Assessing the number of goals in soccer matches

A Master's Thesis
by

Rasmus B. Olesen

Resume

This report documents the research and results made during a master's thesis in Machine Intelligence. The topic of the report is sports betting and the automatic assessment of the total number of goals in soccer matches.
The goal of the project is to develop, examine and evaluate proposed assessors, with regards to determining if it is possible to create a probability assessor which at the minimum can match the bookmakers' assessments on the total number of goals in soccer matches. Secondarily, it has been examined if it is possible using defined betting strategies and probability assessor to bet at bookmakers, and earn a profit.

This project proposes a total of three different probability assessors. The gamblers' approach uses the empirical probability in history matches, to assess the probability of a soccer match will have more or less than 2.5 goals. The Poisson approach uses a calculated expected number of goals for a match as the mean in a Poisson distribution, which forms a probability distribution over the number of goals. The third approach, is that of Dixon-Coles, which in the past has shown good results in predicting the outcome of matches. It utilizes history match data to form offensive and defensive strength measures to determine a probability distribution for the possible results of a match. These three approaches are measured and compared to the assessment of the bookmakers. In this report, formulas have been derived for determining the bookmakers' probability assessment for over or under 2.5 goals, using either the odds for the two total goal outcomes or by combining odds data for other over/under odds lines to derive the needed assessment. The assessors are in turn evaluated based on the scores achieved using an absolute scoring rule, where each assessment is assigned a score of the logarithm of the probability assessed for the observed outcome of the event. An assessors total score is its average log score over a total set of matches.

The secondary part of the project is to evaluate to different betting strategies. The first uses the expected value of a bet, to determine if a bet should be placed, using the known history odds data and a probability assessors assessment of a match. The second approach is a rule-based approach which uses the distance between the expected number of goals and the offered odds to determine if a bet should be placed. The strategies are evaluated on the basis of their ability to generate a profit and the total return of investment over a set of bets.

The parameters for each of the assessors have been tuned using a training data set containing a total of four and a half season of matches. Using the average log score as a measure, the best parameter settings for each of the assessors have been found. These settings were used to evaluate the assessors on a test data set containing a half a season. The results show, that the bookmakers' assessment is better than those of the assessors. Of the three proposed assessors, the gamblers' and Poisson approach was, a bit surprisingly, the better. The Dixon-Coles approach was the worst of the four in the larger part of the tests. In order to establish the statistical significance of the results found, hypothesis testing using the Wilcoxon Signed-Rank Test has been used. These tests showed, that no of the three proposed assessors where significantly better than the bookmaker, nor were any of them better than the other assessors. In one out of three tests, it was determined that the bookmakers' assessments were significantly better than those of the Poisson and Dixon-Coles approach.

The evaluations of the betting strategies gave irregular results. There was no consistent performance by any of the value betting strategies (using the proposed assessors), nor by the threshold strategy. In some of the strategy runs, some of the strategies, primarily the value betting using the Dixon-Coles assessor and the threshold strategy showed very promising results with a very high net profit. However, the inconsistency with very fluctuant net results and return of investments leads to the conclusion than none of the strategies would over a longer period in time be able to create a profit. If an even better probability assessor could be modeled, perhaps the value betting strategy could return a profit.

This report concludes, that it with the approaches taken, was not possible to create probability assessments which were better than those of the bookmakers. However, results show, that it is possible to almost match them. This leads to a possible discussion as to whether a cheap automatic probability assessor with assessments almost as good as a human bookmaker, could replace an expensive human bookmaker. This report proposes possible additions to the assessors evaluated in the project, in order for them to get even closer to the bookmakers' evaluations. Despite it not being possible to fully match and beat the bookmaker, the results of this report show indications of it being possible to create an automatic assessor which possibly with some additions could be used as a replacement of a human bookmaker, when setting the odds on sporting events.

AALBORG UNIVERSITET

Department of Computer Science
Selma Lagerlöfs Vej 300
DK-9220 Aalborg Ø
Telephone: (45) 9635 8080
Telefax: (45) 96359798
http://www.cs.aau.dk

Title:
Assessing the number of goals in soccer matches

Topic:
Machine Intelligence

Project period:
DAT6, spring 2008
February 4th - June 12th

Project group:
d531a

Member of the group:
Rasmus B. Olesen

Supervisor:
Manfred Jaeger

Number of copies: 3

Pages: 75

Synopsis:

This project proposes a number of models, for assessing the number of goals scored in a soccer match. The motivation lies in the challenge of automatizing the task of setting odds for certain type of goal bets in soccer betting. The models proposed uses historic result data solely, each in its own way, to assess the probabilities for the number of goals. The models use empirical probability, Poisson distributions and the concepts of offensive strength and defensive weakness. Each of the proposed assessors are measured with each other and compared to the assessments made by actual bookmakers. The evaluation is based on hypothesis testing using scoring rules. The results of this report conclude that, it is very difficult to create automatic probability assessors which can outperform the bookmakers. However it is possible, through rather simple methods, to create assessments which are very close to those of the bookmakers.

Preface

This project is a master's thesis, made at Aalborg University in the Machine Intelligence department. The topic of the project is automatic assessment of the probabilities of the number of goals occurring in a soccer match.
In the course of this project, knowledge and data have been provided by extern sources. I therefore give great thanks to:

- My supervisor Manfred Jaeger, for providing expert knowledge on data mining, statistics and learning.
- BetXpert.com and Tip-Ex.com for providing result and odds data.
- Klaus Rasmussen for the initial inspiration for the project, through the threshold betting strategy concept.
- Frederik Skov, Scandic Bookmakers, and Søren Hansen, Danbook, for their input on assessing the number of goals in soccer matches.

A CD containing the code and implementation has been enclosed in the report.

This project was made by:

Rasmus Bang Olesen

Contents

Chapter 1

Introduction

Since the betting industry went online in the 1990's, there has been a tremendous growth in all areas. More and more bookmakers emerge, now with a total of several hundred providers. As the turnover grows, and the competition increases, the betting market is constantly evolving and many attempts are made towards lowering costs and maximizing the profit.

1.1 Motivation

The concept of gambling has been around for centennials, evolving from chance games with dice to modern day casinos. In the late 19th century the preliminaries to present day bookmakers saw the light of day, being the first step on the way to the multi billion dollar industry that the betting industry is today. For the remainder of this report, the term betting industry refers to the market of bookmakers with focus on gambling on sporting events, while the term betting is sports gambling. The betting industry rests on the conflicting interest of bookmakers and customers wanting to earn money, respectively. In Chapter 2 the structure of sporting event betting is introduced, and the mathematical conflict of bookmakers and customers are explained in detail. Needles to say is, that customers seek to win money away from bookmakers, while bookmakers want to, in the long run, create a steady return. Both sides will look for ways to improve their possibilities of achieving their goals, which raises a motivation of investigating how this could be done. From a customers point of view, it would most certainly be welcome if a "machine" for finding good bets existed, which could guarantee a constant profit. To believe that it is possible to create such a mechanism, is naive. However it is none the less possible to generate a positive return on investment by betting at the bookmakers (see Appendix A.1). It is interesting to look into formalizing the methods of a professional gambler to see if algorithms for evaluating odds offers to see if a bet should be made. From the bookmakers' point of view, a tool for determining (close to) correct probability distributions and setting odds on sporting events would be a very strong instrument. Bookmakers' have employees which are constantly alert for news of the sports world, to always be a step ahead when offering odds. Being well informed is a time consuming task, and expensive for a bookmaker with a large team of odds setters. If the employees tasks of creating the odds could be eased, then time and money could be saved. If it was possible to create applications strong and good enough, the odds setters could be out of a job. However, it is not as simple as that, but it is interesting to see how far one can go towards creating tools for determining probabilities and setting odds.

1.2 Goals

In sports betting the largest sport is soccer, and the best known bet type is outcome betting: "Who wins the match between Arsenal and Manchester United?". The intense competition between the great number of bookmakers has over the last decades caused the development of new betting types. For soccer a very popular betting type is "Total goals", which is betting on the total number of goals in a specific match, i.e. "Over or under 2.5 goals?". Any soccer team can in some way be said to have a tendency towards the goals scored in a match. For the top teams it can (in most cases) be said that they are good at scoring goals, and keeping the opponents from scoring. For a bottom team the opposite can be said, that they are poor at scoring goals and at defending. If two teams which are good at scoring goals, and mediocre in defending is to play each other, it would be a fair assumption to say that there is a good chance of a lot goals in the match. The goal of this project is to, in a number of different approaches investigate the correlation between a soccer teams history results and the probability distribution for the number of goals in a given soccer match. The examination of soccer result and odds data is to lead to the establishment of a model for assessing the probability of the number of goals scored in a soccer match. The model will use only historical match result data, and has no other prior information on the match which it is assessing. The goal of this project is to put forth a number of candidate assessors and through evaluation establish the possibility of creating an automatized assessment of the number of goals in a soccer match, which can at the least match those of the bookmakers.

1.3 Report Structure

Chapter 2 gives an introduction to sports betting, and introduces basic concepts of betting which are necessary for understanding the remainder of the report. In the chapter a number of probability assessor models are presented, as well as two betting strategies, which are to be modeled, implemented and evaluated in later parts. Chapter 3 introduces and defines theoretical concepts which are used in the implementation and evaluation process. Chapter 4 gives an introduction to the two data sets; a result data set containing the results of soccer matches over the last five years, and an odds data set containing historical odds data for a season of matches. Chapter 5 explains the training and evaluation process used in the project. Chapters 6 through 9 presents the modeling and implementation of the four assessors used in the project, while chapter 10 defines the betting strategies. In all of these chapters, the the training of parameter settings have been made, in order to optimize the performance of the assessors and strategies for the final tests. Chapter 11 presents the results found for both the assessors and betting strategies, and discusses which of these have shown the best performance. Chapter 12 reflects on the project, and put forth possible future additions for the models to improve the predictions, and concludes on the project goals.

Chapter 2

Sports Betting

The goal of this project is primarily to create a prediction model which can assess the probabilities of the number of goals in a soccer match as well as the bookmakers. Secondarily the goal is to devise a betting strategy, which in the long run can minimally break even when betting at bookmakers. In order to create such a model and strategy, one must fully understand the mathematics behind sports betting odds and the mechanisms which influence them. In this chapter the basic theoretical background for understanding sports betting is presented, and various concepts related to this are accounted for. Finally previous works in the subject of odds assessment and betting are discussed, with regards to their possible utilization in this project.

2.1 Calculating Odds

A given sporting event has a finite number of outcomes. For a soccer match, for instance, the number of possible outcomes is three, and can be one of *home*, *draw* or *away*. For any given sporting event with n possible outcomes, $outcome = 1, ..., n$, the probability of the outcome i is $P(outcome = i)$. The outcomes are mutually exclusive, since a match can not have two winners for example. So the following holds:

$$\sum_{i=1}^{n} P(outcome = i) = 1 \tag{2.1}$$

2.1.1 European Odds

When speaking of odds, there are several odds formats used on the market. In America moneyline odds are used, while some British bookmakers use fractional odds. Both types differ from the decimal type, which is most commonly used format in Europe. A decimal odds of 1.80 means, that if you place a wager with a stake s, your winnings will be $s \cdot 1.80$, and you net winnings will be $(s \cdot 1.80) - s = s \cdot 0.80$. The decimal odds type, will be the type used for the remainder of this report.

Bookmakers are basically companies trying to make money through sports wagers. Therefore they operate with a theoretical payback percentage when offering odds to their customers. The theoretical payback is set by the bookmaker and is the percentage of the turnover on a betting event which is expected to be paid back to the customers. The payback is less than 100%, normally around 90-95%. The higher the percentage, the lower the margin of theoretical profit for the bookmaker and the higher the odds. An odds for an outcome, can be calculated as:

$$Odds_i = tpb \cdot \frac{1}{P(Outcome = i)},$$

where $Odds_i$ is the odds for outcome i, and tpb is the bookmakers theoretical payback. An odds calculated with a payback of 1 (100%) is called a fair odd, since there is no theoretical advantage.
For a soccer match with the outcomes home, draw and away, with a probability distribution of 60%, 25% and 15% respectively the odds can be calculated. Using equation 2.2 with a theoretical payback of 100% the odds for the respective outcomes would be 1.67, 4.00 and 6.67. If instead a payback of 92% is used, the odds would be 1.53, 3.68 and 6.13. A rather big difference. If one knows the theoretical payback for an event and the odds of one of the outcomes, the bookmakers assessment of the probability of the outcome can be found:

$$P(Outcome = i) = \cdot \frac{tpb}{Odds_i},$$

If one has the odds for all possible outcomes for an event, the theoretical payback can be calculated:

$$tpb = \frac{1}{\sum_{i=0}^{i} \frac{1}{Odds_i}}$$

In this report the focus is on the total number of goals in a soccer match and the bets possible in this area. A very popular variety, which is offered by almost any bookmaker, on almost any match, is "over/under 2,5 goals". If the total number of goals in a match is 0, 1 or 2, then $outcome$ ="under 2.5" obviously, and if there are three goals or more, then $outcome$ ="over 2.5".

2.1.2 Asian Odds

In the sports betting market, some of the biggest bookmakers reside in Asia. Here it is very common, that for soccer matches the draw result is eliminated from the possible outcomes. The match is assigned a so-called Asian handicap. In a match between A and B, one team is assigned a handicap, for instance -0.50 goals. The opposition is assigned +0.50. When the final result of a match is in, the handicap is added/subtracted from the result and the outcome of the bet is found. If a bet is made on A, with a handicap of -0.50 and the result is 2-1 in the favor of A, the bet result is 1.50-1 and the bet is won. If

the result was 1-1, the bet result is 0.50-1 and the bet is lost. The Asian variety of odds always has two outcomes, and the application of the handicap ensures the possibility of making a more "even" match up. If team A is a very big favorite, with 75% chance and 92% payback percentage giving and odds of 1.23. By instead assigning A with a handicap of -1.5, and setting $P(A-1.5)$ to 48%, $Odd_{A-1.5}$ is 1.77, while $Odd_{B+1.5}$ is 1.92. If the handicap is -0.00, this means that the team has no actual handicap. However the event is still regarded to have only two deciding outcomes. If the match ends in a draw, the bet is paid back as a win with odds 1.00. Similarly, if A has a handicap of -1,00 and they win by exactly one goal, the bet is "voided" and payed back.
A very interesting, and complicated, aspect of the Asian handicaps is the possibility of quarter handicaps. Instead of a handicap -0,5, it is possible to have a handicap of -0,25, which is called a split bet, where the stake is divided into two and placed on the separate bets; one with handicap -0,00 and one with handicap -0,50. When the match result is in, the two bets are evaluated separately.
The Asian variety is also very common in the over/under market. Instead of a line of 2,5 goals, it is not unusual to see 2,25 or 2,00. In the latter case, a score of exactly two goals would result in a void bet, while in the case of two goals in an over 2,25 line bet, the result would be a half loss (refund on the half of the bet on over 2,00 goals, and loss on over 2,50 goals). In an under 2,25 line bet, a result of two goals would yield a half win (refund on under 2,00 goals, and win on under 2,50 goals). Table 2.1 shows a win/loss explanation for the most commonly used lines for the possible goal outcome.

	0-1 goals	2 goals	3 goals	4 goals or more
Over 1.75 (Over 1.50 and 2.00)	Loss	Half Win	Win	Win
Over 2.00	Loss	Void	Win	Win
Over 2.25 (Over 2.00 and 2.50)	Loss	Half Loss	Win	Win
Over 2.50	Loss	Loss	Win	Win
Over 2.75 (Over 2.50 and 3.00)	Loss	Loss	Half Win	Win
Over 3.00	Loss	Loss	Void	Win
Over 3.25 (Over 3.00 and 3.50)	Loss	Loss	Half Loss	Win
Under 1.75 (Under 1.50 and 2.00)	Win	Half Loss	Loss	Loss
Under 2.00	Win	Void	Loss	Loss
Under 2.25 (Under 2.00 and 2.50)	Win	Half Win	Loss	Loss
Under 2.50	Win	Win	Loss	Loss
Under 2.75 (Under 2.50 and 3.00)	Win	Win	Half Loss	Loss
Under 3.00	Win	Win	Void	Loss
Under 3.25 (Under 3.00 and 3.50)	Win	Win	Half Win	Loss

Table 2.1: Explanation of win/loss on Asian line bets.

When placing ones bet and several lines are offered, choosing the "correct" line can be crucial, as can be seen from the table. However more important, is the odds at which you bet. A bet must have value, in order for a gambler to win in the long run.

2.1.3 Betvalue

The term betvalue is often used in discussions about bets between gamblers. The reason for bookmakers making fortunes is, that due to the margin achieved through the theoretical payback, most gamblers place bets which are under value. Meaning, that in the long run the bookmaker wins. At a casino, for instance, all games have rules for how the game plays out and how winnings are won. These rules are carefully set, so that the casino in the long run will make money. A single gambler can very well get lucky and score a big winning, but were he to carry on playing he would in the end have less money than when he started. We define the betvalue for the outcome i, BV_i, as:

$$BV_i = Odds_i \cdot P(outcome = i) \tag{2.2}$$

Here $Odds_i$ is the odds for outcome i, and $P(outcome = i)$ is the assessed probability that the outcome will happen. A bet is then said to have value if, the odds and probability together yields a betvalue, $BV > 1$.
The roulette is a good example of this. The board is a spinning wheel, with 37 numbers on it. 18 are marked as black numbers, 18 are marked as red and one is green. It is the single green number that does the trick. If a gambler plays "red", and the color comes out, he wins an amount equal to the stake placed. He doubles up. If the color comes out black, he loses. If the color comes out green he loses. The chance winning is therefore $\frac{18}{37}$ and the odds is 2,00. The betvalue is calculated:

$$BV = \frac{18}{37} \cdot 2{,}00 = \frac{36}{37}$$

Since $BV < 1$, the bet can be said to be under value, and should not be placed. At least not if the meaning is to win.

2.2 Candidate Assessors

With the problem setting in mind, a total of four candidate assessors is proposed in this section. One of these is the bookmakers' prediction, which bases its assessments on the bookmaker odds. Two of the assessors are, somewhat, naive approaches, which respectively use the number of total goal instances and the average number of goals as a means for predicting. The fourth approach, the Dixon-Coles model, has been selected from the research made in the field of soccer match prediction, due to its nature fitting the problem setting of this project well.

2.2.1 Gamblers' Assessment

An often used approach by gamblers, is a rather naive approach. Here the probability of a match having more or less than 2.5 is assumed to be the empirical probability over prior

matches. If, i.e. a team has played 11 matches, with 7 of them having more than 2.5 goals, the probability is assumed to be $\frac{7}{11} = 0.636$. The gamblers' assessment approach uses results of prior matches, to count the instances of matches with a certain number of goals, and uses the empirical probabilities as the probability assessment of the number of goals in a match.
This very simple assessment is the common approach for identifying over/under bets of a lot of gamblers, hence the name of this approach. The bookmaker is aware of this, and the odds set for a match takes this into account (see Appendix A.2 for interview).

2.2.2 Poisson Distribution Assessment

The Poisson approach is a naive prediction model, which also uses past results to assess the probability of the number of goals in a match. This model assumes, that the number of goals in a soccer match follows a Poisson distribution. For a given match, the two participating teams taken into account, the average number of goals, Avr, in prior matches are calculated. A prediction of the probability distribution of the number of goals is then made, based on a Poisson distribution with mean value Avr.

2.2.3 Dixon-Coles Assessment

Dixon and Coles[DC97] formerly proposed a model for assessing probabilities of soccer match outcomes, based solely on the number of goals scored in previous matches by the two participating teams. The usability of the assessor with regards to predicting the outcome of a soccer match was explored in [CH], with good results. The approach forms an offensive strength α and a defensive weakness β parameter for the participating teams in a match under assessment. It also takes into account the dependencies between certain scores in a match and incorporates the advantage of playing at home. The model utilizes Poisson distributions over the products of the strength and weakness values, to form a probability distribution of the possible results for the match (0-0, 1-0 etc). For this project the Dixon-Coles method will be used to predict the number of goals in a match, and not the actual outcome. The probability of the number of goals, is the sum of the probabilities for results where the total number of goals are the same. The probability of 1 goal in a match is therefore the sum of the probabilities of 1-0 and 0-1.

