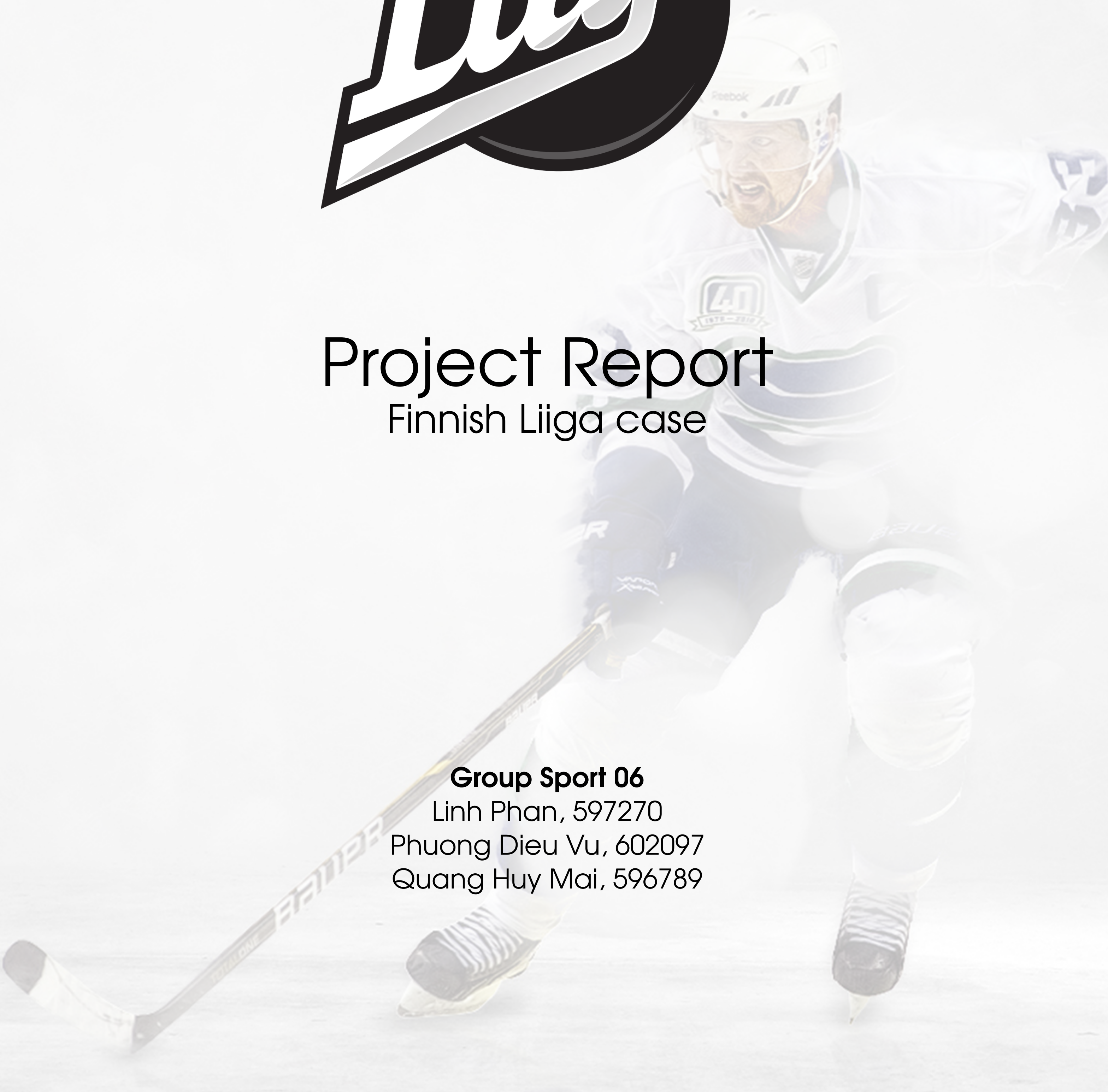

Project Report

Finnish Liiga case

Group Sport 06

Linh Phan, 597270

Phuong Dieu Vu, 602097

Quang Huy Mai, 596789

Contents

Summary

This project aims to solve two problems: (1) estimating winning probabilities of each team in Liiga playoff series, and (2) constructing a betting portfolio to maximize its expected value.
In order to get adequate estimated results for problem (1), we use the Poisson regression to estimate relative strengths and home advantages of each team at the first step. Based on these estimated coefficients, we build a predictive model to predict goals scored in a single game. Then, using Monte Carlo simulation technique, we attempt to estimate the probabilities of each team to win the championship. The simulation was run 1 million times to gain more precise estimations.
In problem 2, we use linear programming technique to solve the optimization of maximum expected results based on winning probabilities of each team. Though it may seem that the numbers are simply the calculated chances multiplied by the budget, the outcomes make sense that proportions on teams that have higher chances to win are larger.
By using abovementioned predictive and prescriptive models, we reach the conclusion that Kärpät has the highest chance to win the title, and the largest proportion of the betting budget should be for Kärpät accordingly (but still satisfy the requirement that no single team receives more than 50% of betting budget). We expect that our result will be closed to playoffs results in real-life.

