. Vi har alltså 40% vinstchans för hemmavinst och antag vidare att oddset på spelmarknaden är 3 ggr pengarna för hemmavinst. Det kommer då således att spelas på hemmavinst då skattad sannolikhet för hemmavinst multiplicerat med oddsen > 100 ($40 \cdot 3 = 120$). Så även om klassificeringen är låg så kan ändå modellen vara vinstgivande ur ett spelperspektiv.

Ett problem med analysen om hur bra modellen anpassar sig till data med testen ovan är att vi har en utomstående faktor, nämligen oddsen. Om modellen fungerar eller inte beror helt på vad oddsen på spelmarknaden satts till. Detta gör att lämplig modell och variabler främst måste väljas från ett teoretiskt perspektiv. Den valda modellen gav högst vinst om den används då resultaten är kända relativt de andra modellerna.

Tabell 6: Resultat om modellen används på spelmarknaden med kända utfall

Spel	Antal spel	Vinst	Förlust	Nettovinst (Vinst-Förlust)
Hemmalaget	99	5072	-2500	2572
Bortalaget	78	7158	-3700	3458
Totalt	180	12230	-6200	6030

Totalt över perioden är det 185 kända observationer. Modellen hittar spelmöjlighet i 180 av dessa. Totalt skulle en vinst erhållas på 6030 kr om 100 kr satsas vid varje speltillfälle på de oddsen som fanns strax innan spelstart. Modellen antas kunna användas för att beräknas sannolikheterna.

5.4 Beräkning av sannolikheter.

Modellen kommer nu att prövas ”i verkligheten” genom att den kommer att användas för att beräkna sannolikheterna för utfallen i varje match. I den första spelomgången i december fanns det sex matcher att spela på. Tyvärr så är inte utfallen kända längre (om vi visste dem så skulle vi vara väldigt rika). Sannolikheterna för utfallen i de sex matcherna beräknas genom de skattade värdena på koefficienterna i modellen samt de värdena på variablerna för respektive lag. Värdet på varje variabel är ett medelvärde av lagets prestationer i de tre senaste matcherna som laget spelat, d.v.s. värdet på exempelvis variabeln *Home_Oturn* är det genomsnittliga andelen offensiva turnovers laget haft de senaste tre matcherna.

För exempelvis matchen mellan Phoenix Suns mot Indiana Pacers, används estimaten från tabell 4 och värdena på variablerna i denna match. Utifrån formel 5 beräknas sannolikheten för hemmavinst.

$$P(\text{Result}=1) = \frac{1}{1 + \exp(-(-11,11 + (-0,029 \times 0,33) + (0,037 \times 1,43) + (0,02 \times 18,24) + \dots + (0,053 \times 58,81))} = 0,66$$

Detta betyder att den skattat vinstchansen för hemmalaget är 66% och därmed 34% för bortalaget. Oddsen på denna match var initialt 1,66 på hemmalaget och 2,37 på bortalaget. Innan matchstart hade oddsen rört sig till 1,60 för hemmalaget och 2,52 för bortalaget. Avkastning för spel på hemmalaget beräknas till $66 \times 1,6 = 109,56$ för startoddset eller 105,6 för slutoddset. Detta värde överstiger 100 och innebär att om det

spelas hemmavinst enligt dessa förutsättningar kommer man få tillbaka 109,56 kr eller 105,6 kr per satsad hundring (vinst på 9,56 kr eller 5,60 kr) om det spelar ett oändligt antal gånger. I detta fall resulterade det dock i en förlust.

När detta har gjorts för de sex matcherna, så ska sannolikheterna för matcherna i nästkommande omgång bestämmas. Utfallen i de senaste sex matcherna är nu kända och observationerna läggs till de tidigare kända observationerna. Modellen skattas om på nytt och de nya värdena på koefficienterna tillsammans med variabelvärdena på matcherna som ska skattas används för att beräkna de nya sannolikheterna. När detta gjorts görs processen om igen och håller på till sista omgången i mitten av april nåtts. Fördelen med att skatta om modellen är dels att vi har mer information än tidigare vilket ger bättre skattningar och dels att vi kan fånga upp förändringar i vad som är viktigt för sannolikheterna i utfallen. Noterbart kan vara att förändringen i koefficienternas värden är väldigt små från en skattning till en annan. Eftersom datamaterialet är väldigt stort kan det inte redovisas i denna uppsats, men kan vid intresse erhållas av författaren. Om modellens skattningar blir bättre (bättre på att nå korrekta sannolikheter) med mer information bör återspeglas i att vinsten ökar med tiden.