2.2.4 Bookmakers' Assessment

The fourth, and final, prediction model is based on the actual odds from the bookmakers. As stated in equation 2.2 the odds for an outcome can be calculated using the bookmakers' payback percentage and the probability estimation. Therefore, by knowing the odds and the payback percentage, the bookmakers' probability assessment can be calculated:

$$P(Outcome = i) = \frac{tpb}{Odds_i}$$

In this way, from a data set containing odds for soccer matches, the bookmakers' probability assessment can be calculated. For this project, the bookmakers' assessment is an important part of the evaluation process, to see if the other proposed probability assessors are better than or match the bookmakers' assessments.

2.2.5 Assessor Comparison

Table 2.2 shows an overview of the proposed assessors, and sums up what parameters are estimated by each of them. The difference in the models can also be seen in the actual outcome predictions, which are shown in the third column of the table. In the fourth column a description on how the predictions by the model is used for over/under 2.5 probability assessment. Common for all of the assessors is, that they are all provided with a data set containing match result data and a data set containing odds data for these matches. The descriptions in the table, is made with regard to a single match under assessment, between a home team, t_h, and away team, t_a, using a match data set, $m_{h,a}$, containing k matches where the home team is equal to t_h and away team equal to t_a.

2.3 Betting Strategies

When deciding to place a bet on a soccer match, it is individual from person to person how the decision is made. Some people put more emphasis on recent form, where others may regard the weather forecast as a more important aspect of deciding to bet or not. Each person has a personal betting strategy for making this decision. For the remainder of this report, a betting strategy is a combination of a probability assessment and a set of rules, which provided with a data set containing soccer match results and bookmaker odds can decide whether to place a bet on an outcome or not. In this report two different betting strategies will be used.

2.3.1 Value Betting Strategy

In gambling communities, the concept of value betting is widely used. In order to generate a positive return on investment, one needs to place bets only if there is a positive expected value in doing so (see Section 3.1 for details). In order to calculate the expected value of a bet, a probability assessment of the possible outcomes are needed, along with the corresponding odds. In this report, the value betting strategy will be evaluated using three different assessors; Gamblers' assessment, Dixon-Coles assessment and the Poisson Distribution assessment. Over a set of matches and corresponding odds, the bet with the highest positive expected value, if any, is selected based on the assessment of the assessor under evaluation. When all matches have been assessed, and bets have been "placed", the bets are payed out, according to the results of the matches. A positive or negative return for each assessor can then be determined and compared.

2.3.2 Threshold Strategy

The second betting strategy is a rule based strategy, which assumes that the average number of goals in previous matches of participating teams is the expected number of goals in a match under assessment. It uses this value as a measure for deciding to bet or not. If the distance between the expected number of goals and the line of the offered odds exceed a threshold, a bet is placed if the odds meets a predetermined level.
The initial proposal by Klaus Rasmussen, used a threshold of 0.25, however no documentation for the choice of this value has been made. Therefore for this strategy an examination of what threshold value maximizes the winnings will be performed, while a further examination of the odds criteria also will be examined.
This betting strategy also finds the single best bet, that meets the requirements, if any, and places a bet. Using the results of the matches under evaluation, the selected bets are payed out and the total return is calculated.

2.4 Summary

The foundations for the project has been set, introducing the basic concepts of bookmaking and sports betting. Based on knowledge about the sports betting industry and research into prior attempts at predicting soccer scores, four assessment approaches have been proposed, and will in turn be examined and evaluated The concept of betting strategies has been accounted for, and two different strategy types have been presented. In the following chapters the approaches will individually be implemented and examined, and the parameters tuned to maximize the performance. Finally the assessors and strategies will be evaluated on a test data set, and compared to establish the better probability assessor and the better betting strategy.

Assessor	**Estimates**	**Predicts**	**Over/under 2.5 Prediction**
Gambler	From $m_{h,a}$ the number of instances where the total number of goals is under 2.5, x_u and the number of instances where it is over 2.5 goals, x_o.	Using the counts, the empirical probability for over and under 2.5 is found: $P(< 2.5) = \frac{x_u}{k}$ and $P(> 2.5) = \frac{x_o}{k}$	Given by initial prediction.
Poisson	From $m_{h,a}$ the average number of goals per match, x_{avr}, is calculated.	Calculates the probability distribution of the number of goals in the match under assessment as the Poisson distribution with mean value x_{avr}.	$P(< 2.5)$ is the sum of the Poisson distributions probabilities where the number of occurred goals is less than 2.5. Similarly for $P(> 2.5)$
DC	Estimates Offensive strength, defensive weakness and home team advantage parameters	Using the product of Poisson distributions, a probability distribution is made over the possible results of the match, e.g. P(result = 0-0).	$P(< 2.5)$ is the sum of the probabilities of the match results where the sum of the goals is less than 2.5, and similarly for $P(> 2.5)$.
Book	(Uses odds data)	Calculates the probabilities of the over and under outcomes using formula 2.2.	Given by initial prediction.

Table 2.2: Overview of the four assessors, showing what estimates are made and how the probability assessment is made for a single match.

Chapter 3

Theoretical Concepts

This chapter serves the purpose of introducing theoretical concepts and terminology used in this project. Firstly the concept of expected value is defined, followed up by a definition of the Poisson distribution, while the rest of this chapter is devoted to the evaluation of predictions by the proposed assessors. Here scoring rules and testing methods are accounted for.

3.1 Expected Value and Loss

The expected value can be calculated as the sum of the probability of all possible outcomes multiplied with the corresponding gain or loss.

Definition 1. *(Expected Value). For an event with discrete outcomes, the expected value is[DS]:*

$$E[X] = \sum p_i x_i,$$

where p_i is the probability of outcome i, and x_i is the is reward given for the outcome i.

An example can be a soccer match between two soccer teams, where the probability of the three possible outcomes $(home, draw, away)$ is $(0.30, 0.31, 0.39)$ and the corresponding odds is (3.00, 3.10, 2.65). The expected values of the three possible bets are:
E[Bet = Home win] $= \sum p_i x_i = 0.30 \cdot 3.00 + 0.31 \cdot 0 + 0.38 \cdot 0 = 0.90$
E[Bet = Draw] $= \sum p_i x_i = 0.30 \cdot 0 + 0.31 \cdot 3.10 + 0.38 \cdot 0 = 0.961$
E[Bet = Away win] $= \sum p_i x_i = 0.30 \cdot 0 + 0.31 \cdot 0 + 0.39 \cdot 2.65 = 1.034$

Of the three possible bets on the outcome, with the mentioned probability assessment, the only bet with a positive expected value is the bet on "Away win".

3.2 Poisson Distribution

The Poisson distribution is a discrete probability distribution, known from probability theory and statistics. It expresses the probability of a number events occurring within a fixed period of time.

Definition 2. *(Poisson Distribution) Let* k *be a number of occurrences and* $\lambda \in \mathbb{R}$ *and* $\lambda > 0$ *be the mean value, then Poisson states that the probability of* k *occurrences is [DS]:*

$$f(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Here $f(k|\lambda) \geq 0$. The restrictions on λ ensures that $e^{-\lambda} \leq 0$, $\lambda^k \leq 0$ and $k! \leq 1$.
Being a discrete probability distribution, it is certain that:

$$\sum_{k=0}^{\infty} f(k|\lambda) = 1$$

3.3 Scoring Assessors

In order to draw conclusion about which probability assessor is the better, a common evaluation method is needed. In this section, concepts for evaluating the quality of a probability assessors' prediction is presented, along with methods for evaluating the results of different betting strategies.

3.3.1 Scoring Rules

In order to decide the quality of a probability assessor, it is necessary to be able to evaluate its assessments compared to the actual outcome of the event assessed. We introduce scoring rules as a measure of quality.
A scoring rule is a function, F, which takes a probability assessors assessment in the form of vector $\vec{P}$ and an observation of the outcome $\vec{D}$ and returns a score. The assessment made by the assessor must be made prior to the observation of the outcome.

$$F : (\vec{P}, \vec{D}) \rightarrow \mathcal{R} \tag{3.1}$$

For this project two scoring rules are taken into consideration; quadratic scoring and logarithmic scoring.
Before presenting the two rules, some notation should be in place. Let E be an event, under assessment, with n mutually exclusive outcomes $(E_1, ..., E_n)$. Let vector $\vec{R} = \langle r_1, ..., r_n \rangle$ be a probability assessors assessment, $\vec{P} = \langle p_1, ..., p_n \rangle$ be the true probability distribution and $\vec{D} = \langle d_1, ..., d_n \rangle$ represent an observation of the event, $d_i = 1$ if E_i occurs, and zero otherwise. Here $r_i \geq 0$, $r_i \geq 0$, $\sum_{i=1}^{n} r_i = 1$ and $\sum_{i=1}^{n} p_i = 1$.

3.3.1.1 True Probability and Properness

Notice the above distinction between the assessors assessment $\vec{R}$ and the true probability distribution of the event $\vec{P}$. For soccer matches, it is very hard to determine a precise (and true) probability distribution, since there are a vast amount of influencing factors and their actual influence is very difficult to decide. Imagine if the same soccer match, under the exact same conditions, was repeatedly played. We then define the probability distribution over the possible outcomes as the frequency of identical outcomes. In the case of the bet type under examination in this report, the count of matches with a total of under 2.5 goals and over 2.5 goals respectively.
Winkler introduces the scoring rule property of properness[WM]. He regards a "perfect" assessor as one, that obeys the postulates of coherence and makes assessments based on true judgement. For some scoring rules (which are not proper) it is possible to maximize the expected score by adjusting the assessment. The expected score is similar to the expected value function introduced in Section 3.1, where instead of using odds as a score, the score from the scoring rule is used. The expected score is:

Definition 3. *(Expected Score). Let $\vec{R}$ be a probability assessors' assessment of an event with i outcomes. Let $\vec{D}_i$ be an observation of the i'th outcome, where the i'th entry is 1 and all other entries 0. The expected score is:*

$$E[S(\vec{R},\vec{D})] = \sum_i P_i \cdot S(\vec{R}_i,\vec{D}_i)$$

With a proper scoring rule, the score is maximized if, and only if, the assessors assessment $\vec{R}$ is set to the true probability distribution $\vec{P}$. We define properness property:

Definition 4. *(Properness). Let $\vec{P}$ be the true probability distribution for the event event with i outcomes, and $\vec{R}$ be a probability assessors' assessment. Let $\vec{D}_i$ be an observation of the i'th outcome, where the i'th entry is 1 and all other entries 0. A scoring rule is said to be proper, if the following holds:*

$$E[S(\vec{P},\vec{D})] \geq E[S(\vec{R},\vec{D})]$$

In the following sections, two scoring rules will be presented, both of them being proper.[WM]

3.3.1.2 Quadratic Scoring

The quadratic scoring rule uses the sum of the squared difference between pairs in $\vec{R}$ and $\vec{D}$, (r_i,d_i) as a measurement:

$$Q(\vec{R},\vec{D}) = 1 - \sum_{i=1}^{n} (r_i - d_i)^2 \tag{3.2}$$

If the j'th outcome is observed, then d_j is 1, and all other entries in $vecD$ is 0. This yields:

$$Q_j(\vec{R},\vec{D}) = 1 - (r_j - 1)^2 - \sum_{i \neq j} (r_i - 0)^2 = 2r_j - \sum_{i \neq j} r_i^2 \tag{3.3}$$

From 3.3 it can be seen, that if r_j is set to 1, this maximizes the score. The expected score for the quadratic scoring rule is:

$$E(Q) = \sum_j p_j(2r_j - \sum_{i \neq j} r_i^2) \tag{3.4}$$

or

$$E(Q) = \sum_j p_j - \sum_{i \neq j} (r_j - p_j)^2 \tag{3.5}$$

The expected score is therefore maximized if $\vec{R}$ is set to $\vec{P}$, and the quadratic scoring rule is therefore proper.
The quadratic scoring rules takes the assessment of all outcomes into consideration when scoring. Two assessors A_1 and A_2 has the probabilities $\langle 0.4, 0.3, 0.3 \rangle$ and $\langle 0.45, 0.5, 0.05 \rangle$ respectively. The first outcome is then observed, yielding a quadratic score of 0.46 for A_1 and 0.445 for A_2. One would expect that A_2 which has the highest probability for the observed outcome. However, the results are due to that the quadratic scoring rule penalizes the unobserved outcomes. If an assessor has an uneven assessment for the unobserved it is penalized more than if it had an even distribution over the unobserved outcomes. A significantly uneven distribution over the unobserved outcomes will blur the ability to evaluate the score without looking at the single assessment.

3.3.1.3 Logarithmic Scoring

The logarithmic scoring rule is different from the quadratic scoring rule in the sense that it only takes into account the probability assessment of the observed outcome.

$$L(\vec{R}, \vec{D}) = \ln \sum_{i=1}^{n} r_i \cdot d_i \tag{3.6}$$

If outcome j is observed, then the logarithmic score is:

$$L_j(\vec{R}, \vec{D}) = \ln(r_j \cdot d_j) + \sum_{i \neq j} (r_i) \cdot d_i = \ln r_j \tag{3.7}$$

The fact that only r_j is taken into account, yields that the higher the probability assessed for the observed outcome the higher the score. Reusing the example from before, the logarithmic scoring rules gives A_1 a score of -0,916 and A_2 a score of -0,799. Here A_2 is, as one could expect, regarded as the better assessor.

3.4 Hypothesis Testing

Hypothesis testing[TSK] is a statistical inference procedure used to determine whether or not a hypothesis should be accepted (or rejected) based on results derived from data.

Among other things, hypothesis testing can be used for validating the significance in the difference in performance between two classification or prediction models.
In hypothesis testing often two contrasting hypothesis are used; the null hypothesis and the alternative hypothesis. The procedure of testing hypothesis can be seen as a four step procedure:

1. Establish the null hypothesis and the alternative hypothesis

2. A test statistic θ is defined, to be used to determine whether or not to accept the null hypothesis.

3. For the data, compute the value of θ, and determine the p-value using the probability distribution of the test statistic.

4. Definition of a significance level, which is used for determining in which range the θ values leads to rejection of the null hypothesis.

It is often used, that the null hypothesis is formulated as an unwanted result, while the alternative hypothesis is actually the result one is seeking for. The objective of the test is then to reject the null hypothesis.
When performing hypothesis testing, there are two types of errors which one can make. A type 1 error is rejecting a true null hypothesis, while a type 2 error is accepted a false null hypothesis. For this project the hypothesis which are to be tested regards one assessors performance against another (see Section 5.1.2). The testing and evaluations in this report must therefore minimize the type 1 and 2 errors. The use of Wilcoxon Signed-Rank Test[Wil] ensures this.

3.4.1 Wilcoxon Signed-Rank Testing

The Wilcoxon Signed-Rank Test is a non-parametric statistical test, usable for measuring and comparing two related samples. The Wilcoxon test uses the difference between pairs in the two samples, to determine if the two samples are the same or to identify differences. Assume the presence of two samples X and Y, which both are of size n. Each member in X has a corresponding member in Y. Together these form a pair. Let $\vec{X} = (x_1, ..., x_n)$ and $\vec{Y} = (y_1, ..., y_n)$, then $\vec{Z} = (z_1 = x_1 - y_1, ..., z_n = x_n - y_n)$ are the differences. It is assumed that the differences are independent continuous variates from symmetrical populations with a common mean μ. If one wishes to test the hypothesis, that the two samples are identical, the null hypothesis can be set to $\mu = 0$, and the alternative hypothesis are $\mu < 0$ and $\mu > 0$. The Wilcoxon Signed-Rank ranks the observed differences in an increasing order of absolute magnitude, and the sum of the ranks is computed for all the differences of the same sign (+ or -). Differences where $z_i = 0$ are not included. If two or more differences have the same magnitude, they are all given the average rank value. If -2, 2 and -2 are three observed differences, and should they have had ranks 5, 6 and 7, they are all given the average rank $\frac{5+6+7}{3} = 6$. In the above example with

null hypothesis $\mu = 0$, the Wilcoxon rank test can be used to determine whether the hypothesis should be rejected or not. By summing the ranks for each sign, a positive rank sum and a negative rank sum is found. On the null hypothesis the two rank sums are expected to be equal. If the positive rank sum is the smaller, the null hypothesis will be rejected at a predetermined level of significance, in favor of the alternative hypothesis $\mu > 0$. If the negative rank sum is the smaller, $\mu < 0$ would be the alternative if the null hypothesis was rejected.
In statistics the confidence of a conclusion is highly related to the estimated significance. If a result is said to be significant, it means that with a high probability the result is not faulty. Meaning that it with a high probability the result is correct. In order to decide it this is so, a significance level, α is used. This value could be 0.05 meaning that the observation is significant, or 0.01 meaning that the observation is highly significant[Kee95]. These are often used significance levels, but normally the α should be chosen appropriate for the test at hand. For determining if a test result is significant, the p-value is found and compared to the α. The p-value is the smallest level of significance for which the null hypothesis would be rejected.
[Kee95] states that, the minimum rank sum, T, is approximately normally distributed with mean:

$$\mu = \frac{N(N+1)}{4},$$

and variance:

$$\sigma^2 = N(N+1)\frac{2N+1}{24}$$

From this, Z can be calculated, being the minimum rank sum, T fitted to a standard normal distribution:

$$Z = \frac{T-\mu}{\sigma}$$

From the Z value, the cumulative standard normal distribution can be used to find the exact probability. Then, the test is said to reject the null hypothesis if $|Z| \geq \Phi^{-1}(1-\alpha)/2)$. The p-value is the smallest value of α, for which this statement is true.

Chapter 4

Data

In the task of creating and examining prediction models for assessing soccer matches in this project, two data sets are used; a data set containing soccer match results, and a data set containing the odds corresponding to the result data set.

4.1 Result Data Set

The result data set contains results from three European soccer leagues; Spanish Segunda Division, Danish Superliga and English Premier League. The reason for choosing these three leagues specifically is, that they are assumed to have different tendencies. The Danish SAS League has a reputation of having a high number of goals, while the Spanish Segunda Division is regarded as a low scoring league. The English Premier League lies in between as a medium scoring league[Bet]. By having three different leagues with different behaviors, the assessors can be examined if they are more or less fit for leagues with relative few or many goals. For each of these leagues the data set contains the result of each match played between July 1st 2002 and June 30th 2007, adding up to a total of five full seasons. The data set has been provided by a Danish sports betting portal, BetXpert[Bet], and has been parsed from comma separated text files to database tables, from where it can be handled.
Figure 4.1 shows an outtake of the result database.

Date	Home Team	Away Team	Home Goals	Away Goals
29-10-2003	AaB	FC København	0	1
29-10-2003	AGF	AB	3	1
29-10-2003	Brøndby	Frem	3	1

Table 4.1: Example of database entries for the Danish SAS League

4.2 Odds Data Set

In order to examine and evaluate the different approaches, it is necessary to have historical data about the offered odds for the matches which are to be predicted upon. There are several companies and web sites, which focus solely on collecting odds, however in most cases this is done in regards to display the present odds for a future match. The odds data is stored, but since in a commercial perspective the present odds are more interesting than historical data, the historical data is in most cases not easy to come by in a format which is easy to work with. Through one of these companies, Tip-Ex[Te], the historical odds data has been available for use in this project. The company has supplied text files containing data from the seasons 2006/07 for each of the leagues mentioned in the previous section. The data set also contains odds data for the present season 2007/08, holding data up until March 20th 2008. The data set contains both the opening and the closing odds.
The data set used for this project contains odds from the major Asian bookmakers, which are regarded as the leading suppliers on the over/under odds market: Pinnacle-Sports.com, 10Bet.com, sbobet.com, 188Bet.com and Mansion88.com. For any given match in the odds data set, the opening and closing odds from each of these bookmakers has been collected, for any line present. In some cases only one line is offered, but in most cases two or more lines are present. For the remainder of this report only the closing odds of the data set will be used.
The text files provided by Tip-Ex have been parsed into a database format, in order to make it easier to work with, with respect to match extraction.