While the process we applied in the project is reasonable, we would recommend that we should adjust the assumptions depending on the practical conditions. For example, more influential elements should be included in the model, such as the effects of playoff round. The assumption of equal home advantage for all teams should also be reconsidered.

1. Introduction

1.1 Motivation

The rise of Data-driven era have been changing almost every aspect of life, and the world of sports is no exception. Today, "the popularity of data-driven decision-making in sports has trickled down to the fans, which are consuming more analytical content than ever" – Forbes (2015). With data freely available and analytical tools in hands, ones can easily investigate the past to gain more insight about the future. Sport gamblers are no doubt among those who benefit most from this new trend.

Inspired by the rise of sport analytics and great things we can experience with it, we decided to conduct this study on estimate Sport results and maximizing betting portfolio.

1.2 Problem Description

The Finnish Ice hockey League *Liiga* is one of the most well-known ice hockey competitions in the world. As a result, gambling on the league has been rising constantly, which makes the ability to predict the teams' performance and result of the league becomes highly demanded. Should one can estimate the probabilities of winning by each teams, one can maximize returns and reduce the risk on betting portfolio. That is also the problem this study aims to resolve.

1.3 Objectives and expected results

The project has two main objectives: to estimate the relative strengths of 15 teams in Liiga, by which one can estimate the chances of wining the title by each of the teams and allocate the 1000 € betting budget to maximize the expected return, while making sure to keep the constrains.
As discussed above, we expect to get the estimated results exact or nearest to the actual results of the series. If our estimations are right, then the betting outcome will be optimized accordingly.

2. Data

2.1 Description of raw data

The given data sets consist of "RegularSeasonData.csv" and "BettingOdds.csv". The first file contains data of all regular season games played by the 10th March 2018, i.e. when the regular season ended. There are six data fields: date, name of the home team, name of the visiting team, number of goals by the home team, number of goals by visiting team and the binary variable of whether the game went to overtime. In total, the data set has 450 results for 450 games during the season in the order of the time the game took place. There were 15 teams playing in the regular season and each team played 60 games, 30 games as the home team and 30 games as the visiting team. The second file contains the betting odds data with team names and decimal odds for each team to win the championship.

2.2 Description of preprocessing

In order to estimate the number of goals scored by teams, we need to estimate α_i (the relative attacking strengths), and β_i (the relative defensive strengths of the teams). First, since some of the matches include over time, during which the ties were broken, the reported goals for these matches include the goals scored in overtimes, and therefore need to be corrected. As the rules require a game to stop immediately after a team scores during ot, and all of the 110 games with OT=1 resulted in one team scored one goal more than the other, we assume that the ties in all of overtimed games were broken during the extra time. We, therefore, set the goals of winning teams in those matches to equal to the goals of the losing teams.

After adjusting the numbers, using each of the matches, we create two observations of four different features: $Attacker$, $Defender$, $Goals$, and $Home$. Each of the participating teams becomes is the $Attacker$ in one of the observations and $Defender$ in the other one, $Goals$ denotes the goals scored by the Attacker in that observation, and $Home$ is a dummy variable denoting whether the Attacker is the home team. With this transformation, we double the original 450 observations to 900 observations, of which 450 have $Home = 1$. There are 15 teams in the league, which convert to 15 categorical values for each of the $Attacker$ and $Defender$ variables. Using one-hot encoder (transforming categorical variables to dummy variables) for these two categorical variables, we attain a total 31 different dummy variables, of which 15 denote $Attacker$, 15 denote $Defender$, and 1 denotes whether the Attacker is the home team. By estimating the coefficients of these dummy variables, we can estimate the relative strengths in attacking and defending of each teams, as well as the home advantage μ.

The dependent variable is the number of goals in each of the observations, which is home team's goals if $Home = 1$ and away team's goals if $Home = 0$. There is no missing value in provided data set.

2.3 Descriptive Statistics

2.3.1 Regular season results

Of the 450 matches, 110 ended in over time, i.e. the two teams had same number of scores in regular time. Tappara was the team that had the most overtimed matches, 21 times, and they were also the one with best performance in OT by winning 14, or 66.7% of their OT-ed matches. On the other end of the string, KooKoo had highest chance to lose in OT, when they only won 5 in their 15 OT-ed matches. The teams' OT performance is displayed in figure 2.1