6. Resultat av studien

6.1 Spel till slutoddsen

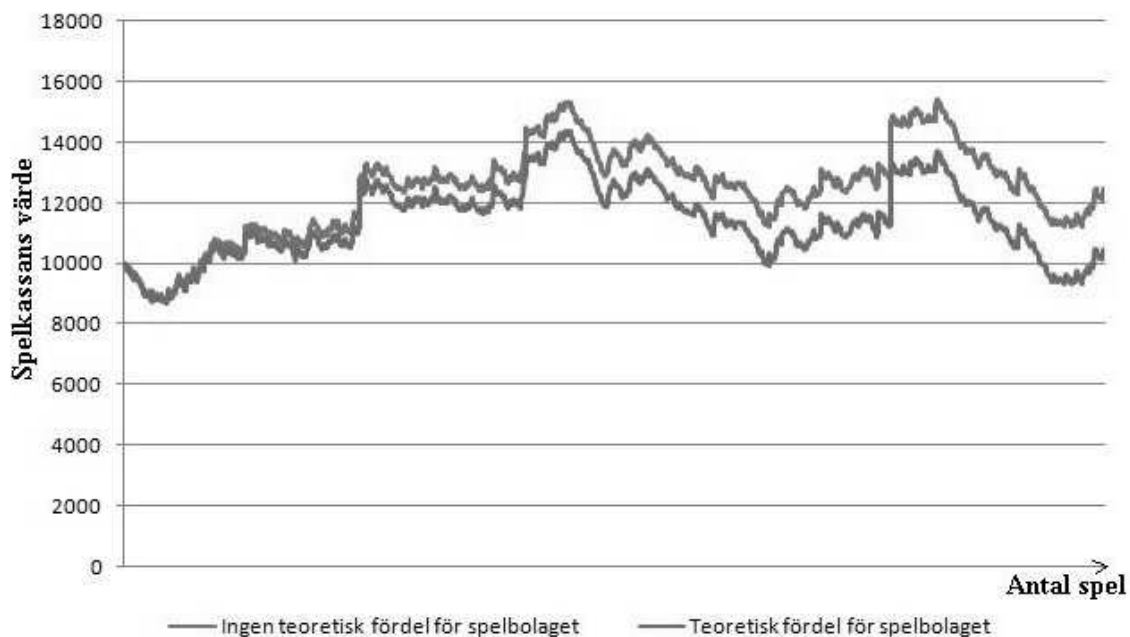
Under studieperioden har det varit möjligt att spela på 944 olika matcher. Med de skattade sannolikheterna från modellen i kombination med oddsen på spelmarknaden så har det funnits spelvärde i 851 av dessa matcher, givet att de estimerade sannolikheterna är de korrekta.

Tabell 7: Resultat på spelmarknaden med oddsen strax innan matchstart

Spel	Vinstspel		Förlustspel		Resultat	
	Antal spel	Belopp (kr)	Antal spel	Belopp (kr)	Antal spel	Belopp (kr)
Hemmalaget	243	18035	176	-17600	419	435
Bortalaget	153	27855	279	-27900	432	-45
Totalt:	396	45890	455	-45500	851	390

Det var i totalt 419 matcher mest fördelaktigt att spela på hemmalaget (tabell 7). Dessa spel resulterade i en vinst på 18035 kr samt en förlust på 17600 kr vilket ger ett sammanlagt resultat på +435 kr. I de 432 matcher där det var överavkastning (hypotetisk sett) på bortalaget blev det totalt en förlust på 45 kr. Sammantaget gav alla matcher som spelades till de oddsen som var strax innan matchstart en vinst på 390 kr. Utvecklingen av spelen illustreras genom att anta att en spelkassa på 10 000 kr avsatts för studien och att 100 kr spelats varje gång.

Figur 2: Utveckling av spelkassan i relation till antalet spel.



