NTNU
Norwegian University of
Science and Technology

Valuing Individual Player Involvements in Norwegian Association Football

Vegard Rødseth Bjertnes
Olav Nørstebø
Eirik Vabo

Industrial Economics and Technology Management
Submission date: June 2016
Supervisor: Stein-Erik Fleten, IØT
Co-supervisor: Magnus Stålhane, IØT
Lars Magnus Hvattum, HiMolde

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Preface

This Master's Thesis is written as a concluding part for achieving a Master of Science at the Norwegian University of Science and Technology. The degree specialisation is Investment, Finance and Financial Management at the Department of Industrial Economics and Technology Management. Two of the authors have technical background in Mechanical Engineering and one author in Energy and Environmental Engineering. All authors have specialised in Empirical and Quantitative Methods in Finance. The study has been conducted over the spring of 2016.

This work started as an initiative from Rosenborg Ballklub, a Norwegian top division association football team. They have been inspired by the focus on statistical analysis in football in recent years, and wants to shift towards making decisions supported by quantitative methods in combination with qualitative considerations. This enquiry from Rosenborg was welcomed at the Department of Industrial Economics and Technology Management and resulted in this thesis.

Trondheim, 10.06.2016

Vegard Bjertnes | Olav Nørstebø | Eirik Vabo

Summary

Evaluation of player performance in association football has to a large extent been limited to subjective opinions and simple, easily observable parameters such as goals, passing accuracy and ball recoveries. In this thesis, three models with the goal of objectively rating players by looking at individual actions are presented and documented.

An expected goals (xG) model is developed by looking at 13,440 shots attempted in 480 football matches in the Norwegian top division, Tippeligaen. The likelihood of scoring is estimated using binary logistic regression with ten explanatory variables. This model is used as a foundation to evaluate the performance of players with regard to their shot efficiency.

Two variations of a zero-sum two-agent Markov game model based on matches from two seasons are developed in order to evaluate other actions than shots. The large state spaces contain three contextual parameters: time period, match status and manpower difference. In addition, different field zones are used in the definition of a state. Reinforcement learning through a Q-function is applied to learn the value of each state and state-action pairs. Players are rated by their impact per 90 minutes played, and results are presented as top 10 lists of players in each position.

The reliability of the models is assessed by looking at correlations across seasons. Validity of the two Markov models are examined through comparisons to two subjective player ratings and one provider of market value estimates of players. The models are also tested out of sample. Areas in which extensions of the three models seem possible or appropriate are addressed and highlighted.

Sammendrag

Evaluering av spillere i fotball har lenge vært begrenset til subjektive meninger og enkle, lett observerbare parametre som skårede mål, pasningssikkerhet og ballgjenvinninger. Tre modeller er utviklet og dokumentert i denne masteroppgaven, med mål om å evaluere spillere objektivt.

En expected goals (xG)-modell er utviklet ved å se på 13,440 skuddforsøk fra 480 fotballkamper spilt i Tippeligaen. Sannsynligheten for å skåre på et skudd er estimert ved hjelp av binær logistisk regresjon med ti forklaringsvariable. Denne modellen er videre brukt til å evaluere spilleres skudd-effektivitet.

To versjoner av en nullsum to-agents Markov game-modell er utviklet basert på kamper fra to sesonger for å evaluere andre involveringer enn skudd. Et stort sett av tilstander inneholder tre kontekstvariable: tidsperiode, kampstatus og forskjell i antall spillere på banen. I tillegg fotballbanen delt inn i soner som også en del av en tilstand. Maskinlæringsteknikken forsterkende læring er brukt gjennom en Q-funksjon til å lære verdien av hver tilstand, og par av tilstander og handlinger. Spillere er rangert etter hvor stor påvirkning de hadde på kamper per 90 spilte minutter, og resultatene er presentert som topp 10-lister for hver posision.

Påliteligheten til modellene er vurdert på grunnlag av korrelasjoner på tvers av sesonger. Validiteten av de to Markov-modellene er vurdert gjennom sammenligning med to subjektive spillerrangeringer og estimater på markedsverdien til spillerene. Modellene er også testet på data utenfor utvalget de er lært på. Områder der de tre modellene kan utvides og forbedres er presentert i slutten av oppgaven.

Acknowledgments

First we want to thank our supervisors Magnus Stålhane, Lars Magnus Hvattum and Stein-Erik Fleten. Their commitment and interest in our work have motivated us and made the progress steady and efficient. Without their quick and thorough feedback this thesis would never been a reality.

We also want to thank Rosenborg Ballklub for providing us the opportunity to work with statistical analysis and problem solving within an area we all have a passion for. The kind welcome received from Rosenborg Ballklub and the close collaboration with Stig Inge Bjørnebye made this project a unique experience.

Finally we want to thank Opta Sports, Kenneth Wilsgård from Norsk Toppfotball and The Planning, Optimization and Decision Support Group at Molde University College for providing us with the data used in this thesis.

Table of Contents

Preface i

Summary iii

Sammendrag iv

Acknowledgments v

Table of Contents ix

List of Tables xiii

List of Figures xvi

1 Introduction 1
- 1.1 Motivation . . . 2
- 1.2 Research Questions . . . 3
- 1.3 Limitations . . . 3
- 1.4 Report Structure . . . 4

2 Literature Review 5
- 2.1 Sports Analytics in Football . . . 5
- 2.2 Expected Goals (xG) . . . 8
- 2.3 Markov Models in Sports Analytics . . . 9
- 2.4 Evaluation of Individual Player Performance . . . 10
- 2.5 Financial Valuation of Football Players . . . 12

3 Basic Theory 15
- 3.1 Binary Logistic Regression . . . 15
- 3.2 Markov Games . . . 16
  - 3.2.1 Definition . . . 17
  - 3.2.2 Policies and Value Functions . . . 17
  - 3.2.3 Value of a Policy . . . 18

4 **Experimental Setup** **19**
4.1 Data 19
4.2 Expected Goals Model 22
4.2.1 Model Building Approach 22
4.2.2 Explanatory Variables 23
4.2.3 Application of the Expected Goals Model 25
4.3 Markov Game Model 1 25
4.3.1 State Space: Context Variables and Field Zones 25
4.3.2 State Transitions: Actions 27
4.3.3 Reward Function and Value Iteration Algorithm 29
4.3.4 Valuing Individual Player Actions 31
4.4 Markov Game Model 2 32
4.4.1 State Space: Context Variables and Field Zones 33
4.4.2 State Space: Actions 33
4.4.3 State Transitions 35
4.4.4 Reward Functions and Value Iteration Algorithm 36
4.4.5 Valuing Individual Player Actions 36
4.5 Model Evaluation 37

5 **Evaluating Results** **39**
5.1 Expected Goals Model 39
5.1.1 Reliability 41
5.1.2 Validity 42
5.1.3 Results 43
5.2 Markov Game Model 1 45
5.2.1 Reliability 46
5.2.2 Validity 46
5.2.3 Results 49
5.3 Markov Game Model 2 55
5.3.1 Reliability 55
5.3.2 Validity 56
5.3.3 Results 58
5.4 Case Studies 62
5.4.1 Mike Jensen 63
5.4.2 Martin Ødegaard 67
5.5 Reliability Out of Sample 69
5.6 Evaluation of Research Questions 71

6 **Conclusion** **75**

7 **Recommendations for Further Research** **77**

Bibliography **79**

Appendix **89**

A Tables from Tippeligaen **89**
A.1 2014 89
A.2 2015 92

B Results from Expected Goals Model **95**
B.1 2014 96
B.2 2015 98

C Results from Markov Game Model 1 **101**
C.1 2014 102
C.2 2015 105

D Results from Markov Game Model 2 **109**
D.1 2014 110
D.2 2015 113

E Markov Game Models: Validation from Team Evaluations **117**

F Case Studies **119**
F.1 Mike Jensen 119
F.2 Martin Ødegaard 121

G Results for 2016 **123**
G.1 Expected Goals Model 123
G.2 Markov Game Model 1 124
G.3 Markov Game Model 2 125

List of Tables

2.1 EA Sports Player Performance Index EPL 2015/2016 11

4.1 Simple descriptive statistics for the dataset 20
4.2 xG Model: Structure of shot data 23
4.3 xG Model: Variables considered in binary logistic regression 24
4.4 Model 1: Context variables 26
4.5 Model 1: $\mathcal{A}$, action set for the team in possession of the ball 27
4.6 Model 1: $\mathcal{O}$, action set for the team not in possession of the ball 27
4.7 Model 1: Structure of data 28
4.8 Model 2: Context variables 33
4.9 Model 2: Overview of the outcome parameter for the actions 34
4.10 Model 2: Pass values for the context $TP = 3, MS = 1, MD = 0, H = 1$ and $Z = 3$ 37

5.1 xG Model: Parameters for regular shots 40
5.2 xG Model: Parameters for free kicks 40
5.3 xG Model: Correlation coefficients of performance measures from 2014 to 2015 42
5.4 xG Model: All players scoring more than 8 goals in Tippeligaen 2015, ranked by G/xG 44
5.5 Model 1: Correlations to VG-børsen, Altomfotball and Transfermarkt in 2014 . 48
5.6 Model 1: Correlations to VG-børsen, Altomfotball and Transfermarkt in 2015 . 49
5.7 Model 1: Top 10 forwards, Tippeligaen 2015. Average total value = 0.1124, Minimum total value = -0.1248 50
5.8 Model 1: Top 10 wingers, Tippeligaen 2015. Average total value = 0.1124, Minimum total value = -0.1248 50
5.9 Model 1: Top 10 attacking midfielders, Tippeligaen 2015. Average total value = 0.1124, Minimum total value = -0.1248 51
5.10 Model 1: Top 10 central midfielders, Tippeligaen 2015. Average total value = 0.1124, Minimum total value = -0.1248 52
5.11 Model 1: Top 10 full backs, Tippeligaen 2015. Average total value = 0.1124, Minimum total value = -0.1248 53
5.12 Model 1: Top 10 centre backs, Tippeligaen 2015. Average total value = 0.1124, Minimum total value = -0.1248 54
5.13 Model 2: Correlations between VG-børsen, Altomfotball, Transfermarkt and Model 1 in 2014 57

5.14 Model 2: Correlations between VG-børsen, Altomfotball, Transfermarkt and Model 1 in 2015 . . . 58
5.15 Model 2: Top 10 forwards, Tippeligaen 2015. Average total value = 0.4608, Minimum total value = 0.0059 . . . 59
5.16 Model 2: Top 10 wingers, Tippeligaen 2015. Average total value = 0.4608, Minimum total value = 0.0059 . . . 59
5.17 Model 2: Top 10 attacking midfielders, Tippeligaen 2015. Average total value = 0.4608, Minimum total value = 0.0059 . . . 60
5.18 Model 2: Top 10 central midfielders, Tippeligaen 2015. Average total value = 0.4608, Minimum total value = 0.0059 . . . 61
5.19 Model 2: Top 10 full backs, Tippeligaen 2015. Average total value = 0.4608, Minimum total value = 0.0059 . . . 61
5.20 Model 2: Top 10 centre backs, Tippeligaen 2015. Average total value = 0.4608, Minimum total value = 0.0059 . . . 62
5.21 Model 1: Impact values for Mike Jensen 2015 . . . 63
5.22 Model 2: Impact values for Mike Jensen 2015 . . . 64
5.23 Model 1: Impact values for Martin Ødegaard 2014 . . . 67
5.24 Model 2: Impact values for Martin Ødegaard 2014 . . . 67

A.1 League Table Tippeligaen 2014 . . . 89
A.2 Top 20 players with respect to goals 2014 . . . 90
A.3 Top 20 players with respect to assist 2014 . . . 91
A.4 League Table Tippeligaen 2015 . . . 92
A.5 Top 20 players with respect to goals 2015 . . . 93
A.6 Top 20 players with respect to assist 2015 . . . 94

B.1 xG Model: Correlation matrix . . . 95
B.2 xG Model: All players scoring more than 8 goals in Tippeligaen 2014, ranked by G/xG . . . 96
B.3 xG Model: Tippeligaen 2014 teams rated by xG difference . . . 97
B.4 xG Model: All players scoring more than 8 goals in Tippeligaen 2015, ranked by G/xG . . . 98
B.5 xG Model: Tippeligaen 2015 teams rated by xG difference . . . 99

C.1 Model 1: Top 10 players in offensive positions, Tippeligaen 2014. Average value = 0.1118, minimum value = -0.2324 . . . 102
C.2 Model 1: Top 10 players in defensive positions, Tippeligaen 2014. Average value = 0.1118, minimum value = -0.2324 . . . 103
C.3 Model 1: League table 2014 based on the model compared to the real table . . . 104
C.4 Model 1: Top 10 players in offensive positions, Tippeligaen 2015. Average value = 0.1124, minimum value = -0.1248 . . . 105
C.5 Model 1: Top 10 players in defensive positions, Tippeligaen 2015. Average value = 0.1124, minimum value = -0.1248 . . . 106
C.6 Model 1: League table 2015 based on the model compared to the real table . . . 107

D.1 Model 2: Top 10 players in offensive positions, Tippeligaen 2014. Average value = 0.4668, minimum value = 0.0118 . . . 110

D.2 Model 2: Top 10 players in defensive positions, Tippeligaen 2014. Average value = 0.4668, minimum value = 0.0118 111
D.3 Model 2: League table 2014 based on the model compared to the real table . . 112
D.4 Model 2: Top 10 players in offensive positions, Tippeligaen 2015. Average value = 0.4608, minimum value = 0.0059 113
D.5 Model 2: Top 10 players in defensive positions, Tippeligaen 2015. Average value = 0.4608, minimum value = 0.0059 114
D.6 Model 2: League table 2015 based on the model compared to the real table . . 115

E.1 Model 1 and Model 2: p-values from the ADF test, Kendall's tau and Spearman's rho in 2015 118

G.1 xG Model: Top 20 players 2016 with respect to G/xG 123
G.2 Model 1: Top 20 players in all positions 2016. Average tota value = 0.0752, minimum total value = -0.4143 124
G.3 Model 2: Top 20 players in all positions 2016. Average total value = 0.4541, minimum total value = 0.0356 125

List of Figures

4.1 Data structure of F24 feed . . . 20
4.2 Illustration of passing data in terms of pass counts and passing accuracies . . . 21
4.3 Illustration of shot data in terms of conversion rates . . . 22
4.4 xG Model: Illustration of angle and length variables for a shot. Red lines show the length and orange arrows show the angle. Angle measured in radians . . . 23
4.5 Model 1: Field zones . . . 26
4.6 Model 1: Definition of a state . . . 27
4.7 Model 1: Illustration of states and transitions. Example from Tromsø - Sarpsborg 08 in 2015 . . . 28
4.8 Model 1: Zone values of state space for illustrative purposes. $TP = 1, MD = 0$ and $GD = 0$ . . . 31
4.9 Model 2: Field zones . . . 33
4.10 Model 2: Definition of a state . . . 34
4.11 Model 2: Illustration of states and transitions. Example from Tromsø - Sarpsborg 08 in 2015 . . . 35

5.1 xG Model: Scatter plot of the G/xG of all players with the correlation coefficient between the 2014 and 2015 season . . . 42
5.2 xG Model: ROC curve for the two shot models . . . 43
5.3 Model 1: Scatter plot and correlation between the two seasons . . . 46
5.4 Model 1: Scatter plots and correlation with Altomfotball . . . 47
5.5 Model 1: Scatter plots and correlation with VG-børsen . . . 47
5.6 Model 1: Scatter plots and correlation with Transfermarkt . . . 48
5.7 Model 2: Scatter plot and correlation between the two seasons . . . 55
5.8 Model 2: Scatter plots and correlation with Altomfotball . . . 56
5.9 Model 2: Scatter plots and correlation with VG-børsen . . . 56
5.10 Model 2: Scatter plots and correlation with Transfermarkt . . . 57
5.11 Model 1: Box plot for Mike Jensen in 2015 excluding shots . . . 63
5.12 Model 2: Box plot for Mike Jensen in 2015 . . . 64
5.13 Model 2: Comparison of Mike Jensen vs league for 2015 . . . 65
5.14 Model 1: Box plot for Martin Ødegaard in 2014 excluding shots . . . 67
5.15 Model 2: Box plot for Martin Ødegaard 2014 . . . 67
5.16 Model 2: Comparison of Martin Ødegaard vs league for 2014 . . . 68
5.17 xG Model: Scatter plot and correlation of G/xG between 2015 and 2016 . . . 69

5.18 Model 1: Scatter plot and correlation between 2015 and 2016 70
5.19 Model 2: Scatter plot and correlation between 2015 and 2016 70

E.1 Model 1 and Model 2: Real table position compared to estimated each matchday 117

F.1 Model 1: Box plot for Mike Jensen 2014 including shots 119
F.2 Model 1: Box plot for Mike Jensen 2014 excluding shots 120
F.3 Model 2: Box plot for Mike Jensen 2014 120
F.4 Model 1: Box plot for Mike Jensen in 2015 including shots 121
F.5 Model 1: Box plot for Martin Ødegaard 2014 including shots 121

Chapter 1

Introduction

Analysing the competition has for a long time been an important task in the world of association football (throughout this thesis referred to as football). Typical applications of such analyses are to get information and insights on opponents before, during and after matches, to keep track of performance of own squad and to scout for talent. Analyses have previously been qualitative in nature. However, quantitative data analysis has gained momentum in recent years, and several companies, like Opta Sports and Prozone, have made a business out of collecting fine grained data during football matches. This data is sold to customers such as football clubs, betting companies, TV broadcasters and newspapers. In combination with the steady increase in computational power, these data sets create new possibilities for statistical analysis in sports, commonly referred to as sports analytics. Football clubs are attempting to lever the potential for valuable information that data and statistical analysis can provide, and several of the biggest clubs in the world try to exploit this potential to create and sustain a competitive edge. The onset of sports analytics is also apparent amongst smaller clubs who are not able to compete financially with the larger ones. They are trying to close the gap in resources through innovation within analysis and player logistics. Good examples of such clubs are Brentford FC in the English Championship and FC Midtjylland in the Danish Superliga. Brentford FC has moved from English League Two (fourth tier in England) in 2009, to almost being promoted to the English Premier League in the 2014/2015 season (losing to Midlesborough in the playoffs). In the 2014/2015 season FC Midtjylland won the Danish Superliga for the first time in club history. Both these clubs claim that sports analytics has been an important part of their culture, and an important factor in their recent success.

Several areas of application are possible for statistical analysis on data from association football. The most popular ones in academic research include identifying key performance indicators (KPIs), analysing the home field advantage, finding the effect of situational variables and investigating set-pieces, crosses, yellow and red cards and referee bias. Such analyses have been performed for several years. More recent analyses focus on fine grained tracking data and evaluation of individual players and their involvements on the field.

Evaluation of player actions on the field has for many years been quite unsophisticated. It has mostly been based on simple parameters, like goals scored, the number of assists and passing accuracy for attacking players and goals conceded, duels won and number of interceptions for defending players. This approach has several drawbacks. First of all, the context of where the players perform actions on the field is not incorporated, and players are evaluated based on absolute measures rather than compared to expectations. It is a better accomplishment to score on a well placed free kick from 25 meters than to convert a penalty kick. Similarly, for defensive players, an interception should be valued higher if it prevents a good scoring opportunity.

In The Numbers Game (Anderson and Sally (2013)), the authors emphasise two other factors that the traditional valuation does not take into account when valuing football players. The first is the importance of the goals. That is, how many points that are ensured by the goals a player scored. A match winning goal is more valuable then scoring goal number five late in the game in a 5-0 victory. Secondly, defensive players seems to be undervalued relative to offensive players. Of the 20 most expensive transfers in football history, only one is a defender (David Luiz, 50 million GBP, from Chelsea to Paris Saint-Germain in 2014). The rest are either offensive midfielders or attackers. Data analysis shows that this is an irrationality and a market inefficiency, as it is stated in Anderson and Sally (2013):

> "Through the ten Premier League seasons from 2001/2002 to 2010/2011 it appears that not conceding any goals during a match ensured almost 2.5 points per match. Compared with scoring one time, which gives approximately one point per match, it is twice as valuable to prevent the opponent scoring a goal. In fact, conceding only one goal during a match gives the team 1.5 points on average, approximately 30 % more than the value of scoring only one goal."

As mentioned, quantitative analyses has gained momentum in recent years. The increasing body of research has to a certain extent rectified several biases and challenged many irrationalities in the game of football. One popular concept that has been useful in this effort is expected goals (xG). This is a concept that incorporates the context of the player by including several variables affecting a shot. It assigns a likelihood for converting a shot into a goal. In this way, the model can provide a more objective foundation for evaluation of players by looking at how they performed compared to expectations. However, xG models only look at shots, and their best application is therefore to evaluate the efficiency of primarily attacking players. xG models in their traditional form are not able to assign values to actions that leads to a shot. For example, the assist to a goal could be a brilliant penetrating pass, but the passing player will not get a share of the xG value in such models. In recent years, efforts have been made in order to evaluate all player involvements, not only shots.

1.1 Motivation

Valuing player involvements in association football can be a cumbersome and difficult task. Besides favouring players in offensive positions, player evaluations are very often influenced by human bias. The main objective in this thesis is to extend the existing body of research on football with models that can objectively evaluate individual player involvements. The obtained values can form a basis for player ratings free from human bias, which could serve several purposes for clubs. A substantial amount of resources is committed in scouting for talent, and the objective player ratings can be used to make short lists of players that are worth a closer examination, both domestically and internationally. This could make the process of scouting more accurate and less time consuming, which can have a big monetary impact in the long run. They can also be used to scout own squads in order to monitor the performance of the players and see which players are performing on the desired level.

Another motivation is to take on challenges regarding modelling of contextual parameters in football and the rating of defensive players. This is, to the authors knowledge, only addressed to a limited extent, and little work has previously been done on Norwegian football.

An objective and data driven player evaluation can also create possibilities for a financial analysis, useful for coaches, directors, owners and other stakeholders in football clubs. They can utilise such models either for evaluating contracts or values of players on the transfer market. They operate in an environment where competition for the best players and the best talent is increasingly fierce. Crucial decisions regarding player logistics and transfers are being made every day. Getting it right more often than competitors can create competitive advantages, and may be an important step to ensure success in the long run.

1.2 Research Questions

The purpose of this thesis can be synthesised as incorporating contextual parameters and situational variables when assessing player performance, developing a player efficiency rating, and rate outfield players based on all involvements. For clarity and guidance, a set of formal research questions (RQs) are developed. They are used to guide the work done in this thesis, and are addressed in the discussion of results and conclusion regarding how successful the works in this thesis have been. The three formal research questions are formulated as:

RQ 1 Is it possible to create a statistically significant xG model that assesses the quality of all shots in Tippeligaen in order to evaluate the efficiency of primarily offensive players?

RQ 2 Is it possible to create a Markov game model for football that is able to evaluate all player involvements in a match and rate players over the course of a season?

RQ 3 Is it possible to reveal undiscovered talent or identify under- and overvalued players based on the evaluation of individual player involvements?

1.3 Limitations

To conduct a relevant and applicable analysis of individual player performance is far from a trivial task. Abilities of clubs to gather the right kind of players with the correct set of skills and complementary characteristics is crucial in order to succeed. In football, as a collective sport, the total ability of a team is greater than the sum of the individual parts. These complementary effects are not accounted for in the models developed in this thesis. Neither are effects that come from team configurations and stability of playing style and strategy. Performing on a high level is possibly easier when the coach is the same across many seasons.

Another limitation in this thesis is that all players are evaluated on the same basis. Obviously, different skills or characteristics are important for a defender and a forward. The nature of the data set used in this thesis is such that it makes it harder to appreciate defensive attributes than offensive. This makes the models biased towards attackers and offensive involvements.

Only data from two seasons of football is used. Although the two seasons include a high number of events, it would have been preferable to have more data in order to evaluate the validity and consistency of the models developed. In addition, some events have been removed by the authors, which has resulted in a further reduction of the data set. These events are mainly events which does not involve the ball.

Because of the novelty of modelling football as a Markov game, the validity of the models in this thesis is hard to assess. The only appropriate comparable sources found, are subjective player ratings and estimated market values. The authors admits that using only correlation coefficients between these benchmarks and the models as a basis for validation, is questionable.

1.4 Report Structure

Chapter 1 serves as an introduction and presents the motivation, three research questions and the limitations considered in this thesis.

Chapter 2 gives an overview of the existing literature on sports analytics, where the main focus is on relevant works regarding xG models, Markovian approaches and player evaluation.

Chapter 3 describes the theoretical background for the models developed in this thesis.

Chapter 4 introduces the data set, and explains the procedures for building the models in this thesis.

Chapter 5 presents the reliability, validity and results from the different models, which are further discussed with regard to the research questions.

Chapter 6 highlights the major findings, and conclusions are drawn regarding the merits of this research.

Chapter 7 suggests extensions of the models, and addresses the potential for further research.

Chapter 2

Literature Review

The focus on sports analytics related to decision making has for many years experienced a significant increase, also in the world of football. Important stakeholders like club owners, managers and analysts have realised the importance of using data driven arguments when analysing one of the most complex and dynamic sports in the world. This is also reflected in the growth of the body of research on football in recent years. Due to the prematurity of this research field, the first part of this chapter gives an overview of important research. The second part investigates the work performed on the concept of expected goals (xG), and the third part looks at Markov game models used within sports analytics. The fourth part focuses on research performed to evaluate individual player involvements and player valuations. Analyses on valuing football players financially is the focus of the fifth and last part of this chapter.

2.1 Sports Analytics in Football

It is hard to define exactly when the field of sports analytics was conceived, but Charles Reep is by many viewed as a pioneer within football analytics. In Reep and Benjamin (1968) and Reep et al. (1971) the passing patterns of English Premier League clubs are analysed, and the research concludes that "It seems, however, that chance does dominate the game and probably most similar ball games". It is now widely recognised that luck is a considerable element during a football match, but also that regularities can be revealed and influenced. From the beginning of sports analytics, a wide array of approaches on research topics and methods has been used in the quest for revealing these regularities, and by examining 140 journals in operations research, statistics, applied mathematics and applied economics, Coleman (2012) defines the field of sports analytics to be "sizable and growing".

One distinctive research topic that is pronounced in the investigated literature is that of performance analysis and finding key performance indicators (KPIs). The main focus of such analyses is team performance, and KPIs like passes, ball possession and ball recoveries are popular subjects. Passing is a cardinal event in a football match, and it is interesting to see that different analyses give different results. While Reep and Benjamin (1968) and Wright et al. (2011) find that short passing sequences are favourable, Hughes and Franks (2005) and McHale and Scarf (2007) find that longer passing sequences are preferable. These findings further emphasise the notion that one should take care when interpreting such analyses. Another KPI that has gotten a lot of attention is ball possession. There are studies performed that indicate ball possession as a statistical significant parameter that affects the outcome of matches (Lago-Ballesteros and

Lago-Peñas (2010), Castellano et al. (2012), Collet (2013), Bradley et al. (2013), Bjertnes et al. (2015)). However, ball possession is a complex parameter, which is affected by many other parameters. Lago and Martín (2007) finds that 66 % of the variation in ball possession in the 2003/2004 La Liga is determined by the situational variables team strength, match location and match status. Two other KPIs that are given some attention are ball recoveries (Barreira et al. (2014)) and penalty area entries (Ruiz-Ruiz et al. (2013)), which both reached statistical significance.