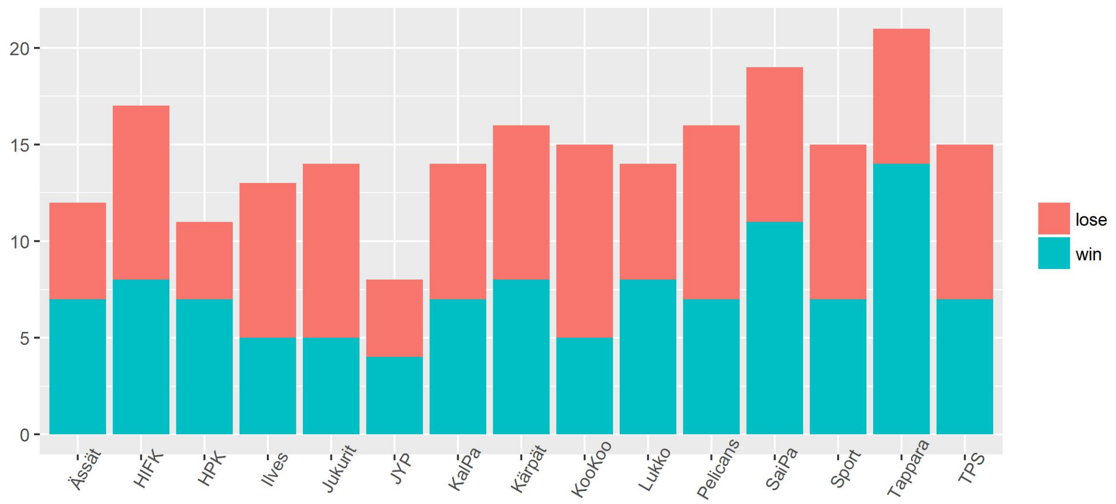

Figure 2.1: Numbers of each team's wins and loses in OT

Of all non-OT matches, Kärpät was the team with most outstanding performance of 33 wins in 44 matches (75% winning), while Sport had the worst result of only 26.7% winning in total 45 non-OT matches. Details of non-OT performance can be found in 5.1 in Appendix.

In terms of home and away performance, Kärpät also had the best performance at home, with 87.5% of their matches as home team resulted with their victories. However, when it comes to performance away from home, TPS is the best with 61.9% winnings in all of their away matches (Figure 5.2). Overall, the home teams won 57.78% of the matches, and the visiting teams only won 42.22%.

Regarding numbers of goals scored at home and away, Kärpät had the highest average number of goals as the home team, which was 3.63 per game, while TPS and JYP had the highest average number of goals when playing as visiting teams, which was 3.07 per game. On average, a home team scored 2.873 goals and a visitor scored 2.456 goals in one game.

When it comes to defending strength, as a home team, Tappara seems to be the best defender, with on average 1.73 goals lost per home game, while the worst one, KooKoo has on average 3.27

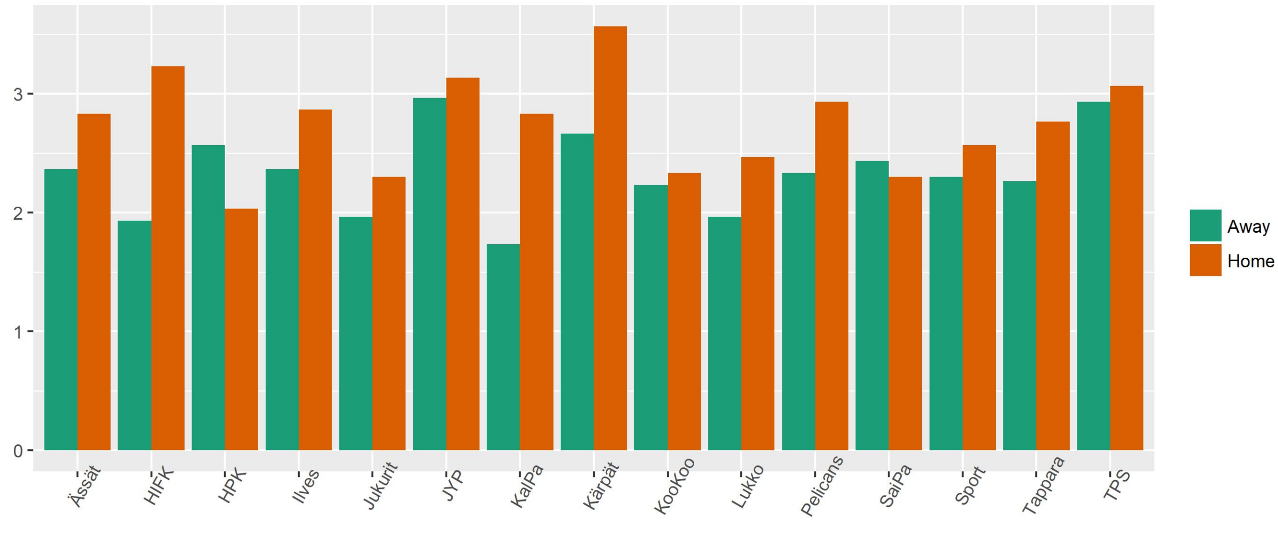

Figure 2.2: Average scores of each team as home team and visitor team

goals lost for each match in their home stadium. As a visitor, Kalpa lost least number of goals to its home team, only 2.2 per game, while Sport received the most: 3.6 goals per game (Figure 2.3)