Den undre (röda) kurvan i figur 2 illustrerar spelkassans utveckling då det spelas till de faktiska oddsen som ges av spelbolaget. I början av studieperioden är utvecklingen negativ och en förlust uppstår. Värdet av spelkassan är som lägsta nere i 8714 kr innan den vänder uppåt och når som mest ett värde av 14333 kr. Noterbart är att spelkassan ligger på rätt sida om 10 000kr-strecket över stora delar av perioden. Spelkassans medianvärde ligger på 11368 kr och medelvärdet på 11421,88 med en standaravvikelse på 1241,67 kr. Mot slutet av studieperioden faller värdet i spelkassan och återigen uppstår en förlust. Värdet på spelkassan vid detta skede är som lägst 9340 kr. Återigen (för tredje gången) vänder kassan upp igen från förlust och avslutar på en liten vinst (+390 kr) jämfört med startläget. Vid tre tillfällen ökar värdet i spelkassan snabbt. Detta sker då spelkassan når över 12 000 kr och över 14 000 kr första gången och slutligen då den stiger från 11300 till 13124 kr. Anledningen till detta är att de sannolikheter som modellen ger i jämförelse med oddsen från spelbolaget indikerar på överavkastning på matcher med höga odds samt att dessa utfall går in. Vid första tillfället prickar modellen in en match till oddset 10,4 (940 kr i vinst), vid andra tillfället sitter två matcher till oddsen 6,9 och 7 gånger pengarna. Slutligen har vi den största stegringen då en match till oddset 19,24 sitter (1824 kr i vinst). Anledningen till att resultatet sjunker mot slutet kan

vara att vi trots allt inte har odds på spelmarknaden som är rättvisa ur en spelarens synvinkel. Spelbolagen ger inte odds i paritet med sannolikheten då de trots allt är vinstmaximerande företag med fördelarna på sin sida. Det är således möjligt att den höga vinsten i ungefär mitten av studieperioden är ett resultat av slumpen (tur) men när antalet spel blir tillräckligt stort blir det faktum att spelaren har sämre möjligheter än spelbolagen mer påtagligt. Detta undersöks genom att matematiskt justera (se avsnittet för odds) för den fördel som spelbolaget har genom att räkna om oddsen sådant att oddsen som ges motsvarar en sannolikhet av 100%. Resultatet med rättvisa odds illustreras av den övre blåa kurvan. Kurvorna följer ett liknande mönster med den skillnad att vid vinst stiger den blåa kurvan snabbare eftersom oddsen blir högre utan den teoretiska fördelen som spelbolaget har. I nedgångsperioder faller givetvis kurvorna lika mycket eftersom värdet av förlusten är lika stort i båda fall. De båda kurvorna divergerar. Det vill säga spelbolagets teoretiska fördel för större och större betydelse då antalet spel ökar. Det är således möjligt att när antalet spel blir ändå större (än de som är med i uppsatsen) kommer denna divergens medföra en förlust för spelaren och en vinst för spelbolaget. Skillnaden i det slutliga resultatet är stort. Med de oddsen som faktiskt ges blir vinsten 390 kr över perioden, medan om det givits odds utan fördel skulle vinsten uppgå till hela 2383 kr. Denna vinst är drygt 6 ggr större och spelkassan skulle ha växt med 23,83% jämfört med de 3,9% under de faktiska förhållandena. Den teoretiska fördelen som spelbolaget har på ungefär två procent får alltså stora konsekvenser på resultatet för spelaren. Utan någon teoretisk fördel hamnar spelkassan enbart på minus under den initiala perioden, efter det ligger den på plus hela tiden. Det verkar som att fördelen för spelbolaget är så pass stor att över ett stort antal spel kommer de vinster som görs på kort sikt ätas upp av oförelaktiga förhållanden.

Tabell 8. Fördelning av spelen

Spel	Vinstspel		Förlustspel		Resultat	
	Antal spel	Belopp (kr)	Antal spel	Belopp (kr)	Antal spel	Belopp (kr)
Hemmafavorit	187	8754	94	-9400	281	-646
Hemmaskräll	56	9281	82	-8200	138	1081
Bortafavorit	65	4087	33	-3300	98	787
Bortaskräll	88	23768	245	-24500	333	-732
Lika	-	-	1	-100	1	-100
Totalt:	396	45890	455	-45500	851	390

Tabell 8, visar hur spelen hamnar. Det finns ingen entydigt val av spel, d.v.s. det spelas ungefär lika mycket på hemmalaget som på bortalaget. Det finns inte heller någon indikation på att modellen enbart spelar favoriter eller skrällar¹¹. Spelen på hemmaskrällar och bortafavoriter var de spel som påverkade resultatet positivt, de övriga spelen hade sammanlagt ett negativt resultat. Intressant att notera att det är hemmaskrällar och bortafavoriter som är de vinstgivande spelen då de är varandras motsatser. Modellen är således relativt bra på att särskilja på när den skall spela på vilket. Det bästa resultatet vore alltså att låta bli att spela på hemmafavoritet och bortaskrällar, då hade vinsten uppgått till 1762 kr istället för det nuvarande resultatet 390 kr.