Situational variables has gotten a lot of attention by the sports analytics community in recent years. Lago-Peñas (2012) sums up the concept of situational variables concisely:

> "Given that soccer is dominated by strategic factors, it is reasonable to suggest that situational variables of match status (i.e. whether the team is winning, losing or drawing), strength of opposition (strong or weak), and match location (i.e. playing at home or away) may somehow influence the teams' and players' activities. These situational variables need to be analysed in depth to understand their influence in team sports."

It is widely recognised in both the football and the sports analytics community that these variables influence other in-match events. Several articles support this, and amongst the parameters that seem to be affected are technical behaviour (Taylor et al. (2008)), distance covered by single players (Lago et al. (2010)), ball possession (Lago-Peñas and Dellal (2010)), attacking patterns (Machado et al. (2014)) and ball recovery dynamics (Almeida et al. (2014)). Thus, there is substantial evidence that situational variables should be accounted for when analysing other parameters. One situational variable that has gotten especially much attention is the home field advantage. Within this narrow scope, most analyses support the notion of home field advantage and its affect on other parameters (Clarke and Norman (1995), Brown Jr et al. (2002), Bray et al. (2003), Carmichael and Thomas (2005)). Such parameters include team strategy, shot statistics and disciplinary behaviour (Seçkin (2006), Seçkin and Pollard (2008)). Goumas (2014) also shows that travelling time and length negatively affect the away team, which enhances the home field advantage. By using six seasons from 72 countries, Pollard et al. (2008) summarises the widespread the home field advantage neatly:

> "In Europe, home advantage in the Balkan countries, especially Bosnia and Albania, is much higher than average. It is generally lower than average in northern Europe, from the Baltic republics, through Scandinavia to the British Isles. In South America, home advantage is high in the Andean countries and lower elsewhere, especially in Uruguay."

The displacement patterns and tactical patterns of a football team are also distinct features which have been analysed by several researchers. Displacement patterns refer to how a team is geometrically organised on the field, while tactical patterns can refer to how a team passes, how they put pressure on the opponent and which playing style they prefer. Much of the analyses on displacement patterns are quite new due to the requirement of advanced tracking data on the players in order to obtain sound analyses. Several analyses find that the organisational shape and structure of the team are significantly different dependent on the type of play (Moura et al. (2012), Duarte et al. (2012), Duarte et al. (2013), Barreira et al. (2013)). Bialkowski et al. (2014a) uses spatiotemporal data (containing details on both space and time) to perform automatic formation analysis to investigate the difference in team behaviour when playing home and away. This is further used in Bialkowski et al. (2014b) to identify the uniqueness in team passing, movement and interaction. Tactical patterns have also been investigated, and the analyses find that attacks differs dependent on whether the team is winning or losing (Jankovic et al. (2011), Niu et al. (2012)). In recent years, some pioneering work within analysing passing strategies have been done. Gyarmati et al. (2014) uses automatic extraction of passing strategies to search for a unique passing style in football, and reveals for instance that Barcelona's tiki-taka "does not consist of uncountable random passes but rather has a precise, finely constructed structure". Also in analyses of tactics, spatiotemporal data is used. Niu et al. (2012) applies trajectory extraction to analyse football tactics, and finds six distinct attack patterns. Bojinov and Bornn (2016) studies team tactics in the English Premier League. This analysis investigates the ability of each team to control the ball in offensive zones, and their ability to disrupt the opponents flow of play, which induce possibilities for tactical considerations by managers.

Surprisingly, little research performed on free kicks and corner kicks was found (Bray and Kerwin (2003), Taylor et al. (2005), Alcock (2010), De Baranda and Lopez-Riquelme (2012)). Penalties on the other hand, being a monumental occurrence in a football match, are a more popular subject for analysis. Lopes et al. (2008) and Jordet and Hartman (2008) investigate the psychological aspect of a penalty and find a link between player behaviour and outcomes of penalties. Other analyses are more quantitative in nature. Bakkerode and Koning (2014) reveals that the number of penalty kicks and the outcome are positively affected by the home field advantage, Misirlisoy and Haggard (2014) finds that the diving direction of the goalkeeper is affected by the preceding penalty and Noël et al. (2015) uncovers three variables predicting the penalty kick strategy: attention to goalkeeper, the run-up and the kicking technique. Other important in-match events that have been analysed for their impact on football match results are crosses (Kumar et al. (2013), Orth et al. (2014), Yamada and Hayashi (2015), Vecer (2014)), yellow cards (Unkelbach et al. (2008), red cards (Ridder et al. (1994), Vecer et al. (2009), Mechtel et al. (2011)) and referee bias (Buraimo et al. (2010), Buraimo et al. (2012), Reilly and Witt (2013), Constantinou et al. (2014), Goumas (2014)).

This thesis utilises data from Tippeligaen, and it is therefore interesting to investigate what research related to Norwegian football that exists. Brillinger (2007) is a piece that looks at the 2003 season of Tippeligaen, and based on the number of goals estimates the likelihood for winning, drawing or loosing. Tenga et al. (2010a) applies logistic regression on team possession from Tippeligaen 2004. The purpose of this study is to examine what effects match location has on playing tactics. The data set used in this article is further exploited in Tenga et al. (2010b), Tenga et al. (2010c) and Tenga et al. (2010d). These three analyses also apply logistic regression. The first one investigates the effect playing tactics has on entries into the a defined zone near the opponents goal, while the second one looks at the effect playing tactics has on goal scoring. The third one assesses the effectiveness of teams by studying the relationship between broader measures (scoring opportunities and offensive zones possessions) and goals scored. Halvorsen et al. (2013) presents a case study including a video sensor system installed at Alfheim stadion. However, this system seems to be more appropriate for qualitative analysis than as a basis for quantitative modelling. Sæbø (2015) looks at several leagues in his analysis, including fixtures from six seasons (2009 - 2014) of Tippeligaen. This analysis uses the financial standings of the clubs as the point of origin, and valuates players based on how they affect the revenue of the club. This work is synthesised in Sæbø and Hvattum (2015).

To the authors knowledge, the specialisation project Bjertnes et al. (2015) is the first analysis that utilises data consisting of descriptive match statistics in order to measure team performance on a quantitative basis. This work applies linear and logistic regression in order to determine which match statistics that were prominent in Tippeligaen 2015, and how they affected the likelihood for winning, drawing or losing a match. Although having a focus on Norwegian football, none of the above analyses examine individual player performance based on in-match performance statistics. Thus, this thesis offers new insights on Norwegian football by focusing on the performance of individual players.

2.2 Expected Goals (xG)

In recent years, the concept of xG has gotten a lot of attention in sports analytics communities. The largest community is blogs and fan pages online, where a variety of models are presented. However, also in academic circles, xG has existed for some years within several sports. In both NBA (National Basketball Association) and NHL (National Hockey League) this concept has been embraced as a useful metric. xG is a metric that incorporates contextual parameters when evaluating a shot. Commonly used parameters are distance to goal, angle on goal, which body part that was used when shooting, if the assist was air borne or not and the number and proximity of defenders. Hence, in contrast to more common metrics like the number of shots or shooting accuracy, the xG metric also evaluates the quality of the shots.

xG has been the topic of several research papers. One that uses the concept explicitly is Macdonald (2012). This article analyses NHL teams and players over three seasons, and develops an xG model using ridge regression with goals, shots, hits, hits against and faceoffs as explanatory variables. When compared to simpler models, this model shows a higher correlation between actual and predicted goals. It also has the lowest mean squared error, and is therefore considered more appropriate for evaluating individual players.

The research paper discussed above uses discrete in-match events, but another possibility is to use fine grained tracking data. The provision of such data has increased in recent years and works have been performed to evaluate the feasibility of utilising such data for football analysis (Kim et al. (2011), Bialkowski et al. (2014c), Gudmundsson and Wolle (2014)), assessing football tactics (Wei et al. (2014), Fernando et al. (2015)) and identifying regularities within football matches (Lucey et al. (2013a), Bialkowski et al. (2014b)). Furthermore, tracking data has been used to generate xG models in other sports than football. Chang et al. (2014) exploits the opportunity of using a tracking data system in the NBA to analyse and quantify shooting. The most popular metric used in NBA for evaluating the shooting ability of players and teams is Effective Field Goal Percentage (eFG%), which is a ratio of the number of field goals to the number of field goal attempts for a player. However, this paper recognise the need for incorporating the quality of a shot into a metric: “If the player is standing still with no one within 10 feet, that is an entirely different shot than if the player is fading off the dribble with two defenders in his face.” Variables like shot distance, shot angle, defender distance, defender angle, player speed and player velocity angle are used as input parameters in the model. Machine learning methods are then utilised on the data set, and the author establishes two new metrics for basketball: Effective Shot Quality (ESQ), which capture the quality of the shot and eFG%+, which is ESQ subtracted from eFG%, and shows how much better than expectation a player or a team shoots.

In the world of football, Lucey et al. (2014) uses a season worth of tracking data with 10,000 shots from a professional league provided by Prozone to quantify the likelihood of scoring by specifying several input parameters. The analysis is complex, dividing the shots into six different match contexts: open play, counter attacks, corners, penalties, free kicks and set pieces.

The shots are analysed in the context of strategic features like defender proximity, interaction of surrounding players and speed of play. Then the authors use logistic regression to estimate the xG for each shot in order to rate teams based on their efficiency.

As can be seen from the works above, several academic pieces apply the concept of xG. It is recognised that xG is a useful measure for evaluating the quality of the shots to better examine if the team or player performed well. No academic works were found that present xG models based on data from Tippeligaen. This thesis presents an xG model, which is developed by applying logistic regression. However, tracking data from Tippeligaen is not available, thus event-based data is utilised. Nonetheless, to the authors knowledge, the xG Model presented in this thesis is the first of its kind for Tippeligaen, and an important step towards developing quantitative models for player performance evaluation in Norway.

2.3 Markov Models in Sports Analytics

As can be seen in the book by Wright (2015), on applications of operational research methods in sports, it is not uncommon to apply Markov processes when analysing sports phenomena. Hirotsu and Wright (2002) presents a continuous Markov process with four states. These states are when team A scores a goal, team A is in possession of the ball, team B is in possession of the ball and when team B scores a goal. As stated in the paper, defining these states and calculating the transition probabilities “makes it possible to estimate the probability distributions of goals scored and the expected number of league points gained, from any position in a match, for any given set of transition probabilities and hence in principle for any match”. By knowing these time dependent parameters, this paper estimates the optimal time for tactical changes during a match, like player substitutions. In Hirotsu and Wright (2003a), a similar model is used to evaluate the offensive and defensive strengths of a football team. These results are then utilised to form a substitution strategy for the team, showing how many of each type of players should start and be substituted during a match. The same Markov model is applied again in Hirotsu and Wright (2003b). This analysis is similar to the former, but focuses on visualising team characteristics like offensive and defensive strength, the home field advantage and relative success against particular teams.

Markov models have also been developed in ice hockey. Thomas (2006) defines 19 states specified by puck possession, takeaways, faceoffs and goals. Using this state space he models ice hockey as a semi-Markov process, which is used to determine the average number of goals scored by a team as a function of the starting state. These values are then used to look at which tactic that should be used in a certain situation. An interesting finding is that giving up puck control in order to get territorial advantage is beneficial both defensively and offensively and, as he states “may explain the prominence of location-based defence throughout professional hockey”.

It can be seen that none of the analyses above are related to individual player performance in order to rate players, but looks to aid teams in tactical decisions. Thomas et al. (2013) marks a start of using Markov process methodology to evaluate individual players. In this analysis, scoring rates for each NHL team are modelled by a semi-Markov model. However, every player will in some way affect this scoring rate, and hazard function models are used for quantifying this effect, enabling player rating within different player positions.

Also in basketball the Markovian approach has been used. Cervone et al. (2014) present an analysis that uses tracking data to value player decisions in real time. The Markovian element in this analysis is the assumption that the decision of the ball handler depends only on the current spatial configuration of the team, disregarding the play history. Statistical methods are used for computing a new metric, expected possession value (EPV). EPV is a quantification of the decision a player makes, and if the decision leads to a higher or lower probability of scoring. This measure is further used to create expected possession value added (EPVA), that is, how much better decision a player takes compared to league average. This is a cutting edge analysis, which utilises complex statistical algorithms on fine grained tracking data.

A thesis highly relevant to this thesis is Routley (2015), which has been a source of inspiration. He looks at multiple seasons in NHL, including 600,000 play sequences and 2.8 million events to model ice hockey dynamics as a Markov game between two teams. A large state space is created, where a state is defined by three context variables (time period, goal differential and manpower differential) and an action history in the form of a play sequence. The sequences are made by defining start and end events, which suits ice hockey dynamics. Every time an action is taken by a player, a state transition is initiated. Multiple reward functions are used to obtain the impact values for goals, wins and penalties. The value iteration algorithm is run over the Markov game state domain to assign a value to each state depending on what is rewarded in the state space. A valuing of individual player involvements is done by looking at the aggregated impact by each player for each season.

In the literature review above no analyses were found on Norwegian football, nor on individual football player performance, that utilises Markovian techniques. Thus, to best of the authors knowledge, the Markov models presented in this thesis are the first in Norway, and probably the first of their kind published in the football analytics community.

2.4 Evaluation of Individual Player Performance

In recent years, focus has shifted towards evaluation of individual players, their involvements and their contribution to the team. Due to the use of simple statistics, several in-match phenomena are not captured, like an inactive defence on a superior team, full backs that need to adjust their roles depending on in match situational variables and the context a player was in when making a choice and performing an action. However, new methods for measuring performance and fine grained data sets now make it possible to incorporate much more information. This enable evaluation of players that perform well, but ”disappear” in such simple statistics.

The EA Sports Player Performance Index is an index that rates players in the two top tiers of football in England - the Premier League and the Championship. This index was developed due to the desire from the football leagues and a news media company to have a rating system that was built on a statistical basis, which enables comparison of players in different positions. In McHale et al. (2012) this index is described, and it consists of six subindices from which individual players get scores: match contributions, winning performance, match appearances, goals scored, assists and clean sheets. Based on these parameters, the performance index is created, and the top 10 players of the English Premier League in the season of 2015/2016 can be seen in Table 2.1. (Premier League (2016))

Although being based on statistical modelling, this index only accounts for simple statistics that does not capture important dynamics of player involvements. Also, the situational context for where the action was carried out is not incorporated.

Another analysis that evaluate players based on simple statistics is that of Santín (2014). He measures ”technical efficiency” of Real Madrid legends by using four variables collected

Table 2.1: EA Sports Player Performance Index EPL 2015/2016

Rank	Player	Club	Position	Index Score
1	Harry Kane	Tottenham Hotspur	Forward	1026
2	Riyad Mahrez	Leicster City	Midfielder	957
3	Jamie Vardy	Leicster City	Forward	938
4	Sergio Aguero	Manchester City	Forward	770
5	Mesut Özil	Arsenal	Midfielder	769
6	Christian Eriksen	Tottenham Hotspur	Midfielder	732
7	Romelu Lukaku	Everton	Forward	711
8	Odion Ighalo	Watford	Forward	700
9	Dimitri Payet	West Ham United	Midfielder	691
10	Alexis Sánchez	Arsenal	Forward	674

from the official Real Madrid website: number of official games played, the number of titles won with their national team, the number of international titles won with Real Madrid and the number of goals scored. The method of Data envelopment analysis (DEA) is applied to find the most efficient Real Madrid legend since 1946 through 2011. This model succeeds in rating players based on their historical performance, but does not provide insights on how to measure and rate players based on in-match choices, actions and events.

In the period after 2010, it seems that large efforts have been made using in-match events to generate statistical models that capture important dynamics. Macdonald (2010) develops a regression based model to evaluate individual ice hockey players. A similar analysis is also performed in the NBA, by Fearnhead and Taylor (2010). These models are called plus-minus models, which enable a player rating based on how the teams perform when the player is on the ice/field as opposed to when he is not. The model from NHL is improved and adjusted in Macdonald (2011a) and Macdonald (2011b), and so far this particular research has ended in an xG model which is described in Macdonald (2012) (see Section 2.2). The output of this model is a list of players and their performance measured in xG per 60 minutes. This analysis is refined in Macdonald et al. (2012), where both goalies and skaters are evaluated, in addition to NHL teams. Another analysis with data set from NHL is Macdonald et al. (2013). In this piece, a playmaker metric is created, which enables the identification and evaluation of playmakers, and their contribution to the team. This is an important and interesting analysis, because earlier works focused mainly on the actual scoring attempt and the quantification of a chance, and the probability of converting it into a goal. This analysis looks at the events that lead up to a goal, that is, quantifies the steps preceding the scoring attempt. This gives valuable insights regarding player contribution besides scoring goals and providing assists. Schuckers and Curro (2013) builds on previous work and creates a model that rates both forwards and defenders. The model is built with on ice events from several NHL seasons, and the method of ridge regression is applied. This methodology ensures a probability assigned to each action event. The per event value is converted into a per season estimate of the number of wins created per player, which is called Total Hockey Rating (THoR).

The models described above are important tools when trying to understand and recreate the complex dynamics of sports. However, there are some limitations that are not accounted for. One important obstacle is to incorporate match status. Pettigrew (2015) bypasses this challenge by creating a model that generates in-game winning probabilities. This is a time line where the probability for one of the teams winning fluctuates dependent on events during the game. As he describes himself:

> "...imagine that one of those players scored all of his goals in overtime, while the other player scored all of his when his team already had a large lead. It is clear that the first player is a more valuable asset. But what about in situations that are not nearly as clear-cut? By using the change in win probability when a goal is scored, we can evaluate how much a player's goal contributions impact their team's chance of winning the game."

Thus, this analysis assesses the offensive efficiency of NHL players, which can be valuable when investigating players with identical goal scoring rates.

Other sports have also been analysed with regards to individual player evaluation. Singh and Ahmad (2015) looks at land hockey, and applies machine learning techniques to predict the better performing player. A recent article that looks at movement patterns in football is Gyarmati and Hefeeda (2016). Here, an event-based data set from Spanish La Liga 2012/2013 is used to extract player movements throughout the season. Vectors are constructed and contain information on start and end of a movement including time stamps, the speed of the movement and whether the player was in possession of the ball or not. These vectors are used for extracting movement characteristics for each player in the league. The uniqueness and consistency of player movements are quantified, which enables a comparison of players.

The above descriptions show that the focus on individual player performance in recent years has increased. The quantitative methods applied on sports leagues like NHL and NBA are increasing in complexity. Furthermore, it seems like the football community has understood the importance of a numerical foundation for aiding decision making regarding individual players. However, in Norwegian football, no analyses were found on quantifiying individual player performance, which is the goal of this thesis.

2.5 Financial Valuation of Football Players

Surprisingly, not much academic work was found on the task of valuing football players financially. Especially notable is the missing research on the link between a players performance on the team and his market value. With the amount of money circulating in the world of football, it is reasonable to believe that there are several stakeholders looking for inefficiencies in the pricing of football players. Why this is the case is hard to determine, but one reason can be the difficulty of pricing intangible assets like human beings. Another possible explanation is that clubs are not disclosing their models to the public.

However, some studies exist. An early work by Amir and Livne (2005) looks at player contracts, and finds no evidence between the investment in player contracts and future benefits for their clubs. In Tunaru et al. (2005) this link is better established by developing a framework for pricing football players based on performance data provided by Opta. Real options theory is used to model the uncertainty surrounding the performance of football players. The framework developed in this paper is also used in Tunaru and Viney (2010), giving a more thorough presentation of the application of the real option pricing formulas.

It can be seen from the above literature review that sports statistical analysis has been through an evolution. The different methods have existed for some time, but the increasing quality of the data sets opens up to novel approaches. Now, advanced methods and models are applied on both tracking data and large event based data sets. These give valuable insight into sports with complex dynamics. This thesis seeks to further develop the field of sports analytics, and inspire to further innovations to improve decision making in the game of football.

Chapter 3

Basic Theory

This chapter presents the basic theory behind the methods applied in this thesis. First, the theory behind binary logistic regression is described, then the definition, notation and dynamics of a two-agent Markov game are introduced. Application of the theory is described in Chapter 4.

3.1 Binary Logistic Regression

Logistic regression is a statistical technique useful when the dependent variable is discrete instead of continuous. The goal is to find the best fitting model to describe the relationship between a dependent (outcome) variable and a set of explanatory (independent) variables. Logistic regression makes it possible to associate likelihoods to outcomes of discrete random variables. The explanatory variables can be dichotomous (taking only two values), polychotomous (taking $k > 2$ values) or continuous. When the dependent variable only has the two outcomes 0 and 1 it is called binary logistic regression. This is appropriate when modelling football phenomena because of dependent variables like match outcome,

$$y_i = \begin{cases} 1, & \text{if a team is winning match } i \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

or the outcome of a shot,

$$y_i = \begin{cases} 1, & \text{if shot } i \text{ resulted in a goal} \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

and to assess what explanatory variables that significantly influence these outcomes.

Following Hosmer and Lemeshow (2000) and Wikipedia (2016), assume a series of N observed data points, where each data point i includes a set of m explanatory variables and an associated binary outcome variable Y_i. Also assume that Y_i is Bernoulli distributed with an unobserved probability π_i of outcome 1. This is specific for the outcome of observation i, but assumed to somehow relate to the explanatory variables. Hence, the expected value of Y_i given the explanatory variables is equal to π_i, or $\mathbb{E}\{Y_i|x_{1,i}, \ldots, x_{m,i}\} = \pi_i$.

The basic idea of logistic regression is the same as for linear regression, which is to model a dependent variable (in this case the probability π_i) using a linear combination of explanatory variables by estimating a set of regression coefficients. The binary logistic regression is run on the logit of the unobserved probabilities, which are defined as

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_m x_{m,i}, \tag{3.3}$$

and the estimated probabilities are calculated by solving Equation (3.3) for π_i, which yields an expression for the inverse of the logit as

$$\begin{aligned} \mathbb{E}\{Y_i | x_{1,i}, \ldots, x_{m,i}\} &= \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \ldots + \hat{\beta}_m x_{m,i}) \\ &= \tfrac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 x_{1,i}+\ldots+\hat{\beta}_m x_{m,i})}} \end{aligned} \tag{3.4}$$

where $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_m$ are the regression coefficients obtained by running the regression.

In order to validate a binary logistic regression, it is possible to look at a classification table. A cutoff or threshold must be specified to determine if an outcome of the dependent variable is classified as 1 or 0, and the hit rate can be calculated by looking at how many observation the model got correct. Following Hosmer and Lemeshow (2000), the sensitivity of the model is defined as the probability of predicting an outcome as positive when $Y = 1$, while the specificity is defined as the probability of predicting negative when $Y = 0$. Only looking at those measures gives limited insight because a threshold has to be specified. A more comprehensive description of classification accuracy is given by the area under the ROC (Receiver Operating Characteristic) curve, which assesses the discrimination abilities of the model. It plots the probability of detecting true observation (sensitivity) and false observation (1-specificity) for an entire range of possible cutoff points.

Hosmer and Lemeshow (2000) also provide a general rule for values of the area under the ROC curve, which is between 0 and 1, as

$$\begin{aligned} ROC = 0.5 &: \text{this suggests no discrimination (that is, one might as well flip a coin)} \\ 0.7 \leq ROC < 0.8 &: \text{this is considered acceptable discrimination} \\ 0.8 \leq ROC < 0.9 &: \text{this is considered excellent discrimination} \\ ROC \geq 0.9 &: \text{this is considered outstanding discrimination.} \end{aligned}$$

3.2 Markov Games

The idea behind a Markov game, also called a stochastic game, is to model two or more decision making agents who are operating in a common state space (environment) with a set of actions available to them. Each action gives a different expected reward in each state for the agents. A single-agent Markov game is called a Markov Decision Process, where the agent interacts with its environment represented as a probabilistic transition function. This means that the environment is fixed in its behaviour. A Markov game is a generalisation of a Markov Decision Process which allows for multiple adaptive agents with interacting or competing goals. For all agents in a Markov game, the goal is to maximise some kind of future reward. This section describes a two-player zero-sum Markov game and its components, which is later used as a framework to model a football match.

3.2.1 Definition

Following Littman (2001), a two-player Markov game is defined by a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, T, \mathcal{R}, \beta \rangle$, where

- $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ is a finite set of game states
- $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$ and $\mathcal{O} = \{o_1, o_2, \ldots, o_l\}$ are finite sets of actions available to the agents ($\mathcal{A}$ is the set of actions for agent 1 and $\mathcal{O}$ is the set of actions for agent 2, the opponent of agent 1)
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{O} \rightarrow \Pi(\mathcal{S})$ is the transition function. It gives, for each state and one action from each agent, a probability distribution over states. $T(s, a, o, s')$ is the probability of ending in state s', given that the agents starts in state s, agent 1 takes action $a \in \mathcal{A}$ and agent 2 takes action $o \in \mathcal{O}$
- $R_i : \mathcal{S} \times \mathcal{A} \times \mathcal{O} \rightarrow \mathcal{R}$ for $i = 1, 2$ is the reward function, giving the expected immediate reward gained by agent i for taking each action in each state. $R_i(s, a, o)$ is the expected reward to agent i in state s when agent 1 chooses $a \in \mathcal{A}$ and agent 2 chooses $o \in \mathcal{O}$
- $0 \leq \beta < 1$ is a discount factor. This describes how the relationship between near-term versus long-term reward is valued. A reward R^i to agent i received j steps in the future is worth $\beta^j R^i_j$ to the agent now. If $\beta = 1$, the Markov game is undiscounted

The agents seek the best actions in each state in such a way to maximise some measure of their long term expected reward received, $\mathbb{E}\{\sum_{j=0}^{\infty} \beta^j R^i_j\}$. The special case where the two agents have diametrically opposite goals allows using only one reward function. This is called a zero-sum Markov game, in which one of the agents is trying to maximise the reward and the other, called the opponent, is trying to minimise.

A technique for estimating values for the states in a Markov game is by using reinforcement algorithms, such as value iteration, to learn a Q-function for each state s and state-action pairs (s, a, o), in the Markov game state domain $\mathcal{S}$.

3.2.2 Policies and Value Functions

A policy is a description of the behaviour of an agent. It specifies a probability distribution over actions to be taken for each state, and is denoted $\pi_i : \mathcal{S} \rightarrow \Pi(\mathcal{A}_i)$. A policy π_i, where $i \in \{1, 2\}$ are the two agents, maps a state $s \in \mathcal{S}$ to a probability distribution $\pi_i(s, a, o)$ from $\mathcal{A}_i$. The probability assigned to actions (a, o) in state s is $\pi_i(s, a, o)$. How good a policy is for an agent can be evaluated by computing the long term value the agent can expect by following the policy. The expected future discounted reward for taking actions a for agent 1 and o for agent 2 in state s and continuing according to policy π_a and π_o thereafter, is denoted $Q^\pi_i(s, a, o)$ for agent i. This can be defined by a set of simultaneous linear equations, one for each state s and agent i,

$$
\begin{aligned}
Q^\pi_i(s, a, o) = R_i(s, a, o) + \beta \sum_{s' \in \mathcal{S}} T(s, a, o, s') \\
\sum_{a' \in \mathcal{A}\, o' \in \mathcal{O}} \pi_a(s', a') \pi_o(s', o') Q^\pi_i(s', a', o').
\end{aligned}
\tag{3.5}
$$

The function Q_i^π is called the Q-function for the policy π_i. Given an initial state s, the agents should execute policies π_a and π_o that maximises

$$\sum_{a,o} \pi_a(s,a)\pi_o(s,o)Q_i^\pi(s,a,o).$$

In most Markov games, the objective is to obtain or learn the optimal policy for the agents acting in the game, in which case the value of the Q-function is said to be learnt off-policy. Contrary, if the Q-function is learnt on-policy, it finds the value of the policies being carried out by the agents in the game. This is the main focus in this thesis, because the policies of the agents are already known from the data set. More specifically, when the Markov game is zero-sum and only one agent is active in each state (the other choose no action each time), the game can be considered a Markov Decision Process. Thus, the mathematics and expressions can be simplified.