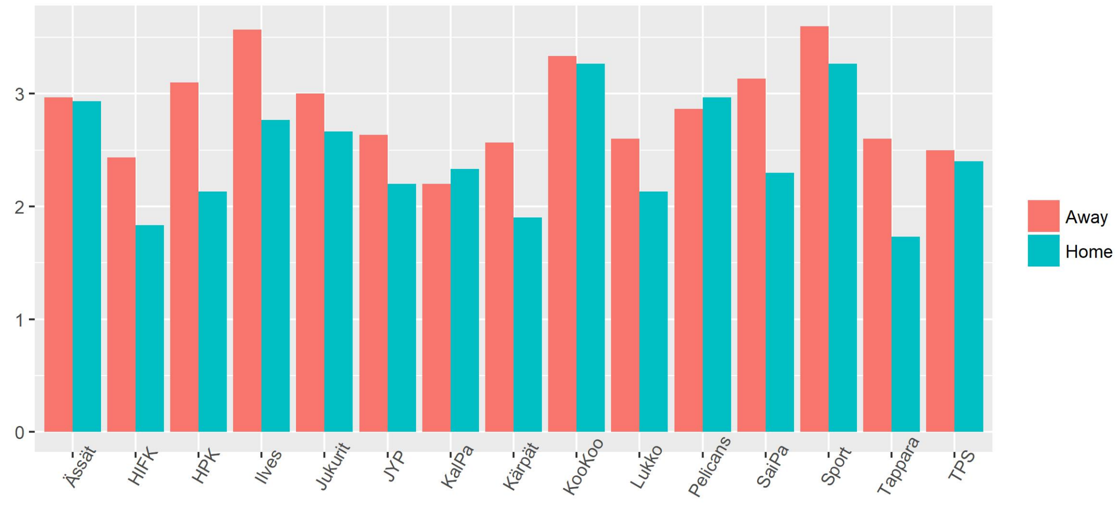

Figure 2.3: Average goals lost per game at home and away

2.3.2 Betting odds

HPK and Ilves are two teams which had the highest betting odds of 700, while Kärpät had the lowest of 3.79. The decimal odds show a large variance between teams (standard deviation approx. 261). As the decimal odds reflect the inverse of the implied success probability, it seems that HPK and Ilves have the least expectation to be the champion. The higher uncertainty, the higher risk and payment.

By the end of the regular season, HPK was eliminated. The payment for a successful bet is the stake multiplied by these decimal odds in case the team wins the championship, so if Ilves wins the champion, the payoff will be large.

3. Predictive model and results

With the data processed, in this section, we started to build the models to solve the main purpose of analyzing this set of data. The source code (in R programming language) of this and all following chapters can be found in the Appendix.

3.1 Model teams' relative strengths and Home advantage

Assuming that the goals scored by the Attacking team follow a Poisson distribution, we estimate the coefficients for each of the 31 dummy variables. In the first try, the Defending coefficient of team TPS turned out to be *N/A*, might have been due to the high colinearity of this variable with some of the other variables. This, however, was not an unexpected situation, and as instructed in the tutorial, we attempted to remove this variable from the model. The result is displayed in table 3.1.

3.2 Predicting goals scored by teams during a single game

The vector of coefficients we attained in last section was incomplete, as β_{TPS}, or the Defending strength of TPS was missing. As described, we assume $\beta_{TPS} = 0$ and add number 0 to the 30th position of the coefficient vector θ

After achieving the coefficients for the parameters, we can model the representation of goals scored by teams in a single game. First, using names of the Home team and Away team, we create two observations of 31 dummy variables, whose value initially set to 0. In the first observation, α of Home team, β of Visitor team, and *Home* are set to 1, while in the second observation, α of Away team and β of Home team are set to 1. Call X is the matrix of size 2×31, with rows represent two aforementioned observations, and θ the coefficients vector of size 31×1. The estimation of goals scored by each team is:

$$Y = e^{X \times \theta} \tag{3.1}$$

In this equation, Y is the result vector of length 2, of which the first element represents home team's goals and the second element represents away team's goals.

Team	Attacking Strength (α)	Defending Strength (β)	Home advantage (μ)
Ässät	0.808	0.199	
HIFK	0.777	-0.174	
HPK	0.682	0.068	
Ilves	0.809	0.259	
Jukurit	0.605	0.119	
JYP	0.953	0.022	
KalPa	0.656	-0.099	
Kärpät	0.971	-0.096	0.163
KooKoo	0.674	0.282	
Lukko	0.623	-0.053	
Pelicans	0.804	0.171	
SaiPa	0.689	0.091	
Sport	0.746	0.337	
Tappara	0.731	-0.140	
TPS	0.931	-	

Table 3.1: Relative strengths and Home advantage estimations

As an example, we estimated goals in a hypothetical match with *HIFK* as the Home team and *SaiPa* as the Away team. The model resulted that the home team would score 2.8 and away team would score 1.67 goals.

3.3 Estimate winning probabilities of individual games

As described, the goals scored by teams are assumed to follow a Poisson distribution, consequently, the Poisson distribution is applied in all of the simulations of this paper.

With the ability to represent goals scored by each team, we can use Monte Carlo estimation to estimate their chances of winning in that particular game. Taking the numbers resulted by model 3.1 as the expected values, we simulate 1000000 (one million) different Poisson values to represent the teams' goals, and compare the results of two teams to see in how many percent of the simulations that one team turns out to have more goals than the other, and in how many percent they end up to have the same number of goals. We also have one option to ignore the drawing possibility by scale up the possibilities of the other two outcomes (home team wins and away team wins) so that their sum equals 1.