Tabell 9: Resultat av konsekventa strategier.

Spel	Vinstspel		Förlustspel		Resultat	
	Antal spel	Belopp (kr)	Antal spel	Belopp (kr)	Antal spel	Belopp (kr)
Hemmafavorit	480	18323	202	-20200	682	-1877
Hemmaskräll	94	15313	167	-16700	261	-1387
Bortafavorit	167	9691	94	-9400	261	291
Bortaskräll	202	45719	480	-48000	682	-2281

Som jämförelsenorm för resultatet av modellen så jämförs resultatet med olika ”konsekventa strategier” (tabell 9). Om vi inte hade tillgång till någon modell och enbart spelade på samtliga hemmafavoriter (och inget annat) eller enbart bortaskrällar skulle vi göra en förlust på 1877 kr respektive 2281 kr. Enbart om det spelas konsekvent på de lag som är bortafavoriter kan en vinst uppnås (+291 kr). Modellen som används gör en större vinst och är således bättre än samtliga av dessa strategier.

¹¹ Favorit definieras här av att oddset är lägre för detta lag än för (skällen). Med andra ord är det en högre sannolikhet att favoriten vinner.

6.2 Spel till startoddsen

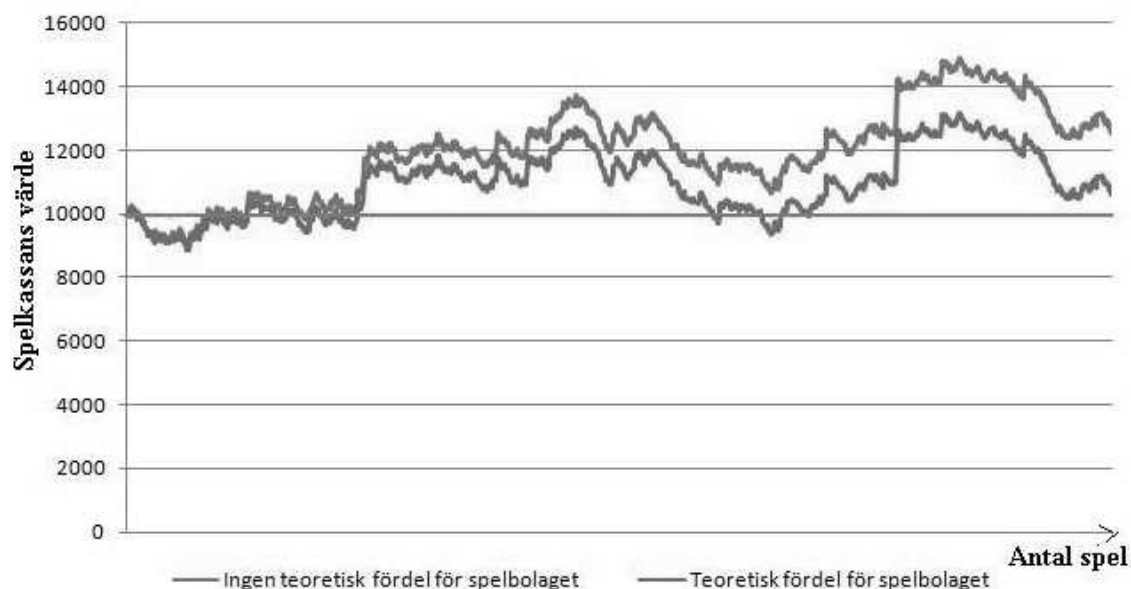
För att undersöka om tidpunkten för vid vilket spelen läggs påverkar resultatet undersöks även resultatet av att spela till de odds som spelbolaget först publicerade då spel på matchen öppnar.