3.2.3 Value of a Policy

Following Poole and Mackworth (2010), consider a Markov Decision Process with a known stationary policy π. A policy is said to be stationary when it assigns a known distribution of actions to each state and does not change with time. If a reward criteria is defined for the Markov Decision Process, the policy has an expected value for every state. Let $V^\pi(s)$ be the expected value of following π in state s. $V^\pi(s)$ denotes the reward the agent can expect to receive from following the policy in that given state, and it is defined as the discounted value of all future rewards with discount factor β,

$$\begin{aligned} V_1 = \mathbb{E}\{\sum_{j=0}^{n} \beta^j R_j\} &= R_0 + \beta R_1 + \beta^2 R_2 + \cdots + \beta^n R_n \\ &= R_0 + \beta(R_1 + \beta R_2 + \cdots + \beta^{n-1} R_n) = R_0 + \beta V_1. \end{aligned}$$

Furthermore, let $Q^\pi(s,a)$ denote the expected value of performing action a in state s, and following policy π thereafter. V^π and Q^π can be defined recursively in terms of each other. Let $R(s,a,s')$ be the immediate reward of performing action a, leading to a transition to state s' when being in state s. An agent who performs actions a in state s can therefore be said to receive the immediate reward $R(s,a,s')$ plus the discounted expected value of state s', $\beta V(s')$. When the expected value averaged over the possible resulting states is used for $Q(s,a)$, the following expression is obtained,

$$Q^\pi(s,a) = \sum_{s' \in \mathcal{S}} T(s,a,s')\Big(R(s,a,s') + \beta V^\pi(s')\Big). \tag{3.6}$$

An expression for V^π, is obtained by doing the actions specified by π as

$$V^\pi(s) = Q^\pi(s,\pi(s)) = \sum_{a \in \pi(s)} \pi(s,a) Q^\pi(s,a). \tag{3.7}$$

Chapter 4

Experimental Setup

This chapter describes the experimental setup in this thesis. First, the data used to build the models is explained. Next, three different models based on theory from Chapter 3 are presented. The first is an expected goals (xG) model, developed by the use binary logistic regression in order to assign values to shots attempted in football. The remaining two models are variations of Markov game models applied to football, which share many of the same features, but are nonetheless clearly separated in order to avoid any confusion. Results from running the models are presented and discussed in Chapter 5.

4.1 Data

The data used in this thesis was collected by Opta (2016), and consists of almost two and a half seasons worth of matches played in Tippeligaen in the F24 football feed format. Opta collects the same type of data from every match in 30 leagues and tournaments worldwide. This type of data is mainly used by businesses and stakeholders in football, such as betting companies and statisticians working at football clubs. Sports analytics firms, newspapers and broadcasters are also potential users of such data. The F24 football feed has been used to some extent in academic work as well, such as in Lucey et al. (2013b) and Kerr (2015). The data is not publicly available, and one season costs several thousand GBP.

Opta collects the data by the use of human annotators. All ball events that occur around the ball during a match are stored chronologically with information on team, player, x- and y-coordinates, outcomes and time stamp. A modified example of the structure of the F24 feed is shown in Figure 4.1. Values can not be shown due to a non disclosure agreement between the authors and Opta. F24 files for every match played in Tippeligaen 2014 and 2015 are the main source of data used in this thesis. With 16 teams competing in the league and 30 match days, 240 matches are played each season. In addition, the matches from the first 13 match days of the 2016 season are used to test the models out of sample. 69 different event types exist, where each event can have multiple additional qualifiers associated with it, depending on the type. For example, a pass has qualifiers like end coordinates, distance, angle and several more, and a shot has qualifiers like which body part was used to shoot, what type of play it came from and information on where the shot ended. The x- and y-coordinates are normalised to 100×100 in the data.

Such a comprehensive list of qualifiers in addition to the list of 69 different event types results in a complex event space. This lead to a quite extensive job for the authors to get familiar

```
▼<Event id="          " event_id="  " type_id=" " period_id=" " min=" " sec=" " player_id="      " team_id="   "
  outcome=" " x="    " y="    " timestamp="                       " last_modified="                "> 
    <Q id="         " qualifier_id="   " value="   "/>
    <Q id="          " qualifier_id="   " value=     "/>
    <Q id="          " qualifier_id="   " value="    "/>
    <Q id="          " qualifier_id="  " value="    "/>
    <Q id="         " qualifier_id="   " value="    "/>
  </Event>
▼<Event id="         " event_id="  " type_id=" " period_id=" " min=" " sec="  " player_id="      " team_id="  "
  outcome=" " x="    " y="    " timestamp="                        " last_modified="                ">
    <Q id="         " qualifier_id="   " value="   "/>
    <Q id="         " qualifier_id="   " value="    "/>
    <Q id="         " qualifier_id="  " value="      "/>
    <Q id="          " qualifier_id="   " value="    "/>
    <Q id="         " qualifier_id="   " value="    "/>
  </Event>
```

Figure 4.1: Data structure of F24 feed

with the data set. A substantial amount of time has been spent watching video tapes comparing the data to the tape, in order to fully understand how Opta characterises the different events. Furthermore, not all the event types in the data set are considered relevant for an approach such as in this thesis. The data set, which is on the form as one XML file for each match, was read in C++ using Xcode for Mac, where objects are created for each match and each event.

In order to make the data applicable for the desired approaches, a restructuring of the data set is carried out. Events such as period start, period end, expected or unexpected player or referee changes and other events independent of the ball, are removed. Moreover, each match is divided into sequences, where a sequence stops when the ball goes out of play for a goal kick or a throw-in, if a goal is scored or if a period is over. For the events in these sequences, some extra features are calculated and added. These features are described further in the forthcoming sections. This lead to a data structure as shown in Table 4.7 for the Markov models. The data used for the expected goals model, is a list of all the shots and goals in the data set, excluding the own goals.

The data only follows the ball and does not contain information on the position of players not acting on the ball, and is thus not spatiotemporal. In addition, no event for carrying the ball across the field is defined by Opta. Because of this, it is possible for apparent jumps between areas on the field where subsequent events are taking place. However, by looking at the previous and current event, inferences can be made about the position of the ball between the events as in Lucey et al. (2013b).

Data from all the matches are collected by hand and are verified at least once after each match. Mistakes that are revealed are corrected or removed continuously, and the data can be considered clean, complete and accurate for all practical purposes. In addition, the aforementioned comparison of video and data by the authors of this thesis did not reveal significant discrepancies between the hand annotated data and what actually occurred. Table 4.1 shows an overview of some simple descriptive statistics for the data set.

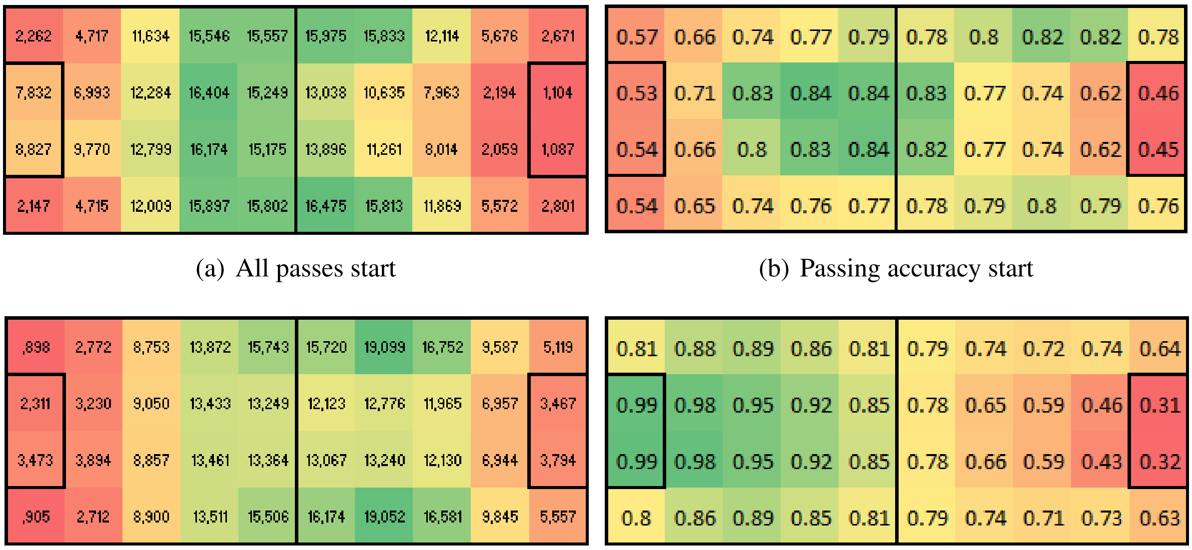
Table 4.1: Simple descriptive statistics for the dataset

Tippeligaen	2014	2015	2016
Matches	240	240	104
Goals*	735	774	261
Number of shots*	6,894	6,598	2,763
Number of passes	199,675	198,168	88,277

* including own goals, 27 in 2014 and 22 in 2015.

From Table 4.1 it can be seen that a pass is the most frequent type of event in the data set. The number of events vary from match to match, depending on the playing style of the teams and other factors. Furthermore, it can be seen that the numbers seem fairly consistent across seasons.

Figures 4.2 and 4.3 illustrate simple examples of what can be calculated from this kind of data.



(a) All passes start

(b) Passing accuracy start

(c) All passes end

(d) Passing accuracy end

Figure 4.2: Illustration of passing data in terms of pass counts and passing accuracies

Figure 4.2 shows four maps of the field with pass counts to the left and passing accuracies to the right. The data is always from the perspective of the attacking team playing from left to right. As one should expect, most passes are hit in the middle of the field, while the passing accuracy is higher on the defensive half.

Figure 4.3 consists of four maps over the attacking half of the field. The thick black lines in the maps illustrate the 18-yard box. Again a couple of unsurprising observations can be made. The shooting accuracies are higher closer to the goal, and the accuracy decreases if the shot was a header or if the shot was assisted by a cross.

(a) Regular shots (b) Headed shots (c) Free kicks (d) Shots from crosses

Figure 4.3: Illustration of shot data in terms of conversion rates

4.2 Expected Goals Model

The method of binary logistic regression, introduced in Section 3.1, is used to develop an xG model in order to rate primarily offensive players in Tippeligaen 2014 and 2015 based on their efficiency in converting shots into goals. This section describes the model building approach and the explanatory variables that are tested in order to obtain an xG model for assigning likelihoods of scoring on shots attempted. This model is denoted as the xG Model hereafter.

4.2.1 Model Building Approach

Three distinctive types of shots are considered in the xG Model. They are free kicks, penalty kicks and regular shots (all shots other than free kicks and penalty kicks). This distinction is chosen due to the different nature of these situations. Free kicks and penalty kicks are static events where the shooters are completely undisturbed by defenders, contrary to regular play where more factors are influencing the outcome of a shot. The general idea is still the same for all three types, namely to assign a likelihood for scoring on each shot.

The outcomes of 12,781 regular shots and 558 free kicks from Tippeligaen 2014 and 2015 are used as the binary dependent variable (1 for goal, 0 for miss) in two separate logistic regression models. All penalty kicks are taken from the same position and therefore have almost no variation to account for. Thus, the likelihood of scoring on a penalty is estimated by the conversion rate observed from the 101 penalty kicks taken over the two seasons. An example of the structure of the shot data is shown in Table 4.2.

Three free kicks that ended in a goal are taken out of the data set. The three goals are Stian Ringstads 1-0 goal for Lillestrøm against Odd in 2014, goalkeeper Håkon Opdals 1-1 goal for Start against Vålerenga in 2015 and Espen Børufsens 3-1 goal for Start in the game against Sarpsborg 08 in 2015. They are considered outliers because the players that took the free kick did not intend to score, and a substantial amount of luck was involved for each of them to end up in a goal.

A large set of explanatory variables are used as a starting point for the binary logistic regressions for regular shots and shots from free kicks. They are further discussed below. The method

Table 4.2: xG Model: Structure of shot data

Var_1	Var_2	Var_3	Free kick	...	Var_n	Goal
21.1	0.1245	1	0	...	1	1
5.1	0.2458	1	1	...	0	0
18.0	0.0459	0	1	...	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
11.4	0.1648	0	0	...	1	1

of backward elimination described in Groebner et al. (2011) is used to determine a set of statistically significant explanatory variables for the two regression models for regular shots and free kicks. Regression coefficients and p-values are estimated using statistical software SSPS which use the method of maximum likelihood.

4.2.2 Explanatory Variables

As mentioned above, a large set of explanatory variables are considered for the two regression models for regular shots and free kicks. They are shown in Table 4.3 along with the type of variable and whether it was calculated or determined by the authors or not.

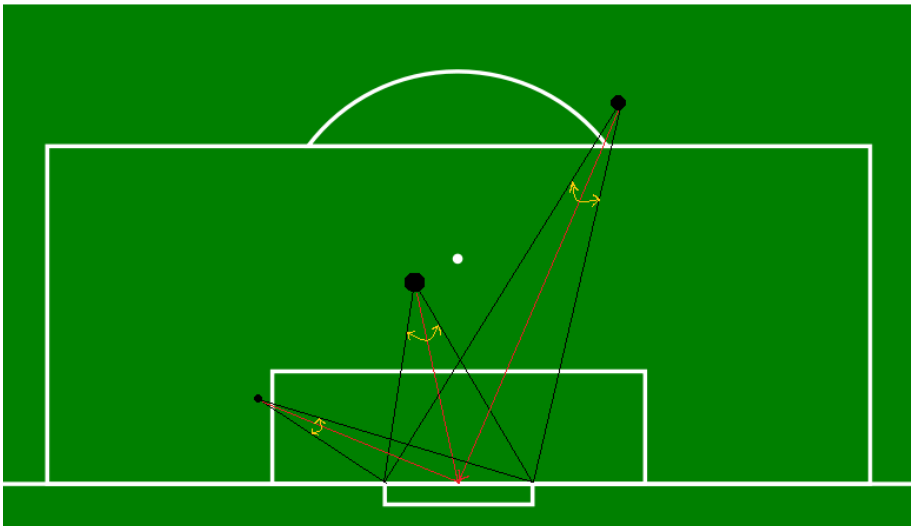


Figure 4.4: xG Model: Illustration of angle and length variables for a shot. Red lines show the length and orange arrows show the angle. Angle measured in radians

The distance to goal and angle to goal (measured in radians) are continuous variables and are illustrated in Figure 4.4. It is believed that a longer distance to goal decrease the likelihood of scoring, because the shooter has to aim more precisely and the goalkeeper has longer time to react. The angle describes how much the shooter can see of the goal, and it is believed that a higher angle is favourable for the outcome. Current match status is the goal difference in the game from the perspective of the player that attempted the shot, and the home or away variable is whether the player played at home or not.

Table 4.3: xG Model: Variables considered in binary logistic regression

Variable description	Variable type	Calculated
Distance to goal	Continuous	Yes
Angle on goal	Continuous	Yes
Current match status	Categorical	Yes
Home- or away game	Binary	No
Difference in team strength	Nominal	Yes
Assisted by pass	Binary	Yes
Assisted by rebound	Binary	Yes
Assisted by long ball	Binary	Yes
Assisted by cross	Binary	Yes
Assisted by throughball	Binary	Yes
Assisted by head pass	Binary	Yes
Assisted by shot	Binary	Yes
Type of play: Regular play	Binary	No
Type of play: Set piece	Binary	No
Type of play: Corner	Binary	No
Type of play: Throw-in	Binary	No
Type of play: Fastbreak	Binary	No
Number of preceding take on	Nominal	Yes
Finished by a header	Binary	No
Finished by other body part	Binary	No

Difference in team strength is assessed using the Elo rating system, described in Hvattum and Arntzen (2010). For the player attempting the shot, the difference in Elo rating between the two teams playing in a match is used on each shot. Players on high quality team are likely better shooters than on lower quality teams. However, the lower quality teams are likely to defend more, which can decrease the likelihood of scoring due to increased defender proximity.

Seven binary variables describe how the shot was assisted, that is how the ball was brought into the shooting position. They are all determined by looking at the event preceding the shot. They are believed to influence the likelihood of scoring differently. For example, if a shot was assisted by a cross the likelihood is believed to decrease because it makes the job of the shooter more difficult, which is supported by Figure 4.3(d). The opposite can be said regarding a shot following a rebound, where it often is a clear chance of goal either from a save by the goalkeeper or a rebound off the post.

Five binary variables describe the type of play the shot came from. They are defined by the human annotators in Opta and are present as a qualifier for each shot in the data set. Such definitions make them subject for human bias. Especially the qualifier for fastbreak is vaguely defined and seldom occur in the data.

The number of preceding dribbles is the number of successful dribbles a player did before shooting. Most likely, the player attempted to move the ball to a better area by dribbling off defenders before a shot, and is therefore believed to influence the likelihood of scoring positively.

The two binary variables in the bottom of the table describe what body part the player making the shot used to finish. They are qualifiers associated to every shot by Opta, and are both believed to negatively influence the likelihood of scoring.

4.2.3 Application of the Expected Goals Model

After the method of backward elimination is applied for regular shots and free kicks separately, one or more statistically significant variables remain. They constitute the xG Model for the two types of shots, and their coefficient values are used as input in Equation (3.4) to assign a likelihood for converting regular shots or free kicks into goals. As mentioned earlier, the likelihood of scoring on a penalty kick is estimated by using the observed conversion rate. These likelihoods are used to rate players in terms of efficiency by looking at how many goals they actually scored, compared to how many they were expected to score by aggregating the xG value from the shots they attempted. The efficiency measure used to rank players from the xG Model is defined as actual goals divided by expected goals and is denoted G/xG throughout this thesis.

It is also possible to aggregate the xG values for each team, both in terms of xG for and xG against by looking at the shots a team attempted and received, respectively. From this, a league table based on the difference in xG (xG for minus xG against) can be generated.

4.3 Markov Game Model 1

A Markov game, formally introduced in Section 3.2, is used to develop a model for evaluating all player involvements, not only shots. Football matches are modelled as a Markov game where the two teams playing represent the agents in the game. The two teams have diametrically opposite goals, which makes it a zero-sum Markov game. In addition, the value of the state and state-action pairs are learnt on policy. As described in Section 3.2.2, these specifications make it possible to consider the game as a Markov Decision Process, which simplify the mathematics. The different parts of this model are described in the following subsections, and the model is referred to as Model 1 in the rest of this thesis.

4.3.1 State Space: Context Variables and Field Zones

The literature review on sports analytics in Section 2.1 suggested that the context variables match status, match location and manpower difference on the field influence the dynamics of the game. These three context variables forms the basis for the state space in this model. Early Markov process models for football, such as Hirotsu and Wright (2002, 2003a), used a simple four-state model with possession and goals constituting the state space. This state space include two of the most fundamental attributes of football, but is still a rough approximation of a highly complex game. In Model 1, the state space is expanded to include context variables and the position on the field where an action is taken.

The context variables provide information on what phase or condition the game is in. The general idea of introducing such variables is that they influence how players on a team make decisions. When a team is leading, they are more likely to defend instead of continuing launching forward to extend their lead. Similarly, if a team gets a player sent off they are more likely to sit back and attempt to launch counter attacks in order to score. Context variables used in the state space are listed in Table 4.4 along with their respective values.

The number of time periods, TP, in a match is chosen to four. This is done to capture the effects the time left to play has on a match. When a game enters its ending phase, scoring a goal can possibly be more valuable than scoring earlier in the match because the opposition have less time to respond. Given the four time periods, TP changes value at 23 and 68 minutes played as

Table 4.4: Model 1: Context variables

Variable	Notation	Values
Time period	TP	$[1, 2, 3, 4]$
Match status	MS	$[-1, 0, 1]$
Manpower difference	MD	$[-2, -1, 0, 1]$

well as at half time. Because of overtime, $TP = 2$ and $TP = 4$ can be slightly longer than the other two. Match status, MS, determines which team is leading a match. If the home (away) team is leading, $MD = 1$ (-1) and $MS = 0$ if the teams are drawing. Manpower difference, MD, is the difference in players on the field between the two teams. $MD = 1$ (-1) if the away (home) team have gotten a player sent off, and $MD = 0$ when they have an equal number of players on the field.

Any combination of the three context variables constitute a pure context state. There are $4 \cdot 3 \cdot 4 = 48$ unique context states, which is derived from multiplying the number of possible values for the three variables. 34 of the possible 48 context states occur at least once in the data set.

It is possible to add more context variables to make the model more detailed. Examples of other context variables that can affect the state of a game are the relative strength between the teams playing a match and the yellow card difference. Introducing more context variables would enlarge the state space, and potentially make the model more realistic. However, by extending the state space there is a chance it could make the number of occurrences of some of the states too low. With data from no more than 480 matches, only the three situational variables considered to have the largest impact on the game are included.

Instead, the state space is extended to include location on the playing field. The football field is split into 15 zones as shown in Figure 4.5, where the attacking team is playing from left to right. If the attacking team lose possession in zone 15, the defending team wins it in their zone 1. As seen from the figure, the grid is not symmetrical about the centre line, but is rougher in the defensive half than the offensive. The difference in value between having the ball in zone 1 by the sidelines or in the centre of the field is likely very small. This can not be said about the difference between zone 14 and 15 in the offensive half where the goal is situated.

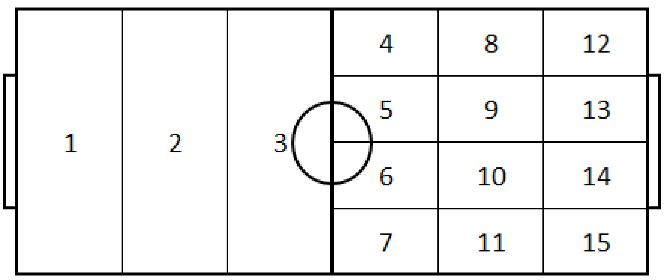


Figure 4.5: Model 1: Field zones

The grid in Figure 4.5 can be changed to become symmetrical and/or include a higher number of zones. It is chosen like this to avoid a too large state space. With the amount of data available, a larger state space could give states with too few observations to build the model from.

In addition to the context variables and the zone of the field, a state also contain information on what team is in possession of the ball. This is done in order to keep track of who is executing the action. Figure 4.6 is an illustration of the information that is included in a state in Model 1. There are $2 \cdot 4 \cdot 3 \cdot 4 \cdot 15 = 1440$ possible states in the state space. Of these, 902 occur at least once in the data set.

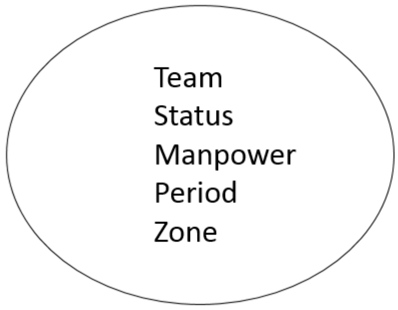


Figure 4.6: Model 1: Definition of a state

4.3.2 State Transitions: Actions

A central part of any Markov game is the transition between states. For each state in the state space, there is a transition function giving probabilities of ending up in a state given one action from each agent in the game. In Model 1, a state transition occurs when the two agents make actions to move the ball to another area of the field (a change of zone) or they make an action that change the context variables (a red card or a goal occurs). The actions available to the team in possession is denoted $\mathcal{A}$ and are shown in Table 4.5. For the defending team, the action space $\mathcal{O}$, shown in Table 4.6 is more limited. In both the sets, the no action is an alternative which is used by the defending team most of the time. For each state transition, one team always performs no action, which make only one agent active. Hence, this can be interpreted as a Markov Decision Process. This reflect how football is played, and is also in accordance with the data set which does not include positional data on players other than the ones acting on the ball. A change of time period occurs automatically independent of the actions.

All actions in Table 4.5 and 4.6 are defined by Opta in the data set, except the Ball carry action which is made by the authors. It is done to avoid jumps from one state to another without an action being made. An example of this is if a player receives a pass in zone 9 and run with the ball to zone 13 and takes a shot. Between the end coordinates of the pass and the shot location, there might be an apparent jump between states without an action. The Ball carry event prevents this from occurring. Caution has been taken when introducing the Ball carry action, and it is only included when there is no doubt that a Ball carry actually occurred.

Table 4.5: Model 1: $\mathcal{A}$, action set for the team in possession of the ball

Pass	Take on	Shot
Aerial duel	Chance missed	Cross
Long pass	Throw-in	Free kick taken
Ball carry	Corner won	No action
Dispossessed	Yellow/red card	

Table 4.6: Model 1: $\mathcal{O}$, action set for the team not in possession of the ball

Foul	Tackle	Interception
Ball recovery	Clearance	Aerial duel
Ball touch	Yellow/red card	No action

The state transitions are calculated from observed play. Table 4.7 shows an example of how the data used in this model is structured. The data consists of a comprehensive list of events, which is iterated through in C++ in order to first create a vector of the occurring states. The occurrences of each state are then counted and denoted $Occ(s)$, as well as the number of transitions from s to s' given actions $a \in \mathcal{A}$ and $o \in \mathcal{O}$ from the two teams, $Occ(s, a, o, s')$. It is possible for a team to have successive actions without the opposing team interfering. The state transition probabilities are estimated by

$$T(s, a, o, s') = \frac{Occ(s, a, o, s')}{Occ(s)}.$$

Table 4.7: Model 1: Structure of data

Match	Sequence	Player	TP	MS	MD	Zone	Team	Action
1	1	21544	1	0	0	3	Home	Pass
1	1	56897	1	0	0	2	Home	Long pass
1	1	26423	1	0	0	1	Away	Aerial duel
1	2	13469	1	1	0	11	Home	Throw-in
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Figure 4.7 shows an example of a play sequence used to build the state space, and how the state transitions occur. It is taken from the match Tromsø - Sarpsborg 08 in 2015. As can be seen from the figure, the away team takes a throw-in from zone 12 which is intercepted by the home team in their zone 2. From here, a take on from the home team takes place, followed by two passes and then a shot from zone 14. The shot can go to one of the two artificial shot outcome states, depending on whether it was a goal or not. In this particular example, the shot did not end up in a goal.

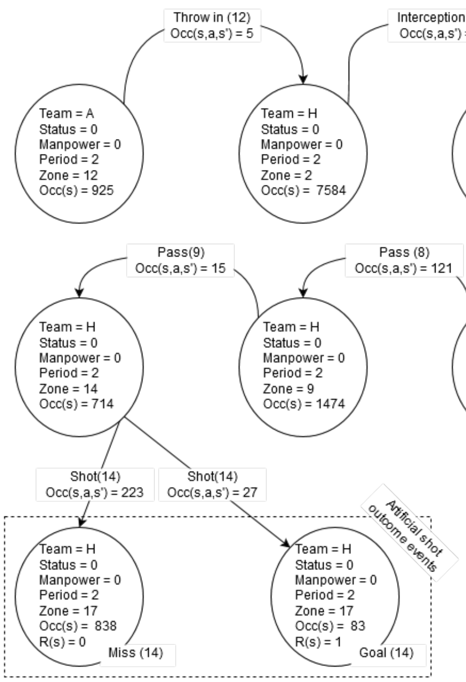


Figure 4.7: Model 1: Illustration of states and transitions. Example from Tromsø - Sarpsborg 08 in 2015

4.3.3 Reward Function and Value Iteration Algorithm

The reward represents the objective of the agents in a Markov game. The ultimate objective in football is to score more goals than your opponent. The objective of your opponent is the direct opposite of yours. Motivated by this, the states in the state space with a goal are given reward, $R(goal) = 1$ if the home team scores, and $R(goal) = -1$ if the away team scores the goal. This way of rewarding states favour the offensive players on the field because they are playing closer to the opposition goal and therefore more often comes in a scoring position. Regardless of that, the defensive players with many ball recoveries and duels won are also rewarded, but most likely with smaller values. It is possible to define rewards to other states as well, depending on what one wants to measure, but the obvious choice for football is goals.