Using the same example as in the previous section, we estimated that the Home team (*HIFK*) has 66% chance to win the match, while *SaiPa* only has 19% chance to win. The match has 15% chance to end up in a tie.

3.4 Likelihood of different outcomes for the entire playoff bracket

Using Monte Carlo simulation technique, in this section we attempt to estimate chances of winning the title for each of the teams in the playoff rounds.

In the first step, we built a function to determine the winner of one single match in one simulation. With model 3.1, we estimated expected goals of the teams, and used Poisson distribution simulation to generate one number of goals for each of the team, by which we can determine which team wins the match. In case the model results in a tie, it will re-generate the number of scores until there is a winner. This is a weakness of the model, as in real life, this situation will be dealt by a three-on-three sudden death, rather than a rematch. *In fact, we intended to use the overtimed matches in Regular round to make better estimate of winner for this situation, but then decided that the benefits are not worth the complexity.* In a later part, we will discuss this issue further.

After having the tool to generate one winner for a match, we can build the function to generate one winner for a playoff round, which consists of 7 single matches, with the teams take turns to be the home team. The round stops right when one team has 4 wins, and that team goes on to the next round. In Liiga, the playoffs have 4 quarter finals, of which four winners enter the two semi-finals, whose winners compete in the final 7 matches, which means after running the best-of-seven function seven times, we will have the name of the champion. By running this same simulation 1000000 (one million) times and keeping track of each team's number of "becoming the champion", we estimated the likelihood that each of the eight final teams would win the title, as in table 3.2.

Team	Chance	Team	Chance
Ässät	0.33%	HIFK	11.46%
JYP	10.42%	KalPa	2.69%
Kärpät	44.64%	SaiPa	0.48%
Tappara	9.63%	TPS	20.35%

Table 3.2: Teams' estimated chances of winning the title

4. Prescriptive model and results

4.1 Solve the allocation of 1000 € budget

The expected return of 1 € bet on one particular team's winning the title equals how likely that team will win times the amount of returns in case that team wins. We fortunately have both of these numbers, so to maximize the expected return of 1000 € betting allocation is simply a linear optimization problem with known parameters. Call E_1 to E_8 the product of each team's chance of winning and its Betting Odd, and x_1 to x_8 the amount of bet on each team, we obtain the following maximization:

$$Max(x_1E_1 + x_2E_2 + x_3E_3 + x_4E_4 + x_5E_5 + x_6E_6 + x_7E_7 + +x_8E_8)$$

Constrains:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + +x_8 &\leq 1000 \\ x_1 &\leq 500 \\ x_2 &\leq 500 \\ x_3 &\leq 500 \\ x_4 &\leq 500 \\ x_5 &\leq 500 \\ x_6 &\leq 500 \\ x_7 &\leq 500 \\ x_8 &\leq 500 \end{aligned}$$

The result returned that maximum expected value is 4562.31 €, with the betting allocation is 500 € each on *Kärpät* and *HIFK*, and 0 € on the rest of the teams.

4.2 Another allocation

From our point of view, the maximum return that we obtained in the last section was too good to be true. That was because the linear optimization put too much weights on teams with high betting odds, while these are the teams that are least likely to win in the bookie's opinion. We, therefore, propose another allocation, in which the maximum proportion of the budget that should be allocated to a team is that team's chances to win the title. With c_1 to c_8 denoting team's chances to win of the teams, the maximization then became:

$$Max(x_1E_1 + x_2E_2 + x_3E_3 + x_4E_4 + x_5E_5 + x_6E_6 + x_7E_7 + +x_8E_8)$$

Constrains:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + +x_8 \leq 1000 \\ x_1 \leq c_1 * 1000 \\ x_2 \leq c_2 * 1000 \\ x_3 \leq c_3 * 1000 \\ x_4 \leq c_4 * 1000 \\ x_5 \leq c_5 * 1000 \\ x_6 \leq c_6 * 1000 \\ x_7 \leq c_7 * 1000 \\ x_8 \leq c_8 * 1000 \end{aligned}$$

The LP optimization resulted a maximized expected value of 1266.405 €, which is 26.6% profit. It is a much more reasonable expected return from a bet than the number we obtained in the last section. The allocation of the budget are in table 4.1. While it seemed that the numbers are simply the calculated chances multiplied by the budget, it is still a reasonable allocation that it puts larger proportions on teams that have higher chances to win.

Team	Bet(€)	Team	Bet(€)
Ässät	3.3	HIFK	114.6
JYP	104.2	KalPa	26.9
Kärpät	446.4	SaiPa	4.8
Tappara	96.3	TPS	203.5

Table 4.1: Teams' estimated chances of winning the title

5. Discussion and managerial implications

5.1 Limitation of the process

Though the process that we have been applying so far in this project is reasonable and practical, the model was developed based on several assumptions, of which the validations may be questionable. In this section, we will describe some of the most noticable assumptions which, in our opinion, might be considered over-optimistic.