Tabell 10: Resultat och fördelning av spelen till startoddsen

Spel	Vinstspel		Förlustspel		Resultat (Vinst-Förlust)	
	Antal spel	Belopp (kr)	Antal spel	Belopp (kr)	Antal spel	Belopp (kr)
Hemmafavorit	213	10015	103	-10300	316	-285
Hemmaskräll	52	8165	70	-7000	122	1165
Bortafavorit	75	4834	39	-3900	114	934
Bortaskräll	76	19362	206	-20600	282	-1238
Lika	1	95	1	-100	2	-5
Totalt:	417	42471	419	-41900	836	571

Totalt hittades spelvärde i 836 matcher och det sammanlagda resultatet är dessa spel blev en vinst på 571 kr, vilket är något högre än vad spelen till "slutoddsen" gav. Återigen är det spelen på hemmaskrällar och bortafavoriter som är de vinstgivande medan spelen på de övriga utfallen ger negativt resultat. Det är en högre andel spel på hemmalagen med även på favoriterna jämfört med spelen till slutoddsen. Detta beror på att de flesta spelare föredrar att spela favoriter eftersom det är störst chans att vinna. Det är troligtvis så att dessa spelare överskattar vinstmöjligheten för dessa favoriter. Således är odds på favoriterna högre initialt och modellen hittar spelvärde i fler hemmamatcher och favoriter. När det spelas till slutoddsen har odds rört sig så att värdet på skällarna har ökat och således blir det fler spel på dessa. Att både spel på favoriter och hemmalagen ökar samtidigt är inte så förvånande då de flesta hemmalagen också är favoriter till att vinna. Slutligen för startoddsen analyseras utvecklingen av en spelkassa på 10 000 kr där 100 kr spelas vid varje spel.

Figur 3: Utveckling av spelkassan om spelen görs till startoddsen



Kurvornas utseende påminner mycket med de för de slutliga oddsen och som lägst är spelkassan nere i 8854 kr och som mest uppe i 13174. Den största skillnaden är främst att spelkassan aldrig når lika höga toppvärden som då det spelas till slutoddsen. Att dessa nu är lägre beror främst på att stegringarna är som störst när matcher till höga odds går in och dessa odds är generellt sätt lägre nu. Massan av spelare har inte ännu hunnit trycka ned favoritoddsen (vilket ökar skrälloddsen). Om vi i samma matcher som vi nu spelar skrällar istället spelade dessa till slutoddsen så hade resultatet av detta blivit en vinst på 1263¹² kr jämför med en förlust på 73 kr (tabell 10) med startoddsen. Det verkar alltså så att i de fall som modellen indikerar spel på favoriter skall man spela tidigt (startodds) och om det istället är skrällar som har spelvärde ska man spela så sent som möjligt (slutoddsen). Återigen visar den övre (blåa) kurvan resultatet med rättvisa odds, vilket skulle givit en vinst på 2519 kr jämfört med nuvarande 571 kr. Kurvorna divergerar återigen när antalet spel ökar. Denna divergens kan tyda på att spelbolagens fördel kommer få en avgörande roll när antalet spel ökar. Vinsten som givits över spelen kan således vara tillfällig och resultatet kan övergå till förlust när antalet spel ökar ytterligare.

¹² Bortaskrällar skulle givit 76 vinstspel till ett värde av 20501kr och 206 förlustspel till -20600 kr, hemmaskrällar ger 52 vinstspel till värdet 8362kr och 70 förlustspel (-7000kr). Sammanlagt resultat blir då $20501 - 20600 + 8362 - 7000 = 1263$ kr.

Spelkassans medianvärde under studieperioden ligger på 10982 kr och medelvärdet på 10944,09 kr med en standardavvikelse på 1022,69 kr. Det slutliga resultatet av att spela till startodds eller slutodds har således ingen större betydelse men det viktiga från jämförelsen mellan dessa är att vi erhållit kunskap om att tidpunkten för spel på favoriter eller skrällar kan ge en betydelse på resultatet.