Date	Home	Away	Bookmaker	Over	Under	Line	Offset	Change Date
02-08-2006	Vejle	OB	sbobet.com	1.68	2.282	2.25	0	02-08-2006
02-08-2006	Vejle	OB	sbobet.com	2	1.9	2.75	13	31-07-2006
02-08-2006	Vejle	OB	Pinnacle	1.971	1.935	2.5	0	02-08-2006
02-08-2006	Vejle	OB	Pinnacle	2.01	1.901	2.75	15	31-07-2006
02-08-2006	Vejle	OB	Mansion88	1.952	1.935	2.5	0	02-08-2006
02-08-2006	Vejle	OB	Mansion88	1.971	1.87	2.75	36	31-07-2006
02-08-2006	Vejle	OB	IBCBET	2	1.9	2.5	0	02-08-2006
02-08-2006	Vejle	OB	IBCBET	2	1.9	2.75	42	31-07-2006

Table 4.2: Example of odds database entries for the SAS League

Table 4.2 shows the entries in the odds data set for a single match. For the match between Vejle and OB, played on August 2nd 2006, four bookmakers offered odds. Notice that each of the four bookmakers have two entries. Each of the entries have an $Offset$ value, which indicates the time at which the data was collected. A value of "0" means that this is the last odds collected, while "13" is an offset used by Tip-Ex to distinguish different collected data. In the case of this project the values will be used to distinguish opening from closing odds. "0" means closing, and all other values mean opening odds. The last attribute in the table, *ChangeDate*, has no importance the use in this project. For each entry in the data set there is a *Line*, and the odds corresponding for over and under this line. For the match in the example each bookmaker has only one line at a given time, and each of the four has lowered the line and changed the odds accordingly.

Chapter 5

Assessor and Strategy Evaluation

The goal of this project is to find a best assessor for determining the probabilities of the number of goals in a soccer match, to see if it is possible to create a probability assessor which can match the bookmakers' assessment with regards to setting odds. The second part is to examine how the proposed assessors and betting strategies perform on actual offered odds. In the following sections the evaluation plan for these two evaluations are presented.

5.1 Assessor Evaluation

In finding the better assessor, each of the four approaches presented in Section 2.2 are evaluated on similar terms, on the data set presented in Section 4.1. Where the bookmakers assessment is based on the odds offered, the three other approaches rely on the historical data. Each of these approaches have parameters, which best values need to learned or determined. The data set is split into two subsets, as can be seen in Figure 5.1. The two "outer" boxes represent the training data set and test data set respectively. Within these sets, the "inner" boxes represent a half season. The training data set consists of nine half seasons from the fall season 2002 to the fall season 2006, while the test data set holds only a single half season, the spring season 2007.
For examining and tuning the parameters, assessments of the matches in the training data set is performed, using the the log score method presented in Section 3.3.1.3 as a measure to determine the best settings. There are two types of parameters, which difference needs to be distinguished. For an assessor, a *global* parameter is a fixed parameter value which is common for all predictions made. This could for example be a fixed fade factor which phases out the influence of older history matches. A *local* parameter is one which is specific to the single match under assessment. An example is the expected number of goals in a match or the home team advantage factor. Table 5.1 shows an overview of the parameters which apply to the individual assessor.
The examination of each assessor and the determination of parameter settings will in turn be presented and accounted for in later chapters of this report.

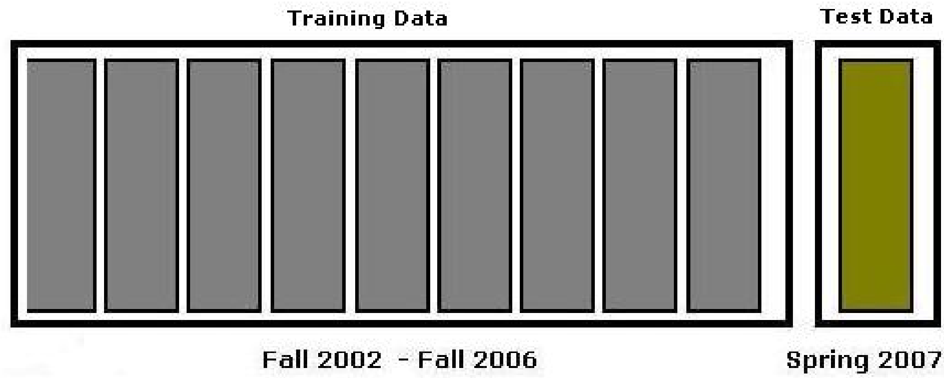


Figure 5.1: Division of the data set for learning and evaluating the probability assessors.

Assessor	Global Parameter	Local Parameter
Gamblers' Assessment	Number of Historic Matches	Count of Over Matches, Count of Under Matches
Poisson Assessment	Number of Historic Matches	Expected Goals
Dixon-Coles Assessment	Fade factor ϵ	Offensive Strengh, Defensive Weakness, Home Team Advantage, Dependency Factor
Bookmakers' Assessment		Match odds

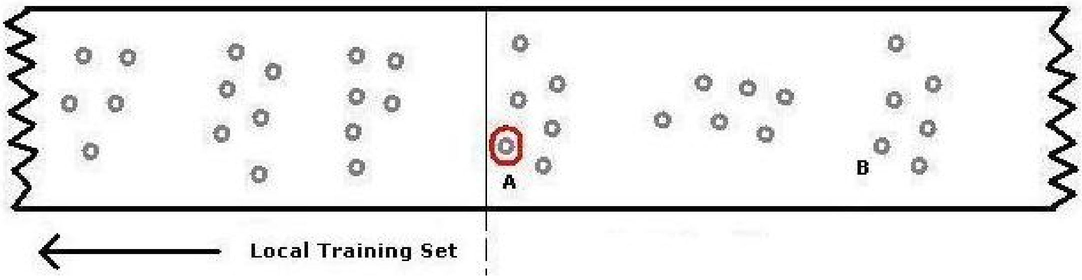
Table 5.1: The average log score of the bookmakers' assessment over the fall season of 2006, for each of the leagues.

The evaluation approach bears resemblance to leave-one-out cross-validation (LOOCV), where the data set is split into two subsets. In LOOCV the test set contains a single observation, and the rest of the set serves as the training set. For a set containing k instances, k validations are made, with each instance acting as the test set.

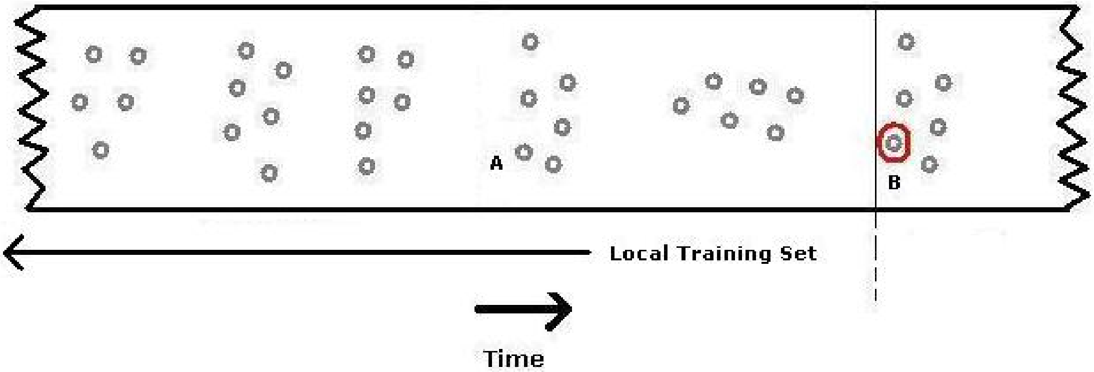
The evaluations performed in this project in similar in the sense that the local training set and local test set is different from assessment to assessment. By local sets, it is meant that for an assessment of a single match, the local training data set is a set of matches used for training the local parameters for the assessment.

The process of parameter tuning can be used for further explanation of the evaluation process. The set of matches used for tuning, the validation set, is the full (global) training set. An assessment of each match will be made, in order to find the log score for the match under the given settings.I actuality, since all matches will be evaluated, the global training is also the global test set for the parameter tuning process.

Regard Figure 5.2 as an outtake from the middle of the the validation set. The dots represent matches, which are ordered chronologically. In (a) the match A is now under assessment, and can be regarded as a local test set (consisting of one match). For this



(a) a



(b) b

Figure 5.2: Two individual matches from the test set under evaluation. In (a) the line before match A indicates the border of the training data matches. In (b) match B is under evaluation, again with the line drawing the training set border. Notice that match A has crossed over, from the test set to the training set.

assessment, the local training set is used for calculating the local parameters. In the assessments made in this project, the local training set will always be a subset of the global training set, where the match date is prior to that of the match under assessment. In (b) the match B is now under assessment and constitutes the local test set. Notice now the local training set, now containing the match A. When a match has been assessed in the evaluation process, it simply crosses over and becomes a part of the local training set for later matches.
For the final tests in this report, the spring 2007 season is the global test set, and the training set in Figure 5.1 is the global training set. In Figure 5.3 an assessment scenario from the final test is shown. Here the local test set is match C, and the local training set is the global training set and the matches from the test set which have crossed over after assessment.

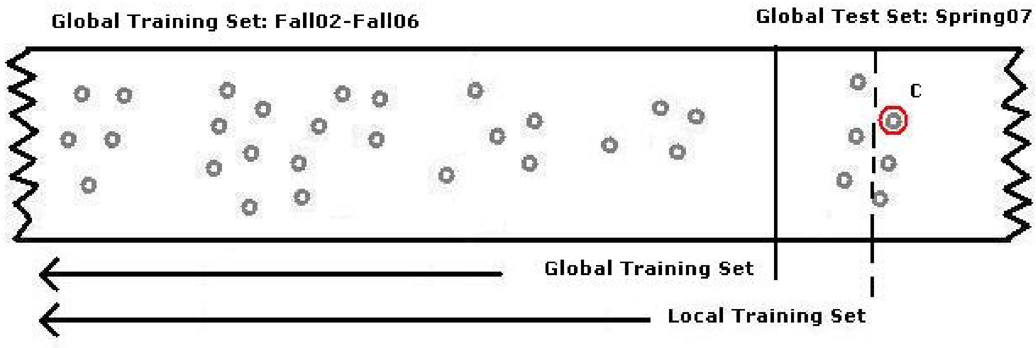


Figure 5.3: Test scenario with a global test set containing the spring 2007 season, and a global training set containing fall 2002 to fall 2006. Notice how the already assessed matches in the test set becomes a part of the local training set for later assessments.

5.1.1 Overfitting

When dealing with probability assessor evaluation using a data set partition into training and test data, there are certain things one must be aware of. Training data is used for adjusting the parameters of an assessor to create the best probability assessments. A probability assessor, however, must not only fit the training data well, but also the test data. [TSK].
See Figure 5.4. Two graphs are plotted to show the score of a probability assessor, which is evaluated on the number of historic matches used for prediction. Notice with a low number of matches, the assessor performs poorly over both the training and test data. A concept known as underfitting. As more matches are used, the performance over the training data improves, as does the performance over the test data. However, if the number of matches is increased even more, the performance over the test data becomes worse. Mitchell[Mit97] speaks about overfitting in respect to classifying instances of observations, and defines overfitting as follows:

Definition 5. *(Overfitting) Given a hypethesis space H, a hypothesis h ∈ H is said to overfit the training data if there exists some alternavitve hypothesis h' ∈ H, such that h has smaller error than h' over the training examples, but h' has a smaller error than h over the entire distribution of instances.*

In the setting of soccer match prediction, the concept of overfitting is relevant. In this project, the probability assessors base their predictions on historical data, observations of soccer match results. However, the relevance of a soccer match in 1980, can not be said to bear as much influence on a prediction on a match in present time, as a match played last week. An investigation of the parameters for some of the assessors need to be made. For the Dixon-Coles and the bookmakers' assessments, the aspect of over/underfitting is not deemed relevant and will not be examined. However it is for the

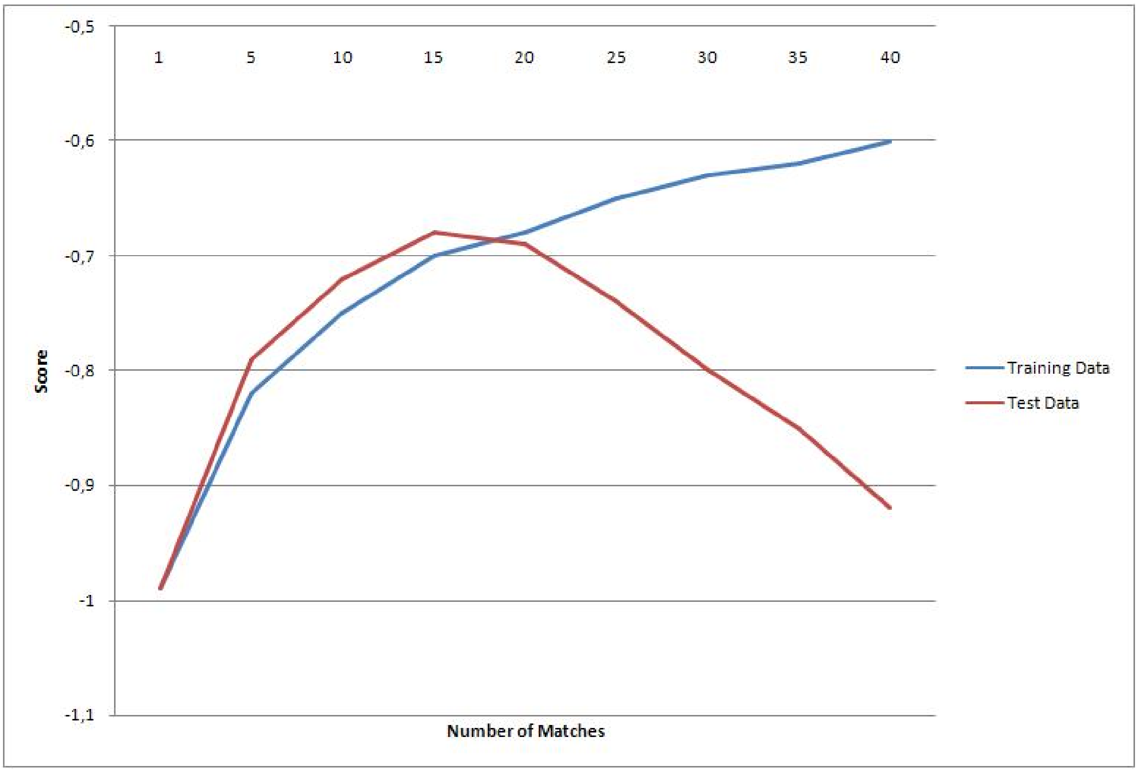


Figure 5.4: The error rate of a probability assessor over the training and test data respectively. The error rate is plotted against the number of parameters.

two naive approaches, the Poisson and gamblers' assessment the number of prior matches the prediction is to be based upon need to be examined, to find a suiting number. Using too low a number of matches to base the prediction upon will make the predictions too adapt to the training data, overfitting the model. Instead, using too high a number of matches will adjust the prediction specifically to the training data, and will not allow a single new observation a significant impact. It will be too overly general. In example a Poisson prediction over 150 soccer matches with an average of 2.51 goals, will not differ much from the prediction using 151 matches, where a new match with three goals have been added to the training data. The model is said to be underfit. For both approaches, a best number of matches used will be examined. The results of this examination can be found in Sections 8.1.2 and 7.3.

5.1.2 Test Result Evaluation

The results found by evaluating the assessors on the test data set, needs to be itself evaluated, in order to say if the results are in fact plausible. For this purpose, hypothesis testing, using Wilcoxon Signed-Rank Test presented in Section 3.4 will be used. The null hypothesis is defined as follows:

> Probability assessor A performs at least as good as probability assessor B, measured using the log score of the assessments.

In order to say, that a probability assessor is better than the bookmakers assessment, the null hypothesis needs to be rejected, within a specified significance level. The significance level α chosen for the tests is 0.05, based on [Kee95], which states that a 0.05 significance level enables the conclusion of a difference being significant.

5.2 Betting Strategy Evaluation

To determine which betting strategy is the better, the different approaches are evaluated individually. For this purpose a data set containing actual bookmaker odds for several thousand matches has been acquired. See Section 4.2 for details about the data. In this project two betting strategies are to be examined; value betting strategy and the threshold strategy. The value betting strategy, takes the approach of deciding to bet or not based on a calculation of the expected value of a bet. For a given match, provided with the probability assessment for the outcomes from a probability assessor and the odds data for the outcomes, the strategy will place a 1 unit bet on the bet with the highest positive expected value. The expected value strategy will be used together with three probability assessors; Gamblers' assessment, Poisson assessment and Dixon-Coles assessment.

The second strategy, the threshold strategy does not utilize a probability assessment for choosing its bets. Instead, as assessor, it uses the expected number of goals for a given match (see Section 6 for further details). The expected number of goals is calculated based on prior results of the participating teams, and compared with lines of the offered odds in the odds data set. A set of rules are then used to decide if the strategy should place a 1 unit bet or not bet at all on a given match. By letting each of the four betting strategies evaluate each match contained in the validation set, and place bets accordingly, a total set of bets can be found for each of the strategies.

By using the observed results of every match, the bets can be settled and the win or loss of each strategy can be calculated and compared. The betting strategies will then be compared on their ability to generate a positive return and hold a stable return on investment. The evaluation of the betting strategies and the results of this can be found in Chapter 11.

Chapter 6

Bookmakers Prediction

When trying to establish the quality of a probability assessor, the log score is a good method for measurement. However, it is not enough to say, that the assessor with the highest average log score is a good assessor for creating soccer match odds. If an assessor is to be used for setting odds, it must at the very least almost match the quality of the bookmakers assessments. In this chapter, the bookmakers odds data are examined and used to establish the bookmakers prediction.

6.1 Calculating Probabilities

It is fairly simple to calculate a bookmakers assessment of a sporting event, if there is a fixed number of outcomes and the odds are known, at least for the European odds variety. See Section 2.1 for details. However, the odds data set used for this project, contains odds for several lines for over/under, and for a given match it is not certain that "clean" lines as over/under 2.5 and 3.5 are present, and since the assessor comparison is made on probability assessments for over/under 2.5 goals, these probabilities are needed for every match. For a given match in the odds data set, there are in almost all cases two lines present. If over/under 2.50 is not present, it should be possible to derive the odds for over/under 2.50 goals using the odds for either lines 2.00 and 2.25 or 2.75 and 3.00. In order to do so, one must understand how asian line goals bets work.

6.1.1 Asian Lines

As shown in Table 2.1, several lines can be used for the over/under bet type. The Asian market bookmakers shift the lines for several reasons: To have even odds, and to have several markets on the same match. Lines such as over/under 2.75 is a combination of two other betting lines: 2.50 and 3.00. A bet on over 2.75 goals, is in reality a split bet, where the half of the wager is placed on over 2.50 and the other half on over 3.00, as described in Table 2.1 in Section 2.1.2. As [Han08] states in the interview in Section A.2.2, the different lines can be made using the probability distribution for the number of goals. Where it is quite simple to calculate the odds for over/under 2.5 goals, a combination of two bets are needed to calculate odds for over/under 2,75. We will use under 2.75 as an example, which, as mentioned, is a combination of under 2.50 and under 3.00. (6.1) shows the calculation of the two lines respectively.

$$\begin{aligned} Odds_{Under2.50} &= tpb \cdot \tfrac{1}{P(\leq 2)} \\ Odds_{Under3.00} &= tpb \cdot \tfrac{1-P(3)}{P(\leq 2)} \end{aligned} \tag{6.1}$$

In the case of exactly three goals, the under 3.00 bet will be void and the bet is refunded. Therefore the probability of exactly three goals is subtracted from the numerator. The odds for under 2.75 can now be made, as the average of these two:

$$Odds_{Under2.75} = \frac{1}{2}(tpb \cdot \frac{1}{P(\leq 2)} + tpb \cdot \frac{1-P(3)}{P(\leq 2)}) = tpb \cdot \frac{2-P(3)}{2 \cdot P(\leq 2)} = tpb \cdot \frac{1-\frac{1}{2}P(3)}{P(\leq 2)}$$

In the same way the odds equations can be found for all other combination lines. The following table denotes the equations for calculating the odds for different lines based on the probability distribution over the number of goals. Let *tpb* be the theoretical payback, and $P(x)$ the assessed probability for x goals scored in the match, then the odds for the lines can be calculated as such:

Under 2,00	$\frac{tpb \cdot (1-P(2))}{P(0)+P(1)}$
Over 2,00	$\frac{tpb \cdot (1-P(2))}{P(\geq 3)}$
Under 2,25	$\frac{tpb \cdot (1-\frac{1}{2}P(2))}{P(0)+P(1)+\frac{1}{2}P(2))}$
Over 2,25	$\frac{tpb \cdot (1-\frac{1}{2}P(2))}{P(\geq 3)}$
Under 2,75	$\frac{tpb \cdot (1-\frac{1}{2}P(3))}{P(\leq 2)}$
Over 2,75	$\frac{tpb \cdot (1-\frac{1}{2}P(3))}{P(\geq 4)+\frac{1}{2}P(3)}$
Under 3,00	$\frac{tpb \cdot (1-P(3))}{P(\leq 2)}$
Over 3,00	$\frac{tpb \cdot (1-P(3))}{P(\geq 4))}$

Table 6.1: Using a probability assessment and a theoretical payback one can set odds for asian line over/under bets.