5.1.1 Assumption of equal home advantages for all teams

In a Finnish Liiga game, the advantages that the home team gets is clearly visible; however, it might be unrealistic to consider the advantage that every team in the league gets as the home team in a game the same fixed amount. In fact, based on the team's history, the stadium's capacity, the enthusiasm of citizens in the home city, ect., the home advantages of the teams may vary. For that reason, it might be more reasonable to consider the advantage that every team gets as the home team a different measure than to estimate one single advantage for all teams as we did in previous sections.

5.1.2 Not taking into account the "playoff element"

The regular season's matches and playoff rounds' matches are, by definition, very different. There are 60 rounds to determine the positions of the teams in the league, so as any other long-term competitions, teams must be very careful with their strategies, especially when the benefits of ending up in the first, instead of sixth, or seventh, instead of tenth position, is relatively small. In several cases, losing a game is necessary for a team to recover or protect their best players from injury, avoid facing a stronger team in the playoff, or simply because they have earned enough points (or have no chance to earn enough) to have a place in the playoff round. The playoff round, on the other hand, is very different. Every game is a real fight, into which both the teams have to put all they have in order to win. There is no more "strategic loss", as losing a game increases, significantly, the probability to lose the round and go home with nothing. For these reasons, it might not be very rational not taking the nature of playoffs into considerations.

Also related to the "playoff element", it will be worth restating that in the simulation, we dealt with the draws (i.e. situations, in which the simulated goals of two teams are equal) by re-simulating the goals, which is not a correct illustration of the Liiga laws. As mentioned in an earlier part, we were thinking of using the data from matches that ended in OT to better model this situation, but turned out that would not help much. For one reason, since all of the OT time only has at most 1 goal scored, we are unable to use the same Poisson distribution model we used earlier to estimate the winning chances of teams. Applying a binary algorithm like stochastic regression might help, but as we tried several algorithms, the data size (110 matches) seemed to be too small to result in good models. Also, as during a Regular round's match, the winner of an OT only gets 1 point more than the loser, the OT-ed matches in Regular round certainly does not reflect the nature of OT-ed matches in playoffs.

5.1.3 Use of linear optimization

To solve the optimization of betting allocation, we applied linear optimization method with the betting odds issued by the bookmaker, and maximize the expected returns without concerning the risks. In reality, the bookie normally put unreasonably high betting odds on teams that have very low chances to win, so while our linear algorithm normally favor those teams with much higher betting odds and not too lower winning likelihood over teams that are highly likely to win, the risks of betting on teams that have high betting odds are very high.

In the next section, we will focus on trying to improve the model by resolving the issues we have mentioned.

5.2 Suggested changes for the model

This section is dedicated to discuss a more effective model by "fixing" the issues that have been stated in the previous part. After that, we will also attempt to partly apply this new model into the existing data.

5.2.1 The new studying process

As an effort to solve the issue of equal home advantage for every team, we suggest to include, instead of one, 15 μ values for 15 teams in the league. As a result, the model that we aim to solve in the first place will consist of 15 dummy variables of attackers, 15 dummy variables of defenders and 15 dummy variables of home teams. Since otherwise would double the effects of home advantage, in a match between home team A and visitor team B, we only set $Home_A = 1$ in the observation that $Attack_A = 1$, not when $Defend_A = 1$.

We also include the data of previous seasons into calculation, to measure the effects of "playoff element". By estimating the relative attacking and defending strengths of the teams in regular seasons and in playoff rounds through the seasons, we can test and estimate the effects of "playoff elements", using A/B test, event study or linear regression. This would help adjusting the estimated relative strengths to suit more with this season's playoff rounds.

For the betting allocation issue, to prevent the model from allocating too much on teams that have low chances to win, we will use the constrain that we proposed earlier: the maximum proportion of budget to bet on one team is the chances that team wins the title.

5.2.2 Attempting the new method

Since accounting for the "playoff element" requires the use of previous seasons' data, which is unavailable for us, we could only attempt to incorporate every team's own home advantage into the model. The in-hands data currently has 450 matches, which equal to 900 observations, and that should be enough to estimate 45 coefficients.

We applied the same method of the original model in data preparation process, but instead of using one single *Home* dummy variable, which equals 1 for all observations with home team being the attacking team, we use a different *Home* dummy variable for each home team. As described in the previous part, we do not set the $Home_A$ variable to equal 1 in observations with team A is the defending team, to avoid double effects. The full source code of the analysis can be found in the appendix.