In order to assign a value to each state-action pair, a Q-function which updates the values in each iteration can be used. Since the objective is to find the values of a known policy, let Equation (3.6) from Section 3.2 be the starting point to define a Q-function

$$Q^{\pi}(s,a) = \sum_{s' \in \mathcal{S}} T(s,a,s')\Big(R(s,a,s') + \beta V^{\pi}(s')\Big), \tag{3.6}$$

while Equation (3.7) is the starting point when defining an expression for the value of a state

$$V^{\pi}(s) = Q^{\pi}(s,\pi(s)) = \sum_{a \in \pi(s)} \pi(s,a) Q^{\pi}(s,a). \tag{3.7}$$

The policy for a state in this model, $\pi(s)$, can be interpreted as the probability of performing the different actions a in state s, which for each state and each action is denoted by

$$\pi(s,a) = \frac{Occ(s,a)}{Occ(s)}.$$

From this, the expected value of being in state s' for the home team can be written as

$$V_H(s') = \frac{1}{Occ(s')} \cdot \sum_{a'_H \in \mathcal{A} \cup \mathcal{O}} \Big(Occ(s', a'_H) \cdot Q_H(s', a'_H)\Big). \tag{4.1}$$

Furthermore, the probability of ending in state s' when performing action a, is denoted

$$T(s,a,s') = \frac{Occ(s,a,s')}{Occ(s,a)}.$$

Hence, Equation (3.6) can be written as the following when the home team (H) is performing action a

$$Q_H(s,a_H) = \frac{1}{Occ(s,a_H)} \sum_{s' \in \mathcal{S}} \Big(Occ(s,a_H,s') \cdot \big(R_H(s,a_H,s') + \beta \cdot V_H(s')\big)\Big). \tag{4.2}$$

Equation (4.2) is the Q-function which is applied in the value iteration algorithm in order to learn the values of the state-action pairs in Model 1. The values of each state, $V_H(s')$, are calculated by Equation (4.1).

The input to the value iteration algorithm is the vector containing all the states, including information of number of occurrences and state transitions. The algorithm shown in Algorithm

1 is run in C++ using Xcode for Mac, and runs until convergence or until the maximum number of iterations is reached. A relative convergence criteria of 0.0001 and a maximum number of iterations of 10,000 is used. The use of relative convergence implies that the algorithm will terminate only when the Q-values for the state-action pairs are being increased by a small amount. Conversely, the algorithm will continue when the Q-values increases with larger amounts.

Require: Markov game model, convergence criterion c, maximum number of iterations M
$lastValue = 0$
$currentValue = 0$
$converged = false$
for $i = 1; i \leq M; i \leftarrow i + 1$ **do**
  for all state-action pairs (s, a) in the Markov game model **do**
    if $converged == false$ **then**
      $Q_{i+1}(s,a) = \frac{1}{Occ(s,a)} \sum_{(s') \in \mathcal{S}} (Occ(s,a,s') \cdot (R_i(s,a,s') + V_i(s')))$
      $currentValue = currentValue + |Q_{i+1}(s,a)|$
    end if
  end for
  if $converged == false$ **then**
    if $\frac{currentValue - lastValue}{currentValue} < c$ **then**
      $converged = true$
    end if
  end if
  $lastValue = currentValue$
  $currentValue = 0$
end for

Algorithm 1: Dynamic Programming for Value Iteration

The Q-function is not discounted ($\beta = 1$), implying that the players in the Markov game are indifferent whether a goal is scored after a long or a short sequence of actions. It is possible to argue that goals scored after fewer moves could be more worth than long build-ups ending in a goal. However, goals are scarce in football and it is not an unreasonable assumption that all goals should be valued the same regardless of how they came about.

Again, it is worth mentioning that the goal of Model 1 is not to find the best possible policy or strategy for the agents in each state, as in many other Markov games. This is due to the known transition probabilities, which are estimated from what is observed in the data. Instead, the goal is to find the value of the policy that is being carried out by the agents. However, it is worth mentioning that the optimal action to perform in each state, can be calculated based on the immediate reward generated by the action. This can be done by the formula

$$\pi^*_H(s) = \operatorname*{arg\,max}_{a_H \in \mathcal{A} \cup \mathcal{O}} Q(s, a_H). \tag{4.3}$$

Since the purpose of the model is to assign a value of a given policy, this action can not be considered an optimal policy. Instead it can be referred to as an optimal action in a state. These formulae are also applied on the other agent in the Markov game, the away team.

4.3.4 Valuing Individual Player Actions

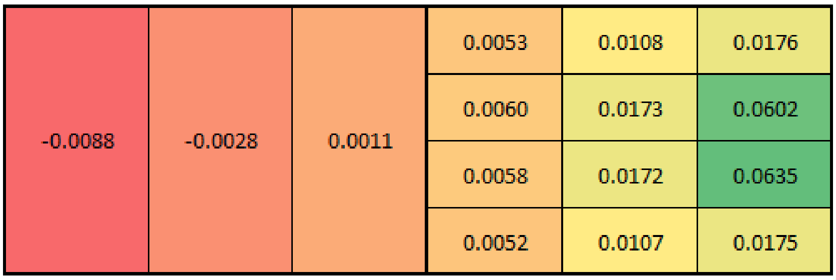
After the value iteration algorithm has converged, all states and state-action pairs have been assigned a value. On the basis of this, each player involvement can be assigned a value based on its impact on the game. Several functions for the impact can be used. For this model, especially two impact-functions are considered. They are

$$I_H(s, a_H, a_A, s') = V_H(s') - Q_H(s, a) \tag{4.4}$$

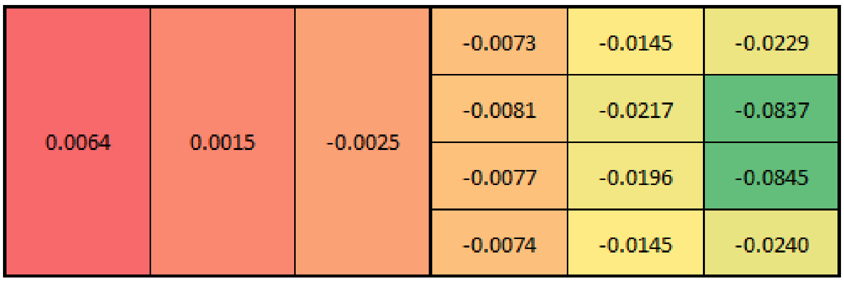
and

$$I_H(s, a, s', a') = Q_H(s', a') - Q_H(s, a). \tag{4.5}$$

With the former, Equation (4.4), it can be said that the performing player receives the difference between the value of state s' the action ended in and the value of performing action a in state s. With the latter, Equation (4.5), players receive the difference between the value of performing action a' in state s' and the value of performing action a in state s. With both these formulae, the model captures the value of how the action affected the state the match was in. The latter accounts for what the given action actually lead to with the average impact of this action in the given state, while the former compares the value of performing the action with the average value of being in state s'. Based on this, the latter is chosen as the impact function in order to capture the effect of what the action actually resulted in. An example and further discussion of this effect are given below. Positive impacts are in favour of the home team while negative impacts are in favour of the away team. As mentioned above, the output from the value iteration algorithm are values for each state and state-action pairs. Figure 4.8 shows an illustrative example of state values for both the home and away teams in the simplest context available: $TP = 1, MD = 0$ and $GD = 0$.



(a) Home team zone values



(b) Away team zone values

Figure 4.8: Model 1: Zone values of state space for illustrative purposes. $TP = 1, MD = 0$ and $GD = 0$

From Figure 4.8(a) it can be seen that the state in the bottom right corner for the home team (zone 15) has the value $V_H(15) = 0.0175$ and the adjacent state in front of the goal (zone 14) has the value $V_H(14) = 0.0635$. For simplicity, assume that a cross from zone 15 in this context has the same value as the state, $Q_H(15, Cross) = 0.0175$ and that a shot from zone 14 in this context has the value $Q(14, Shot) = 0.1250$. If a player on the home team crosses the ball from zone 15 to zone 14 in this context and the following action was a shot from zone 14, he would gain an impact of

$$I_H = Q_H(14, Shot) - Q_H(15, Cross) = 0.1250 - 0.0175 = 0.1075.$$

On the other hand, assume that a pass from zone 14 has a value of $Q(14, Pass) = 0.05$. A cross from zone 15 that lead to a pass in zone 14 would then only receive an impact of

$$I_H = Q_H(14, Pass) - Q_H(15, Cross) = 0.050 - 0.0175 = 0.0325.$$

From this it can be seen that the impact value of an action can be highly dependent on the next action when using Equation (4.5). If Equation (4.4) is used as the impact function, the player would have received an impact of

$$I_H = V_H(14) - Q_H(15, Cross) = 0.0635 - 0.0175 = 0.0460,$$

independent of which type of action is performed next.

From this example it is possible to argue that the Equation (4.5) captures how the player involvement affected the game in a better way than Equation (4.4). From another point of view, it might be unfair that a player is punished if his teammate, which he passed the ball to, loses the ball. In the long run, if a player performs the same pass (a pass between the same two zones) multiple times, his total impact should converge to a value that is righteous for the player. If a given player is considered a good passer, it is more likely that a teammate that receives a pass might be in a better position to do something useful with the ball, compared to if the pass was from a player that is considered a worse passer. Based on this, it is believed that the latter impact function is fair over the course of a whole season. However, this impact function leads to a suspicion that the model might favor players on the better teams because of better teammates.

Values from the impact function in Equation (4.5) can be aggregated for each player in every game over the course of a season. This enables an evaluation of a players total performance over a season. In addition, team performances can also be evaluated. In each game, the values for both the home and away team can be aggregated and compared to determine which team performed the best. For instance, if the home teams impact totalled 3.5 and the away teams 2.5, the home team receives a value of 1 and the away team a value of -1 for that match. Again, these values can be aggregated over a season, from which a league table can be created.

4.4 Markov Game Model 2

As mentioned in the introduction to this chapter, two Markov game models are developed for evaluating all player involvements, not only shots. This section describes the second Markov game model, hereby referred to as Model 2. It is built upon the basic principles from Section 3.2. The same specifications as for Model 1, which make it possible to consider it as a Markov Decision Process, apply. The main difference between the two models is the amount of information included in a state, which now includes what action was made. With this definition, after the states have been assigned a value, the value of the state actually represents the value of that specific action. An important drawback with this model is that the role of the agents choosing an action in each state becomes less apparent, making the state transitions somewhat more abstract. The different parts of Model 2 are described in the following subsections.

4.4.1 State Space: Context Variables and Field Zones

Following Section 4.3.1, it can be said that context variables such as match status, match location and manpower difference influence how the players on the field make their decisions. The context variables used in Model 2 are the same ones as in Model 1 and are repeated in Table 4.8.

Table 4.8: Model 2: Context variables

Variable	Notation	Range
Time period	TP	$[1, 2, 3, 4]$
Match status	MS	$[-1, 0, 1]$
Manpower difference	MD	$[-2, -1, 0, 1]$

The next similarity with Model 1 is the extension of the state space to include the location on the field, repeated in Figure 4.9. As seen earlier, the field is divided in 15 zones which in combination with the context variables leads to a state space of potentially 1440 states.

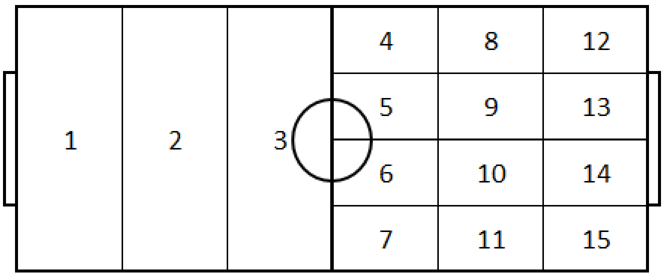


Figure 4.9: Model 2: Field zones

4.4.2 State Space: Actions

The actions available to the two agents are the same as for Model 1. In Model 1, the effect of the different actions is captured through the impact function. For Model 2, the type of action is included in the state definition, in order to separate the different types. In addition, the zone which an action ended in is also included in the state definition. This is done in order to capture the difference between especially passing events, for instance a pass from zone 2 to zone 3 and a pass from zone 2 to zone 5. As previously mentioned, Opta delivers the end-coordinates for passing events making this possible.

Another feature which might separate an action from another, is whether it was successful or not. Opta captures this through their outcome parameter. Table 4.9 shows the definitions of the two outcomes of each type of action. In order to separate a successful action from an unsuccessful one, the outcome of each action is also included in the definition of a state. Figure

4.10 shows a summary of what kind of information that is included in the definition of a state in Model 2.

Table 4.9: Model 2: Overview of the outcome parameter for the actions

Action	Outcome = 1	Outcome = 0
Pass	Successful	Unsuccessful
Take on	Successful	Unsuccessful
Foul	Foul won	Foul conceded
Corner won	Always set to 1	-
Tackle	Wins possession	The other team retains possession of the ball
Interception	Always set to 1	-
Clearance	Always set to 1	-
Shot	Always set to 1	-
Yellow card	Always set to 1	-
Aerial duel	Duel won	Duel lost
Ball recovery	Always set to 1	-
Dispossessed	Always set to 1	-
Chance missed	-	Always set to 0
Ball touch	Ball simply hit the player	Unsuccessful control of the ball, possession lost
Cross	Successful	Unsuccessful
Long pass	Successful	Unsuccessful
Throw-in	Successful	Unsuccessful
Free kick taken	Successful	Unsuccessful
Ball carry	Always set to 1	-

The inclusion of all these features lead to a very large and complex state space. Since the ball can go out of play or into one of the goals, an action can end in 17 different zones. 19 different actions can be carried out, with outcome equal to 0 or 1. This, in combination with the context variables, yields a number of 930,240 possible states, but the majority of them are infeasible due to the nature of the action being carried out. For instance, it is impossible to hit a cross from zone 3 to 5. In the data set, only 34,697 states occur at least once. These numbers are indeed very high compared to the number of events in the data set, but the main idea behind this model is that when a state has been assigned a value, this value is actually the value of that specific involvement.

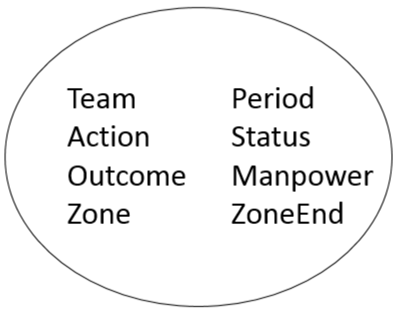


Figure 4.10: Model 2: Definition of a state

4.4.3 State Transitions

The state transitions become less intuitive in Model 2 because almost all information is stored in the states. Moreover, with the action type and in which zone the action ended stored in the definition of a state, the decision of which action to perform for the agents is not visible at first sight. The structure of the data used in Model 2 is the same as for Model 1 (see Figure 4.7), and again the comprehensive list of events is iterated through in C++ in order to create the state space. The occurrences of each state are counted and denoted $Occ(s)$, in addition to the number of transitions between state s and s', $Occ(s, s')$. From this, the transition probabilities are calculated by

$$T(s, s') = \frac{Occ(s, s')}{Occ(s)}.$$

The rationale behind state transitions like this, is more in terms of what the performed action lead to, as opposed to the change of context or state because of an action as for Model 1. With this is mind, the role of the agents in Model 2 vanishes to some extent. When a player has the ball in a given zone, the action he performs, to which zone and the outcome of the action, decides which state the game is in. The following state is decided by the subsequent action. Hence, the role of the players as agents is not as visible as for Model 1. Figure 4.11 illustrate an example sequence modelled as described above, and is the same sequence used for Model 1 in Figure 4.7.

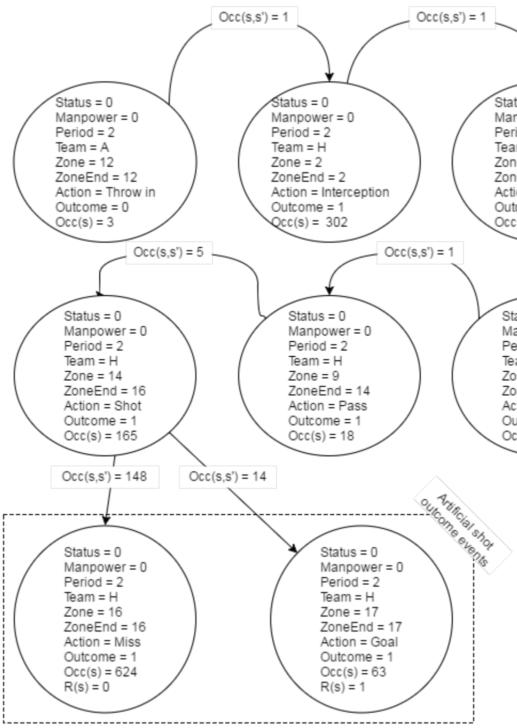


Figure 4.11: Model 2: Illustration of states and transitions. Example from Tromsø - Sarpsborg 08 in 2015

From Figure 4.11, it should be evident that a state includes all the information except the transitions. The first state in the figure is an unsuccessful (outcome is equal to 0) throw-in for the away team from zone 12 to zone 12. Since this throw-in was unsuccessful, the other team (the home team) has the ball, and therefore the next state is and interception for the home team, which again is followed by a take on. The next actions are passes from zone 8 to 9, and from zone 9 to 14, followed by a shot from zone 14. As mentioned before, this particular shot did not end up in a goal.

4.4.4 Reward Functions and Value Iteration Algorithm

The reward function and the value iteration algorithm are the same as for the previous model. Reward is given for states where a goal occurs, and states are rewarded 1 when the home team scores and -1 when the away team scores. These rewards are given to the artificial shot outcome-events, as found in the $R(s)$ values in Figure 4.11. The Q-function used in this model is almost the same as for the previous one, since this model also is on-policy. As Equation (4.6) shows, the only difference is the replacement of the values for the state-action pair with Q-values for the states. This is mainly due to the absence of actions as state transitions. The Q-function for Markov Game Model 2 is defined as

$$Q_H(s) = \frac{1}{Occ(s)} \cdot \sum_{(s,s') \in S} \Big(Occ(s, s') \cdot \big(R_H(s, s') + \beta \cdot Q_H(s') \big) \Big). \tag{4.6}$$

This Q-function is also undiscounted ($\beta = 1$) of the same reasons as for the previous model. As for Model 1, the input to the value iteration algorithm is the vector containing all the states, including information regarding occurrences and state transitions. The algorithm is then run in C++ using Xcode for Mac.

When again considering Figure 4.11, the unsuccessful throw-in exemplifies how the outcome parameter captures an important feature in football. Since the throw-in is unsuccessful, the following states consist of events performed by the other team, which leads to a non-favourable contribution to the value of the state. On the other hand, if the throw-in was successful, it should in most cases be followed by an action performed by the same team, which yields a favourable contribution.

4.4.5 Valuing Individual Player Actions

After the value iteration algorithm has converged, each state has been assigned a value. Once again, since all the information is included in the states, the value of a state equals the value of the involvement described by the state properties. This leads to the following function for the impact of an action, defined simply as

$$I_H(s) = Q_H(s). \tag{4.7}$$

As for Model 1, these values can be aggregated over a match or a whole season for each team or for each player. An important artifact with an impact function like this, is that the model prefers offensive involvements to defensive ones, due to them being closer to the objective of the game. It is also worth mentioning that a shot is valued on the basis of the value of taking a shot in the given context, $Q(s)$, not whether the shot ended up in a goal or not. Thus, a player that misses a shot receives the same value as a player that scored on the same shot. The motivation behind this is to be able to evaluate players on a basis where the outcome from shots is reduced. Hence, it can be said that the players are being evaluated in terms of their impact on creating scoring opportunities, where the player receives a high value if the involvement creates a scoring opportunity that is likely to result in a goal.

The output of the model can be utilised to examine which action is the best in a given context. If one wishes to find the best action in the context of for instance $TP = 3,\ MS = 1,\ MD = 0,\ H = 1$ and $Z = 3$, one can compare the different state values which has these properties. Table 4.10 shows an example of values for a pass in this context for illustrative purposes.

Table 4.10: Model 2: Pass values for the context $TP = 3, MS = 1, MD = 0, H = 1$ and $Z = 3$

Type	ZoneEnd	Outcome	Goal	StateId	Occurrences	Value
Pass	1	1	0	2856	41	0.0013
Pass	2	1	0	466	469	0.0023
Pass	3	1	0	465	1219	0.0045
Pass	4	1	0	6204	198	0.0056
Pass	5	1	0	562	132	0.0082
Pass	6	1	0	528	161	0.0079
Pass	7	1	0	514	186	0.0065
Pass	8	1	0	7185	25	0.0091
Pass	9	1	0	25416	7	0.0407
Pass	10	1	0	17777	14	0.0074
Pass	11	1	0	9984	20	0.0108
Pass	2	0	0	17049	13	-0.0066
Pass	3	0	0	1382	153	-0.0005
Pass	4	0	0	1385	53	0.0002
Pass	5	0	0	6224	36	0.0005
Pass	6	0	0	6189	43	-0.0002
Pass	7	0	0	539	49	0.0003
Pass	8	0	0	10235	19	0.0012
Pass	9	0	0	8440	16	0.0023
Pass	10	0	0	21362	21	0.0012
Pass	11	0	0	24731	14	0.0020

Hypothetically, if a pass was the only option for a player being in this context, a successful pass to zone 9 is optimal. This can be seen in Table 4.10, with a successful pass to zone 9 having a value of 0.0407, which is the largest value in the rightmost column. Furthermore, it can be seen from the table that a successful pass is better than the corresponding unsuccessful one. It is worth noting that zone 12, 13, 14 and 15 are unreachable from zone 3 with a regular pass. A pass from zone 3 to one of these zones is considered a long pass by Opta

4.5 Model Evaluation

To evaluate the performance of the three models developed in this thesis, their reliability and validity are investigated individually.

The reliability of the xG Model is assessed by looking at the correlation coefficients between 2014 and 2015 for the performance measures used to rate players. High correlations would indicate that the model is able to rate players consistently across the seasons, and can give insight on how reliable the model is for decision making. To make inferences on the validity of the xG Model, the area under the ROC curve, introduced in Section 3.1, for the two shot models are assessed. A high area under the ROC curve is desirable, as it indicates that the models have good discrimination abilities.

The reliability of the two Markov game models are assessed by looking at the correlation between the aggregated impact value of the players per 90 minutes played across the two seasons. In a game like football, players are believed to perform at a relatively stable level across two seasons, which is the motivation for exploring the correlation between seasons as a measure of reliability. A player can improve or get worse from one season to another, but the difference should not be too large. Positive correlation coefficients between the two seasons would indicate that the models are able to assign high positive values to good players in both seasons and not to players that occasionally did something extraordinarily good. Only players that have played at least 900 minutes (one third of a season) in both 2014 and 2015 are considered in the assessment of the reliability. It is believed that the effects of randomness should be limited if a player has played at least 900 minutes.

Tests of the validity of the two Markov game models are done by looking at correlations to three benchmarks. No objective player ratings from Tippeligaen are known to the authors, which makes this part of the model evaluations especially challenging. However, possible indicators of the level of a player might be their market value, salary or subjective player ratings by journalists and pundits. Some of the largest Norwegian newspapers and TV stations provide player ratings after each match solely based on subjective opinions. They form the basis for an average value of a players performance during a season. Two of the best known providers of such ratings are the broadcaster TV2, who owns the TV rights of Tippeligaen, through their statistical web page Altomfotball(Altomfotball (2014) and Altomfotball (2015)) and the largest newspaper in Norway, VG (VG (2014) and VG (2015)). Investigating the correlation coefficients between the ratings of the two Markov game models and the two subjective player ratings can give insight on whether players that are considered the best by journalists, also are considered good players by the models. The ratings of the journalists pundits are subjective, thus observing a positive correlation coefficient between the ratings and the developed models would be desirable, but is not a goal in itself. The purpose of making such models is to ensure objective rating of football players free from human bias to measure performance and identify talent.

In addition, the results from the models are compared to the market value of the players provided by Transfermarkt (2016). For each season, the market value of the players are extracted from the website of Transfermarkt at year end 2014 and 2015. The values provide a more objective benchmark for the results from the Markov game models than the journalist ratings. Correlation coefficients are again the measure of how valid the results are. Positive correlation coefficients between values from Transfermarkt and the models would indicate that players with high market value also are the best in the rating from the models. These correlation tests are not done on the xG Model, since the purpose of the xG Model is to measure the efficiency, G/xG, of mainly offensive players. The measure of efficiency is not as important for a defender as for a forward, and hence the values are not directly comparable with the ratings from Altomfotball and VG, as well as the market values from Transfermarkt.

Chapter 5

Evaluating Results

In this chapter, the results from implementing the models described in Chapter 4 are presented and discussed. The variables included in the xG Model are first presented and interpreted. Further, the reliability and validity of the xG Model is evaluated, before the results in the form of a player rating are presented and discussed. For the two Markov game models, the reliability and validity are discussed before the evaluation of the results, which are presented as top 10 lists for the different player positions. Players that have been sold to foreign clubs, presumably due to a noticeable performance, are marked with an asterisk in all tables. In the last part of this chapter, the results are evaluated with regard to the research questions from Section 1.2.

Results from Tippeligaen 2015 are presented in this section, while the results from Tippeligaen 2014 can be found in Appendix B.1, C.1 and D.1 for the xG Model, Model 1 and 2, respectively.

5.1 Expected Goals Model

As mentioned in Section 4.2, the xG Model consists of three distinctive types of shots: regular shots, free kicks and penalties. First, the significant explanatory variables for regular shots and free kicks are presented and interpreted, before the reliability and validity of the two shot models are assessed. Ratings based on the obtained values from the xG Model are presented and discussed in the last part of this section.

Players are evaluated on five different performance measures: the number of goals (G), total xG, the number of shots (S), xG per shot (xG/S) and the number of goals divided by total xG (G/xG). xG/S assesses the average quality of each shot a player attempted, and is referred to as the quality of a shot hereafter. It is important to clarify that this performance measure has nothing to do with the quality of the actual executions of the shots, but is rather a measure of the average likelihood of scoring on the shots attempted. G/xG is a measure of how efficiently a player converted his shots into goals.

Variables in the xG Model

Following the model building approach described in Section 4.2.2, the significant variables and coefficients for the two types of shots are shown in Tables 5.1 and 5.2 for regular shots and free kicks, respectively. The correlation coefficients between the variables can be seen in Table B.1 in Appendix B.