For a given match, if the line 2.50 is not present, the odds for either of the line pairs 2.75/3.00 or 2.00/2.25 can be used to calculate the probabilities for over/under 2.50. If the lines 2.75 and 3.00 are offered, the probabilities can be calculated for $P(3)$ using the formula for under 2.75 and under 3.00. Notice, that the theoretical payback for the two offered lines are not necessarily the same, and needs to be calculated individually. See Section 2.1 for how to do so. In the following $P(\leq 2)$ means $P(0) + P(1) + P(2)$:

$$Odds_{under2.75} = \frac{tpb_{Under2.75} \cdot (1 - \frac{1}{2}P(3))}{P(\leq 2)} \Leftrightarrow P(\leq 2) = \frac{tpb_{Under2.75} \cdot (1 - \frac{1}{2}P(3))}{Odds_{under2.75}} \tag{6.2}$$

$$Odds_{under3.00} = \frac{tpb_{Under3.00} \cdot (1 - P(3))}{P(\leq 2)} \Leftrightarrow P(\leq 2) = \frac{tpb_{Under3.00} \cdot (1 - P(3))}{Odds_{under3.00}} \tag{6.3}$$

By isolating $P(\leq 2)$ in each equation, Equation 6.2 and 6.3 can be set equal to each other. By doing so, $P(3)$ can be found:

$$\frac{tpb_{Under2.75} \cdot (1 - \frac{1}{2}P(3))}{Odds_{under2.75}} = \frac{tpb_{Under3.00} \cdot (1 - P(3))}{Odds_{under3.00}}$$

$$\Updownarrow$$

$$P(3) = \frac{Odds_{Under3.00} \cdot tpb_{Under2.75} - Odds_{Under2.75} \cdot tpb_{Under3.00}}{\frac{1}{2} \cdot Odds_{Under3.00} \cdot tpb_{Under2.75} - Odds_{Under2.75} \cdot tpb_{Under3.00}} \tag{6.4}$$

By inserting 6.4 into 6.2 or 6.3 the probability $P(\leq 2)$ can be found:

$$P(\leq 2) = \frac{\frac{\frac{1}{2} \cdot Odds_{Under3.00} \cdot tpb_{Under2.75}}{Odds_{Under2.75} \cdot tpb_{Under3.00} - \frac{1}{2} \cdot Odds_{Under3.00} \cdot tpb_{Under2.75}}}{Odds_{Under3.00}} \tag{6.5}$$

Here $P(\leq 2)$ is the probability of under 2.5 goals. The probability for over 2.5 goals can be found by $P(\geq 3) = 1 - P(\leq 2)$. In the same way, using the odds for over 2.00 and over 2.00, $P(\geq 3)$ can be found:

$$P(\geq 3) = \frac{\frac{\frac{1}{2} \cdot Odds_{Over2.00} \cdot tpb_{Over2.25}}{Odds_{Over2.25} \cdot tpb_{Over2.00} - \frac{1}{2} \cdot Odds_{Over2.00} \cdot tpb_{Over2.25}}}{Odds_{Over2.00}}$$

If the over/under odds for 2.50 are not present, and there is no pair of 2.00/2.25 or 2.75/3.00 odds present, then it is not possible to calculate the bookmakers assessment, and the match can therefore not be included in the evaluation.

6.1.2 Verifying Formulas

Since the above approach to calculating the probabilities for over/under 2.50 is not provided directly from a bookmaker, and since no of the contacted bookmakers will reveal how they create their odds based on assessments, it is necessary to verify that the formulas used in this project can in fact be used.
Taking an example match from the odds data set, where the lines 3.00 and 2.75 are present, as well as 2.50 it should be possible to make the calculation for over/under 2.50 goals based on the odds, and comparing it to the actual offered odds from the bookmaker.
The odds for over/under for the match between Brøndby and FC Nordsjælland on July 21st 2007 are shown in Table 6.2

Line	Over	Under	tpb
O/U 3.00	2.11	1.80	0.9714
O/U 2.75	1.84	2.06	0.9719
O/U 2.50	1.69	2.23	0.9614

Table 6.2: The bookmaker IBCBets odds for over/under bets on Brøndby-FC Nordsjælland on July 21st 2007, along with the corresponding theoretical payback, for use in the calculations

By inserting the odds and theoretical paybacks into Formula 6.5, a value for the probability for under 2.50 goals is found to be $P(\leq 2) = 0.43143$. Using Formula 2.2 the bookmakers odds for under 2.5 can be found using the found probability and the theoretical payback: $Odds_{Under2.50} = \frac{tpb_{2.50}}{P(\leq 2)} = \frac{0.9614}{0.43143} = 2.228$ According to the odds data set, the odds for under 2.5 goals is 2.23, which corresponds to the found value.

6.2 Bookmaker Scoring

The evaluation of the bookmakers' assessment is made using the log scoring method, on the evaluation data set of matches in the spring season of 2007, as described in Section 5. The algorithm for evaluating is shown in Algorithm 1.
The algorithm shows the pseudo code for calculating the average log score for the bookmakers' assessor over a set of matches, using the corresponding odds data. Here the *CalculateOver* and *CalculateUnder* functions utilize the formulas presented in the previous section for determining the probabilities for over and under 2.5 goals.

6.3 Test Results

The result data set reaches farther back than the odds data, which makes it impossible to examine the bookmaker assessments as extensively as the three other assessors, which are evaluated over match data from fall 2002 to fall 2006. Since the odds data set only

Algorithm 1 Evaluation algorithm for bookmaker assessment.

```
Function BookmakerLogScore(OddsData, MatchData)
TotalLog = 0
NoOfMatches = 0
for all Match ∈ MatchData do
   goals = Match.NoOfGoals
   P(> 2.5) = Match.CalculateOver(OddsData)
   P(< 2.5) = Match.CalculateUnder(OddsData)
   if (goals > 2.5) then
      MatchLogScore = ln(P(> 2.5))
   else if goals < 2.5 then
      MatchLogScore = ln(P(< 2.5))
   end if
   TotalLog = TotalLog + MatchLogScore
   NoOfMatches++
end for
AvrLogScore = TotalLog / NoOfMatches
Return AvrLogScore
```

hold data ranging from fall 2006 to spring 2008, and due to the fact that the spring 2007 data is to be used for testing, it is only possible to examine the bookmakers' assessment on the fall season of 2006.

For each of the three leagues, a single bookmaker has been chosen from the odds data set. The chosen bookmaker has been picked based on the odds provided for the given league. For the SAS League the bookmaker chosen is 188bet, which is the only bookmaker with at least two lines for all matches in the SAS League season 2006-07. For the Segunda Division and Premier League, the bookmaker 10Bet has been chosen. When any of the leagues assessors are compared in the remainder of this report, the chosen bookmaker for the respective league will be referred to as the bookmaker. Since the results presented

League	Avr. Log Score
SAS League	-0.68408
Premier League	-0.68526
Segunda Division	-0.69281

Table 6.3: The average log score of the bookmakers' assessment over the fall season of 2006, for each of the leagues.

in Table 6.3 are only for the fall season of 2006 it can not be used in direct comparison with the results for the entire training data (2002 to 2006). The average log score for a half season is not necessarily representative for the several seasons, since results can vary. The results found in this section can however be used as an indication of the bookmakers' level. In order to enable comparison, average log scores will be found for the fall season of 2006 for the other assessors.

Chapter 7

Gamblers' Approach

The gamblers' approach is the empirical probability of the number of goals in a match, being higher or lower than a specified line. With the result of prior matches known, a count of the instances of over 2.5 and under 2.5 can be made, in order to establish the probability of the match ending with a low or high score. In this chapter, the gamblers' approach will be examined and evaluated. First the approach is explained in detail, and then followed by the determination of the best parameter setting. Lastly, preliminary results for the assessor is made based on the training data set.

7.1 Gamblers' Assessment

The gamblers' assessment is presented in Algorithm 2. It takes a match under assessment, m_{assess}, a match data set $Matches$, an integer k, being the number of historical matches to base the prediction on. Lastly a *Line* for over/under assessment is also required. In this chapter, the line value is fixed at 2.5, and the match data set are matches from all three leagues from fall 2002 to fall 2006. In the algorithm, in lines 2 through 5, the k latest home and away matches for the home and away teams are selected, and the instances of matches with less than *Line* goals are counted. In lines 6 to 8 the probabilities P($>$ *Line*) and P($<$ *Line*) are calculated as the frequencies of over and under matches. The result is returned in line 10.

7.2 Evaluating Gamblers Assessment

The evaluation of the gamblers' assessments is made using the log score method. The training set of matches is to be evaluated, in order to determine the optimal number of history matches. For a prediction made for a single match, all matches prior to the date of the match is regarded as the local training data. Matches which were played after the date of the match, do not count towards the prediction. For each match assessed, the predicted probabilities are used along with the observed result to assign the assessment

Algorithm 2 Gamblers' assessment algorithm

1: **function** GamblerAssess(Match m_{assess}, $Matches$, k, $Line$)
2: **Select** the latest k home matches for the home team from $Matches$
3: Count the number of matches, $x_{homeunder}$ where the number of goals is less than $Line$

4: **Select** the latest k away matches for the away team from $Matches$
5: Count the number of matches, $x_{awayunder}$ where the number of goals is less than $Line$

6: The probability of m_{assess} having under or over $Line$ goals is
7: $\mathrm{P}(< Line) = \frac{x_{homeunder}+x_{awayunder}}{2k}$
8: $\mathrm{P}(> Line)$= 1 - $\mathrm{P}(< Line)$
9: **return** $\mathrm{P}(> Line)$, $\mathrm{P}(< Line)$

a score, just as it was the case in Algorithm 1 with the bookmakers' assessment scoring. The calculation of the score for the gamblers' assessment is made in the same way, only with the minor change that the $CalculateOver()$ and $CalculateUnder()$ functions are replaced by the $GamblersAssess()$ function from Algorithm 2.

7.3 Optimal Number of Games

When having a vast amount of result data, it is possible to find empirical probabilities for over/under using hundreds or thousands of games. As discussed in Section 5, the issue of overfitting (or underfitting) needs to be addressed. Using result data from, say, thirty year old soccer matches might not give a clear view of how a match in 2008 will progress. Using too large a number of matches might underfit the model, whilst using too few might overfit it. The optimal number of matches to base the probability upon needs to be examined. As described in Section 5, the matches from the fall of 2002 to fall 2006 are used for training the parameters. For the Danish SAS League, a team has 72 to 77 home and away matches in the training data period. For Premier League and Segunda Division this number is 85 and 95 respectively. The matches in the training data set are evaluated using the log score method, using value in specified intervals for the k variable. For all three leagues, an evaluation is made by calculating the average log score of the gamblers' assessment approach, based on a varying number of history matches, ranging from 2 to 100. When evaluating with k equal to 30, for example, the match only contributes to the average log score if, there exists 30 history matches for each of the participating teams. Therefore, with high values for k, there are fewer contributing matches than with lower values. Figure 7.1 depict graphs of the average log score plotted against the number of history matches used, for each of the three leagues. For the SAS League, Figure 7.1 , from 1 to 10 matches, the performance is not good, showing a low average log score. This is due to the low number of matches used, where there is a good chance of a lack of tendency in the results. From 69 to 75, the performance

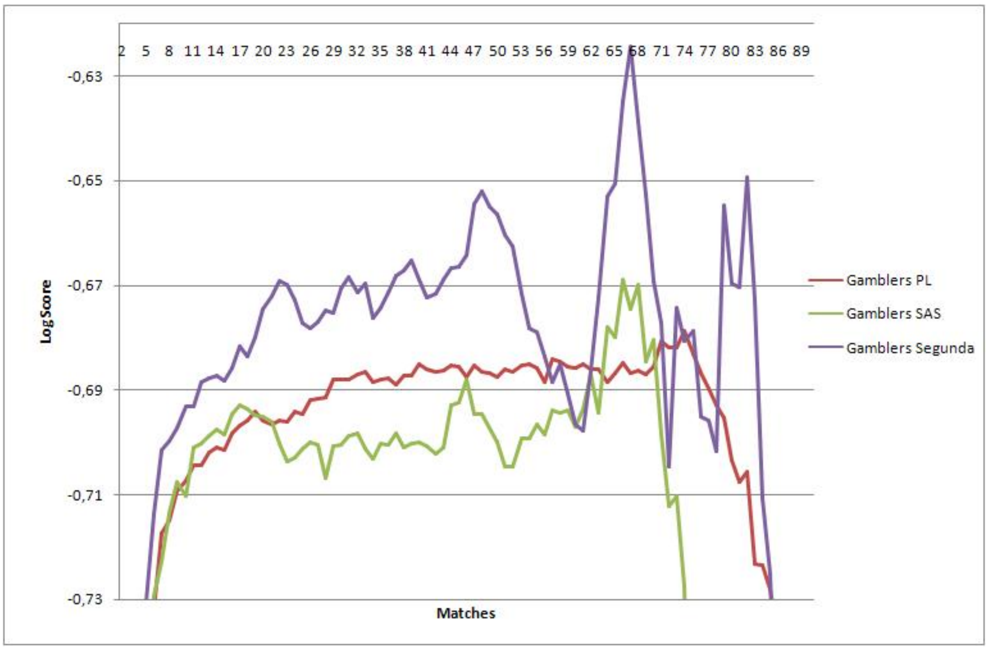


Figure 7.1: The log score of the gamblers approach, on all three league training data sets, plotted against the number of matches used to create the assessment.

again is poor, showing low average log scores. This is due to a low number of contributing matches, due to a high number of history matches used. For m=75, there are in fact only two matches in the training set, which contribute to the log score. Notice that the curve is relatively smooth in the interval from 1 to 60. For larger values the curve becomes more irregular. This is due to the lack of contributing matches, why the values over 60 do not come in to consideration as possible optimal values.
For the Premier League the curve is more smooth over the entire interval. It does not become irregular with high values for history matches. However, it does have the similar behavior with a low log score for values higher than 72.
For the Segunda Division the curve becomes very irregular for values higher than 55. There is a very differing log score in the interval from 55 to 90, why these values can not come into consideration as optimal values.

7.3.1 Test Results

Based on the curves above, a guess towards the optimal values for the number of history matches to base the gamblers' assessment on can be made. Table 7.1 presents the results. It is impossible to say, if a larger training set would have provided even better log scores, and therefore also better optimal values. The best log scores seem to be at values, very close to the point where the curves began showing irregular behavior. This leads to believe, that using an even larger number of history matches would provide an even

League	Best No Of Matches	Avr. Log Score Fall 2002-Fall 2006	Avr. Log Score Fall 2006 only
SAS League	46	-0.68811	-0.69185
Premier League	74	-0.67867	-0.68706
Segunda Division	48	-0.65192	-0.70603

Table 7.1: The optimal value for the number of history matches to base the Gamblers' assessment upon. For each league the optimal number and corresponding log score is presented.

better result. The result data set does, however, not hold more matches than the used, and it has therefore not been possible to evaluate for larger numbers. For the comparison tests performed in later chapters of this report, the values presented in the above table has been used.

Chapter 8

Poisson Assessment

As mentioned in Section 2.2.2, this approach was inspired by an initial betting strategy introduced by [Ras08]. The idea of using an expected number of goals for a match, calculated from the average goals in historic matches, inspired the investigation of the distribution of the number of goals in a soccer match. Dixon-Coles introduced the use of a Poisson distribution as a part of predicting the probability of a given result of a match, based on the participating teams offensive and defensive skills. Instead of using the individual skill measures, the expected number of goals is viewed as a representation of the combination of the two teams skills. It is therefore examined if the number of goals can be used in a Poisson distribution to predict the number of goals. Before doing so, the expected number of goals is defined:

Definition 6. *(Expected number of goals) In a match between home team* i *and away team* j*, the expected number of goals, Avr, based on i's last n home matches and j's last n away matches is:*

$$Avr(i,j,n) = \frac{\sum_{k=1}^{n} Goals_{i,k} + \sum_{k=1}^{n} Goals_{j,k}}{2n}$$

, where $Goals_{i,k}$ and $Goals_{j,k}$ is the total number of goals in the k'th home and away match respectively.

8.1 Goal Histograms

To establish the plausibility of using a Poisson distribution of the expected number of goals, the result data set is examined. If this approach is plausible, then for a set of soccer matches, the distribution of the matches over the number of goals must follow the Poisson distribution with a mean value equal to the average number of goals scored per match in the set of matches.
Figure 8.1 shows two distributions. The x-axis is the number of goals, while the y-axis is the number of matches. The blue columns show the distribution of matches over the

number of goals for the SAS League 2006/07 season. The red columns show the Poisson distribution with mean value 2.80, which is the average number of goals scored per match in the 2006/07 season.

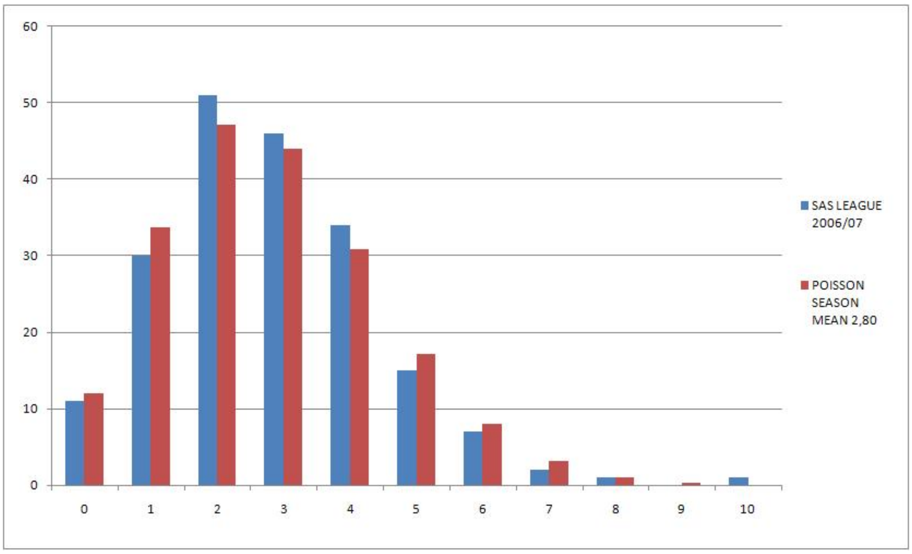


Figure 8.1: Plot of histogram and poisson distribution for the SAS League season 2006/07

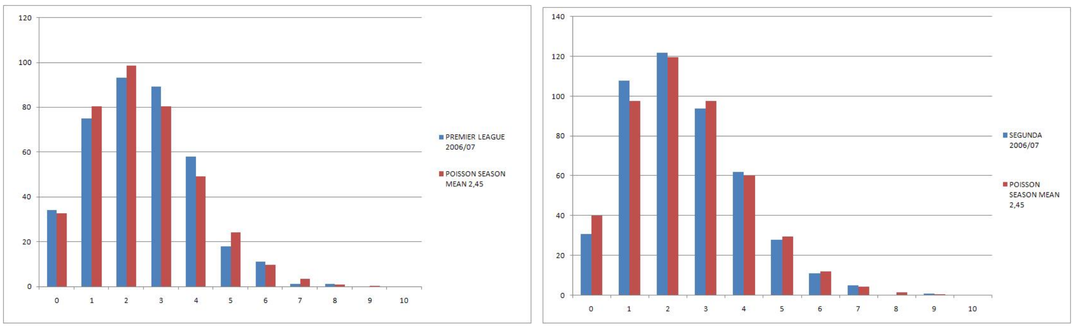


Figure 8.2: Plot of histogram and poisson distribution for the Premier League and Segunda Division season 2006/07

There is a strong resemblance of the two distributions for the SAS League season. This is also the case for the English Premier League and the Spanish Segunda Division, see Figure 8.2. The use of the expected number of goals as the mean of a Poisson distribution therefore seems to be a good approximation to the distribution of goals in soccer matches, and therefore can be viewed as a candidate for assessing the number of goals in a given soccer match.