In our first try, the coefficient of the variable *Defender_TPS* turned out to be *N/A*, just like in the original model. We once again removed that variable from the model, and added the coefficient 0 in its position of the estimation model, just as we did before. The estimated coefficients of the other 44 dummy variables are in table 5.1

Team	Attacking Strength (α)	Defending Strength (β)	Home advantage (μ)
Ässät	0.7988798	0.1993654	0.1799714
HIFK	0.5721497	-0.1744723	0.514268
HPK	0.8768435	0.06781017	-0.2329316
Ilves	0.7935453	0.2592747	0.1916674
Jukurit	0.6084836	0.1189073	0.1565691
JYP	1.009879	0.02167827	0.05465841
KalPa	0.461827	-0.09859855	0.4914075
Kärpät	0.9000703	-0.09560925	0.2908022
KooKoo	0.7371105	0.2823684	0.04380262
Lukko	0.588498	-0.05286361	0.2265276
Pelicans	0.7678637	0.170504	0.2288416
SaiPa	0.8014581	0.09066661	-0.05635294
Sport	0.7748334	0.3371994	0.1096989
Tappara	0.7112578	-0.1403626	0.1993329
TPS	0.9930246	-	0.04445176

Table 5.1: Relative strengths and Home advantages estimations

With the estimated relative attacking strengths, defending strengths and home advantages of all the teams, we can write our functions to predict goals of each team in a single game, calculate probabilities of winning for each team in a game, estimate teams' chances to win the title, and solve the maximizing profits of the one-thousand-euro-budget, with the same method we used in original model (We still used 1000000 loops in the simulation). The table 5.2 showed the new estimations of teams' winning chances and the new optimal betting portfolio.

The betting allocation maximization resulted table 5.3, with the expected value of 1281.03 €.

Team	Chance	Team	Chance
Ässät	0.35%	HIFK	9.405%
JYP	11.27%	KalPa	1.983%
Kärpät	45.86%	SaiPa	0.55%
Tappara	10.2%	TPS	20.38%

Table 5.2: Teams' estimated chances of winning the title

Team	Bet(€)	Team	Bet(€)
Ässät	3.5	HIFK	94.05
JYP	112.7	KalPa	19.8
Kärpät	458.6	SaiPa	5.5
Tappara	102	TPS	203.8

Table 5.3: Teams' betting allocation

Appendix

Descriptive graphs

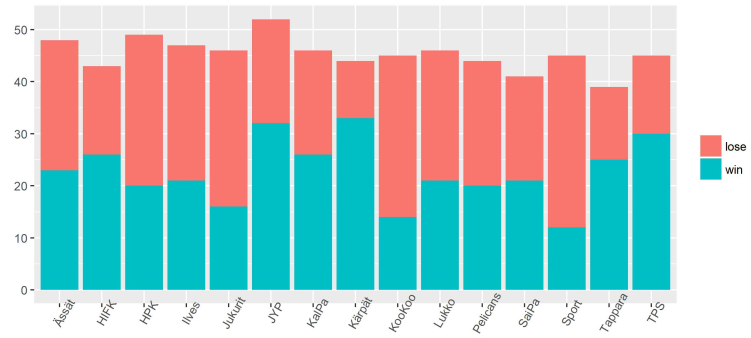

Figure 5.1: Numbers of each team's wins and loses in non-OT matches

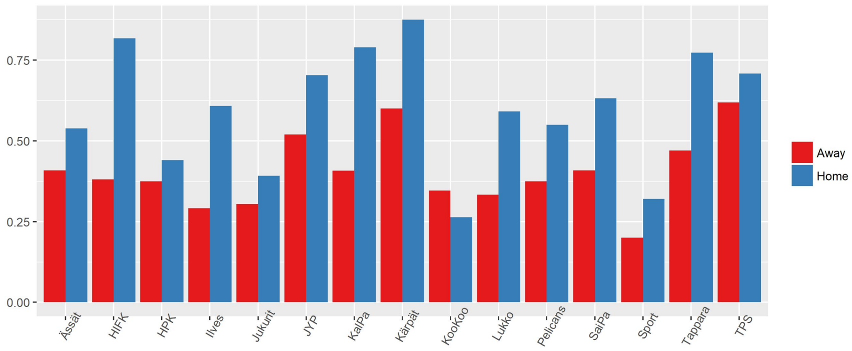

Figure 5.2: Teams' proportion of winnings at home and away

Source code

Model teams' relative strengths

```
library(dplyr)
df <- read.csv('RegularSeasonData.csv', fileEncoding = 'latin1')
teams = sort(unique(df$home))

df<- df%>% mutate(h = ifelse(ot==1,pmin(h,v),h), v = ifelse(ot==1,pmin(h,v),v))

df1 = df %>% select(home, visitor, h) %>%
rename(Attacker = home, Defender = visitor, Goals = h) %>% mutate(Home = 1)
df2 = df %>% select(home, visitor, v) %>%
rename(Attacker = visitor, Defender = home, Goals = v) %>% mutate(Home = 0)
data = df1 %>% rbind(df2)
library(dummies)
data2 = dummy.data.frame(data,names = c("Attacker","Defender"), sep="_")
glm_model = glm(data=data2, formula = Goals ~ -1+.-Defender_TPS, family=poisson())
```

Predicting goals scored by teams

```
coefs = glm_fixed$coefficients
coefs = c(coefs[1:29],0,coefs[30])
coefs = matrix(coefs)
predict_goals = function(home_team, away_team){
  home_id = which(teams==home_team)
  away_id = which(teams==away_team)
  new_match = matrix(rep(0,62),ncol=31)
  new_match[1,home_id] = 1
  new_match[1,away_id+15] = 1
  new_match[1,31] = 1
  new_match[2,away_id] = 1
  new_match[2,home_id+15] = 1
  predicted_goals = exp(new_match%*%coefs)
  return(predicted_goals)
}
```