7. Slutsats

Syftet med uppsatsen var att med en logistisk modell skatta sannolikheter för utfallen i basketmatcher och undersöka om dessa kan användas på spelmarknaden för att erhålla vinster. Svaret på detta blir över perioden att det går även om vinsten inte blev så stor, +390 kr om det spelades till oddsen strax innan matchstart eller +571 kr till startoddsen. Men det har visats att det finns möjligheter att öka denna vinst genom att antingen selektivt välja typen av spel eller tidpunkten för vid vilket spelet läggs. Att enbart spela på de bortafavoriter eller hemmaskrällar på matcher som anses ha spelvärde skulle givit en vinst på 1181 kr med slutoddsen eller 2099 kr med startoddsen. Dessa vinster hade varit ungefär 4,8 respektive 3,7 gånger större än resultatet av samtliga "värdespel". En skattefri ökning av spelkassan på 11,81% eller 21% måste anses tillfredställande. Ett annat sätt att öka vinsten var att spela favoriter tidigt och skrällar sent. Hade alla spelen gjorts där det ansågs vara spelvärde i men med den justeringen att favoriter spelas tidigt och favoriter sent skulle vinsten blivit 1907 kr. Återigen en stor skillnad i resultatet. I uppsatsen har enbart ett spelbolag används men det är givetvis konkurrens mellan spelbolagen om spelarnas pengar. En konkurrensfaktor är således oddset. Det är alltså möjligt att på vissa matcher hitta högre odds hos andra spelbolag och därmed öka vinsten ytterligare, vilket har bortsetts ifrån i denna uppsats.

Samtidigt har det även visats att spelbolagets teoretiska fördel på ungefär två procent får stor betydelse på resultatet. Denna faktor kan få ändå större verkan när antalet spel blir ändå större vilket kan ses i att kurvorna för spelkassan utveckling med de faktiska oddsen och i den med rättvisa odds konvergerar. Det är således inte säkert att den positiva vinst

som erhöjls över denna säsong kommer att kvarstå när antalet spel ökar ytterligare. Visionen när jag började med denna uppsats var att den skulle innefatta flera säsonger, detta blev dock inte möjligt då datainsamlingen för enbart denna säsong tog mycket längre tid än jag trott. Det går alltså inte att ge rekommendationen att använda modellen för att göra vinster, trots det positiva resultatet. Risken för framtida förluster måste anses för stor relativt möjligheterna till konsekventa positiva vinster. Med uppsatsen har jag lärt mig att det är möjligt att använda en logistisk regression för att prognosera sannolikheter för utfall. I modellen kombineras expertbedömningar som variabeln *Sannolikhet* representerar i kombination med objektiva beräkningar som den logistiska modellen ger. Det har varit en lärorik tid med både framgångar och motgångar.

Litteraturförteckning:

Aiken Leona S, Jacob Cohen, Patricia Cohen och Stephen G. West (2002). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd Edition. Lawrence Erlbaum Associates. Inc, USA.

Afifi Abdelmonom, Virginia A. Clark & Sussane May (2004). *Computer-Aided Multivariate Analysis*. 4th Edition. Chapman & Hall, USA

Databasebasebasketball:

<http://www.databasebasketball.com/leagues/leagueyear.htm?lg=N&yr=2003>

Gamst Glenn, A.J. Guarino och Lawrence S. Meyer (2005). *Applied Multivariate Research: Design and Interpretation*. Sage Publications, USA

Hoopstats:

<http://www.hoopstats.com/basketball/fantasy/nba/detroitpistons/team/profile/04/7/1-1-1-eff>

Field Andy (2005). *Discovering Statistics using SPSS*. 2nd Edition. Sage Publications, USA.

Jackson Stu (2008). *Official rules of the National Basketball Association 2008-2009*. Basketball Operations department of NBA, USA. <http://www.nba.com/media/2008-09-NBARuleBook.pdf>

NBA Media Ventures, LLC. USA, New York.

http://www.nba.com/canada/Basketball_U_on_Western_ConferCanada_Generic_Article-18063.html

http://www.nba.com/canada/Basketball_U_on_Eastern_Confer-Canada_Generic_Article-18064.html

Oliver Dean (2003). *Basketball on Paper: Rules and Tools for Performance Analysis*. Protomac Books, USA.

Sharma Subhash (1996). *Applied Multivariate Techniques*. John Wiley & Sons Inc, USA.

Sportbookreview.com <http://www.sportsbookreview.com/Reviews/Pinnacle/default.aspx>

Tip-Ex International Limited. England. www.tip-ex.com.

Vaughan-Williams Leighton (2003). *The Economics of Gambling*. Routledge, USA