8.1.1 Poisson Assessment

The actual Poisson assessment of a single match is made using Algorithm 3. It takes a match under assessment, a set of matches and a number k indicating the number of history matches to be used for calculating the expected number of goals.

Algorithm 3 Poisson assessment algorithm

1: **function** GamblerAssess($Matchm_{assess}$, $Matches$, k, $Line$)
2: **Select** the latest k home matches for the home team from $Matches$
3: Calculate the sum of goals in these matches, x_{home}.
4: **Select** the latest k away matches for the away team from $Matches$
5: Calculate the sum of goals in these matches, x_{away}.
6: Calculate the expected number of goals for m_{assess}:
7: $\lambda = \frac{x_{home}+x_{away}}{2k}$
8: P(< $Line$) = 0
9: **for** int i = 0; i < $Line$; i++ **do**
10:     P(< $Line$) = P(< $Line$) + $\frac{\lambda^i \cdot e^{\lambda}}{i!}$
11: **end for**
12: P(> $Line$) = 1 - P(< $Line$)
13: **return** P(> $Line$), P(< $Line$)

In lines 2-5 the total number of goals scored in the latest k home matches for the home team and k away matches for the away team are found. In line 6-7 these are used for calculating the expected number of goals in m_{assess}. Lines 8-12 calculates the probabilities of P(> $Line$) and P(< $Line$) using a Poisson distribution with a mean value equal to the expected number of goals. The result is returned in line 13.

8.1.2 Optimal Number of Matches

For each of the three leagues, a deterministic search for the optimal number of games used to calculate the expected number of goals has been made. As was the case with the search for the optimal number for the gamblers' assessment, an interval from 1 to 100 has been used. The figures plot the average log score against the number of matches used for the assessment.

In Figure 8.3 the plots for the three leagues are shown. For the SAS League the plot is very similar to the plot for the gamblers' approach, the curve shows smooth behavior from 1 up to about 50. For values higher than 50, the curve becomes more irregular, indicating that the number of matches which can be evaluated with 50 or higher history matches becomes insufficient. Therefore any observations for 50 history matches or more are not considered.

The Premier League data shows a more steady curve, and is slightly more smooth over the entire interval. For values over 60 however, there are slight irregularities in the curve, disqualifying values higher than 60 from being candidates.

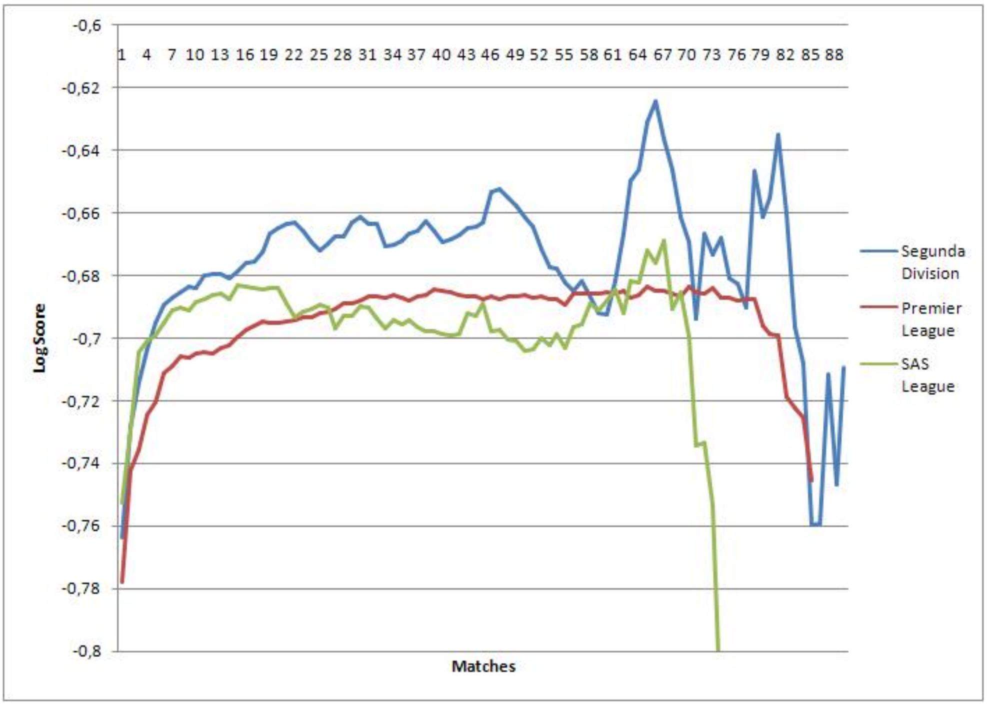


Figure 8.3: The log score of the Poisson assessment, for all three leagues, plotted against the number of matches used to create the assessment.

The Segunda Division data shows a more irregular behavior over the entire interval. Notable irregularities are present from 60 and upwards. In the interval from 1 to 60, the curve is smoother, however it is not steady, as was the case with the SAS and Premier League.

8.2 Test Results

There does not seem to be a telling consistence between the results of the separate leagues. The found optimal number of matches do not reveal a common resemblance. The curves are very similar to those of the gamblers' approach, however there is only a vague indication towards the best number being the highest possible. The SAS League has its best performance at 16 history matches, while Premier League and Segunda Division reaches there maximum at 40 and 48 respectively. Both of these indicate, that a high number of matches used yields the best results.
An interesting observation made, is to see that the plots of average log score for the gamblers' and Poisson approach show similar behavior. Plotting both the curve of the gamblers' approach and the Poisson assessment shows this.
Notice the pairwise curves for each of the leagues, and how they have similar development. There seems to be a correlation between the curves for the individual leagues. This

League	Best No Of Matches	Avr. Log Score Fall 2002-Fall 2006	Avr. Log Score Fall 2006 only
SAS League	16	-0.68333	-0.67347
Premier League	40	-0.68467	-0.69065
Segunda Division	48	-0.65234	-0.72505

Table 8.1: The optimal value for the number of history matches to base the Poissont assessment upon. For each league the optimal number and corresponding log score is presented.

is somewhat expected, since a higher frequency of over 2.5 matches in the gamblers' approach will influence the expected number of goals to be higher. With a high number of history matches with more than 2.5 goals, the chance of the expected number of goals being relatively high is larger. This could give reason to believe that the gamblers' and Poisson approach will give rather similar predictions.

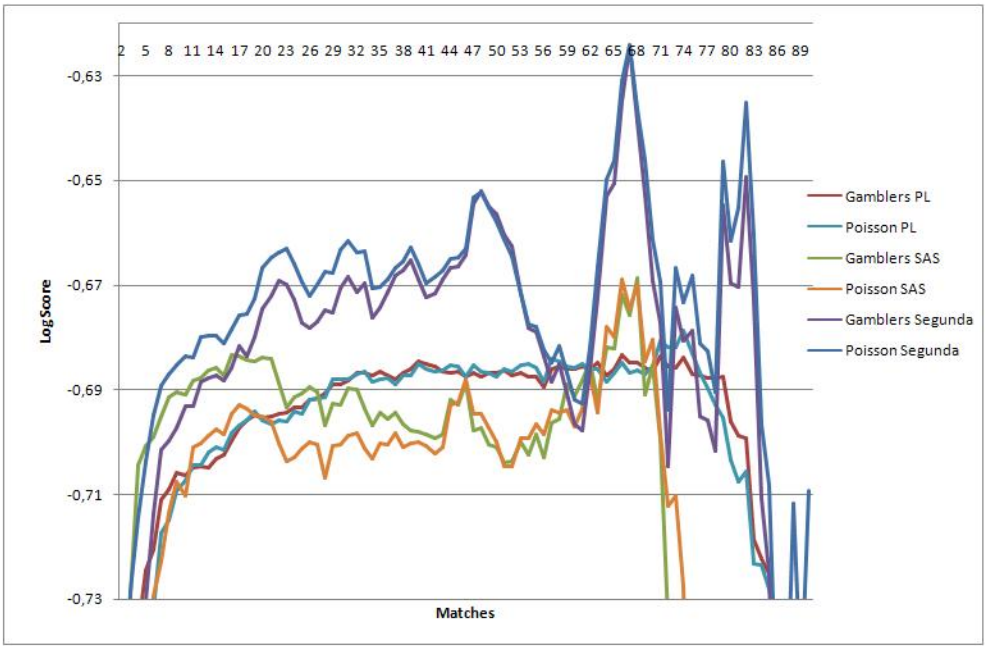


Figure 8.4: The log score of both the Poisson assessment and the gamblers' assessment, for all three leagues, plotted against the number of matches used to create the assessment.

Chapter 9

Dixon-Coles Approach

The Dixon-Coles approach [DC97] is a predictive model, which uses only the goals scored in previous matches to predict the probability of scores of a soccer match. Historic results are considered as a measure for the teams offensive and defensive qualities, since a team that scores a lot of goals are assumed to be offensively potent and a team which concedes a lot of goals are considered defensively weak. This chapter presents the fundamentals of the Dixon-Coles approach, and fits it to the problem domain of this project. The model originally is used for predicting match outcomes, while this project seeks to predict the total number of goals. The model settings are presented and the parameters determined, which are to be used for evaluating the Dixon-Coles performance. Conclusively the results for the assessor on the training set is presented.

9.1 Dixon-Coles Assessment

We consider two teams i and j, home and away team respectively. α_i and α_j are the offensive strengths of the two teams, and β_i and β_j are the defensive weaknesses. We let $X_{i,j}$ and $Y_{i,j}$ be the number of goals scored by the two teams in the match. [DC97] states, that the number of goals scored by the two teams respectively can be modeled by a Poisson distribution over the product of the offensive strength and defensive weakness:

$$X_{i,j} \sim Poisson(\alpha_i \beta_j \gamma)$$
$$Y_{i,j} \sim Poisson(\alpha_j \beta_i)$$

In soccer, the home team often has an advantage of playing games at their home field. The support from the crowd and familiar surroundings give an advantage, which is clear by viewing any soccer league results. In the 2006/07 season in the SAS League, 43% of the matches ended in a home win, while 24% and 33% ended in draw and away win respectively [Bet]. The model implements this advantage, by introducing the home team

advantage factor, γ, which is multiplied to $\alpha_i\beta_j$ when calculating the mean value for the Poisson distribution for the home team goals.
In the above $X_{i,j}$ and $Y_{i,j}$ are independent and $\alpha, \beta > 0$. The independency gives us, that the probability of a match result is given by the product of the probability of the home team goals and the away team goals:

$$P(X_{i,j}, Y_{i,j}) = P(X_{i,j})P(Y_{i,j})$$

As stated in [DC97], the number of goals can be modeled by a Poisson distribution, wherefore the probability of a given result of a match between two teams can be calculated as follows:

$$P(X_{i,j} = x, Y_{i,j} = y) = \frac{\lambda^x exp(-\lambda)}{x!}\frac{\mu^y exp(-\mu)}{y!},$$

where

$$\lambda = \alpha_i\beta_j\gamma$$
$$\mu = \alpha_j\beta_i$$

According to [DC97] the number of goals scored by the two teams are not completely independent. A dependency between home and away team goals was identified for low scoring games. The following modification was therefore imposed:

$$\tau_{\lambda,\mu}(x,y) = \begin{cases} 1 - \lambda\mu\rho & if x = y = 0 \\ 1 + \lambda\rho & if x = 0, y = 1 \\ 1 + \mu\rho & if x = 1, y = 0 \\ 1 - \rho & if x = y = 1 \\ 1 & otherwise \end{cases}$$

Here ρ is a dependency factor. If $\rho = 0$, the scores are independent, and τ is 1. For scores where x and y ≤ 1, the dependency factor changes the value of τ and raises or lowers the probability for the result. The final equation for calculating probabilities of the number of goals, proposed by [DC97] is therefore:

$$P(X_{i,j} = x, Y_{i,j} = y) = \tau_{\lambda,\mu}(x,y)\frac{\lambda^x exp(-\lambda)}{x!}\frac{\mu^y exp(-\mu)}{y!}, \tag{9.2}$$

The Dixon-Coles approach can be used for assessing probabilities of outcome of a soccer match, based solely on statistical data on scores of previous matches. By estimating the probability of all results, the probabilities of home win, draw and away win can be found. For example, the probability of a home win is:

$$P(X_{i,j} = x, Y_{i,j} = y | x > y) = \sum_{x>y} \tau_{\lambda,\mu}(x,y)\frac{\lambda^x exp(-\lambda)}{x!}\frac{\mu^y exp(-\mu)}{y!},$$

In this project the goal is to assess the probability of the total number of goals in a match, and not the actual outcome. The winner of the match is not important to bets on over/under 2.5 goals. So instead of summing the probabilities for outcomes where the number of home goals is larger than the number of away goals to find the probability of a home win, the probabilities where the sum of the number of home and away goals is less or greater than 2.5 are summed to find the probabilities for under and over respectively. For over 2.5 goals, the equation is:

$$P(X_{i,j} = x, Y_{i,j} = y | x + y > 2.5) = \sum_{x+y>2.5} \tau_{\lambda,\mu}(x, y) \frac{\lambda^x exp(-\lambda)}{x!} \frac{\mu^y exp(-\mu)}{y!},$$

In order to create the assessments for a given match, the parameters must be determined. The global parameter, the fade factor ϵ, and the local parameters, the home advantage γ, the dependency factor ρ and the offensive strength and defensive weakness α and β.

9.2 Parameter Calculation

With a total of n teams, there are attack parameters $\alpha_1, ..., \alpha_1$ and defense parameters $\beta_1, ..., \beta_1$, the home parameter γ and the dependence parameter ρ which need to be estimated. In order for the model not to be over-parameterized a constraint upon the attack parameters is imposed:

$$\frac{1}{n} \sum_{i=1}^{n} \alpha_i = 1$$

For the Danish SAS League there are 12 teams, giving a total of 26 parameters which need to be estimated. For Premier League there are 42 parameters, and for Segunda Division there are 46. For a single assessment, all parameters are estimated. To estimate the parameters, the likelihood function is used. For a set of matches, $k = 1, ..., N$, with scores for each match (x_k, y_k):

$$\mathrm{L}(\alpha_1,...,\alpha_n,\beta_1,...,\beta_n,\gamma,\rho) = \prod_{k=1}^{N} \tau_{\lambda_k,\mu_k}\ (x_k, y_k) e^{-\lambda_k} \lambda_k^{x_k} e^{-\mu_k} \mu_k^{y_k} \tag{9.3}$$

where

$$\lambda_k = \alpha_{i(k)} \beta_{j(k)} \gamma,$$
$$\mu_k = \alpha_{j(k)} \beta_{i(k)},$$

Here $i(k)$ and $j(k)$ respectively denotes the indices of the home and away team in match k, while x_k and y_k denote the number of goals scored each team in the match. By making a maximum likelihood estimate of (9.3), the local parameters can be found.
In soccer, teams change over time. Players move around, and teams can hit winning or losing streaks. A lot of factors affect a teams quality, and in general recent form is one of the most important factors when assessing a soccer match. The above approach does not take these changes into account, and weight all matches used in the estimation as equal. A modification is made to (9.3), introducing a fade factor, downgrading the importance of older matches:

$$\mathrm{L}(\alpha_1,\dots,\alpha_n,\beta_1,\dots,\beta_n,\gamma,\rho)=\prod_{k=1}^{N} \tau_{\lambda_k,\mu_k}\ ((x_k,y_k)e^{-\lambda_k}\lambda_k^{x_k}e^{-\mu_k}\mu_k^{y_k})^{\phi(t-t_k)}$$

Here t is the time at which the assessment is made, and t_k is the time at which match k was played. The fade function should yield a smaller value, the farther apart t and t_k are. In this way, older matches are given a smaller significance, while more recent matches are given a higher significance. The function ϕ can be chosen in many ways, and several choices can be used for this. Dixon-Coles have examined some of the possibilities and suggest the use of:

$$\phi(t-t_k) = e^{-\epsilon(t-t_k)}$$

Using this ϕ will downgrade the history matches exponentially. With $\epsilon = 0$ all matches will be weighted equally, while increasing the ϵ value will weight recent matches higher. The nature of the fade function makes it impossible optimize it using the maximum likelihood measure. Instead it will be estimated deterministically, with regards to the assessments made by the model on over/under outcomes. The estimation is presented later in this chapter.

9.2.1 Optimizing Local Parameters

In order to find the optimal local parameters, the likelihood function must be maximized. The method for doing so, is to use gradient descent to find the maximum likelihood, as was done by [CH].
The likelihood function presented in equation 9.5 must be derived to find the partial derivatives which are needed to perform the gradient descent. As the function is in 9.5, it is possible to simplify this derivation, by using the log-likelihood instead of the likelihood.

$$
\begin{aligned}
&\mathrm{LL}(\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_n,\gamma,\rho)= \\
&\ln\prod_{k=1}^{N}(\tau_{\lambda_k,\mu_k}\,(x_k,y_k)e^{-\lambda_k}\lambda_k^{x_k}e^{-\mu_k}\mu_k^{y_k})^{\phi(t-t_k)}= \\
&\sum_{k=1}^{N}\phi(t-t_k)\ln(\tau_{\lambda_k,\mu_k}\,(x_k,y_k)e^{-\lambda_k}\lambda_k^{x_k}e^{-\mu_k}\mu_k^{y_k})= \\
&\sum_{k=1}^{N}\phi(t-t_k)(\ln(\tau_{\lambda_k,\mu_k}\,(x_k,y_k))-\lambda_k+x_k\ln(\lambda_k)-\mu_k+y_k\ln(\mu_k))
\end{aligned}
\tag{9.5}
$$

Remember that $\lambda_k = \alpha_{i(k)}\beta_{j(k)}\gamma$ and $\mu_k = \alpha_{j(k)}\beta_{i(k)}$, where $i(k)$ and $j(k)$ are id's for the home and away teams of match k. In order to find the gradient descent, it is necessary to find the partial derivatives of all variables. The value vector, $\vec{values}$, holds values which need to be found to maximize the likelihood:

$$
\begin{pmatrix}
\alpha_1 \\ \vdots \\ \alpha_n \\ \beta_1 \\ \vdots \\ \beta_n \\ \gamma \\ \rho
\end{pmatrix}
\tag{9.6}
$$

It is necessary to find the partial derivative of the offensive strength α_i and the defensive weakness β_i for any team i of the total n teams, along with the home advantage factor γ and the dependency factor ρ. The partial derivative for the offensive strength of team i is:

$$
\frac{\partial LL}{\partial \alpha_i} = \sum_{k=1}^{N}\phi(t-t_k)
\begin{cases}
0 & \text{if i} \neq \text{i(k) and i} \neq \text{j(k)} \\
\frac{-\beta_{j(k)\mu_k\gamma\rho}}{1-\lambda_k\mu_k\rho} - \beta_{j(k)\gamma} + \frac{x_k}{\alpha_{i(k)}} & \text{if i = i(k) and } x_k\text{=0 and } y_k = 0 \\
\frac{\beta_{j(k)\gamma\rho}}{1+\lambda_k\rho} - \beta_{j(k)\gamma} + \frac{x_k}{\alpha_{i(k)}} & \text{if i = i(k) and } x_k\text{=0 and } y_k = 1 \\
-\beta_{j(k)\gamma} + \frac{x_k}{\alpha_{i(k)}} & \text{if i = i(k) and } x_k \neq 0 \text{ and } y_k \neq 0\text{,1} \\
\frac{-\beta_{i(k)\lambda_k\rho}}{1-\lambda_k\mu_k\rho} - \beta_{i(k)\gamma} + \frac{y_k}{\alpha_{j(k)}} & \text{if i = j(k) and } x_k\text{=0 and } y_k = 0 \\
\frac{\beta_{i(k)\rho}}{1+\mu_k\rho} - \beta_{i(k)} + \frac{y_k}{\alpha_{j(k)}} & \text{if i = j(k) and } x_k\text{=1 and } y_k = 0 \\
-\beta_{i(k)\gamma} + \frac{y_k}{\alpha_{j(k)}} & \text{if i = j(k) and } x_k \neq 0\text{,1 and } y_k \neq 0
\end{cases}
\tag{9.7}
$$