Table 5.1: xG Model: Parameters for regular shots

Parameter	Description of Parameter	Coefficient Value	p-value	Parameter type
Constant		-0.7417	<0.01	
X_1	Distance to goal	-0.1161	<0.01	Continuous
X_2	Angle on goal	$+0.7416$	<0.01	Continuous
X_3	If shot occur after fast break	$+0.4982$	<0.01	Binary
X_4	Number of take-ons before shot	$+0.4340$	<0.01	Discrete
X_5	If shot is executed by a header	-0.6883	<0.01	Binary
X_6	If shot is assisted by cross	-0.5211	<0.01	Binary
X_7	If shot is assisted by through ball	$+0.9684$	<0.01	Binary
X_{MS}	Current match status	$+1.3361$	<0.01	Categorical
$X_{\text{H/A}}$	Match location (home or away)	-0.2700	<0.01	Binary
X_{Elo}	Difference in team strength	-0.0008	<0.01	Continuous

Table 5.2: xG Model: Parameters for free kicks

Parameter	Description of Parameter	Coefficient Value	p-value	Parameter type
Constant		$+9.1943$	<0.01	
X_1	Distance to goal	-0.2730	<0.01	Continous
X_2	Angle on goal	-11.8080	<0.01	Continous

These regression coefficients are inserted into Equation (3.4), estimating the likelihoods of scoring on regular shots and free kicks, given a set of contextual variables. The two models, in addition to the penalty conversion rate, which is equal to 0.7822, constitute the complete xG Model for Tippeligaen. All shots attempted in 2014 and 2015 are evaluated and assigned a likelihood of ending in a goal.

Consider the regression coefficients for regular shots in Table 5.1. In a football perspective it makes sense that the distance, angle (measured in radians) and how a goal is executed are parameters that are important for the conversion rate of shots, thereby supporting the values on the coefficients of X_1, X_2, X_3, X_5, X_6 and X_7. Distance and angle are highly negatively correlated, and a short distance gives the keeper shorter time to react, while a large angle gives a larger target to hit. A fast break is a situation where the defending team often do not manage to reestablish balance, which might result in many attackers against fewer defenders than for regular play. This is a reasonable explanation for the increase in the likelihood of scoring after a fast break. A cross as assist or an execution by a header are both considered to be factors that increase the difficulty of scoring a goal, yielding negative coefficients. Receiving a through ball before taking the shot often results in fewer defenders between the player and the goal, which increases the likelihood of scoring.

Some of the estimated coefficients from Table 5.1 are not easy to interpret at first look. It might not be intuitive that the number of take ons (X_4) before a shot increases the likelihood of scoring. However, these are succeeded take ons, which might reduce the number of defenders between the ball and the goal. The coefficients of current match status (X_{MS}), match location ($X_{\text{H/A}}$) and difference in team strength (X_{Elo}) also needs some explanation. The coefficient of X_{MS} is positive, hence, leading the match might increase the likelihood of converting a shot. If a team is losing, they are likely to move the team further up the field, thereby possibly leaving more space for the other team to attack. Furthermore, match location and difference in team strength have a similar impact on team tactics, but these variables have a negative coefficient. For match location this might be because the home team on average has a more offensive strategy, which can leave more space defensively. In addition, stronger teams have a tendency to put pressure on the opponent, forcing them to a defensive stance. This can lead to more defenders between the ball and the goal, decreasing the likelihood of converting a shot. A comment is appropriate on the small value of the coefficient of team strength (X_{Elo}) of 0.0008. It is by far the smallest in magnitude, but the Elo difference for the teams in Tippeligaen ranges from -350 to 350, and can therefore make a considerable impact on the shot values despite its small coefficient.

As can be seen from Table 5.2, the significant variables in the model for free kicks are distance to goal and angle on goal, which are both negative. Hence, an increase in distance or angle decrease the likelihood of scoring. This is not surprising for distance to goal, but may seem counter intuitive at first glance for the angle on goal. A possible explanation of the high negative coefficient value can be seen in Figure 4.3(c), where it seems that a large share of the free kicks attempted from the sides of the field, resulted in goals. However, a known theory in football is that taking a free kick slightly off-centre is better than straight in front of it. Having an angle could make it easier for the player to curve the ball when placing the free kick, and it often forces the goalkeeper to choose a side to cover in the goal. The distance to goal and the angle are negatively correlated, which means that an increase in distance reduces the angle on goal, as seen in Figure 4.4. Because of the negative signs of the coefficients, this leads to opposite impacts on the likelihood of scoring when moving further away from the goal. However, it is believed that the negative coefficient of the distance can make the impact from the angle on goal redundant when the angle becomes small enough. Therefore, the negative coefficient of the distance can make the value of the angle on goal redundant, thus decreasing the probability of scoring when moving the free kick away from the goal. After a thorough consideration of these features of the model, the negative sign of the angle on goal coefficient was accepted in the model.

When evaluating the two distinctive shot models, it is important comment on the data for which the coefficient estimates are based on. Several of the parameters are subject to human bias. Fast breaks, take ons and through balls are examples of parameters that are exposed to subjective considerations, and it is a possibility that human bias might introduce error into the coefficients. In addition, some of the parameters has been specified or calculated by the authors based on the data set from Opta (see Table 4.3). This introduces another source of error in the form of possible miscalculations.

5.1.1 Reliability

Table 5.3 shows the correlation coefficients of all five performance measures across the two seasons. Only outfield players who played more than 900 minutes in both 2014 and 2015 are considered, and the values are normalised per 90 minutes to evaluate players on equal terms. S/90 min has the highest coefficient of 0.8811, and shows that players attempted a relatively stable amount of shots per 90 minutes played across seasons. The relatively low correlation coefficient of xG/S of 0.3570 indicates that the average quality of the shots a player attempted across seasons was not necessarily the same.

Figure 5.1 shows the scatter plot of G/xG across the two seasons. This is the performance measure used to rank players from the xG Model. It has the lowest correlation coefficient of all five performance measures of 0.1067, which means that the consistency across the two seasons

Table 5.3: xG Model: Correlation coefficients of performance measures from 2014 to 2015

G/90 min	S/90 min	xG/90 min	xG/S	G/xG
0.6631	0.8811	0.8494	0.3570	0.1067

was very limited. It indicates that players are not able to replicate their efficiency rates across seasons. For comparison, players are much more consistent in both xG and in scoring goals, which is reflected in the correlation coefficients of xG/90 min and G/90 min of 0.8494 and 0.6631, respectively.

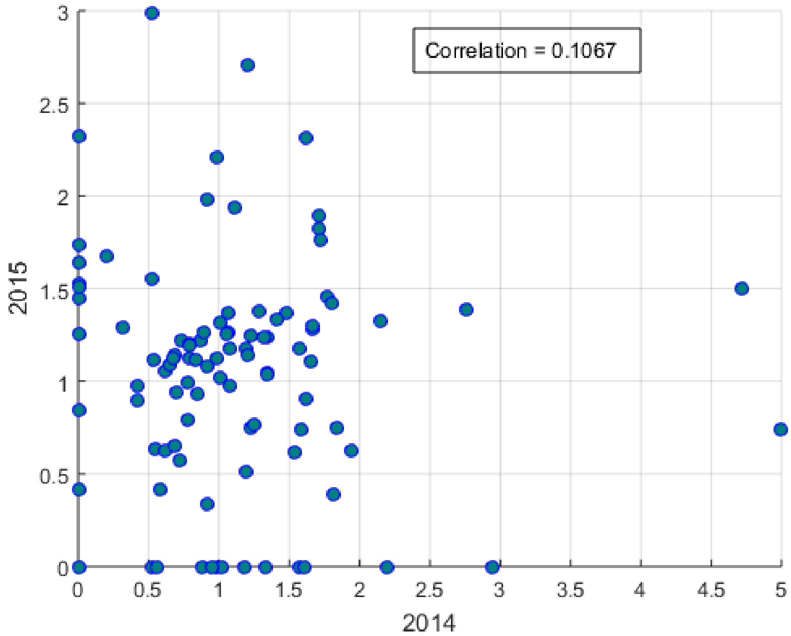
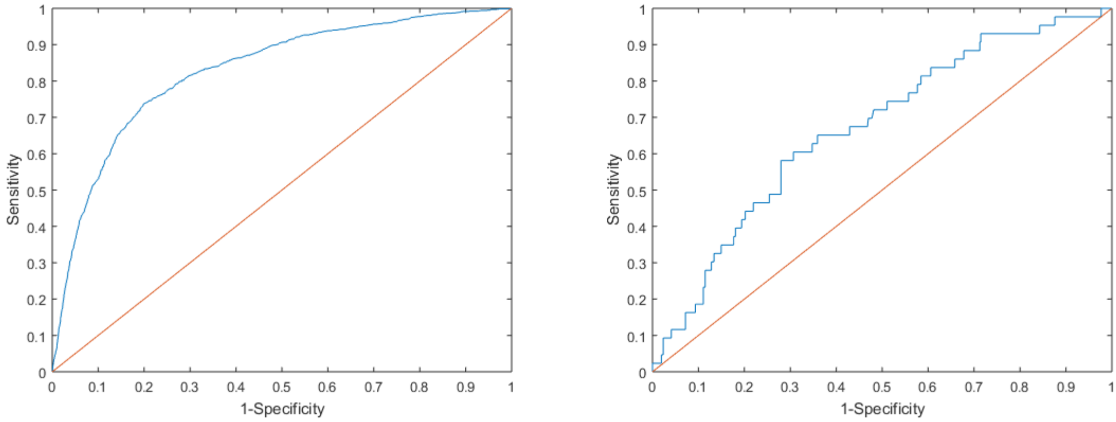


Figure 5.1: xG Model: Scatter plot of the G/xG of all players with the correlation coefficient between the 2014 and 2015 season

5.1.2 Validity

To assess the validity of the logistic regression models, consider the ROC curves for the two shot models shown in Figure 5.2. For regular shots, the area under the ROC curve is equal to 0.84, as seen in Figure 5.2(a). This value is classified from the general rule introduced in Section 3.1 as excellent discrimination ability. Thus, the model for regular shots is good at assigning observations with $Y = 1$ high likelihoods, and likewise low likelihoods to observations with $Y = 0$. For the free kicks, the area under the curve is $= 0.67$, which is classified as right below acceptable discrimination and can be seen in Figure 5.2(b).



(a) Regular shots, area under ROC curve equal to 0.84 (b) Free kicks, area under ROC curve equal to 0.67

Figure 5.2: xG Model: ROC curve for the two shot models

This limited ability to discriminate between the outcomes of free kicks might suggest that there is more to a free kick than only distance and angle. Only 558 free kicks are present in the data set, and another season of data could possibly help to improve the discrimination abilities of the model. Furthermore, a free kick is considered to be more dependent on the ability of the player than for regular shots, for which no variables are tested. Despite these weaknesses, it is believed that this model makes the likelihood estimation of the shots more accurate as opposed to a simple free kick conversion rate. In addition, free kicks only constitute 4.3 % of the total shots, and are therefore believed to have limited impact on the ratings for other players than regular free kick takers.

5.1.3 Results

Table 5.4 shows all players scoring more than eight goals in Tippeligaen 2015, evaluated on the five performance measures. The players in the table are ranked by G/xG. A value of G/xG greater than one means that the player performed well, scoring more goals than expected according to the model. Conversely, if the value is below one, the player performed below expectations. For reference, 14 of the players in the table have been sold to foreign clubs during or after the 2015 season.

The use of G/xG for rating players by efficiency is not unproblematic due to the sources of error that exist. Most importantly, defender proximity is not accounted for, a variable deemed significant by Lucey et al. (2014). Frequent goal scorers are possibly closer marked by defenders which could make their shots harder to convert. The free kick model introduces another possible source of error to the xG values, due to its limited discrimination ability. Of the players in the table, Trond Olsen, Pål Alexander Kirkevold and Pål André Helland have the highest share of shots coming from free kicks with 12 %, 9 % and 20 %, respectively. None of the remaining players have a higher share than 5 %. Another important source of error is the randomness that exists in football. Goals are a rare events, and a single goal can affect the G/xG value significantly. Extreme cases of players who scored few goals and seldom attempted shots can also occur. An example is goalkeeper Håkon Opdal, who would have had a G/xG value of around 2,400 because of his strike of luck on his only shot in 2015. However, this is to some extent rectified by only including players scoring eight goals or more.

Table 5.4: xG Model: All players scoring more than 8 goals in Tippeligaen 2015, ranked by G/xG

Name	Team	Position	Minutes played	Goals	xG	Shots	xG/S	G/xG
Trond Olsen	Bodø/Glimt	Winger	2510	13	6.55	74	0.09	1.98
Simon Diedhiou*	Haugesund	Forward	1945	9	4.90	52	0.09	1.84
Veton Berisha*	Viking	Forward	1256	11	6.57	47	0.14	1.67
Pål Alexander Kirkevold*	Sandefjord	Forward	1881	8	5.35	66	0.08	1.50
Kristoffer Ajer*	Start	Central midfielder	2581	8	5.43	43	0.13	1.47
Marcus Pedersen	Strømsgodset	Forward	835	11	8.20	34	0.24	1.34
Alexander Sørloth*	Bodø/Glimt	Forward	1776	13	9.89	58	0.17	1.31
Luc Kassi	Stabæk	Forward	2130	8	6.26	54	0.12	1.28
Alexander Søderlund*	Rosenborg	Forward	2242	22	17.44	87	0.20	1.26
Tobias Mikkelsen*	Rosenborg	Winger	1931	8	6.44	69	0.09	1.24
Ernest Asante	Stabæk	Winger	2643	10	8.06	70	0.12	1.24
Adama Diomandé*	Stabæk	Forward	1850	17	13.92	79	0.18	1.22
Zdeněk Ondrášek*	Tromsø	Forward	2366	9	7.57	74	0.10	1.19
Tommy Høiland	Molde	Forward	1103	9	7.93	36	0.22	1.13
Matthías Vilhjálmsson	Start/Rosenborg	Forward	2087	9	7.99	46	0.17	1.13
Mohamed Elyounoussi	Molde	Winger	2275	12	10.69	84	0.13	1.12
Erling Knudtzon	Lillestrøm	Forward	2595	10	8.91	51	0.17	1.12
Fredrik Nordkvelle	Odd	Attacking midfielder	1891	9	8.06	48	0.17	1.12
Iver Fossum*	Strømsgodset	Attacking midfielder	2653	11	10.07	69	0.15	1.09
Pål André Helland	Rosenborg	Winger	1502	13	12.07	89	0.14	1.08
Suleiman Abdullahi	Viking	Forward	1882	8	7.77	70	0.11	1.03
Christian Gytkjær	Haugesund	Forward	2545	10	9.84	57	0.17	1.02
Sander Svendsen	Molde	Forward	1720	8	8.13	60	0.14	0.98
Fred Friday	Lillestrøm	Forward	1770	11	11.22	67	0.17	0.98
Ola Kamara*	Molde	Forward	2384	14	14.69	89	0.17	0.95
Bentley	Odd	Winger	2464	8	8.53	67	0.13	0.94
Leke James*	Aalesund	Forward	2610	13	13.97	91	0.15	0.93
Jón Dadi Bödvarsson*	Viking	Forward	2067	9	10.06	58	0.17	0.89
Gustav Wikheim*	Strømsgodset	Winger	2413	9	11.35	68	0.17	0.79
Olivier Occéan	Odd	Forward	2208	15	20.23	89	0.23	0.74

* sold to foreign club during or after the 2015 season   ** on loan from a foreign club in 2015

Bodø/Glimt player Trond Olsen features on the top of the list for the 2015 season. He scored 13 goals, while the model expected him to score a mere 6.55 goals, which gives him a G/xG value of 1.98. Nine of his 69 shots were from free kicks, which could introduce uncertainty to his xG value. When examining only his regular shots, his G/xG value was 2.08. Simon Diedhiou of Haugesund is ranked second in efficiency with a G/xG value of 1.84. Furthermore, both Olsen and Diedhiou have an xG/S value of 0.09, which is the second lowest value of all. For Olsen, this might be because he plays as a winger, thus might often attempt shots with low angle on goal. The value of Diedhiou, on the other hand, is surprisingly low compared to his striker colleague Christian Gytkjær. A closer look shows that the goals of Olsen had an average xG of 0.15 and that he scored four goals with an xG below 0.04. The goals Diedhiou scored had an average xG of 0.21, and two of his nine goals had xG below 0.10. This indicate that these players converted low quality shots that influence their G/xG significantly. However, since the correlation coefficient of G/xG is low across seasons, it is believed that Olsen and Diedhiou might have a tough job replicating their efficiency in the future.

Other noticeable names on the list are Alexander Søderlund and Adama Diomandé, who were number one and two on the top scorer list for 2015, respectively (see Table A.5). Søderlund has slightly higher values in both xG/S and G/xG, as seen from the table. Further examination showed that the goals of Søderlund had an average xG of 0.37, while Diomandés goals had an average xG of 0.35. Thus, according to the model, the two forwards performed on a relatively even level. However, these two players are considered good candidates of forwards that possibly attract much attention from defenders, with both being Norwegian internationals. Because the defender proximity variable is not accounted for in the model, their shots might be estimated as easier than they actually were. Amongst the midfielders, the highest rated player is youngster Kristoffer Ajer who played for Start, who finished 14th in the 2015 season (see Table A.4). He is rated fifth in the model with a G/xG value of 1.47. When also taking into consideration his position as a central midfielder, this efficiency rate is especially noticeable.

Among the players in the bottom of the table are familiar names like Olivier Occéan and Gustav Wikheim. Occéan was the third highest goal scorer in 2015 with 15 goals, seemingly performing on a high level for Odd. However, his efficiency rate was poor according to the model, ranking him the least efficient player scoring eight goals or more, with G/xG equal to 0.74. This might indicate that he attempted a high number of good quality shots, but did not convert them efficiently. Further examination of Occéan showed that his goals on average had an xG value of 0.51, which means that many of his goals were considered to come from high quality shots. Former Strømsgodset winger Gustav Wikheim impressed many in the 2015 season, scoring 9 goals from his position as a winger. However, according to the model, his efficiency in 2015 was relatively poor, with a G/xG of only 0.79. Furthermore, he had an xG/S of 0.17, which is the highest among the wingers in Table 5.4. In addition, his goals had an average xG of 0.40. However, these players should not be deemed inferior based on these numbers. G/xG and xG/S shows low reliability across seasons, while xG/90 min has shown to be more stable. Having high xG/90 min values compared to other players in the same positions, it seems that these two players managed to create good scoring opportunities, and may yet be valuable assets for their teams in the future.

The performance measure xG/S deserves a final comment. From Table 5.4, it seems that the forwards on the teams that finished high on the table have the highest xG/S values. It might be reasonable to suggest that teams with higher strength are more likely to get into favourable positions for high quality shots, despite that the Elo variable accounting for team strength is included in the model.

As mentioned in Section 4.2, a league table can be obtained by looking at the xG values aggregated for each team. The league table obtained by looking at the difference in xG for all teams is shown in Table B.5. Rosenborg won the league in 2015, and also tops the list in xG difference, while Molde who finished sixth in the actual league table should have come second had they performed as expected by the model.

5.2 Markov Game Model 1

Results from Model 1 are presented and discussed in this section. As described in Section 4.5, the reliability and validity of the model are discussed first, then the results from running the model are presented and discussed. Top 10 lists for players in the following positions are presented: forwards, wingers, attacking midfielders, central midfielders, full backs and centre backs. The positions of all players are obtained from Transfermarkt (2016). Players are rated on the basis of their average impact per 90 minutes played, and only outfield players who played at least 900 minutes are considered. Goalkeepers are excluded because they are not suited for evaluation by this model. Player ratings from the 2014 season for the same positions can be found in Appendix C.

5.2.1 Reliability

A scatter plot of the players who played over 900 minutes in both 2014 and 2015, a total of 116 players, and their respective values in the two years are shown in Figure 5.3.

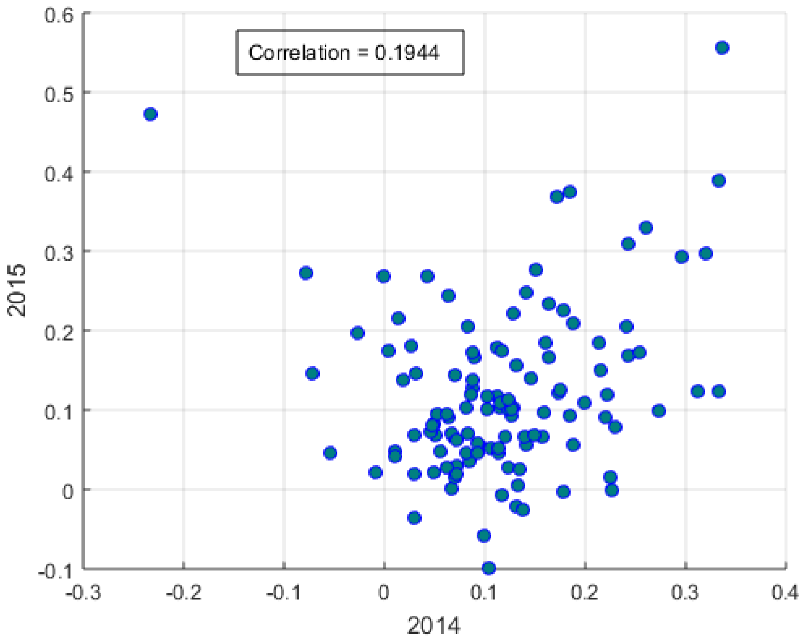


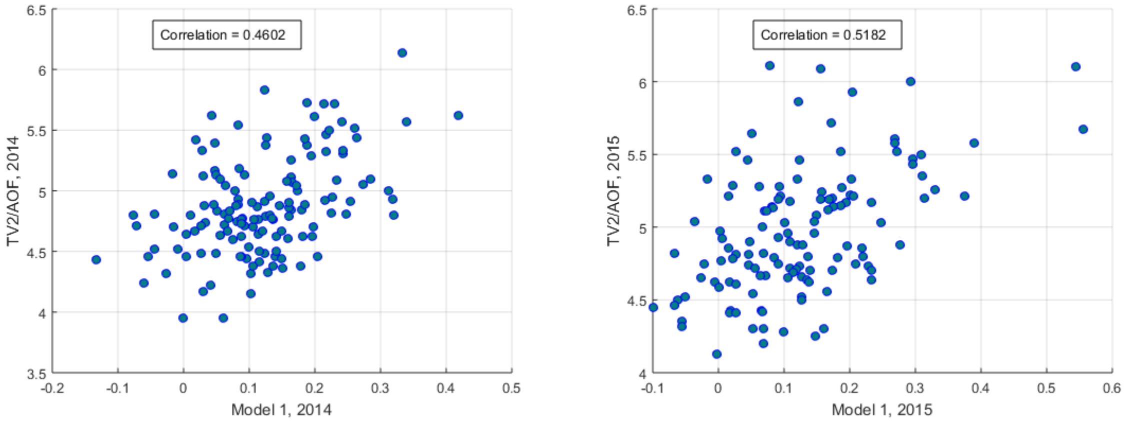
Figure 5.3: Model 1: Scatter plot and correlation between the two seasons

As can be seen from Figure 5.3, the correlation coefficient between the values of each player for the two seasons is positive. This points towards that the model might give high positive values to the better players and not to players who occasionally did something extraordinary. However, the correlation coefficient of 0.1944 might indicate that the reliability across the two seasons is weak. Compared to the G/xG performance measure, the correlation for Model 1 across seasons is slightly higher. Further assessment of the correlation coefficient is difficult because no similar models exists.

As can be seen from the scatter plot in Figure 5.3, some outliers are present in the data. One of them is Veton Berisha, who is represented by the dot in the upper left corner. He went from being rated as one of the worst players to one of the best in only one year. A closer look revealed that the numerous shots he took in 2014 did not result in many goals (G/xG = 0.21), and he was therefore punished by the model. In 2015, however, his efficiency was a lot better (G/xG = 1.67) which is impacting his value positively.

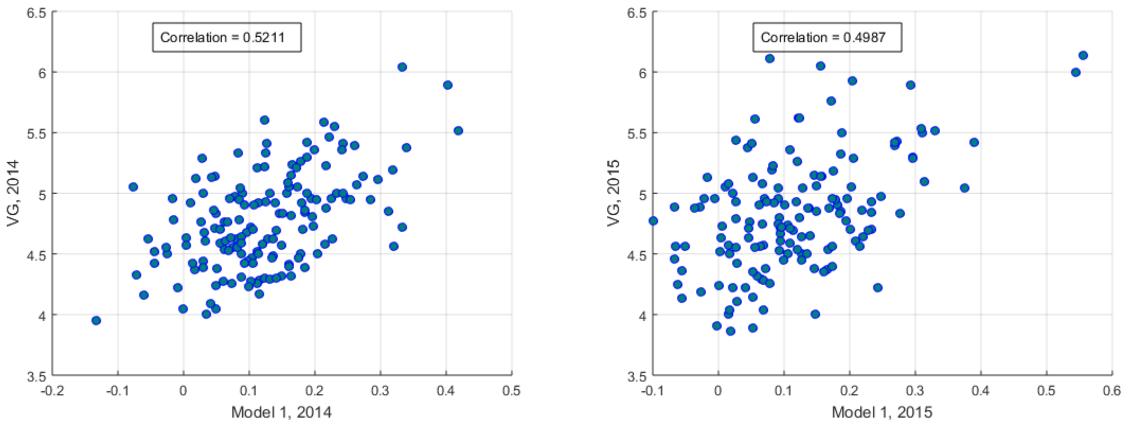
5.2.2 Validity

Figures 5.4 and 5.5 show scatter plots of the value of each outfield player with their respective rating from Altomfotball and VG. Altomfotball requires that a player has received ratings in at least 20 matches to be included in the final list. Similarly, VG requires ratings from at least 18 matches. This lead to a comparison of 137 and 126 players with Altomfotball and 157 and 149 with VG in 2014 and 2015, respectively.



(a) 2014 (b) 2015

Figure 5.4: Model 1: Scatter plots and correlation with Altomfotball

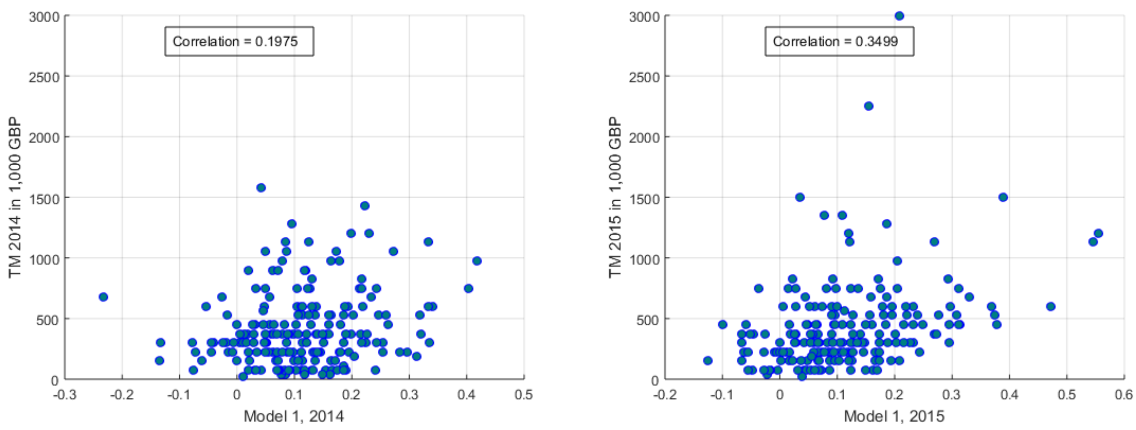


(a) 2014 (b) 2015

Figure 5.5: Model 1: Scatter plots and correlation with VG-børsen

From Figures 5.4 and 5.5 it can be seen that the correlation coefficient is positive and larger than 0.46 in all four cases. This indicates that the players that were considered the best by the journalists (that is, scores a high average rating across a season), also were considered good players by the model by delivering a high impact value per 90 minutes played. Some interesting outliers are evident in the figures. In 2015, Mike Jensen was rated above 6 by both VG and TV2, while scoring a value below 0.1 in the model, and is the upper left dot in both the figures for 2015. A closer examination of Mike Jensen is provided in Section 5.4.1. Another interesting observation is the rightmost dot in 2015, which is Pål André Helland. Helland was rated as the best player by the model and VG, but was only number 8 on the rankings provided by Altomfotball.