Estimate chances of winning in a game

```
match_result = function(home_team, away_team, n_simu = 1000000, playoff=T){
  p_goals = predict_goals(home_team,away_team)
  hgoals = rpois(n_simu, p_goals[1])
  vgoals = rpois(n_simu, p_goals[2])
  hwin = sum(hgoals>vgoals)/n_simu
  vwin = sum(hgoals<vgoals)/n_simu
  if (playoff){
    wplayoff = hwin+vwin
    hwin = hwin/wplayoff
    vwin = vwin/wplayoff
    return(c(hwin,vwin,0))
  }
  draw = sum(hgoals==vgoals)/n_simu
```

```
  return(c(hwin,vwin,draw))
}
```

Likelihood of different outcomes for the entire playoff bracket

```
best_of_seven = function(high_team,low_team){
  pgoals = cbind(predict_goals(high_team,low_team),
   rev(predict_goals(low_team, high_team)))
  high_wins = 0
  low_wins = 0
  i = 1
  while (high_wins<4 & low_wins<4) {
    g1 = 0
    g2 = 0
    while (g1 == g2){
      g1 = rpois(1,pgoals[1,i])
      g2 = rpois(1,pgoals[2,i])
    }
    if (g1>g2){
      high_wins = high_wins + 1
    } else {
      low_wins = low_wins + 1
    }
    i = i + 1
    if (i == 3) {i=1}
  }
  if (high_wins >=4) {
    return(high_team)
  } else {
    return(low_team)
  }
}
the_champion = function(teams){
  q1_winner = best_of_seven(teams[1],teams[2])
  q2_winner = best_of_seven(teams[3],teams[4])
  q3_winner = best_of_seven(teams[5],teams[6])
  q4_winner = best_of_seven(teams[7],teams[8])
  s1_winner = best_of_seven(q1_winner,q4_winner)
  s2_winner = best_of_seven(q2_winner,q3_winner)
  champion = best_of_seven(s1_winner,s2_winner)
  return(champion)
}
final_teams = teams[c(8,1,15,12,14,7,6,2)]
scores = rep(0,8)
n_simu = 1000000
for (i in (1:n_simu)){
  new_champion = the_champion(final_teams)
  champ_index = which(final_teams==new_champion)
  temp = scores[champ_index] + 1
  scores[champ_index] = temp
```

```
}
chances = scores/n_simu
}
```

Solve 1000 € budget allocation

```
betting_odds = read.csv('BettingOdds.csv')
bet_odds = c()
for (team in final_teams){
  odd = betting_odds[which(betting_odds$Team==team),2]
  bet_odds = c(bet_odds,odd)
}
budget = 1000
max_porp = 0.5
nteam = length(final_teams)
A = matrix(0,nrow=nteam,ncol=nteam)
for (i in (1:nteam)){
  A[i,i] = 1
}
A = rbind(rep(1,nteam),A)
b = c(budget,rep(max_porp*budget,nteam))
f = chances*bet_odds
library(lpSolveAPI)
lp = make.lp(nrow(A),ncol(A))
for (c in (1:ncol(A))){
  set.column(lp, c, A[,c])
}
set.constr.type(lp,rep("<=",nteam+1))
set.rhs(lp,b)
set.objfn(lp,f)
lp.control(lp,sense='max')
solve(lp)
OptimalSolution <- get.variables(lp)
maxValue = get.objective(lp)
```

Another allocation

```
A = matrix(0,nrow=nteam,ncol=nteam)
for (i in (1:nteam)){
  A[i,i] = 1
}
A = rbind(rep(1,nteam),A)
b = c(budget,chances*budget)
f = chances*bet_odds
library(lpSolveAPI)
lp = make.lp(nrow(A),ncol(A))
for (c in (1:ncol(A))){
  set.column(lp, c, A[,c])
}
set.constr.type(lp,rep("<=",nteam+1))
set.rhs(lp,b)
```

```
set.objfn(lp,f)
lp.control(lp,sense='max')
solve(lp)
OptimalSolution <- get.variables(lp)
maxValue = get.objective(lp)
```

New model with different Home advantages

```
library(dplyr)
df.new <- read.csv('RegularSeasonData.csv', fileEncoding = 'latin1')
teams = sort(unique(df.new$home))

df.new<- df.new%>%
mutate(h = ifelse(ot==1,pmin(h,v),h), v = ifelse(ot==1,pmin(h,v),v))

df.new1 = df.new %>% select(home, visitor, h) %>%
rename(Attacker = home, Defender = visitor, Goals = h) %>% mutate(Home = Attacker)
df.new2 = df.new %>% select(home, visitor, v) %>%
rename(Attacker = visitor, Defender = home, Goals = v) %>% mutate(Home = "N/A")
data.new = df.new1 %>% rbind(df.new2)
library(dummies)
data.new = dummy.data.frame(data.new,names = c("Attacker","Defender","Home"), sep="_")
data.new = data.new[-47]
new_glm_model = glm(data=data.new, formula = Goals ~ -1+.-Defender_TPS, family = poisson())
new_glm_model
```