Notice how the τ function imposes constraints on the number of goals scored in the k'th match. Therefore low scoring games do not contribute to the derivative in the same way as high scoring matches. The derivative takes the dependence into account. Similarly to the offensive strength, the partial derivative of the defensive weakness β_i is:

$$\frac{\partial LL}{\partial \beta_i} = \sum_{k=1}^{N} \phi(t-t_k) \begin{cases} 0 & \text{if i} \neq \text{i(k) and i} \neq \text{j(k)} \\ \frac{-\lambda_k \alpha_{j(k)}\rho}{1-\lambda_k\mu_k\rho} - \alpha_{j(k)} + \frac{y_k}{\beta_{i(k)}} & \text{if i = i(k) and } x_k\text{=0 and } y_k = 0 \\ \frac{\alpha_{j(k)}\rho}{1+\mu_k\rho} - \alpha_{j(k)} + \frac{y_k}{\beta_{i(k)}} & \text{if i = i(k) and } x_k\text{=1 and } y_k = 0 \\ -\alpha_{j(k)} + \frac{y_k}{\beta_{i(k)}} & \text{if i = i(k) and } x_k \neq \text{0,1 and } y_k \neq 0 \\ \frac{-\alpha_{i(k)}\mu_k\gamma\rho}{1-\lambda_k\mu_k\rho} - \alpha_{i(k)}\gamma + \frac{x_k}{\beta_{j(k)}} & \text{if i = j(k) and } x_k\text{=0 and } y_k = 0 \\ \frac{\alpha_{i(k)}\gamma\rho}{1+\lambda_k\rho} - \alpha_{i(k)}\gamma + \frac{x_k}{\beta_{j(k)}} & \text{if i = j(k) and } x_k\text{=0 and } y_k = 1 \\ -\alpha_{i(k)}\gamma + \frac{x_k}{\beta_{j(k)}} & \text{if i = j(k) and } x_k \neq 0 \text{ and } y_k \neq \text{0,1} \end{cases} \tag{9.8}$$

For the home advantage factor γ, the partial derivative is:

$$\frac{\partial LL}{\partial \gamma} = \sum_{k=1}^{N} \phi(t-t_k) \begin{cases} \frac{-\alpha_{i(k)}\beta_{j(k)}\mu_k\rho}{1-\lambda_k\mu_k\rho} - \alpha_k\beta_k + \frac{x_k}{\gamma} & \text{if } x_k\text{=0 and } y_k = 0 \\ \frac{\alpha_{i(k)}\beta_{j(k)}\rho}{1+\lambda_k\rho} - \alpha_k\beta_k + \frac{x_k}{\gamma} & \text{if } x_k\text{=0 and } y_k = 1 \\ -\alpha_k\beta_k + \frac{x_k}{\gamma} & \text{if } x_k \neq 0 \text{ and } y_k \neq \text{0,1} \end{cases} \tag{9.9}$$

The dependency factor used to infer dependence in low scoring games is:

$$\frac{\partial LL}{\partial \rho} = \sum_{k=1}^{N} \phi(t-t_k) \begin{cases} \frac{-\lambda_k\mu_k}{1-\lambda_k\mu_k\rho} & \text{if } x_k\text{=0 and } y_k = 0 \\ \frac{\lambda_k}{1+\lambda_k\rho} & \text{if } x_k\text{=0 and } y_k = 1 \\ \frac{\mu_k}{1+\mu_k\rho} & \text{if } x_k\text{=1 and } y_k = 0 \\ \frac{-1}{1-\rho} & \text{if } x_k\text{=1 and } y_k = 1 \\ 0 & \text{if } x_k \neq \text{0,1 and } y_k \neq \text{0,1} \end{cases} \tag{9.10}$$

By setting the value vector presented in equation 9.6 to a starting point, the calculation of the derivatives based on the initial values will give a vector which points toward the maximum likelihood. In order to carry out these calculations, a vector class has been implemented in .NET C#, so it was possible to make the necessary vector operations.
Algorithm 4 shows how the parameter optimization has been implemented according to the Dixon-Coles approach.
In lines 2-3 the necessary vectors are initialized and set to values recommended by [DC97]. Lines 4 to 26 is a while loop, which breaks when no further improvements are made. In this loop, the gradient vector is calculated using the functions described previously. From the starting point (the initial settings for the $\vec{values}$ vector, steps are taken along the gradient vector until a point is reached where no improvements are made to the value returned by the likelihood function. When the best point is found, the process is repeated, calculating a new gradient vector and finding a new best value. When the best values have been found, the vector $\vec{values}$ is returned.

Algorithm 4 Parameter optimization

function Optimize(Matches, noOfTeams, ϵ)
Create vectors $\vec{values}$, $\vec{gradient}$, $\vec{normal}$, $\vec{lastvalues}$ of size [noOfTeams $\cdot$ 2 + 2]
Set all entries in $\vec{values}$ to 1, except the last entry which is set to 1.4.
while $\mid \vec{lastvalues} - \vec{values} \mid < 0.001$ **do**
  $\vec{lastvalues} = \vec{values}$
  for int team_id = 1 to noOfTeams **do**
    $\vec{gradient}$[team_id] $= \frac{\partial LL}{\alpha_{team\_id}}$
    $\vec{gradient}$[team_id + noOfTeams] $= \frac{\partial LL}{\beta_{team\_id}}$
  end for
  $\vec{gradient}$[noOfTeams $\cdot$ 2 + 1] $= \frac{\partial LL}{\gamma}$
  $\vec{gradient}$[noOfTeams $\cdot$ 2 + 2] $= \frac{\partial LL}{\rho}$
  for int i = 1 to noOfTeams **do**
    $\vec{normal}$[i] = Average α value in $\vec{values}$
    $\vec{normal}$[i + noOfTeams] = 0
  end for
  $\vec{normal}$[noOfTeams $\cdot$ 2 + 1] = 0
  $\vec{normal}$[noOfTeams $\cdot$ 2 + 2] = 0
  $\vec{gradient} = \vec{gradient}$ - $\vec{normal}$ {α values are averaged to 1}
  $\vec{PresentPoint} = \vec{values}$
  $\vec{StepToPoint} = \vec{values} + \vec{gradient}$
  while LL($\vec{StepToPoint}$) > LL($\vec{PresentPoint}$) **do**
    $\vec{PresentPoint} = \vec{StepToPoint}$
    $\vec{StepToPoint} = \vec{StepToPoint} + \vec{gradient}$
  end while
  Using the LL Value, find the vector between $\vec{PresentPoint}$ and $\vec{StepToPoint}$ which maximizes LL and assign this to $\vec{values}$
end while
return $\vec{values}$

9.2.2 Fade-out Factor

Due to the nature of the fade factor, ϵ, it is not possible to optimize its value as it is done in the parameter optimization in the previous section. The fade factor is a global parameter, unlike the above which are local and specific to the local training set and match under assessment. The fade factor is the factor used to scale the relevance of a history match with regards to a match under assessment. In order to establish a best fade factor, an examination of the Dixon-Coles approach is made over the training data set using a interval of ϵ values. In [DC97] the time unit used for the t and t_k values are halfweeks, why this is also the choice for this project. The ϵ value is examined in the interval 0.001 to 0.02, as did [DC97].

Algorithm 5 Finding the best ϵ value

function FindBestϵ(Matches, minimum, maximum, increment)
Bestϵ-score = -∞
Bestϵ = -∞
for ϵ = minimum; ϵ < maximum, $\epsilon = \epsilon$ + increment **do**
    ϵscore = 0
    for all *match* in *matches* **do**
        localtraining = $match_{train}$ in *matches* — $match_{train}.date < match.date$
        assessment = DixonColesAssessment(match, ϵ)
        ϵ-score = ϵ-score + log($assessment.outcome_{observed}$)
    end for
    if ϵ-score > Bestϵ-score **then**
        Bestϵ-score = ϵ-score
        Bestϵ = ϵ
    end if
end for
return Bestϵ, Bestϵ-score

Algorithm 5 presents the approach for finding the best ϵ value for a set of matches. Provided with a *minimum*, *maximum* and stepsize *increment* (and a set of matches), the best ϵ is returned. In order to find the best ϵ for use on the test data in the tests later in this report, an examination is made on all three leagues using the above algorithm. The algorithm simply creates an assessment for each match in the match data set, and calculates the log score, based on the assessment and the observed result of the match. The best ϵ is the one with the highest total log score over the entire training set.

9.3 Test Results

For the Dixon-Coles approach, a single global parameter exists, namely the fade factor ϵ. The algorithm presented in the previous section was used for the purpose of determining the best ϵ for the SAS League. The estimation of the ϵ values has been made deterministically, in accordance with the settings proposed by [DC97]. The time unit used is half weeks, and the search has been narrowed to the interval 0.00 to 0.02 for the ϵ values. Using values higher than 0.02 has shown worse results than using values within the interval. By examining the interval, and finding average log scores for every 0.001, an initial 20 runs of the algorithm has been made for each league. Based on the initial runs, a curve for the average log score as a function of ϵ has been created for the SAS League.

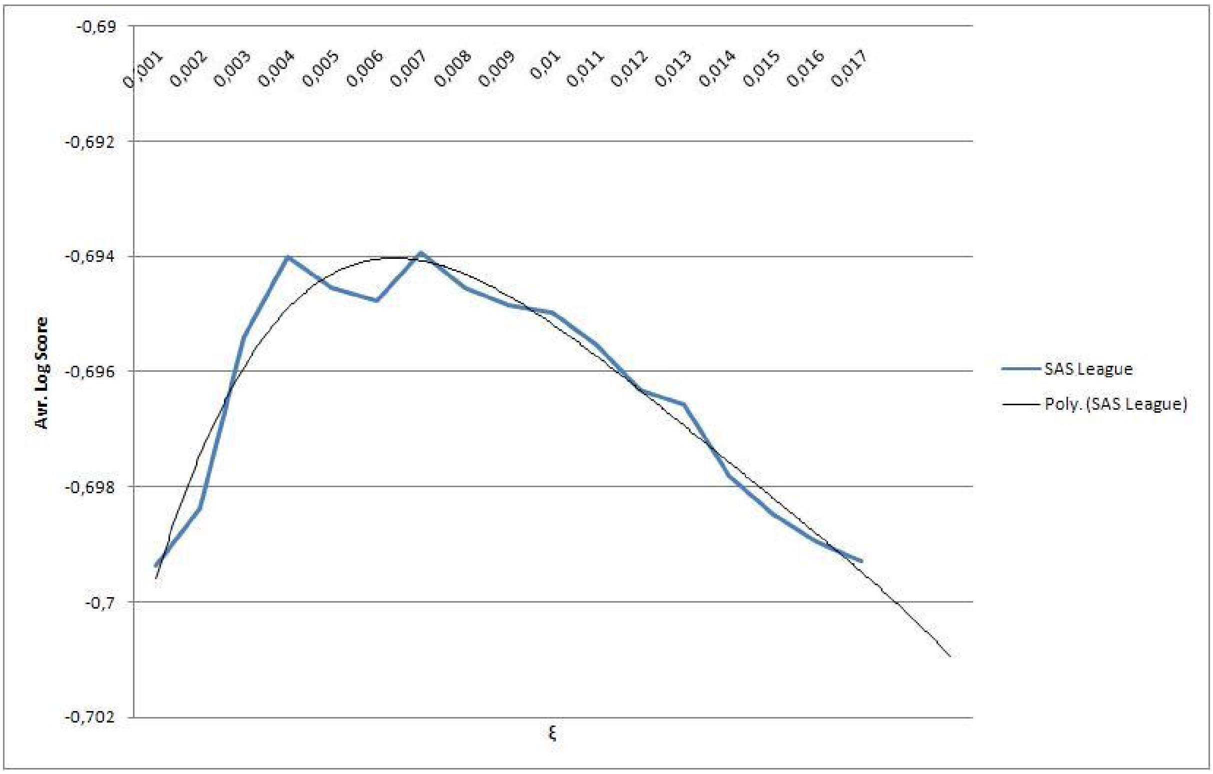


Figure 9.1: Plot of the average log score as a function of ϵ

Figure 9.1 shows the plot of the average log score as a function of the ϵ values. The figure shows a graph over the connected points, as well as a fitted function. The irregularity in the graph indicates that it seems random that any value in the 0.004-0.007 interval is better than any of the already measured. A further increase in granularity has therefore not been made. Instead the fitted function is used. It indicates, that 0.007 is a suitable ϵ value, why this is chosen. The estimation of best ϵ values for each separate league has not been made. This due the fact, that the estimation is very time costly, with a single run taking 10-15 hours, dependant on the number of matches in the data set. This also due to, that the difference between using an ϵ value of 0.004 instead of 0.005 does not seem to yield that large a change in the average log score. Therefore, the ϵ value of 0.007 is used for all leagues in the Dixon-Coles assessments for the remainder of this report. Table 9.1 shows the Dixon-Coles estimation of the team offensive strength and defensive weaknesses. These are presented along with the final position in the league for season 2005/06 and the goals scored and conceded by each team. At the top of the league Recreativo is the team with the best offensive strength with a value of 1.15. They are also the team with the lowest defensive weakness at a value of 0.85. This in accordance with the actual results, since they are the most scoring team and the team with the least conceded goals. Malaga B in 21st place is the team with the most conceded goals, and also with the highest defensive weakness. There is a clear correlation between the number of goals a team scores and concedes and the strength values.

Position	Team	α	β	Scored	Conceded
1	Recreativo	1.15	0.85	67	32
2	Gimnastic	1.00	0.89	48	38
3	Levante	1.04	0.90	53	39
4	Ciudad Murcia	1.04	0.93	53	42
5	Lorca	1.06	0.91	56	39
6	Almeria	1.05	0.94	54	43
7	Xerez	1.09	0.97	60	46
8	Numancia	1.02	1.04	50	55
9	Gijon	0.94	0.87	41	34
10	Valladolid	1.05	1.03	54	54
11	Real Madrid B	1.05	0.99	55	50
12	Castellon	0.99	0.99	46	50
13	Albacete	0.97	1.05	44	57
14	Elche	0.99	1.02	47	54
15	Poli Ejido	0.96	0.99	43	50
16	Murcia	0.95	0.92	41	40
17	Hercules	0.92	0.99	39	49
18	Tenerife	1.04	1.07	53	60
19	Lleida	0.96	1.03	43	53
20	Ferrol	0.97	1.10	44	63
21	Malaga B	0.96	1.15	42	68
22	Eibar	0.83	0.95	28	45

Table 9.1: Using the Dixon-Coles approach the offensive strengths and defensive weaknesses have been made for Segunda Division, at the end of the 2005/06 season.

Over the entire season there was scored over 2.5 goals in 40.58% of the matches, and under 2.5 goals in 59.52%, and in average 2.30 goals was scored per match [Bet]. Looking at a match between Valladolid and Tenerife, both being teams which have strength values in the top half of the table, would be expected as a match up where there is a high probability of over 2.5 goals (in comparison with any other match in the league). A Dixon-Coles assessment predicts a 52% chance of under 2.5 goals and a 48% chance of over 2.5 goals. Despite these probabilities not indicating a match with many goals, the probability of over 2.5 goals is higher than the league frequency. For a league with relatively few goals, the probability assessment seems plausible.

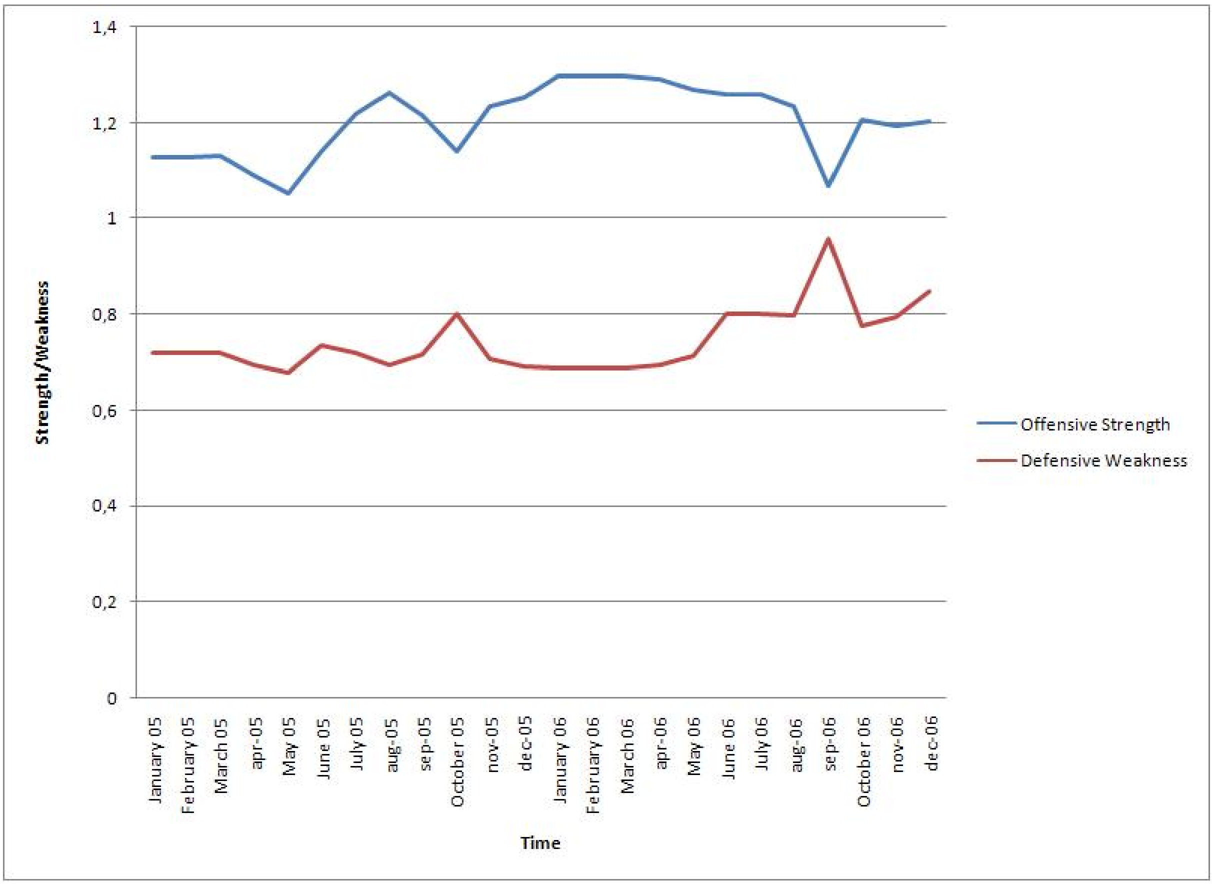


Figure 9.2: Brøndbys offensive strength and defensive weakness in the period January 2005 to December 2006.

Table 9.2 shows the offensive strength and defensive weakness for the Danish team Brøndby over time. The period of time is from January 2005 to December 2006, thus covering the half of the 04/05 season, the whole 05/06 season and the half of the 06/07 season. It is clear to see, that the Dixon-Coles adjusts the offense and defense parameters over time. It is worth noticing, that the two parameters do not follow the same curve, meaning that a team can have an improvement in offense, without it influencing the defensive qualities. Looking at the two curves, two points are the most interesting. At July-August of 2005, which is just after the season, the offensive strength is high. This

is natural due to this being the season where Brøndby won the championship. Another interesting thing to notice is the high stable level of offensive strength and low defensive weakness from late 2005 to mid 2006. In this season Brøndbys performance was high, scoring 60 goals and conceding 34 that season, coming in second place.
Table 9.2 shows the average log scores for the Dixon-Coles approach for each of the three leagues, for the fall 2006 season. The suggested ϵ value of 0.007 has been used.