Figure 5.6 show scatter plots with the Transfermarkt values, which are given in the figure as 1,000 GBP. The players used for comparison consist of outfield players with over 900 minutes on the field. This lead to a sample of 204 and 202 players in 2014 and 2015, respectively.



(a) 2014 (b) 2015

Figure 5.6: Model 1: Scatter plots and correlation with Transfermarkt

Again, from the figure it can be seen that the correlation coefficients are positive, but for 2014 it is as low as 0.1975. However, there is more to the market value of a player than his performance over the last season. Factors such as age and nationality also influence the market value to some extent (see for instance Sæbø (2015)). Moreover, the model has a bias towards offensive players, but this is sometimes the case for market values as well (see Chapter 1). Two outliers in the figure for 2015 are Papa Alioune Ndiaye and Ole Kristian Selnæs, who are the two most valuable players at 3 and 2.25 million GBP, respectively. Veton Berisha, who is only valued at 600 thousand GBP, is rated as number three by the model and represents the bottom-right dot for 2015. For 2014, Berisha is the leftmost dot.

Table 5.5: Model 1: Correlations to VG-børsen, Altomfotball and Transfermarkt in 2014

	VG-børsen	Altomfotball	Transfermarkt	Model 1
VG-børsen	1			
Altomfotball	0.8111	1		
Transfermarkt	0.4478	0.4345	1	
Model 1	0.5211	0.4602	0.1975	1

Tables 5.5 and 5.6 show the correlation matrix for the VG and Altomfotball ratings, Transfermarkt and Model 1. As can be seen from Table 5.6, the correlations between Altomfotball and VG-børsen is as high as 0.8847 for the 2015 season. Model 1 has lower correlation with Transfermarkt than with VG-børsen and Altomfotball in both seasons. This result is considered to be expected, due to the numerous factors that most likely influence the market values from Transfermarkt. In general, the correlations are higher for the 2015 season than for 2014.

However, Transfermarkt is considered the most objective benchmark of the three. Therefore it is interesting to compare the models correlation with Transfermarkt, with the correlations between Transfermarkt and the ratings by VG and Altomfotball. For both 2014 and 2015 the ratings provided by VG and Altomfotball show a higher correlation with Transfermarkt than the model. A portion of this is believed to be due to a larger bias towards offensive players in the model compared with VG and Altomfotball.

Table 5.6: Model 1: Correlations to VG-børsen, Altomfotball and Transfermarkt in 2015

	VG-børsen	Altomfotball	Transfermarkt	Model 1
VG-børsen	1			
Altomfotball	0.8847	1		
Transfermarkt	0.6693	0.6684	1	
Model 1	0.4987	0.5182	0.3499	1

As a closing comment, the fact that Model 1 shows positive correlations with all three benchmarks is considered as favourable for the validity, despite none of them being a fully comparable source. All three are believed to capture player performance to some extent, which is the aim of Model 1. Players that have been sold to foreign clubs after the 2015 season are highlighted for reference, not to serve as validation but rather as observations of what players attracted attention from clubs outside Norway.

5.2.3 Results

The way Model 1 assigns values to shots is important to keep in mind for the results presented below. As described in Section 4.3, two types of artificial shot events are created in order to keep track of the outcomes of the shots. This means that for each state, the state-action pair corresponding to a shot in this state, $Q(s, Shot)$, is assigned a value equal to the likelihood of scoring when being in this state. With the impact function in Equation (4.5), a player can receive two different values when shooting from a given state. If he scores he would receive $I = 1 - Q(s, Shot)$, and if he misses he would receive $I = (\approx 0) - Q(s, Shot)$. The first number in the latter expression is not equal to zero, since there is some value to missing a shot due to the possibility of a rebound. For instance, if the likelihood of scoring in a given state is 0.15, a player would receive an impact of $I = 1 - 0.15 = 0.85$ if he scores and $I \approx -0.15$ if he misses. This makes shots and goals heavy contributors to the value of players in this model. However, if a player takes a lot of shots and rarely scores he will often get punished with negative values and thereby a lower rating. Hence, the values of the shots resemble G/xG. This is not an unreasonable way of being evaluated for forwards, but for players in other positions, where scoring goals is not the primary interest, it might be unfair.

Top 10 lists for the different positions are presented in Tables 5.7 to 5.12. Similar for all the tables are player names, minutes played and total impact value per 90 minutes. In addition, they show the total impact from shots and the five highest valued actions for the respective positions on average. When referring to the average value in the discussion below, it is always regarding the average value of the top 10 players for the position in question.

Consider the top 10 forwards shown in Table 5.7. The primary task of a forward is to score goals, and they are normally not as involved in the build up play as players further back on the field. Passes and other non-shot involvements are therefore less numerous, and the number of shots attempted is higher.

Table 5.7: Model 1: Top 10 forwards, Tippeligaen 2015.
Average total value = 0.1124, Minimum total value = -0.1248

Player	Team	Minutes	Total	Shot	Pass	Take on	Foul won	Cross	Carry
Adama Diomandé*	Stabæk	1850	0.5448	0.2736	0.0252	0.0287	0.0317	0.0454	0.0519
Veton Berisha*	Viking	1256	0.4721	0.3702	0.0475	-0.0167	0.0063	0.0097	0.0562
Alexander Søderlund*	Rosenborg	2242	0.3896	0.3674	-0.0244	0.0003	0.0094	0.0047	0.0154
Tommy Høiland	Molde	1103	0.3771	0.2945	0.0077	0.0175	0.0045	0.0090	0.0099
Fred Friday	Lillestrøm	1770	0.3130	0.1145	-0.0011	0.0574	0.0207	0.0034	0.0954
Alexander Sørloth*	Bodø/Glimt	1776	0.3108	0.2304	-0.0481	0.0258	0.0154	0.0091	0.0250
Luc Kassi	Stabæk	2130	0.2759	0.1064	0.1279	0.0156	0.0140	0.0051	-0.0036
Erling Knudtzon	Lilestrøm	2595	0.2481	0.1051	0.0665	0.0139	0.0118	0.0030	0.0374
Simon Diédhiou*	Haugesund	1945	0.2327	0.1287	0.0097	0.0285	0.0129	-0.0224	0.0431
Matthías Vilhjálmsson	Start/Rosenborg	2087	0.1961	0.0941	0.0450	0.0047	0.0060	0.0146	0.0317
Top 10 average			0.3360	0.2085	0.0256	0.0175	0.0133	0.0082	0.0362

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

According to Model 1, Adama Diomandé was the best forward in Tippeligaen 2015. He obtained most of his total value from accurate shooting, which resulted in 17 goals in 2015, in addition to scoring above average in five out of six categories. Similar observations can be made regarding the shot value for most of the other forwards, especially top scorer Alexander Søderlund, who seems to have limited impact on games beside scoring goals. This illustrates that these two forwards, who were the top two goalscorers, seem to be quite different players. An exception to the observation regarding the impact of shots, is Fred Friday. As seen, his value for shots was approximately the same as for his ball carries. He also had the highest impact from take ons for the players in the table. Another exception is Luc Kassi, who had a higher impact from passes than for shots. These observations show the importance of shots in Model 1, but it seems that forwards with good passing abilities and dribbling skills also are rewarded. One noticeable player that is missing in the table is Marcus Pedersen, who was bought by Strømsgodset in August 2015. Pedersen scored 11 goals on his 10 games that fall, and would have been at the top of the list of forwards if he had played enough minutes (he only played 835 minutes).

Table 5.8: Model 1: Top 10 wingers, Tippeligaen 2015.
Average total value = 0.1124, Minimum total value = -0.1248

Player	Team	Minutes	Total	Shot	Pass	Take on	Foul won	Cross	Carry
Pål André Helland	Rosenborg	1502	0.5553	0.2431	0.0381	0.0055	0.0357	0.0021	0.1153
Trond Olsen	Bodø/Glimt	2510	0.3749	0.1932	0.0374	0.0244	0.0110	0.0503	0.0503
Moryké Fofana*	Lillestrøm	1300	0.3683	0.1777	0.0728	0.0400	0.0109	-0.0013	0.0098
Zymer Bytyqi	Viking	1258	0.3111	0.0527	0.0236	0.0130	0.0043	0.1469	0.0315
Ernest Asante	Stabæk	2643	0.2961	0.0470	0.0850	0.0226	0.0172	0.0206	0.0741
Gustav Wikheim*	Strømsgodset	2413	0.2935	-0.0084	0.1016	0.0708	0.0037	0.0528	0.0439
Espen Børufsen	Start	2154	0.2209	0.0815	-0.0075	0.0107	0.0084	0.0619	0.0145
Ole Jørgen Halvorsen	Odd	1393	0.2160	0.0144	0.0047	-0.0057	0.0038	0.1064	0.0295
Magnus Andersen	Tromsø	2663	0.1950	0.0428	0.0321	-0.0030	0.0036	0.0343	0.0374
Mohamed Elyounoussi	Molde	2275	0.1855	0.0892	0.0513	-0.0051	0.0185	-0.0146	0.0165
Top 10 average			0.3017	0.0933	0.0439	0.0173	0.0117	0.0459	0.0423

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

Table 5.8 shows the top 10 wingers. Wingers play on the sides of the field and are normally heavily involved in crossing and should be good in one-on-one situations to be able to get into good crossing positions. The impact function used to value the player involvements, Equation (4.5), takes into account what a given action lead to. For the wingers, this has at least one important consequence to be aware of. A cross that resulted in a shot from a teammate is valued higher than a cross that resulted in for example an interception by the opposing team. This means that a cross can be perfectly hit into the area from the winger without him getting full reward if the finisher did not do a good job. As pointed out in Secion 4.3.3, this way of rewarding is believed to be fair in the long run.

Pål André Helland tops the list of wingers, and he is in fact the highest rated player of all in Model 1. Helland also tops the player ratings from VG (2015). His total impact value comes primarily from shots and ball carries, where he has the two highest values in the table. On the other hand, it is surprising to see his low impact value from crosses. Trond Olsen, considered by the xG Model as the most efficient player, is second on the list. Besides from shots, most of his impact came from crosses and ball carries which are important for players in his position. The two players with the highest impact from crosses were Ole Jørgen Halvorsen and Zymer Bytyqi. They also had below average impact from shots, and their crossing abilities seemed to be their primary contribution. The only winger with negative impact from shots is Gustav Wikheim, but his total value came primarily from passes and take ons where he has the highest impact values in the table. His high impact value from passes might imply that he chooses to pass more often than the others.

As for the forwards, shots are still a heavy contributor to the total impact value of the wingers. However, it seems that Model 1 is able to give value to wingers with good crossing abilities and to players that are good at passing and take ons.

Now consider the top 10 attacking midfielders shown in Table 5.9. Attacking midfielders are expected to deliver key passes and assists to the forwards, in addition to being a threat to the opposition goal from long range shooting. They can have fewer defensive tasks than central midfielders, and are often technically gifted players with good passing and dribbling abilities.

Table 5.9: Model 1: Top 10 attacking midfielders, Tippeligaen 2015.
Average total value = 0.1124, Minimum total value = -0.1248

Player	Team	Minutes	Total	Shot	Pass	Cross	Corner	Carry	FK
Fredrik Nordkvelle	Odd	1891	0.2719	0.1954	0.0041	0.0174	-0.0011	-0.0069	-0.0006
Daniel Fredheim Holm	Vålerenga	1974	0.2687	0.1749	0.0404	0.0097	0.0026	-0.0116	0.0021
Michael Barrantes*	Aalesund	1000	0.2261	0.0503	0.0358	-0.0020	0.0122	0.0397	0.0297
Eirik Hestad	Molde	941	0.2161	0.0232	0.0963	0.0041	0.0195	-0.0072	0.0381
Papa Alioune Ndiaye*	Bodø/Glimt	1286	0.2087	-0.0279	0.0383	0.0189	0.0098	0.0321	-0.0010
Ghayas Zahid	Vålerenga	2383	0.2055	0.0082	0.0851	-0.0040	-0.0007	0.0479	-
Iver Fossum*	Strømsgodset	2653	0.2045	0.0744	0.0371	0.0174	-0.0003	0.0226	-0.0030
Gjermund Åsen	Tromsø	1709	0.2026	-0.0156	0.0095	0.0334	0.0814	0.0355	0.0277
Henrik Furebotn	Bodø/Glimt	2157	0.1731	0.0769	-0.0008	0.0395	0.0030	0.0047	0.0241
Thomas Kind Bendiksen*	Molde	931	0.1407	-0.0503	0.0233	0.0156	0.0900	0.0084	0.0332
Top 10 average			0.2118	0.0509	0.0369	0.0150	0.0216	0.0165	0.0150

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

For the two players on top, Fredrik Nordkvelle and Daniel Fredheim Holm, the total impact value came primarily from shots. They scored nine and seven goals respectively, and had by far the highest impact value from shots of all the attacking midfielders. Third placed Michael Barrantes had a much lower impact value from shots, but obtained a significant portion of his impact value from ball carries and by executing free kicks accurately. Eirik Hestad had the highest impact value from passes, in close race with Ghayas Zahid, where the latter also seems to have performed well on ball carries. Fifth place on the list, Papa Alioune Ndiaye, is a player with no particularly large impact values. However, he was above average in many of the categories not shown in the table like long passes, take ons and ball recoveries. This is a nice example of how Model 1 is able to give a high total value to players performing well on a variety of actions, like seen earlier for Fred Friday and Gustav Wikheim.

Next, consider the top 10 central midfielders in Table 5.10. Playing as a central midfielder requires a wide set of skills, including passing, duel play, tackling and strong physical condition. They are less likely to get into scoring positions than the attacking midfielders, which might influence the impact values of these players. A more defensive role also means that they more seldom play passes that result in shots or other goal scoring opportunities. Nonetheless, the central midfielders who regularly wins tackles, intercepts, wins duels and is accurate in his passing, should be rewarded in the model.

Table 5.10: Model 1: Top 10 central midfielders, Tippeligaen 2015.
Average total value = 0.1124, Minimum total value = -0.1248

Player	Team	Minutes	Total	Shot	Pass	Tackle	Ball rec	Cross	Corner
Christian Grindheim	Vålerenga	2675	0.3087	0.0985	0.1239	0.0069	0.0105	0.0040	0.0125
Malaury Martin	Lillestrøm	975	0.2947	0.1446	0.0453	0.0045	0.0088	0.0103	0.0336
Giorgi Gorozia	Stabæk	1467	0.2326	-0.0578	0.1369	0.0065	0.0241	0.0063	0.0858
Kristoffer Ajer*	Start	2581	0.2187	0.1016	0.0243	0.0077	0.0122	0.0066	-
Bismark Adjei-Boateng**	Strømsgodset	1293	0.1847	0.0590	0.0192	0.0214	0.0026	0.0440	-0.0002
Morten Konradsen	Bodø/Glimt	1362	0.1740	0.1439	-0.0058	0.0044	0.0071	0.0202	-0.0009
Kamal Issah	Stabæk	1846	0.1733	0.0274	0.0838	0.0143	0.0163	0.0082	-
Fredrik Midtsjø	Rosenborg	2396	0.1718	0.0227	0.0833	0.0152	0.0118	0.0010	-
Johan Andersson	Lillestrøm	973	0.1664	0.1054	0.0429	0.0046	0.0117	-0.0056	-0.0014
Ole Kristian Selnæs*	Rosenborg	1951	0.1553	0.0157	0.0782	0.0133	0.0217	0.0226	-0.0025
Top 10 average			0.2080	0.0661	0.0632	0.0099	0.0127	0.0118	0.0127

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

Christian Grindheim of Vålerenga is the highest rated central midfielder. He had an above average shot value, and his impact from passes is the second highest in the table. This is not surprising considering his five goals and eleven assists in 2015. Third place on the list, Giorgi Gorozia actually had a negative impact value from shots. However, he had the highest impact value for passes, ball recoveries and corners of the players in the table. Bismark Adjei-Boateng had the highest impact from tackles and also was the best to hit crosses amongst the central midfielders. As was the case for the attacking midfielders, players with no exceptionally high values are rated high, like Rosenborg players Fredrik Midtsjø and Ole Kristian Selnæs, are also appreciated by Model 1. The latter had the second highest impact value from ball recoveries, for which he was praised by pundits and journalists throughout the season. Good examples of players that are on the list due to their impact from shots, are Morten Konradsen and Johan Andersson.

A noticeable player that is missing on the list of central midfielders is RBK player Mike Jensen. He was considered by many pundits and journalists as the best player in 2015. Altomfotball (2015) regarded him as the best player, while according to VG (2015) he was second only to Pål André Helland. A closer investigation on why he is not present on the list revealed that he took a lot of shots and scored few goals. He attempted 81 shots, seventh overall in Tippeligaen and scored only three goals, which lead to a G/xG equal to 0.51. As described earlier, this is very unfavourable with this model specification and makes him drop well out of top 10. Not even his 13 assist are enough to get him on the list. To illustrate how significant an impact his shots had on his ranking; if shots were excluded from the model, he would have been the second best central midfielder. A case study of the involvements Mike Jensen had in 2014 and 2015 is shown in Section 5.4.1.

Table 5.11 shows the top 10 full backs in Tippeligaen 2015. Model 1 has some flaws when it comes to valuing players in defensive positions. Due to how the data is built with events around the ball, it is not possible to capture when a full back or centre back should have been in position to interfere but was not. This makes it harder to punish defenders for mistakes. In addition, they often get punished for clearing the ball to corner and throw-ins, typically where state transitions to the other team occur. However, when players in the same positions are compared, it is possible to say something about the relative performance of players.

A modern full back has both defensive and offensive responsibilities. In defence, they should be good in one-on-one situations, tackling and positional play setting up possible interceptions. Solid passing ability is also an important aspect for the full backs, especially if his team plays a possession-oriented style. In attacking play, the full backs are often combining with the wingers and attacking midfielders to get into good crossing positions.

Table 5.11: Model 1: Top 10 full backs, Tippeligaen 2015.
Average total value = 0.1124, Minimum total value = -0.1248

Player	Team	Minutes	Total	Shot	Pass	Cross	Ball rec	Corner	FK
Per-Egil Flo	Molde	1998	0.3290	0.0396	0.1003	0.0327	0.0092	0.0700	0.0087
Espen Ruud	Odd	2476	0.2953	0.0457	0.0925	0.0831	0.0070	0.0087	0.0204
Lars-Christopher Vilsvik	Strømsgodset	2126	0.2692	0.0542	0.0197	0.1462	0.0134	0.0153	-0.0021
Jo Nymo Matland*	Aalesund	1425	0.2429	0.0605	-0.0015	0.0426	0.0049	0.0580	0.0501
André Danielsen	Viking	2700	0.2329	0.0995	0.0428	0.0459	0.0155	0.0179	0.0050
Kent-Are Antonsen	Tromsø	2248	0.2275	0.0557	0.0029	0.0277	0.0160	0.0102	0.0261
Joachim Olsen Solberg	Mjøndalen	2486	0.2087	-0.0588	0.0230	0.0192	0.0143	0.0508	0.1031
Zarek Chase Valentin*	Bodø/Glimt	2074	0.2006	0.0056	0.0112	0.0498	0.0114	-0.0025	0.0014
Jørgen Skjelvik	Rosenborg	1812	0.1855	0.0154	0.0801	0.0449	0.0120	0.0070	0.0006
Mikael Dorsin	Rosenborg	1582	0.1492	0.0365	0.0294	0.0210	0.0106	-0.0011	-0.0011
Top 10 average			0.2341	0.0354	0.0400	0.0513	0.0114	0.0234	0.0212

* sold to foreign club during or after the 2015 season   ** on loan from a foreign club in 2015

Per-Egil Flo is the highest rated full back in Tippeligaen 2015. As can be seen in Table D.2, Flo came out second in 2014, which shows he has been performing consistently on a high level. He had the highest impact value from passes of the players in the table, and his corner kicks also contributed significantly to his total value. Interestingly, his value from crossing is below average. Number two on the list, Espen Ruud also had a high impact from passing, but unlike Flo had an above average value from crosses amongst the full backs. Lars-Christopher Vilsvik is the full back with the best impact value from crosses by a wide margin.

All of the top three full backs are considered good offensive players, and this is supported by the results from Model 1. As one might expect, none of the full backs, except André Danielsen, got a high impact from shots. Additionally, a noticeable player missing is Rosenborg player Jonas Svensson. He was rated 8th and 6th in VG (2015) and Altomfotball (2015) respectively, but he is not in the list for the same reason as for Mike Jensen. He did not score any goals in the 2015 season, while attempting 25 shots.

Finally, consider the centre backs in Table 5.12, which are the core of the defence in football teams. They should be good at tackling, aerial duels and positional play in order to stop the attacks coming from the opposition. Again, the way the impact function is specified, actions such as interceptions, clearance and aerial duels are rewarded on the basis of what they lead to.

Table 5.12: Model 1: Top 10 centre backs, Tippeligaen 2015.
Average total value = 0.1124, Minimum total value = -0.1248

Player	Team	Minutes	Total	Shot	Pass	Clearance	Ball rec	Cross	Long pass
Johan Bjørdal	Rosenborg	1093	0.2398	0.0461	0.1655	-0.0001	0.0107	-0.0026	0.0012
Rhett Bernstein*	Mjøndalen	1234	0.1927	0.1214	0.0301	-0.0068	0.0077	0.0002	-0.0038
Joona Toivio	Molde	1706	0.1781	-0.0190	0.0665	0.0274	0.0093	0.0650	0.0125
Andreas Nordvik*	Sarpsborg 08	1846	0.1667	0.0532	0.0586	0.0168	0.0065	0.0036	0.0403
Brede Moe	Bodø/Glimt	2408	0.1657	0.0544	0.0545	0.0304	0.0072	-0.0016	-0.0082
Lars-Kristian Eriksen	Odd	2340	0.1456	0.0235	0.0746	0.0132	0.0084	0.0003	0.0110
Ole Christoffer Heieren Hansen	Sarpsborg 08	2059	0.1383	0.0634	0.0578	0.0337	0.0051	0.0000	-0.0025
Vegard Forren	Molde	2416	0.1196	-0.0375	0.0902	0.0199	0.0069	0.0052	0.0256
Morten Sundli	Mjøndalen	2018	0.1050	0.0314	0.0381	-0.0137	0.0092	0.0029	-0.0034
Steffen Hagen	Odd	2610	0.1000	0.0174	0.0813	-0.0033	0.0089	-0.0002	0.0079
Top 10 average			0.1551	0.0354	0.0717	0.0117	0.0080	0.0073	0.0081

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

RBK defender Johan Bjørdal was the best central defender according to Model 1. Surprisingly, he had the highest impact value from passes of all players in Tippeligaen 2015, and is well clear of his centre back colleagues in this area. His value from ball recoveries also contributed to his total impact value. Rhett Bernstein was the highest goal scorer amongst the centre backs, which is reflected in his impact value for shots. Andreas Nordvik had the highest impact value from long passes, which might be an important attribute of a centre back, while Ole Christoffer Heieren Hansen had the highest value for clearances. For the defenders, it seems that Model 1 is to some extent able to appreciate defensive attributes. Furthermore, the impact values are smaller for the defensive actions, hence, making shots, passes and other more offensive actions heavier contributors.

As mentioned in Section 4.3.4, it is possible to form a league table based on the results from the Markov game models. Results from doing this for Model 1 are shown in Appendix C, Tables C.3 and C.6 for 2014 and 2015, respectively. In both the tables, the estimated table and the real table does not differ much. This is not very surprising due to the heavy effect shots have on the values. However, Table C.6 can once again indicate that Molde underachieved compared to what could be expected in 2015.

5.3 Markov Game Model 2

The results from Markov Game Model 2 are presented and discussed in this section. The structure of the section is the same as for Model 1 in the previous section. Again, only outfield players who played at least 900 minutes are considered. Player ratings from the 2014 season can be found in Appendix D.

5.3.1 Reliability

A scatter plot of the players who played over 900 minutes in both 2014 and 2015, a total of 116 players, and their respective values in the two years are shown in Figure 5.7.

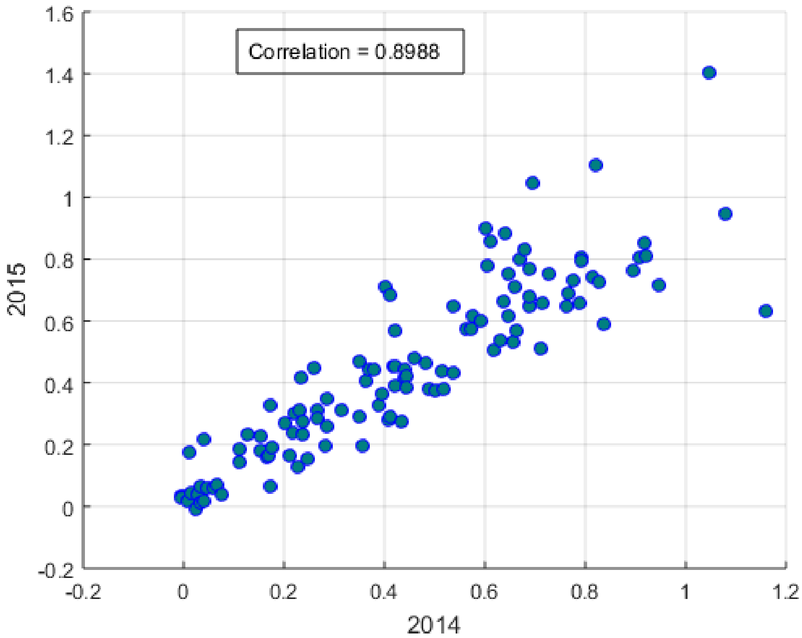
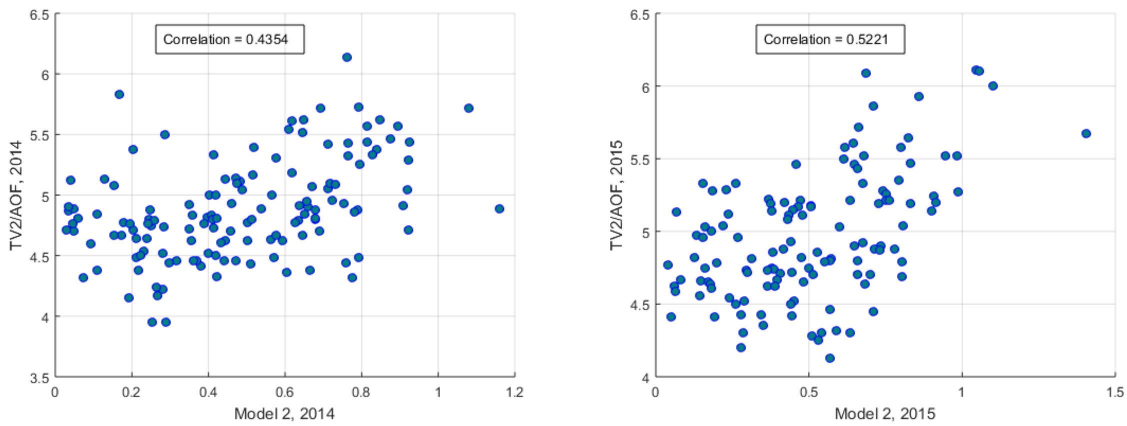


Figure 5.7: Model 2: Scatter plot and correlation between the two seasons

From Figure 5.7 it can be seen that the correlation coefficient is positive, like for Model 1. However, with a value of 0.8988, the coefficient is substantially higher for Model 2. The coefficient shows excellent correlation across the two seasons, and thus the model seems to be very stable. As mentioned earlier, this might indicate that the model is favoring the best players instead of assigning high values to players that occasionally have a big impact on the matches. Like for Model 1, no similar models are available for comparison. The dot in the upper right corner on the figure represents Pål André Helland, who performed very well in both seasons according to the model. The rightmost dot is Trond Olsen, who in this model seemed to have a very good season in 2014, while 2015 was more average. It is worth mentioning that the results for Trond Olsen in Model 1 is opposite, his 2014 season was average while he was very good in 2015.

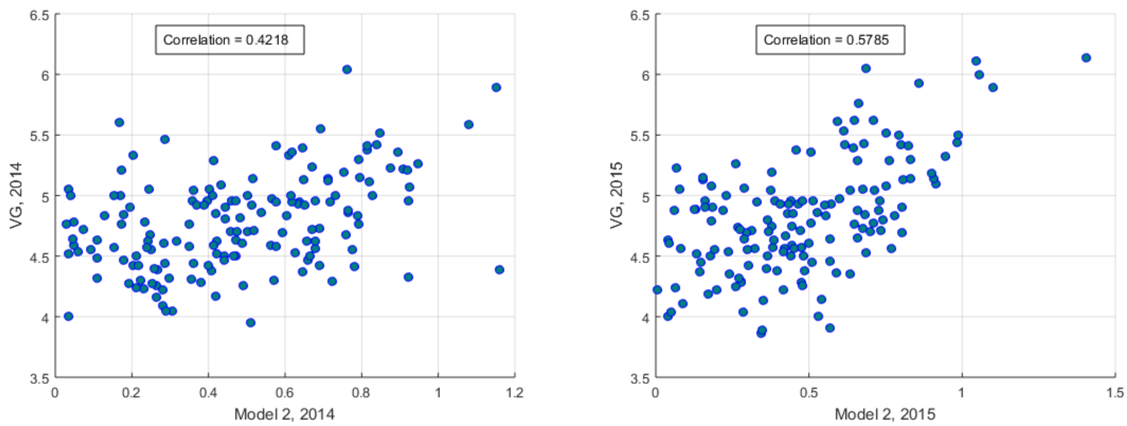
5.3.2 Validity

Figures 5.8 and 5.9 shows scatter plots of the value of each player with their respective rating from Altomfotball and VG. The players considered in these figures, are the same as for Model 1.



(a) 2014 (b) 2015

Figure 5.8: Model 2: Scatter plots and correlation with Altomfotball

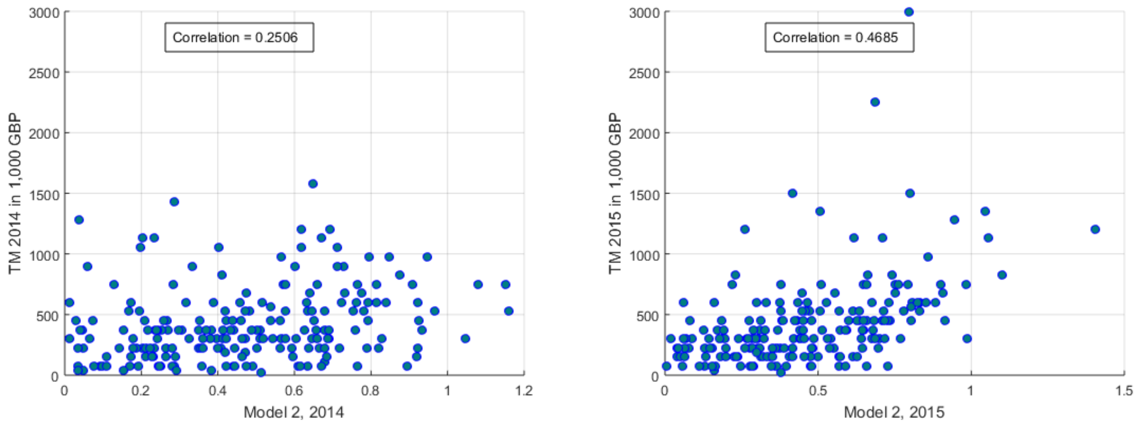


(a) 2014 (b) 2015

Figure 5.9: Model 2: Scatter plots and correlation with VG-børsen

From Figures 5.8 and 5.9 it is evident that the correlation is positive in all four cases. The correlation coefficients are a bit lower in two out of four cases compared to Model 1. All the coefficients have values above 0.42 nonetheless, which yields the same conclusion as for the previous model. Players that are considered best by the journalists also tend to be considered good by the model. Some outliers in the figures are worth mentioning. The rightmost dot in both Figure 5.8(a) and Figure 5.9(a) is Trond Olsen, who performed very well according to the model while only achieving average ratings by the media. As previously pointed out, Altomfotball rated Pål André Helland as number 8 in 2015, while VG had him as number one. This can be seen in the figures for 2015, where Helland is the rightmost dot. One last player that is worth pointing out is Martin Ødegaard who represents the upper right dot in Figure 5.9(a). Ødegaard did not play enough to be included in the final list by Altomfotball, and is therefore not included in Figure 5.8(a).

The scatter plots and comparison with Transfermarkt values can be seen in Figure 5.10.



(a) 2014 (b) 2015

Figure 5.10: Model 2: Scatter plots and correlation with Transfermarkt

It can be seen from Figure 5.10 that the correlation coefficients are positive for both seasons. Model 2 shows improvements compared to Model 1, yiedling higher correlation coefficients for both seasons. As pointed out earlier, there is more to the market value of a player than their performance in a given season. Again, Papa Alioune Ndiaye and Ole Kristian Selnæs are two conspicuous outliers with their high market values in 2015. The two highest market values in 2014, were defenders Lars-Christopher Vilsvik and Vegard Forren. It seems that Model 2 assigned average impact values to the two, and this is possibly due to the models bias towards offensive involvements and players, which is discussed further in the next section.

Table 5.13: Model 2: Correlations between VG-børsen, Altomfotball, Transfermarkt and Model 1 in 2014

	VG-børsen	Altomfotball	Transfermarkt	Model 1	Model 2
VG-børsen	1				
Altomfotball	0.8111	1			
Transfermarkt	0.4478	0.4345	1		
Model 1	0.5211	0.4602	0.1975	1	
Model 2	0.4218	0.4354	0.2506	0.3597	1

As for Model 1, the correlation matrix for Model 2 to the three benchmarks is shown in Table 5.13 and 5.14. Again in general, the correlations in 2014 are lower than in 2015. This is also the case for the correlations between the two models. The biggest difference from Model 1 is the correlations with the Transfermarkt values. As was the case for Model 1, VG and Altomfotball achieves a higher correlation with Transfermarkt than Model 2. Again it is worth mentioning that this is believed to be mainly due to a larger bias towards offensive players in the models than in the ratings by VG and Altomfotball. The positive correlations between Model 2 and the three benchmarks, is considered as favourable for the validity. All three benchmarks are believed to capture player performance to some extent, which is also the aim of Model 2.

Table 5.14: Model 2: Correlations between VG-børsen, Altomfotball, Transfermarkt and Model 1 in 2015

	VG-børsen	Altomfotball	Transfermarkt	Model 1	Model 2
VG-børsen	1				
Altomfotball	0.8847	1			
Transfermarkt	0.6693	0.6684	1		
Model 1	0.4987	0.5182	0.3499	1	
Model 2	0.5785	0.5221	0.4685	0.5567	1

Like for Model 1, players that have been sold to foreign clubs after the 2015 season are highlighted for reference only.

5.3.3 Results

The way Model 2 assigns values to players is important to keep in mind when interpreting the results presented below. Because of how a state is defined for Model 2, described in Section 4.4, the value of a state is actually the value of that specific involvement. This has some implications for how players are given impact, and especially for shots. By the definition of the impact function, $I(s) = Q(s)$, a player is assigned the value of attempting a shot, regardless of whether it ended in a goal or not. Hence, the values for shots in this model resembles that of xG to a larger extent than G/xG. This results in almost entirely positive shot values, and hence, the punishment for missing shots, which was evident in Model 1, is limited.

Like for Model 1, top 10 lists for the different positions are presented in Tables 5.15 to 5.20. Similar for all the tables are player names, minutes played and total impact value per 90 minutes. In addition, they show the total impact from shots and the five highest valued actions for the respective positions on average. When referring to the average value in the discussion below, it is always regarding the average value of the top 10 players for the position in question.

Table 5.15 shows the top 10 lists for forwards in Tippeligaen 2015. Adama Diomandé is on top of the list, just as he was in Model 1. As can be seen from the table, he obtained most of his total impact value from shots, passes and ball carries. His value for take ons is also above average, which can indicate that Diomandé is a forward who performed well on a number of important attributes in his position. Olivier Occéan is number two on the list. He had a higher impact value from his shots than Diomandé (reflected in his high xG value in Table 5.4) and his impact value from aerial duels is the highest for the forwards in the table. Third placed Fred Friday had the highest impact value from both take ons and ball carries, which indicate he is good at bringing the ball into favourable positions and has the ability to dribble past defenders. Top scorer Alexander Søderlund is ninth on the list. He had a high impact value from shots, but is below average in all other categories except aerial duels. This further supports the suspicion that he did not contribute much in the build up play, but focused on scoring goals. In total, five out of top 10 are the same as for Model 1.

The top 10 wingers are shown in Table 5.16, where Pål André Helland is rated on top. In fact he is considered the best player of all in Model 2. As can be seen from the table, he had

Table 5.15: Model 2: Top 10 forwards, Tippeligaen 2015.
Average total value = 0.4608, Minimum total value = 0.0059

Player	Team	Minutes	Total	Shot	Pass	Take on	Foul won	Aerial	Ball carry
Adama Diomandé*	Stabæk	1850	1.0543	0.5001	0.2130	0.0656	0.0333	0.0312	0.1160
Olivier Occéan	Odd	2208	0.9867	0.5049	0.2432	0.0128	0.0407	0.0888	0.0616
Fred Friday	Lillestrøm	1770	0.9133	0.4026	0.1939	0.0740	0.0303	0.0115	0.1419
Ola Kamara*	Molde	2384	0.9085	0.4298	0.2208	0.0143	0.0246	0.0210	0.1069
Veton Berisha*	Viking	1256	0.8831	0.3770	0.2317	0.0165	0.0232	0.0258	0.1247
Sander Svendsen	Molde	1720	0.8309	0.3897	0.1975	0.0471	0.0101	0.0047	0.1174
Jón Dadi Bödvarsson*	Viking	2067	0.8080	0.3123	0.1932	0.0326	0.0273	0.0365	0.1328
Leke James*	Aalesund	2610	0.8047	0.3676	0.1920	0.0487	0.0212	0.0609	0.0723
Alexander Søderlund*	Rosenborg	2242	0.7993	0.4564	0.1677	0.0131	0.0279	0.0491	0.0550
Alexander Sørloth*	Bodø/Glimt	1776	0.7916	0.3829	0.1769	0.0233	0.0265	0.0719	0.0762
Top 10 average			0.8781	0.4123	0.2030	0.0348	0.0265	0.0401	0.1005

* sold to foreign club during or after the 2015 season   ** on loan from a foreign club in 2015

the highest and third highest impact values from shots and passes, respectively. He also had the highest value from ball carries in the list. This might show why he was considered a vital part in the Rosenborg team winning a domestic double in 2015. Second on the list, Gustav Wikheim had a below average impact value from shots, and the highest impact value from passes and take ons. In addition, he had an above average value for ball carries, which was also observed in the results for Model 1. Ninth on the list is Odd player Bentley, who had the highest impact value from crosses of the players in the list, just ahead of his teammate Ole Jørgen Halvorsen. This corresponds well with the fact that their teammate Olivier Occéan had the highest impact value from aerial duels among the forwards. These two are below average on all other attributes, which is reflected in their position on the list.

Table 5.16: Model 2: Top 10 wingers, Tippeligaen 2015.
Average total value = 0.4608, Minimum total value = 0.0059

Player	Team	Minutes	Total	Shot	Pass	Take on	Corner won	Cross	Ball carry
Pål André Helland	Rosenborg	1502	1.4047	0.4942	0.2508	0.1021	0.0379	0.0729	0.1821
Gustav Wikheim*	Strømsgodset	2413	1.1020	0.3125	0.3692	0.1477	0.0246	0.0381	0.1562
Yassine El Ghanassy*	Stabæk	1879	0.9840	0.2875	0.2355	0.1028	0.0242	0.0531	0.1261
Mohamed Elyounoussi	Molde	2275	0.9445	0.3526	0.2887	0.0509	0.0213	0.0255	0.1259
Tobias Mikkelsen*	Rosenborg	1931	0.9015	0.3444	0.2291	0.0446	0.0197	0.0630	0.1587
Moryké Fofana*	Lillestrøm	1300	0.8512	0.2796	0.2384	0.1325	0.0241	0.0176	0.0852
Ernest Asante	Stabæk	2643	0.8326	0.2662	0.2453	0.0436	0.0195	0.0683	0.1435
Samuel Adegbenro	Viking	1778	0.8245	0.3356	0.1451	0.0648	0.0212	0.0554	0.1234
Bentley	Odd	2464	0.8027	0.2561	0.1901	0.0436	0.0294	0.1292	0.0829
Ole Jørgen Halvorsen	Odd	1393	0.7688	0.2770	0.1771	0.0272	0.0271	0.1195	0.0823
Top 10 average			0.9416	0.3206	0.2369	0.0760	0.0249	0.0643	0.1266

* sold to foreign club during or after the 2015 season   ** on loan from a foreign club in 2015

Examples of players that did not obtain any particularly high impact values, are third and fifth placed Yassine El Ghanassy and Tobias Mikkelsen. This can show that Model 2 is able to value players with a wide spread of abilities to a certain extent, also to players rated high on the lists. It can be seen that these two players have impact values around the average in all six categories. Six of the wingers in Table 5.16 are the same as for Model 1. A noticeable player missing is Trond Olsen, who was considered the most efficient player by the xG Model and rated second among the wingers in Model 1. Although he was efficient, his xG value was not particularly high, which makes him drop out of the list for Model 2.

The top 10 list for attacking midfielders can be found in Table 5.17. Iver Fossum is rated on top. He had the highest and second highest impact value of the players in the list in passes and ball carries, respectively. In addition, his value from shots is also above average. Papa Alioune Ndiaye has the second highest shot value, and the best value for take ons. As previously seen, he had the highest market value from Transfermarkt at year end 2015, at 3 million GBP. Tenth on the list, youngster Eirik Hestad is an example of a player who obtained most of his impact value from other actions than shots. He has by far the lowest value in this action amongst the players in the list, and he earned most of his value by passing, corner kicks and ball carries. He is possibly a player to watch in the future, considering the fact that he is only 20 years old. Eight of the attacking midfielders in Table 5.17 also appeared on the top 10 list from Model 1.

Table 5.17: Model 2: Top 10 attacking midfielders, Tippeligaen 2015.
Average total value = 0.4608, Minimum total value = 0.0059

Player	Team	Minutes	Total	Shot	Pass	Take on	Long pass	Corner	Ball carry
Iver Fossum*	Strømsgodset	2653	0.8587	0.2702	0.3297	0.0242	0.0253	0.0004	0.1227
Papa Alioune Ndiaye*	Bodø/Glimt	1286	0.7967	0.2872	0.1726	0.0670	0.0585	0.0119	0.0929
Ghayas Zahid	Vålerenga	2383	0.7618	0.2373	0.2650	0.0449	0.0048	0.0036	0.1352
Michael Barrantes*	Aalesund	1000	0.7167	0.2097	0.1888	0.0051	0.0689	0.0703	0.0817
Fredrik Nordkvelle	Odd	1891	0.6806	0.2128	0.2292	0.0198	0.0071	0.0009	0.1100
Gjermund Åsen	Tromsø	1709	0.6745	0.1602	0.1151	0.0125	0.0021	0.1190	0.0678
Bojan Zajić	Sarpsborg 08	1760	0.6574	0.2360	0.2166	0.0304	0.0140	0.0221	0.0550
Aron Elís Trándarson	Aalesund	1152	0.6464	0.2980	0.1567	0.0286	0.0163	0.0095	0.0726
Daniel Fredheim Holm	Vålerenga	1974	0.6449	0.1309	0.2683	0.0439	0.0124	0.0110	0.0978
Eirik Hestad	Molde	941	0.6258	0.0661	0.2771	0.0212	0.0201	0.0846	0.0704
Top 10 average			0.7064	0.2108	0.2219	0.0298	0.0229	0.0333	0.0906

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

Now consider the top 10 central midfielders shown in Table 5.18. Mike Jensen was the best central midfielder in 2015 according to Model 2 by a wide margin, in sharp contrast to in Model 1 where he did not even make the top 10 list. This is a good example of how the difference in rewarding shots has big implications for the lists. In fact, this is also evident for second and third placed Etzaz Hussain and Harmeet Singh, who also were absent on the top 10 list in Model 1. Jensen had the highest impact value in shots, crosses and ball carries. A closer look on his impact values both from Model 1 and Model 2 is given in Section 5.4.1. Third placed Harmeet Singh had a low impact value from shots, and the highest from both passes and long passes among the players in the table. This can indicate that he is a good passing player, both short and long. Ole Kristian Selnæs is another example of this, and his impact values show above average passing abilities in addition to crosses and corner kicks.

As for the list from Model 1, both Malaury Martin and Giorgi Gorozia features among the central midfielders. The former was considered a big talent and captained every France national youth team from U-17s through U-21s. The latter is an example of a player that was not appreciated in the journalist ratings, with a rating of 99th and 101st in VG (2015) and Altomfotball (2015), respectively. This makes him a candidate of unappreciated talent, and considering he was born in 1995 he might be a player to watch in the future. Martin did not play enough to obtain a rating in neither VG nor Altomfotball. In total, six of the players in the top 10 list for central midfielders also appeared in the list from Model 1.

Table 5.18: Model 2: Top 10 central midfielders, Tippeligaen 2015.
Average total value = 0.4608, Minimum total value = 0.0059

Player	Team	Minutes	Total	Shot	Pass	Cross	Long pass	Corner	Ball carry
Mike Jensen	Rosenborg	2466	1.0463	0.2562	0.3341	0.0737	0.0285	0.0929	0.1235
Etzaz Hussain*	Molde	1956	0.7513	0.1343	0.3162	0.0099	0.0458	0.0275	0.1208
Harmeet Singh	Molde	2212	0.7429	0.0985	0.3671	0.0203	0.0847	0.0128	0.0946
Giorgi Gorozia	Stabæk	1467	0.6988	0.1110	0.2833	0.0304	0.0205	0.0891	0.0612
Ole Kristian Selnæs*	Rosenborg	1951	0.6865	0.0757	0.3080	0.0319	0.0675	0.0511	0.0875
Fredrik Midtsjø	Rosenborg	2396	0.6609	0.1205	0.3048	0.0143	0.0246	-	0.0885
Jone Samuelsen	Odd	2292	0.6473	0.1435	0.2440	0.0476	0.0114	0.0024	0.0960
Malaury Martin	Lillestrøm	975	0.6427	0.1252	0.2152	0.0290	0.0442	0.1375	0.0372
Christian Grindheim	Vålerenga	2675	0.6139	0.0649	0.2832	0.0138	0.0285	0.0600	0.0567
Bismark Adjei-Boateng**	Strømsgodset	1293	0.5535	0.1999	0.2271	0.0348	0.0149	0.0007	0.0433
Top 10 average			0.7044	0.1330	0.2883	0.0306	0.0371	0.0474	0.0809

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

A general observation is that the impact values from shots are decreasing with respect to playing position on the field. The same can not be said about other values such as for passes which seem more stable across playing positions. As players further back on the field are considered, a larger portion of their values was due to passes than for the more offensive players.

Table 5.19 shows the top 10 list for full backs. Per-Egil Flo is on top of the list, and had high impact values in passes, corner kicks and crosses but none of which are significantly higher than for the rest. This can show that he had good overall performance in 2015, and Model 2 is able to identify that. In fact, Flo was considered the best full back in both 2014 (see Table C.2) and 2015 by Model 1, in addition to being second placed in 2014 (see Table D.2) by Model 2, just shy of Lars-Christopher Vilsvik.

Table 5.19: Model 2: Top 10 full backs, Tippeligaen 2015.
Average total value = 0.4608, Minimum total value = 0.0059

Player	Team	Minutes	Total	Shot	Pass	Cross	Corner	Throw-in	Ball carry
Per-Egil Flo	Molde	1998	0.7505	0.0849	0.2954	0.1053	0.1187	0.0333	0.0470
Jonas Svensson	Rosenborg	2610	0.7108	0.1126	0.3389	0.0730	-	0.0412	0.0709
Espen Ruud	Odd	2476	0.6584	0.0881	0.1954	0.1233	0.0518	0.0496	0.0471
Lars-Christopher Vilsvik	Strømsgodset	2126	0.6179	0.1047	0.2076	0.1542	0.0237	0.0371	0.0593
Martin Linnes*	Molde	2439	0.5079	0.0545	0.1931	0.0908	0.0071	0.0459	0.0430
Joachim Olsen Solberg	Mjøndalen	2486	0.4978	0.0895	0.0798	0.0517	0.1017	0.0222	0.0208
André Danielsen	Viking	2700	0.4634	0.0612	0.1425	0.0892	0.0351	0.0383	0.0523
Birger Meling	Stabæk	2263	0.4592	0.1000	0.1665	0.0279	0.0484	0.0301	0.0585
Jørgen Skjelvik	Rosenborg	1812	0.4468	0.0762	0.1873	0.0587	0.0030	0.0256	0.0620
Akeem Latifu*	Aalesund	2564	0.4454	0.0657	0.1206	0.1373	-	0.0363	0.0558
Top 10 average			0.5558	0.0837	0.1927	0.0911	0.0389	0.0360	0.0517

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

Second placed Jonas Svensson was rated top 10 by both VG and Altomfotball, but was not on the list for Model 1 due to his numerous missed shots. He has the highest value in the table for both passes and ball carries. In addition, it seems that he is good at finding teammates on throw-ins, which are often executed by the full backs. Lars-Christoper Vilsvik is the full back with the highest impact value from crosses, and he also has a high shot value. Amongst the top four full backs are the top three from Model 1, and in total six out of ten are the same. In general, it can be seen from the impact value of crosses in the table, that the model is able to value crosses as an important attribute for the full backs.

Finally, consider the top 10 list for the centre backs in Table 5.20, where six names are familiar from Model 1. As for Model 1, Johan Bjørdal is the highest rated centre back mostly due to his high impact value from passing. He had below average impact values for aerial duels, which is considered important for a player in his position. However, he had above average values for both long passes and ball carries. Stefan Strandberg and Hólmar Örn Eyjólfsson, number two and three on the list respectively, have roughly the same impact values in all categories, but the former seems to have hit better long passes in 2015. All top three centre backs played for Rosenborg in 2015. It is likely that Rosenborg had a higher than average possession, which can lead to the centre backs seeing a lot of the ball and influence their impact value from passing and ball carries positively. Next on the list, Rhett Bernstein has the highest impact value from shots and aerial duels, who constitute a large part of his total impact value.

Table 5.20: Model 2: Top 10 centre backs, Tippeligaen 2015.
Average total value = 0.4608, Minimum total value = 0.0059

Player	Team	Minutes	Total	Shot	Pass	Aerial	Cross	Long pass	Ball carry
Johan Bjørdal	Rosenborg	1093	0.4473	0.1057	0.2274	0.0002	0.0018	0.0490	0.0803
Stefan Strandberg*	Rosenborg	1151	0.4149	0.1064	0.1797	0.0215	0.0057	0.0613	0.0883
Hólmar Örn Eyjólfsson	Rosenborg	2357	0.3758	0.0974	0.1835	0.0124	0.0005	0.0260	0.0742
Rhett Bernstein*	Mjøndalen	1234	0.3516	0.2193	0.1001	0.0650	0.0034	0.0046	0.0067
Joona Toivio	Molde	1706	0.3293	0.1184	0.1363	0.0107	0.0193	0.0259	0.0290
Jørgen Horn*	Strømsgodset	1260	0.3004	0.0737	0.1623	-0.0008	0.0125	0.0335	0.0423
Oddbjørn Lie	Aalesund	1549	0.2715	0.0495	0.1161	-0.0024	0.0381	0.0198	0.0311
Morten Sundli	Mjøndalen	2018	0.2690	0.1332	0.0907	0.0136	0.0106	0.0180	0.0205
Vegard Forren	Molde	2416	0.2614	0.0402	0.1300	0.0067	0.0123	0.0700	0.0353
Andreas Nordvik*	Sarpsborg 08	1846	0.2366	0.0830	0.0950	0.0037	0.0072	0.0630	0.0229
Top 10 average			0.3258	0.1027	0.1421	0.0131	0.0111	0.0317	0.0430

* sold to foreign club during or after the 2015 season ** on loan from a foreign club in 2015

For the players in defence (full backs and centre backs), few of the actions that measure defensive contributions, like tackles, interceptions and ball recoveries are present in the tables. This is because these actions have too low values to become a significant contributor to the total impact value. This might indicate that Model 2 has limited ability to appreciate defensive involvements.

As for Model 1, a league table can be obtained by aggregating the values for each team. The obtained tables for Model 2 can be found in Appendix D, Tables D.3 and D.6 for 2014 and 2015 respectively. The table for 2014 suggests that Brann substantially underachieved during the season according to Model 2, and they did possibly not deserve to have been in the relegation battle. Once again, it can be seen from Table D.6 that Molde was inefficient in the 2015 season. A closer look on the team evaluations from both Model 1 and Model 2 is found in Appendix E.

5.4 Case Studies

For a more in-depth description of how the models evaluate players, two of the most interesting cases are investigated further. Mike Jensen was valued very differently across the two models and seasons. Was he actually as good as people believe? Martin Ødegaard comes out as the best player in his position in both models (see Appendixes C and D) and is considered one of the most talented young footballers in the world (see for instance Mirror (2015), Gazetta (2014)). He was wanted by many of the largest clubs in Europe after his performance in Tippeligaen 2014, and ended up signing for Spanish top club Real Madrid. What did he do better than the rest of the players?

5.4.1 Mike Jensen

Figure 5.11 shows a box plot of some of the different actions performed by Mike Jensen from Model 1, in order to examine his performance more accurately. The shots are excluded from the figure, due to being an order of magnitude larger than the other involvements. A similar box plot including the shots can be seen in Figure F.4 in Appendix F.1. In these figures, different types of actions are on the x-axis and the impact value is on the y-axis. The box plots illustrate the variation in each sample of data (that is, the variation in the impact by each type of action), without making any assumption of the statistical distribution of the sample.