League	Avr. Log Score Fall 2006
SAS League	-0.68429
Premier	-0.70045
Segunda Division	-0.736748

Table 9.2: The average log score of the Dixon-Coles assessor using the estimated ϵ value.

The average log scores for the Dixon-Coles approach are not as good as the bookmakers' log scores presented in Section 6.3. On all three leagues, the Dixon-Coles performs worse than the bookmaker, however the average is not that far behind. The preliminary results show indications of assessments not far from the bookmakers.

Chapter 10

Betting Strategy Evaluation

As introduced in Section 2.3, two different betting strategies are under evaluation in this project. One is the value betting strategy, which based on a prediction by an assessor and offered odds will decide to bet or not. The other betting strategy is the threshold betting strategy, which as the name implies uses a threshold to decide if the distance between the offered odds line and the expected number of goals in a match is sufficiently large for a bet to be placed. This chapter maps out the details of the two betting strategies, and examines and presents the settings to be used in the final tests.

10.1 Value Betting Strategy

The value betting strategy uses the expected value, explained in Section 3.1, to determine if an offered odds is eligible for a bet. Taking a match and an assessment of the number of goals in the match, the function presented in Algorithm 6 evaluates the assessment against the odds offered for the match. If the highest valued bet surpasses the threshold of a predetermined minimum value, a bet can be placed. Regularly, a bet with an expected value higher than 1 is eligible for a bet, however it is interesting to regulate the least accepted value to see if better results can be obtained.
For the value betting strategy, three assessors will be evaluated, where the $Assessment_{Match}$ given as an argument to the function, can be any of the following assessors;

- Gamblers' assessment
- Poisson assessment
- Dixon-Coles assessment

In the algorithm, all odds offered for the assessed match are evaluated with regards to the expected value. The $CalcValue()$ function takes a bet variety in the form of the $Odds.Outcome$ attribute. This could for instance be "Under 2.75". The expected value for these split bets, or over/under 3.00 can not be calculated in the same way as the value for over/under 2.5 for instance. Due to the nature of the different types of over/under,

Algorithm 6 Value betting strategy

function BestValueBet(Match, $Assessment_{Match}$, OddsData, MinimumValue)
BestBet = "No Bet"
BestValue = 0
for all (Odds ∈ OddsData where Odds.Match = Match) **do**
  Value = CalcValue(Odds.Outcome, Odds, $Assessment_{Match}$)
  if ((Value > MinimumValue) & (Value > BestValue)) **then**
    BestBet = "Odds.Outcome, Odds"
    BestValue = Value
  end if
end for
Return BestBet

several formulas for calculating the value is needed. In a similar way as the construction of the formulas in Section 6.1.1, is made and can be seen in Table 10.1.
For all three leagues in the match and odds data sets, a value betting strategy run will be made for each of the three assessors on the final test data. A strategy run on the test data takes all matches in the spring season of 2007, and simulates the placement of bets, if the $BestValueBet()$ function returns a bet for the match. Before performing strategy runs on the test data, a parameter tuning is made on the minimum value parameter setting. Each assessor will be run on each league using three different settings for the minimum value parameter: 1.00, 1.10 and 1.20. The parameter tuning is performed on the fall season of 2006. This is done in order to determine if raising the demands to the expected value will yield better betting results. Both for parameter tuning and final testing, the global parameter settings found in Chapters 7, 8 and 9 are used for the respective assessors. In the parameter tuning, each league will be submitted to a total of fifteen strategy runs. This being with three different minimum value settings for each of the three assessors, and additional three different minimum value settings for two of the assessor ran on a limited data set. For the final test strategy runs, the best minimum value setting is used.
For a strategy run, a number of bets are found and placed in simulation. The result data set is then used to pay out the simulated bets using the observed results. By paying out all bets found in a betting strategy run, an overall result can be found. By comparing these results, it should be possible to draw conclusions about which of the three assessors are the best bettor.

10.1.1 Minimum Value Parameter

The parameter under examination is the minimum value, which is the expected value of a bet, as described in Section 6. If the assessments are correct, and the sample is of sufficient size, an increase of the minimum value should also increase the return on investment. As the sample size N $\rightarrow \infty$ the average expected value and the return on

Under 1.75	$Odds_{Under1.75} \cdot P(\leq 1) + \frac{1}{2} \cdot P(2)$
Over 1.75	$Odds_{Over1.75} \cdot P(\geq 2) + \frac{1}{2} \cdot Odds_{Over1.75} \cdot P(2)$
Under 2.00	$Odds_{Over2.00} \cdot P(\leq 1) + P(2)$
Over 2.00	$Odds_{Over2.00} \cdot P(\geq 3) + P(2)$
Under 2.25	$Odds_{Under2.25} \cdot P(\leq 2) + \frac{1}{2} \cdot Odds_{Under2.25} \cdot P(2)$
Over 2.25	$Odds_{Over2.25} \cdot P(\geq 2) + \frac{1}{2} \cdot P(2)$
Under 2.50	$Odds_{Under2.50} \cdot P(\leq 2)$
Over 2.50	$Odds_{Over2.50} \cdot P(\geq 3)$
Under 2.75	$Odds_{Under2.75} \cdot P(\leq 2) + \frac{1}{2} \cdot P(3)$
Over 2.75	$Odds_{Over2.75} \cdot P(\geq 3) + \frac{1}{2} \cdot Odds_{Over2.75} \cdot P(3)$
Under 3.00	$Odds_{Over3.00} \cdot P(\leq 2) + P(2)$
Over 3.00	$Odds_{Over3.00} \cdot P(\geq 4) + P(3)$
Under 3.25	$Odds_{Under3.25} \cdot P(\leq 3) + \frac{1}{2} \cdot Odds_{Under3.25} \cdot P(3)$
Over 3.25	$Odds_{Over3.25} \cdot P(\geq 4) + \frac{1}{2} \cdot P(3)$

Table 10.1: Formulas for calculating the expected value of any over/under bet encountered in the odds data used for this project.

investment should be very close, if the assessments are correct.
The purpose of tuning of the minimum value parameter is to find a setting, which is best minimum value when the strategy is to place bets on future events. For both the tuning and the tests, an odds data set containing odds from all present bookmakers are used. The test data set consists of only half a season, which is also the case for the data set used for tuning. Since this tuning data set is limited in size, there is a chance of overfitting the model with regard to the minimum value. The results in this section can therefore not be said to be the best parameter settings over all matches, and conclusions made need to be taken with care.
Table 10.2 shows the result for the SAS League. In all five different strategies have been examined. The gamblers' assessor can only predict if a match will be over or under 2.5 goals, which causes the model to only be usable on odds data sets with odds for over/under 2.5 goals. All odds on other lines are not taken into account by the model. They are, however, by the Poisson and Dixon-Coles. In order to be able to compare the three assessors results in the value betting strategy, the Poisson and Dixon-Coles have been run using both a data set containing all odds and a data set only containing the over/under 2.5 odds. In the tables, the number of bets and the net result is presented on the upper line, while the return of investment is presented in the second line. For each assessor, the best net result and the best return of investment is bold faced, for visibility

	1.00	1.10	1.20
Gamblers (2.5 only)	63 bets, -14,55 **(76,9%)**	**51 bets, -13,057** (74,4%)	34 bets, -13,88 (59,17%)
Poisson (2.5 only)	50 bets, +2,59 (105,2%)	**17 bets, +4,19** (124,6%)	4 bets, +1,19 **(129,8%)**
Dixon-Coles (2.5 only)	**70 bets, -1,98 (97,2%)**	**70 bets, -1,98 (97,2%)**	58 bets, -4,303 (92,6%)
Poisson	84 bets, -6,748 (91,97%)	55 bets, +0,653 (101,19%)	**36 bets, +1,98 (105,5%)**
Dixon-Coles	**91 bets, +5,0235 (105,52%)**	**91 bets, +5,0235 (105,52%)**	79 bets, +2,70 (103,42%)

Table 10.2: The betting results for the SAS League, fall season 2006, containing a total of 90 matches.

purposes. For the SAS League, the Dixon-Coles approach using the full odds data set has the best performance, while it actually has a negative return on limited odds data set. Looking at the return on investment, the Poisson assessor is the best, presenting a very high percentage on both the 1.10 and 1.20 setting. In all (4 out of 5), the 1.10 setting shows the best results with regards to the net result, while the 1.00 setting shows the best return on investment (3 out of 5).

	1.00	1.10	1.20
Gamblers (2.5 only)	**158 bets, +18** (111,4%)	57 bets, +8,57 (115,0%)	10 bets, +5,621 **(156,2%)**
Poisson (2.5 only)	154 bets, -17,63 (88,6%)	61 bets, -4,708 **(92,3%)**	**12 bets, -3,12** (74,0%)
Dixon-Coles (2.5 only)	149 bets, -3,77 (97,47%)	**54 bets, +13,92 (125,78%)**	9 bets, -0,001 (100,00%)
Poisson	197 bets, -1,321 (99,3%)	85 bets, -13,166 (84,5%)	**22 bets, +0,224 (101,0%)**
Dixon-Coles	197 bets, +0,19 (100,1%)	**88 bets, +16,91 (119,22%)**	13 bets, +1,69 (113,0%)

Table 10.3: The betting results for the Premier League, fall season 2006.

Comparing Table 10.2 with Tables 10.3 and 10.4, there is no consistency across the leagues. For the Premier League, the 1.10 minimum value seems to yield the best results, with very good results for both the Dixon-Coles based strategies. The 1.10 minimum value also yields good results for the Segunda Division, however here the gamblers' and Poisson assessment are showing good results, and the Dixon-Coles not so much.
In general, it is interesting to see, that the increase of the minimum value does not show an increase in the return on investment, nor in the net return. This raises a suspicion

	1.00	1.10	1.20
Gamblers (2.5 only)	48 bets, -4,04 (91,6%)	**13 bets, +1,08** (108,3%)	1 bet, +0,83 **(183%)**
Poisson (2.5 only)	48 bets, -4,48 (90,7%)	**9 bets, +3,218** (135,8%)	3 bets, +1,30 **(143,3%)**
Dixon-Coles (2.5 only)	**65 bets, +3,59** (105,52%)	37 bets, +3,05 (108,24%)	13 bets, +1,18 **(109,08%)**
Poisson	165 bets, -2,82 (98,3%)	**43 bets, +3,08** **(107,2%)**	11 bets, -1,38 (87,5%)
Dixon-Coles	185 bets, -1,17 **(99,4%)**	113 bets, -8,15 (92,8%)	**47 bets, -0,96** (97,96%)

Table 10.4: The betting results for the Segunda Division, fall season 2006.

towards, what can be regarded as borderline values. For the parameter setting of 1.00, it is hard to say, if a bet with a value of an accepted 1.01, in reality is a value of 99 and therefore should have been discarded. On the other hand, bets with a value higher than 1.20 perhaps are, in some sense, overestimated, leading to too high values. Perhaps the best bets in fact are found in a middle interval of 1.10 to 1.20, which could explain that, the 1.10 minimum value shows the best performance. Any further investigation of the minimum value parameter is left for future work, which could be a part of further tuning the model parameters. For the remainder of this project, the value betting strategy will use a minimum value of 1.10.

10.2 Threshold Betting Strategy

Looking at the statistics for a soccer match, can give some indications about the properties of the participating teams and the outcome of the match. If a team has won ten matches in a row plays a team with ten losses in a row, there is a clear indication of this team winning, if at least their last ten matches has been played against teams on the same level. The same can be said about the number of goals which can be expected in the match. If a team in most matches score and concede a lot of goals, it is likely that this behavior will repeat itself. The expected average number of goals used in the following, is as defined in Equation 6 in Section 8

The distance between the expected number of goals and the line if the offered bets is the used as a measure in deciding to bet, or not to bet. Algorithm 7 shows the threshold betting strategy.

The algorithm is provided with a match and odds data, as well as the calculated expected number of goals. Based on a threshold, which marks the minimum accepted distance between the expected goals and the odds lines, and a least accepted odds, the strategy decides to bet or not. The setting values for the threshold and the least accepted odds were proposed by [Ras08] to be 0.25 and 1.70. It is possible, that these two parameters

Algorithm 7 Threshold betting strategy

```
Function BestThresholdBet(Match, OddsData, ExpectedAverage, Threshold, AcceptedOdds)
BestBet = "No Bet"
for all (Odds ∈ OfferedOdds where Odds.Match = Match) do
  distance = oddsOffer.Line - ExpectedAverage
  PossibleBet = "No Bet"
  if distance ≥ Threshold then
    if oddsOffer.UnderOdds ≥ AcceptedOdds then
      PossibleBet = oddsOffer.Under
    end if
  else if distance ≤ -Threshold then
    if oddsOffer.OverOdds ≥ AcceptedOdds then
      PossibleBet = Odds.Over
    end if
    if Bet.Distance ≥ BestBet.Distance then
      BestBet = "Odds.Outcome, Odds"
    end if
  end if
end for
return BestBet
```

could be tuned to achieve a better strategy result. Three test runs have been made on the 2006 fall season matches, with three different threshold settings. The minimum odds is fixed at 1.70.

League	0.20	0.25	0.30
SAS League	69 bets, +8,2 (111,9%)	**58 bets, +10,104 (117,4%)**	51 bets, +5,61 (111,0%)
Premier League	181 bets, -16,14 (91,1%)	**159 bets, -12,80 (91,9%)**	140 bets, -17,31 (87,6%)
Segunda Division	158 bets, +3,39 (102,1%)	**126 bets, +6,64** (105,3%)	105 bets, +5,01 **(110,0%)**

Table 10.5: The number of bets places and the net result of the threshold betting strategy on the fall 2006 season for the three leagues.

Table 10.5 shows the average log scores for the threshold betting strategy, using the three settings for the threshold parameter. For both the SAS League and the Segunda Division the strategy shows a positive return, and good returns of investment. For the Premier League the results are a negative return on all settings, however the threshold value of 0.25 performs least bad. A reason for the positive results for the SAS League and the Segunda Division, can either be due to random behavior or to the differences in the league behavior. The SAS League is regarded as a high scoring league, while the Segunda Division is regarded as a low scoring league. Therefore the line for the SAS League is often higher than 2.5 and for the Segunda often lower. Looking closer at the bets placed by the threshold strategy, it was noticed, that a big part of the profit came from bet placed on over 1.75 and over 2.00 for the Segunda, and under 3.25, 3.00 and 2.75 for the SAS League. Without it being possible to draw conclusions, it is interesting to notice that the trend for the leagues are not as significant as the bookmakers' odds suggest. The SAS League is not as high scoring as one might think, and the Segunda not as goal-less. All in all the best value for the threshold is 0.25, giving the best net result for all three leagues, and the best return of investment on two of the three.

Chapter 11

Results

In this chapter, firstly the proposed assessors performances are examined and evaluated. The evaluations are based on tests made with a test data set, containing the matches for spring 2007 for the respective leagues. Secondly the betting strategies are run on an odds data set, also for the spring 2007 season. In the first section the test settings are presented along with the means for evaluating the significance of the results. This is followed by the results and evaluation for the assessors, and finally the results for the betting strategies.

11.1 Assessor Evaluation

The proposed assessors are all tested on each league. In order to establish which of the assessors is the best, the average log score is found. This along with testing the statistical significance of the log score results will be used to draw conclusion as to which assessor is the better.

11.1.1 Test Settings

For tests performed in this chapter, the parameter settings found in previous chapters have been used. For the respective leagues, for both the Poisson and the gamblers' approach, the best number of matches parameter has been set to those presented in Sections 8.2 and 7.3.1. For the Dixon-Coles approach the fade factor ϵ is set to 0.007, as established in Section 9.3, while the local parameters, such as strength and weaknesses are found as described in Algorithm 4. For the bookmakers' assessment, the two bookmakers 10Bet and 188bet are used, to find the bookmakers assessment. 10Bet is used for the Segunda Division and the Premier League, while 188bet is used for the SAS League.

11.1.2 Preliminary Results

For all of the four assessors, the average log score has been found, to enable comparison over (part of) the training data. For the gamblers', Poisson and Dixon-Coles approach the average log score was found for both the entire training data and limited data set, which was used for the preliminary the bookmaker assessments. The average log score results for the limited set, containing matches only for the fall season of 2006 is presented in Table 11.1, which summarizes the results presented at the end of each of the assessor chapters. The log scores were found partly to be able to compare the assessors, but also to compare the performance on the fall 2006 season to the results on the test data set of the spring 2007 season.

	Bookmaker	Gamblers	Poisson	Dixon-Coles
SAS League	-0.68408	-0.69185	-0.67347	-0.69250
Premier League	-0.685264	-0.68706	-0.69065	-0.70045
Segunda Division	-0.69281	-0.70603	-0.72505	-0.73675

Table 11.1: Preliminary results for each of the assessors on the three leagues. The table shows the average log score on the matches in the fall season of 2006

As it could be expected, the bookmaker shows the highest average log score. However on the SAS League, the Poisson assessor is the best, where the bookmaker is the second best. It is interesting to see, that on all three leagues, the Dixon-Coles assessor has the lowest average log score.

11.1.3 Assessor Scores

Each of the four assessors have been tested on the spring 2007 season for each league. The average log score results are presented in 11.2. The performance of the individual assessors differs quite a bit from league to league. On average, the bookmakers' assessments scores better, having the highest value for the SAS League and the Segunda Division. On the Premier League, both the gamblers' and the Poisson assessor performs marginally better.

	Bookmaker	Gamblers	Poisson	Dixon-Coles
SAS League	-0.69281	0.70603	-0.725056	-0.73674
Premier League	-0.69206	-0.68863	-0.69067	-0.70680
Segunda Division	-0.67581	-0.70822	-0.70741	-0.68787

Table 11.2: For all four assessors, the average log score is shown for each league for the spring season of 2007.

The Dixon-Coles assessor has the lowest average log score for the SAS League and Premier League, but is the assessor which comes closest to the bookmaker on the Segunda Division. In all, the bookmaker shows the best average log score.

The three leagues have different result behavior, which makes it interesting to examine the assessors predictions respective to the single league. Perhaps the Poisson approach is best for the Premier League, and the gamblers' approach best suited for the SAS League. Table 11.3 shows the average assessment for over/under 2.5, corresponding to the log scores above.

	Bookmaker	Gamblers	Poisson	Dixon-Coles
SAS League	0.520/0.480	0.544/0.456	0.548/0.452	0.474/0.526
Premier League	0.479/0.521	0.473/0.527	0.465/0.535	0.433/0.567
Segunda Division	0.483/0.517	0.408/0.592	0.421/0.579	0.437/0.563

Table 11.3: For all four assessors, the average log score is shown for each league for the spring season of 2007.

It is interesting to notice, that both the Poisson and the gamblers' approach seem very adapt to the known behavior of the leagues. For the high-scoring SAS League, they have over as a clear favorite, and for the low-scoring Segunda they have under as a clear favorite. For the medium-scoring Premier League under is the slight favorite. For the bookmaker it is noticed, that for all leagues, the average prediction is closer to 0.5/0.5 than the other assessors. Most interesting is it to notice, that the Dixon-Coles approach hold under as the favorite for all three leagues. Even for the SAS League, where the other assessors have over as the favorite. This explains the poor performance on the SAS League and Premier League, and the rather good performance on the Segunda Division. This raises the question if the Dixon-Coles is indeed fit as an assessor of over/under outcomes. Remembering that the Dixon-Coles model was initially designed to assess the probability of results, and has shown good performance in predicting the outcome of soccer matches. In this report, the assumption was made, that if the model has shown good performance in predicting outcome of a match (which is simply a combination of predicting the number of goals by each team) the model would also be plausible for predicting the total number of goals in a match. With the above results, this assumption does not seem to hold, except for leagues with a low average number of goals.

11.1.4 Significance Test

The evaluation of the quality of the assessors is based on hypothesis testing using the Wilcoxon Signed-Rank Test. As stated in Section 5.1.2, the null hypothesis used in this project is:

> Probability assessor A performs at least as good as probability assessor B, measured using the log score of the assessments.

The alternative hypothesis would be, that B is the better assessor. For the tests made in this chapter, the significance level used is $\alpha = 0.05$. In the Tables 11.4, 11.5 and 11.6 the test results are shown for each league respectively. The tables shows the result of each assessor versus each of the other assessors. The assessor in the horizontal header is assessor A, and the assessor in the vertical header is assessor B. The Wilcoxon test has been made by, for each match in the test set, subtracting the log score for the assessment of A from the log score of the assessment of B. The null hypothesis would the be, that the mean μ of the distribution of the differences is less than or equal to 0. The difference in scores, have then been sorted by order of absolute values, and assigned ranks. The ranks of the positive ranks are summed, as are the negative ranks. The goal is to reject the null hypothesis, so the positive rank sum is used to calculate the p-value. The positive rank sum is used, because if the p-value for the positive rank sum is lower than the significance level α, the null hypothesis can be rejected. If the null hypothesis is rejected, it would mean that the assessor in the vertical header is significantly better than the on in the horizontal header.

	Bookmaker	Gambler	Poisson	Dixon-Coles
Bookmaker	-	0.214	0.0317	0.0317
Gamblers	0.785	-	0.118	0.132
Poisson	0.968	0.880	-	0.168
Dixon-Coles	0.968	0.868	0.832	-

Table 11.4: The Wilcoxon Signed-Rank Test p-values for each combination of the assessors on the SAS League 2007 spring season.

Table 11.4 shows the p-values found for all combinations of assessors. The average log scores indicated, that the bookmaker was the best of the four assessors, with the gamblers' approach not far behind. The p-values indicate, that the bookmaker is not significantly better than the gamblers' approach. However, with a p-value of 0.0317, the bookmaker can be said to be significantly better than both the Poisson and Dixon-Coles approach at significant level 0.05.

	Bookmaker	Gambler	Poisson	Dixon-Coles
Bookmaker	-	0.170	0.138	0.494
Gamblers	0.831	0.426	-	0.774
Poisson	0.863	-	0.5765	0.863
Dixon-Coles	0.509	0.228	0.138	-

Table 11.5: The Wilcoxon Signed-Rank Test p-values for each combination of the assessors, on the Segunda Division 2007 spring season.

For the Segunda Division, the bookmaker is again the best assessor. However, it is not significantly better than any of the other assessors. It is better than the gamblers' approach at a significance level 0.17, and has its best performance against the Poisson approach, but a p-value of 0.138 can not be said to be significant, since there is a probability of 0.138 that the results are given by chance.

For the Premier League, the Poisson and the gamblers' approach are better than the bookmaker and the Dixon-Coles. However, the difference is not significant with respect to the bookmaker, while the Dixon-Coles shows poor performance. The gamblers' approach is in fact significantly better than the Dixon-Coles at a significance level 0.10, which does not pass test of 0.05 significance level.

	Bookmaker	Gambler	Poisson	Dixon-Coles
Bookmaker	-	0.723	0.592	0.190
Gamblers	0.277	-	0.417	0.093
Poisson	0.409	0.584	-	0.136
Dixon-Coles	0.810	0.907	0.864	-

Table 11.6: The Wilcoxon Signed-Rank Test p-values for each combination of the assessors on the Premier League 2007 spring season.

In all, none of the assessors are significantly better than the bookmaker. In general the bookmaker shows the best performance, but is only significantly better than the Poisson and Dixon-Coles approach on the SAS League. The gamblers' and Poisson approach show an accepted performance on all three of the leagues, being closest to the predictions of the bookmaker. The Dixon-Coles approach fails the test, with rather poor performance. On the SAS League it is significantly worse than the bookmakers. The reason for this seems to be an overestimation of the probability for under 2.5 goals, leading to the Dixon-Coles having its best performance on the low-scoring league, the Segunda Division.

11.2 Betting Strategy Evaluation

In Chapter 10, the two betting strategies were examined on the fall 2006 season, with the goal of finding best settings for the global parameter values. For the strategy tests performed in this chapter, the minimum value parameter has been set to 1.10 for the value betting strategy, while the threshold and minimum odds parameters have been set to 0.25 and 1.70 respectively for the threshold strategy. For the assessors used in the value betting strategy, the parameter settings used has been the same as in the previous section.
In all, seven strategy runs have been made for each of the leagues. In Table 11.7, four strategy results are shown. These results are based on a data set containing only odds with the over/under 2.5 line, in order to enable comparison, because of the gamblers' approach being limited to this single line. In Table 11.8 the results using a full odds data set are presented. Here there are two value betting strategies using Poisson and Dixon-Coles, and a single threshold strategy.

11.2.1 Betting Strategy Results

In the spring season of 2007, the SAS League data contains 90 matches, the Premier League contains 172 matches and the Segunda Division contains 264 matches.
Table 11.7 shows the betting strategy results for the odds data, only containing odds for the over/under 2.5 line. Since a bookmaker not always has odds offered for this line, only a fraction of the matches are actually eligible for a bet. The table shows the net result for each betting strategy on each of the leagues, along with its total net results.

	SAS League	Premier League	Segunda Division	Total
Value, Gamblers	14 bets, +7.55	18 bets, -3.763	17 bets, -13.187	49 bets, -9.40
Value, Poisson	17 bets, -9.14	20 bets, +6.827	16 bets, +1.836	53 bets, -0.477
Value, DC	60 bets, -1.985	8 bets, +5.56	5 bets, -1.14	73 bets, +2.435
Threshold	37 bets, -5.73	8 bets, -0,09	6 bets, -0.296	51 bets, -6,116

Table 11.7: Betting strategy results for the 2007 spring season, using an odds data set containing only over/under 2.5 odds.

For the limited data set with only 2.5 line odds, the net results are very differing. The value betting strategy using the gamblers' assessment shows a very high profit of +7,55 on only 14 bets on the SAS League. However it suffers a large loss on especially the Segunda Division. The opposite is the case for the value betting strategy using the Poisson assessment. It performs with a positive return on the Premier League and Segunda Division, but suffers a big loss on the SAS League. Of all the strategies ran on the limited data set, the value betting strategy which performs with a positive net result in total, is the one using the Dixon-Coles approach. However, the profit gained is solely from the Premier League matches, while the other two leagues yield a loss. There is no sign of consistency in the performance results. The threshold strategy yields a loss on all three leagues. However, this approach is initially intended for use on a odds data containing multiple lines.

	SAS League	Premier League	Segunda Division	Total
Poisson	29 bets, -8.755	49 bets, +9.737	65 bets, -5.21	143 bets, -4.228
Dixon-Coles	90 bets, -1.985	56 bets, +9.672	89 bets, +21.29	235 bets, +28.97
Threshold	59 bets, -3.3	65 bets, +6.71	94 bets, +0.86	218 bets, +4.27

Table 11.8: Betting strategy results for the 2007 spring season, using a full odds data set.

On the full odds data set, the performance is similar. Here the value betting strategy using the gamblers' assessment has been left out, since it can not decide to bet on lines other than 2.5. The Poisson approach again shows differing performance with profit on the Premier League and loss on the two other leagues. The Dixon-Coles approach shows

rather promising results, with a very high profit for both the Premier League and the Segunda Division. Of the total 526 matches, bets have been placed on 235 of these for a net result of +28.97 units, which is a 112,3% return of investment. A result which would impress any bettor. The threshold approach also shows a profit over all three leagues, with 218 bets for a profit of 4.27 units, being 102% return of investment.
Of the four strategies, the value betting using the Poisson and the gamblers' assessments, shows too little stability in the results and too large a loss, in order for one to conclude that the strategy could be used for betting. This is both the case for the limited and the full odds data set. The threshold betting strategy performs better on the full odds data set, and yields a slight profit for the Premier League and Segunda Division. Remembering that that these leagues are not leagues with a high average number of goals, it would interesting to test the threshold strategy on other low- or medium-scoring leagues. The value betting strategy using Dixon-Coles showed the best betting results of all. Both for the limited and the full odds data set, a profit was attained over the total matches. The best results were seen for Premier League and Segunda Division for the full odds data set. Remembering that the Dixon-Coles assessment were prone to over-estimate the probability of a low number of goals, raises suspicion as to which it can be used for placing bets. However, the results attained for the Premier League and Segunda Division indicates that the value betting strategy using the Dixon-Coles assessor might be good for placing bets on low-scoring leagues. However, this is at this unsubstantiated, and would call for further research and tests.

Chapter 12

Conclusion

Having implemented and evaluated the assessors proposed in the report, it is now possible to draw conclusions about the results. The goal of the project was to examine if it is possible to create automatic probability assessments which can at least match those made by a human bookmaker.

12.1 Project Evaluation

As stated in Section 1.2 the goal of this project has been two-fold. Primarily the goals has been the establishment and implementation of a number of candidate assessors for predicting the number of goals in soccer matches, at a level which could compete with that of the bookmakers. Secondarily the goals was to set up a test environment, in which the predictions made by the assessors could be compared and evaluated, in order to establish which of these has shown the bets performance in tests. For the test phase, the choice of a logarithmic scoring rule was made, which fits the problem setting and purpose well. The scoring rule was combined with hypothesis testing, using the Wilcoxon Signed-Rank Test, enabling the establishment of statistical significance of the results.

12.1.1 Assessor Performance

The expectations towards the assessors was, that the bookmaker with good chance would come out as the best assessor. The Dixon-Coles model had shown good results in predicting match outcome probabilities, why the expectations was that the model also could reach a level in predicting the total number of goals, which could almost match the bookmaker. The two more "naive" approaches were initially expected to perform worse than both the bookmaker and the Dixon-Coles model. However, the test an evaluation has shown, that the Dixon-Coles model is not as well suited for predicting the number of goals as it is at predicting the actual outcome. The over-estimation of the probability of low-scoring games, caused the model to show the worst prediction scores of the four assessors. Of the to naive approaches, which both show good performance, the gamblers' approach showed the best performance, being that of the proposed assessors coming closest to both the predictions and the average log scores of the bookmaker. Using the Wilcoxon test, it was determined that none of the proposed assessors showed significantly better performance than the bookmaker. However only in one case, on the SAS League, did the bookmaker show significantly better performance than the Poisson and Dixon-Coles approach. It can therefore not be concluded that any of the assessors proposed was better than the bookmaker, nor can it be said that they in general are significantly worse. It has been shown, that the assessor can not beat the bookmakers, but almost match them with regards to predicting the number of goals.

12.1.2 Betting Strategy Results

For both the value betting strategy and the threshold betting strategy, it was not possible to conclude that any of the two is able to produce a stable income on the leagues in question. At least not with the assessors used in this project. It is natural to assume, that if an assessor exists which makes better predictions than the bookmakers it would be possible to generate a profit. However, it is indeed a difficult task to beat the bookmakers as a gambler. Not only is it necessary to have better predictions, it is also necessary that the predictions are significantly better in order to beat the margin the bookmaker gains by the theoretical payback percentage. The tests made in this project was made on only half a season. This leaves a lot of things to chance. Perhaps this specific half season saw results which was very much different than the results usually are in the league. In this case, all of the assessors proposed in this report would be overfit to the historical data used for training. The sparse amount of test data in mind, it is not possible to dismiss the betting strategies and the fact that it could be possible that they could generate a profit over a larger number of matches. However, since the none of the assessors actually perform significantly better than the bookmaker, this does not seem likely.

12.1.3 Assessors as Bookmakers

Despite not being able to create better predictions than the bookmakers, the automatic models still have their justification. Having a human bookmaker sit and survey the news and happenings in a large number of soccer leagues is costly. It is not uncommon that a full time odds-setter only covers a few leagues at the time. If a bookmaker has perhaps 100 different leagues in various sports, the number of employees and the total salary expenses are high. The use of automatic assessors is therefore a very popular area of research within bookmakers. If costs can be lowered, any initiative will be taken into consideration. The idea of automatic assessors replacing human bookmakers is therefore not far fetched. If an automatic assessor can create a prediction not significantly worse than a human bookmakers, they can be used for setting the odds. By having an automatized system, it would be possible for a bookmaker to offer odds very early, perhaps even several days in advance to the competition (who are investigating the league and market). By monitoring the odds, and the stakes placed, changes in odds can be made accordingly to minimize the risk on an event. In time the odds will adjust to market and find its natural level, which it also would if the odds was from the beginning compiled by a human bookmaker. In this sense, the assessors proposed in this project would be candidates for such a system. A refinement of the Dixon-Coles model would, it is assumed, also be suited for assessing the result at half time or perhaps the probability of a team winning with a larger margin. A such model would be able to create several markets for a single match, which would be desirable for any bookmaker.

12.2 Future Work

In the results chapter it became clear, that the Dixon-Coles model over-estimated the probability of low scores, leading to too high predicted probabilities for under 2.5 goals. The model initially uses a dependency function which modifies the probability of outcomes 0-0, 0-1, 1-0 and 1-1 which are all outcomes corresponding to the under 2.5 outcome for a total goal prediction. While this dependency function is well suited for the use for predicting the outcome of a match, it is not certain that it in fact is well suited when predicting the total number of goals. For future work, it would be interesting to look closer at redefining the dependency function, by examining the result data set closer. The Dixon-Coles model only takes into account the final result of previous games, and totally disregards any other circumstances surrounding the match. A suggestion for two additions to the Dixon-Coles model is proposed. Firstly the implementation of weather measures. The interviewed bookmakers have both confirmed that the weather forecast has influence on their predictions for the total number of goals in a match. This would, however, call for an extensive investigation into the weather conditions for previous matches played, which could be a very time consuming task. By collecting the data from this point in time, this addition can be implemented when sufficient data has been collected. The second suggestion proposed, is the utilization of the match goal data, such as first half goals, second half goals and the time of goals. Teams could have tendencies for starting slow in a match, or starting with very aggressive tactics. These conditions have influence of the probability of the number of goals, and through data mining of old match data, factors can perhaps be found which can be used in the Dixon-Coles Poisson distribution. Despite the gamblers' and Poisson approach being rather "naive" approaches, their performance indicates, that they could be used as, if not as a whole then as a part of, a odds assessment model. This could also be an interesting angle towards improving the Dixon-Coles models ability to predict the total number of goals.

Appendix A

Interviews

To gain information on how probabilities and odds for over/under goals on soccer matches are made, two bookmakers and one professional gambler has been consulted. In the following their views on the matter is accounted for.

A.1 Gambler Interview

What is your background in betting? My "betting career" started in the early 90ies where my interest in odds was started. In the beginning i lost more than I won, but at that time I was not aware of the mathematics and the potential in betting. It was first about ten years later that I started to have interest in the pure mathematics in betting, such as value betting and optimal stake. Also this was the point where I realized that if I was to make a profit on betting I needed to be really well informed. Even better informed than the bookmakers. Since then I have put a lot of attention towards tennis. The matches are somewhat easier to assess than soccer matches or others. There are only two possible outcomes, and there are less factors. For the last five or six years I have been living on my winnings from betting, and I always seek to find new niches to investigate and perhaps find new ways of winning. Here horse race trading and over/under in soccer matches are in my interest.

When assessing a soccer match with regards to an over/under bet, what do you take into consideration? There are a lot of factors to have in mind. First of all, there are the leagues and teams general statistics and tendencies. For example the French second division have very few goals, while Dutch first division have a lot of goals. In each leagues there are defensive teams and offensive teams. These facts are crucial when looking at a bet. These I use to nominate possible bet for each round, finding just a small group of matches I will take a closer look at. Then I look at the team news, tendencies within prior matches between the teams. Is it a local derby, or is it a very important clash at the bottom of the league deciding relegation. This together with injuries and suspensions are the key aspects for me.

But investigating takes a lot of time, and in many cases the game is a no bet. I want a high bet value before I place a bet. The time spent on investigating is often too much compared to the over all winnings. Therefore I have been thinking about trying to quantify the decision to bet, to save time. I do not expect that it is possible to get as great a return on investment as by doing it manually, but if it is possible to make just a small profit by having an automatic system, a lot of time is saved which can be spent else where, on areas where larger winnings are possible. I have been testing an approach where I for a match take the two teams, and calculate the average number of goals in their matches so far in the season. Then I take the average number and compare to the market odds. If the average is more than 0.25 lower than, for example under 2.75 goals, it is a possible bet. I have found that odds should be 1.70 or higher to be a possible bet. I have been testing this on several leagues, in the spring season, and for some of the leagues there could be some interesting areas. Primarily bets in leagues where the average number of goals is very low, like 2.28 in the French league for example, there are often good bets on over 2.00 and over 2.25 goals. Some lines are better than others, but as of now it is not possible to say if I have been lucky or there actually is a possibility of making money here.

A.2 Bookmaker Interview

Two bookmakers with no relation have been asked about how they assess odds for over/under on soccer matches, and which factors they see relevant when doing so. The interviews have been carried out through email correspondence, and the following are outtakes found relevant for this report.

A.2.1 Frederik Skov, Scandic Bookmakers

When assessing a soccer match, which factors do you take into account when setting odds for over/under goals? The largest factor we look at when deciding the odds for over/under is the latest results. If a teams has played eight over matches in the last ten, this will indeed give a higher probability for over. This assessment is made for both teams and the responsible odds setter decides their significance to the total over/under assessment. Also we look at the team news, and adjust if teams are missing signficant players in offence or defence. Most of our turnover on this betting type is on over. Therefore we tend to lower the odds on over outcomes, to minimize our total risk on the event. However, if it is winter or late fall, the weather plays a role in the setting of odds, pulling in the direction of under.

A.2.2 Søren Hansen, Danbook.com

When assessing a soccer match, which factors do you take into account when setting odds for over/under goals? First of all, we look at the statistical material.

If a team normally plays matches with significantly few or many goals. Here we also look at previous matches between the two opposing teams - are they prone to play 0-0 or do they explode in goals. A relevant factor is of course the player material in a team. If a, normally, high scoring team is missing its best attacker, the chance of them scoring goals is lower than if he plays. The same can be said for defensive players and conceding goals. Another relevant factor is the weather, in fact. If it is snowing or raining heavily, it becomes harder for the teams to create good attacking football and the chance of many goals is lowered. The last thing we look at, is if there are any circumstances surrounding the match that can affect the result. For example in a cup game, where the teams play both home and away, and in the second match the away team can progress to the next round if they keep the opponent from scoring. Then they concentrate on defending and attacking becomes secondary. These special circumstances are normally only present in cup games or at the end of a season.

OB and FCK are playing this weekend (March 2008). Can you describe how you have assessed this match? For both teams it is the case that they are very strong in keeping the opponents from scoring. At the same time they are not teams that score a lot of goals. They are teams "that get the job done". They go for the 1-0 win. In their past mutual games the tendency has been low scoring games, perhaps partly due to the matches being important top games, where neither team can afford to lose. Besides that, we are in the late winter, at the start of the spring season, and the pitch is likely to be a bit hard due to frost. This game is likely to end with few goals, and under 2.5 goals is a clear favorite here. Our odds are 1.72 for under and 2.12 for over.

How do you create multiple line over/under odds for a match? In our case, we have created a model that takes the input of our employees on the above mentioned factors, which returns a probability distribution for 0, 1, 2 and so on goals. From these it is simple to create over/under 2.5 goals and over/under 2.75 and so on. Alternatively it is possible to make probabilities for over/under 1.5, 2.5, 3.5 and combine these to make over/under 2.75 for example.

References

[Bet] BetXpert.com.

[CH] Tobias Christensen and Rasmus Hansen. Odds assessment on football matches. Master's thesis, Aalborg University.

[DC97] Mark J. Dixon and Stuart G. Coles. Modelling association football scores and inefficiencies in the football betting market. *Journal of Royal Statistical Society series, Series C. Vol. 46, No. 2, p. 265-280*, 1997.

[DS] Morris H. Degroot and Mark J. Scherwish. *Probability and Statistics*. Addison-Wesley.

[Han08] Søren Hansen, 2008. Interview.

[Kee95] E.S. Keeping. *Introduction to Statistical Inference*. Dover Publications, 1995.

[Mit97] Tom M. Mitchell. *Machine Learning*. MIT Press, 1997.

[Ras08] Klaus Rasmussen, 2008. Interview.

[Te] Tip-ex.com.

[TSK] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Pearson Education.

[Wil] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin, Vol 1, No 6, p. 80-83*.

[WM] Robert L. Winkler and Allan H. Murphy. Good probability assessors. *Journal of Applied Meterology, Vol 7, No 4, p.751-758.*