In a box plot, the blue boxes contain 50% of the observations, where the upper and lower edges are the 75th percentile and the 25th percentile, respectively. The average value is in the middle of the boxes, while the median is marked as a red line. Red crosses are observations that are characterised as outliers, lying outside the whiskers which illustrate the 0.35th and the 99.65th percentile. Practically, an outlier illustrate either a very good or poor involvement by the player, compared with his average impact of performing this action. Box plots, with and without shots, for Mike Jensen in 2014 are shown in Appendix F.1. Table 5.21 shows the impact value per 90 minutes from Model 1 for the same actions. These values are calculated by aggregating all the observations (which are illustrated in the figure) for each action and further normalise them to a per 90 minutes basis.

Table 5.21: Model 1: Impact values for Mike Jensen 2015

Action	Impact
Pass	0.0750
Take on	0.0168
Foul won	0.0120
Corner won	0.0110
Tackle	0.0060
Interception	0.0017
Clearance	0.0010
Shots	-0.1608
Aerial	-0.0026
Ball rec.	0.0073
Cross	0.0365
Long	0.0082
Corner	0.0370
Ball carry	0.0257
Free kick	0.0062
Total	0.0783

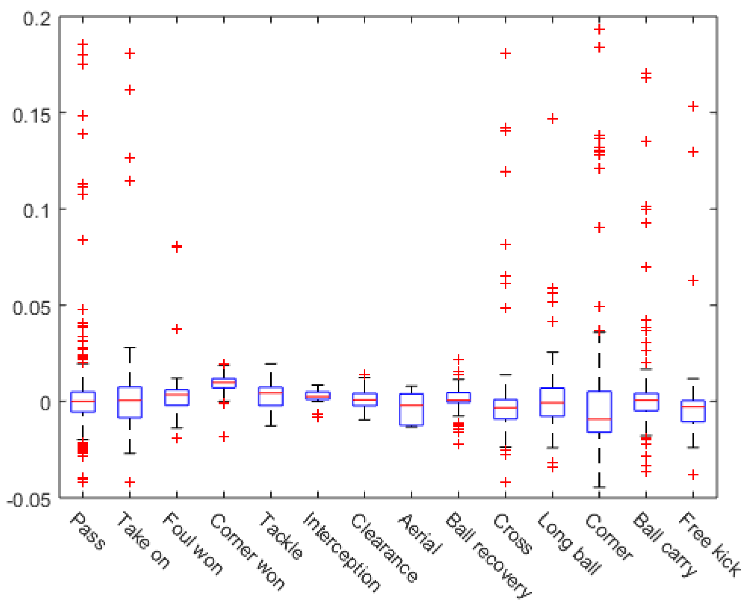


Figure 5.11: Model 1: Box plot for Mike Jensen in 2015 excluding shots

As mentioned above, Mike Jensen was not appreciated by Model 1 as a top 10 player in his position in 2015, although pundits seemed to agree he was one of the best of all players.

In 2014, as can be seen from Table C.2, he was rated fifth in his position. In addition, Model 2 rates him as the best central midfielder in 2015, and the fourth best in 2014 (see Table D.2). Why is this the case?

In Table 5.21, it can be seen that the total impact per 90 minutes from the Mike Jensens shots is equal to -0.1608, which heavily influences the total impact value. Nine out of the top 10 central midfielders in Model 1 had a shot impact larger than zero. If the impact from his shots instead was equal to zero, his total impact value would have been 0.2391, which would have lead to Jensen being ranked as the third best central midfielder. Hence, this is a very good example of how the shots influence the total impact value in Model 1.

The influence from shots is different in Model 2, where the player receives an impact equal to the likelihood of scoring in the given state. Table 5.22 shows the impact value per 90 minutes of the same actions for Model 2, while Figure 5.12 illustrates a box plot of these actions. In this figure, it can be seen that the impact value from a shot always is greater than zero, and significantly larger compared with the other actions. Hence, the total impact from shots is positive for Jensen, as well as for most of the other players. The magnitude of the values in Table 5.22 can not be directly compared with the values from Model 1 in Table 5.21. Instead, the values of the different actions should be compared internally within each table, and with the total impact value per 90 minutes shown at the bottom of each table.

Table 5.22: Model 2: Impact values for Mike Jensen 2015

Action	Impact
Pass	0.3341
Take on	0.0352
Foul won	0.0204
Corner won	0.0301
Tackle	0.0044
Interception	0.0010
Clearance	-0.0042
Shots	0.2562
Aerial	0.0031
Ball rec.	0.0258
Cross	0.0737
Long	0.0285
Corner	0.0929
Ball carry	0.1235
Free kick	0.0141
Total	1.0463

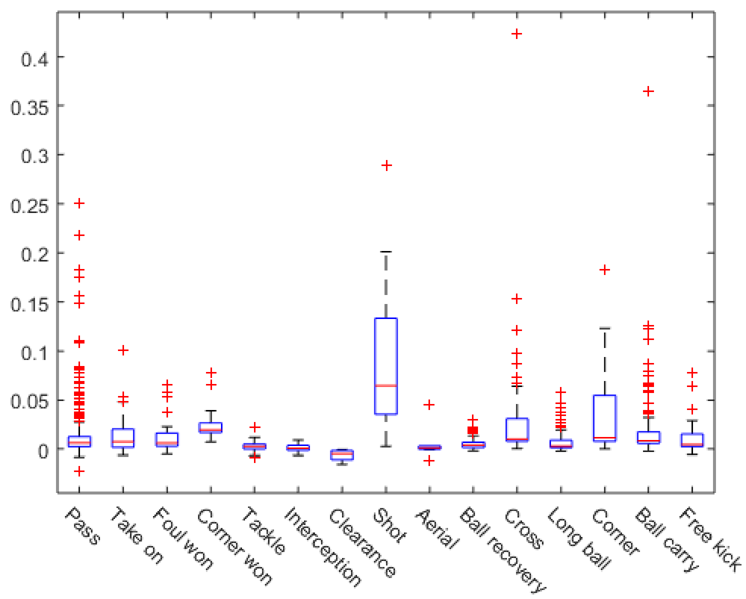


Figure 5.12: Model 2: Box plot for Mike Jensen in 2015

When comparing the plots and tables from the two models, some hypotheses regarding the abilities of Mike Jensen can be made. It seems that he is a good passing player, which can be seen from the large number of positive outliers in Figure 5.12, in addition to the high values in Tables 5.21 and 5.22. Furthermore, it seems that he is a good carrier of the ball, in addition to be able to take on opposing players. The figures and tables also indicate that Mike Jensen is a good crosser and hits accurate corners. The total impact per 90 minutes from these actions are positive in both models, although the median of both actions are below zero in Figure 5.11. However, this is not surprising since both corners and crosses are usually cleared away by the defenders.

In order to evaluate these hypotheses, the samples of values for Jensen in these actions can be compared with the samples containing values for all players in Tippeligaen 2015. Figure 5.13 shows a comparison between Jensen and all players in Tippeligaen from Model 2 for the 2015 season without the outliers. The outliers are excluded because of their magnitude in the samples for Tippeligaen as a whole. Since Jensen is a central midfielder, some of the more defensive attributes are also included in the comparison. Amongst the more defensive actions, it especially seems that Jensen is good at ball recoveries from the two tables.

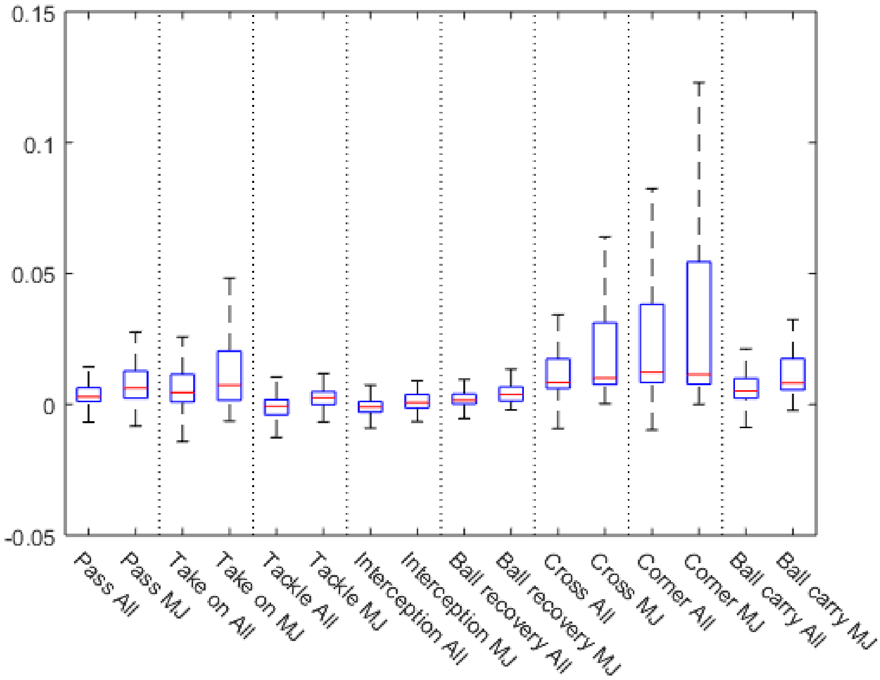


Figure 5.13: Model 2: Comparison of Mike Jensen vs league for 2015

In all five offensive categories except for corners the medians of Jensen lie higher than the median for Tippeligaen. Furthermore, the top of the box and the upper whisker are located higher for Jensen in all five offensive categories. Hence, from the figure it seems that Mike Jensen is better compared with the league as a whole in the following five categories: passing, take ons, crosses, corners and ball carries. Jensen primarily plays to the right in a central midfield of three players, where the player in the middle has more defensive duties. This allows Jensen to support the attackers and contribute offensively. Based on this, the five categories above can be considered as important capabilities for a player in the role of Jensen. In addition, if the values for these actions from the two models, in Table 5.21 and 5.22, are aggregated, it can be seen that they constitute a major part of the total impact by Jensen.

In addition to his offensive capabilities, Jensenis also known as a high-intensity player who does not back down from a duel. From Figure 5.13 it can be seen that he is also an effective player in some of the more defensive parts of the game. His median lies above the median of Tippeligaen in both tackling, interceptions and ball recoveries. Furthermore, his upper whiskers are also located a bit higher than for Tippeligaen as a whole. As previously indicated, it seems that his impact from especially ball recoveries is significant. Hence, Jensen can also be considered a good duel player that wins the ball in favourable positions.

Overall, this case study has shown that Mike Jensen is a very good player in Tippeligaen. He has shown good abilities, especially offensively, but also defensively to some extent. However, his willingness to take a lot of shots in 2015 lead to a total impact value per 90 minutes slightly below average in Model 1. On the other hand, Model 2 rates him as the best central midfielder and fourth overall among the players with over 900 minutes played in 2015.

5.4.2 Martin Ødegaard

To reveal what wonderkid Martin Ødegaard did better than the rest in Tippeligaen before completing a 35 million NOK transfer to Spanish club Real Madrid, an examination of the impact made by Ødegaard in the 2014 season is done. Figures 5.14 and 5.15 show box plots containing selected actions of Ødegaards involvements from Model 1 and 2 respectively. Tables 5.23 and 5.24 show the total impact value per 90 minutes from the same actions. A box plot from Model 1 including the impact from his shots can be seen in Figure F.5.

Table 5.23: Model 1: Impact values for Martin Ødegaard 2014

Action	Impact
Pass	0.1340
Take on	0.0038
Foul won	0.0130
Corner won	0.0053
Tackle	0.0083
Interception	-0.0008
Clearance	-0.0002
Shot	0.1472
Aerial	-0.0007
Ball rec.	0.0100
Cross	0.0112
Long	0.0068
Corner	0.0309
Ball carry	0.0154
Free kick	-0.0052
Total	0.4030

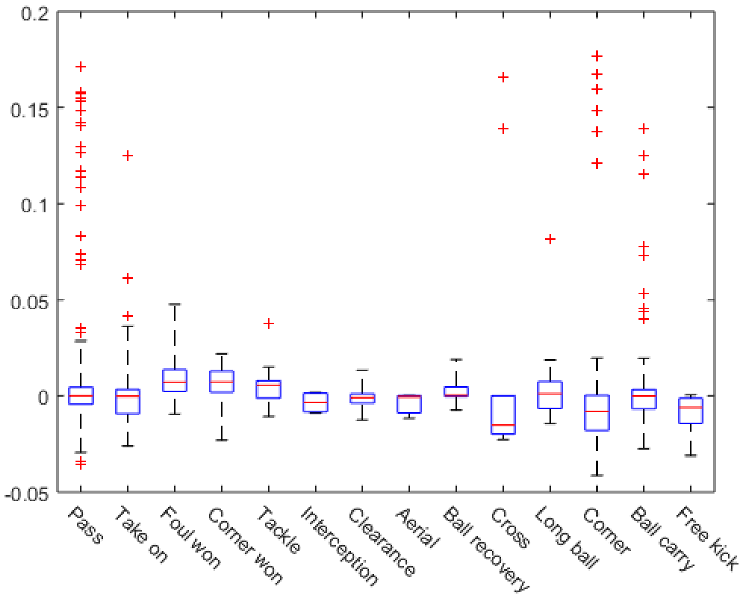


Figure 5.14: Model 1: Box plot for Martin Ødegaard in 2014 excluding shots

Table 5.24: Model 2: Impact values for Martin Ødegaard 2014

Action	Impact
Pass	0.5469
Take on	0.0594
Foul won	0.0300
Corner won	0.0201
Tackle	0.0013
Interception	0.0002
Clearance	-0.0022
Shot	0.1496
Aerial	0.0008
Ball rec.	0.0180
Cross	0.0117
Long	0.0208
Corner	0.0552
Ball carry	0.2207
Free kick	0,0045
Total	1.1504

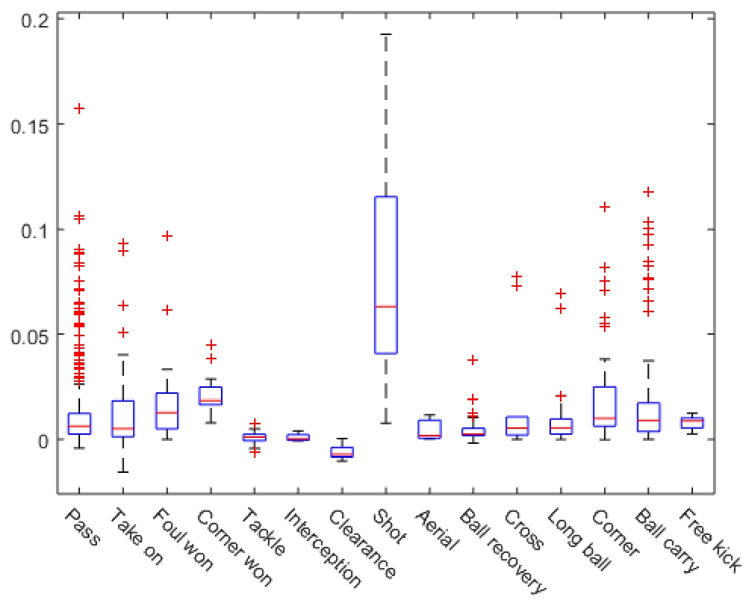


Figure 5.15: Model 2: Box plot for Martin Ødegaard 2014

From the figures it seems that Ødegaard has positive involvements from many of the different actions. In one or both of the models all actions seem to have a positive impact, except interceptions, clearances and aerial duels. Surprisingly, Ødegaard playing as an attacking midfielder, he seems to contribute positively through the more defensive actions tackles and ball recoveries. Nonetheless, his contribution through interceptions and clearances seems to be more limited. A closer examination of his actions, shows that Ødegaard only won 15 corners, attempted ten crosses and took nine free kicks. The samples containing these actions are therefore considered too small, despite showing positive impact values in one or both of the two tables. Furthermore, one or both tables indicate that passes, take ons, shots, corners and ball carries are most influential among the actions. Especially the values for passing seem to heavily influence the total impact value in both the models. In addition, the large value for ball carry in Model 2 is also worth noting. Figure 5.16 shows a comparison of the impact made by Ødegaard and all players in Tippeligaen within a selection of actions based on the above discussion.

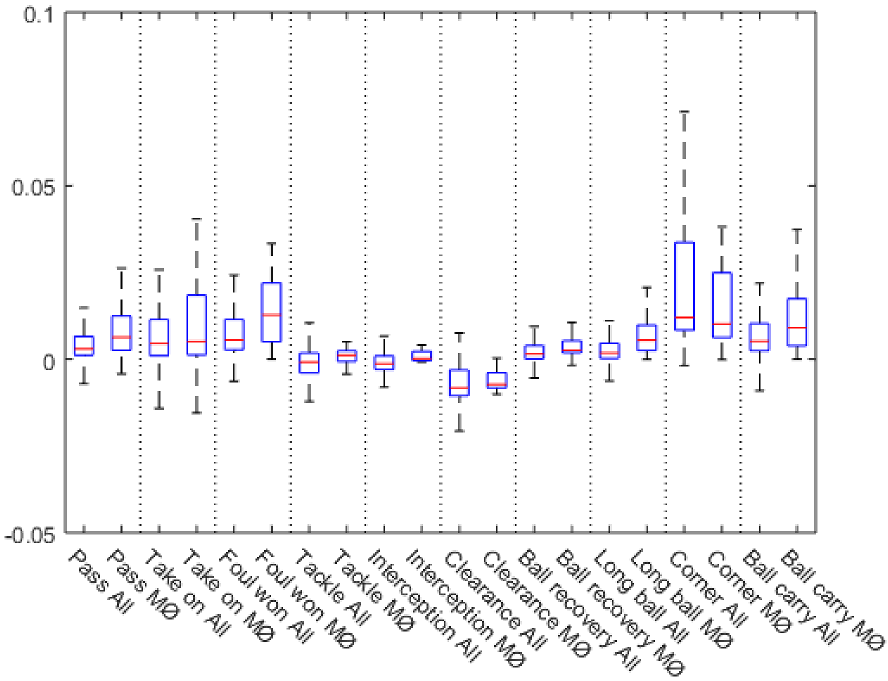


Figure 5.16: Model 2: Comparison of Martin Ødegaard vs league for 2014

Based on the last box plot, it seems that Martin Ødegaard performed better compared to the samples of all the players on at least six types of actions. His passing ability is good, he is capable to take on a player, gets fouled in good positions, wins the ball in favourable positions, can pick out a long pass and he is a good carrier of the ball. Martin Ødegaard plays primarily as an attacking midfielder, and these abilities are believed to be important for such a player. In addition, this is correspondence with the praise he has received for his passing and technical abilities. Moreover, it seems like his defensive contributions are quite limited, but he had some good tackles and interceptions. Further examinations revealed that he took 41 corners, which seems to be a potential area of improvement for Ødegaard when looking at Figure 5.16, despite having a positive impact on his total impact value for both models.

5.5 Reliability Out of Sample

In order to further investigate the performance of the models, they have been tested out of sample using data from the first 13 match days of the 2016 season. Reliability tests are done by comparing the values of players who have played more than 390 minutes (corresponds to one third of the season so far) in the 2016 season and over 900 minutes in 2015. Figure 5.17 shows the results when these players are compared across seasons with respect to their G/xG value. In Appendix G top lists for 2016 from the three models are presented.

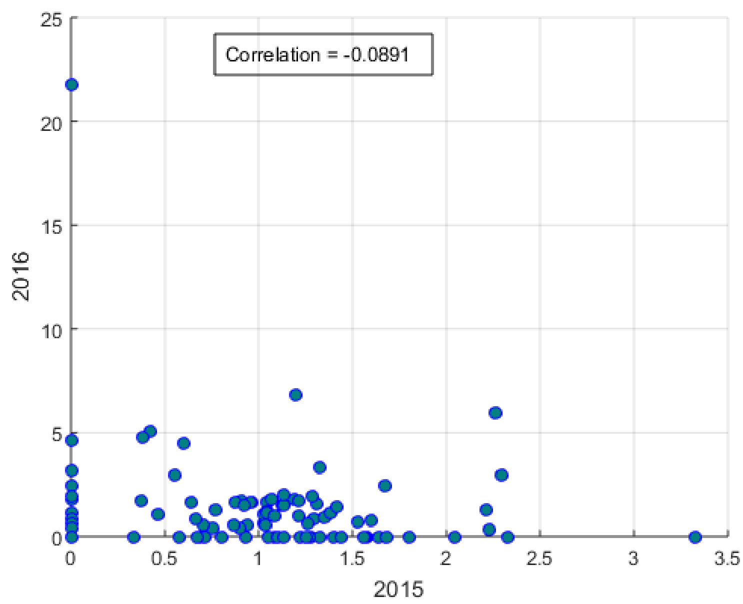


Figure 5.17: xG Model: Scatter plot and correlation of G/xG between 2015 and 2016

As can be seen in Figure 5.17, the correlation coefficient between 2015 and 2016 is -0.0891. This clearly indicates that the players do not manage to maintain their efficiencies across seasons, which was also seen when examining 2014 to 2015. However, the correlation coefficient between 2015 and 2016 is 0.7565, when comparing the xG/90min values. Hence, the players evaluated manage to maintain their respective xG/90min value across the two seasons. This indication was also evident when evaluating in sample data.

Figures 5.18 and 5.19 show the results when the players are compared across seasons based on their average impact value per 90 minutes from the two Markov models.

The scatter plot in Figure 5.18 illustrates the values from Model 1 for the two seasons. From the plot it seems that the players does not manage to maintain their respective values from 2015 to 2016, and the correlation coefficient is 0.0932, which is somewhat lower than in sample. However, as previously pointed out, the values from shots heavily influence the value of each player. As illustrated by the low correlation of G/xG, the players are often not capable of maintaining their effectiveness over seasons, which can impact the value of a player in Model 1 negatively. Thus, the reliability of Model 1 seems to be limited due to the heavy influence by the outcome of shots.

In Figure 5.19 the scatter plot of the player values from Model 2 is shown. The correlation coefficient is as high as 0.9006, which is almost exactly equal to the coefficient when comparing the 2014 and 2015 seasons. This indicates that Model 2 is reliable in assessing the impact on

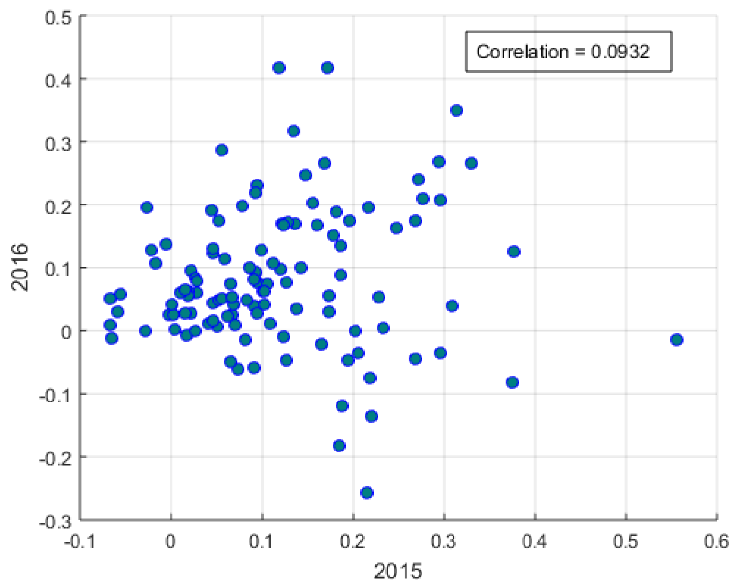


Figure 5.18: Model 1: Scatter plot and correlation between 2015 and 2016

the field by each player, also when assessing out of sample data.

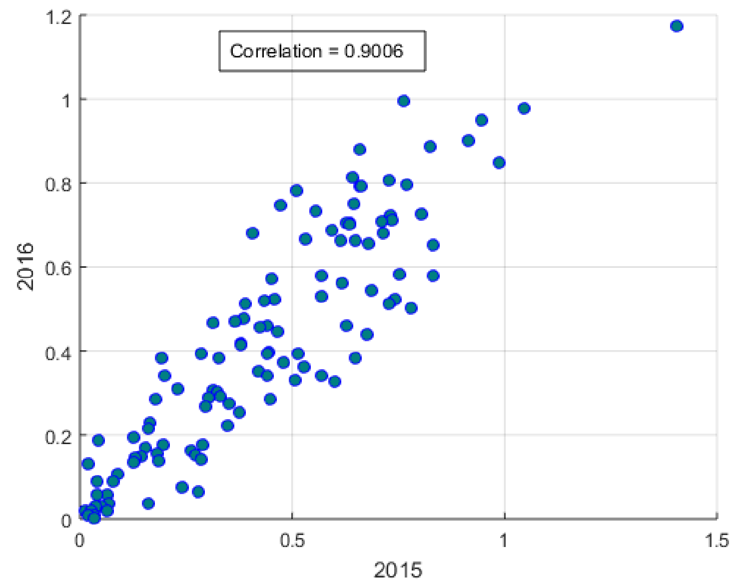


Figure 5.19: Model 2: Scatter plot and correlation between 2015 and 2016

5.6 Evaluation of Research Questions

In this section, answers to the three research questions, introduced in Section 1.2, are provided and discussed on the basis of the previously presented results.

RQ 1 Is it possible to create a statistically significant xG model that assesses the quality of all shots in Tippeligaen in order to evaluate the efficiency of primarily offensive players?

An xG model for Tippeligaen has been developed in this thesis. The results from the logistic regression show that it is possible to create a model with several statistically significant in-match variables that affect the likelihood of converting a shot into a goal. The explanatory variables include multiple aspects that influence the likelihood of scoring, and their signs can be argued for from a footballing point of view. The estimated xG/90 min values for individuals playing more than a third of the season are fairly consistent across seasons for in-sample data, and the xG/90 min also shows to be reliable on a reasonable level across seasons when testing on out of sample data.

However, the xG Model has some important limitations. The one considered to be the most prominent is the inability to model defender proximity. Defenders that cover attacking players and block their shots are an important aspect that has been shown by other works to have an impact on the likelihood of scoring. The data set utilised in this thesis does not support such an analysis, which has to be a subject for further research. In addition, the free kick model exhibited low discrimination, which also impacts the likelihood of the shots for some players. Furthermore, the data set used in building the xG Model is also a source of error. Human annotators are collecting the data, which gives rise to bias for some variables. In addition, some variables are calculated by the authors, which introduces the possibility of miscalculation.

The player evaluations based on efficiency measured by G/xG is to some extent as what should be expected. However, the G/xG performance measure exhibits poor reliability across seasons. Thus, the xG Model assigns likelihoods that enables a G/xG player rating only for the season in question. Due to low correlation coefficients across seasons both in and out of sample, few inferences can be made on the consistency of the performance measure G/xG of players.

When assessing players across seasons it is more interesting to look at their xG/90 min value. The xG/90 min assigned to each player has a higher reliability compared to G/xG across seasons. The total xG/90 min is a product of the number of shots and the quality of the shot, indicating that a player with high xG/90 min shoots a lot, creates high quality shots or is responsible for a favourable mix of these two factors. By using xG/90 min as the primary performance measure, one can make more reliable inferences on player performance across seasons.

RQ 2 Is it possible to create a Markov game model for football that is able to evaluate all player involvements in a match and rate players over the course of a season?

Two variations of a Markov game model applied to football have been developed in this thesis. To the best of the authors knowledge, neither of the two approaches have been attempted on data from football before. The results show that it is possible to create models that are able to assign values to individual player involvements with such approaches. The Markov game models also have the ability to investigate individual players and evaluate both offensive and defensive contributions, and show promising results when comparing them to three benchmarks. However, due to the nature of the game and the model specifications, the values of offensive contributions are often orders of magnitude larger than the defensive. This has implications on the player ratings for the defensive positions, where the results show that they are primarily rewarded for their offensive contributions and not, as would have been preferable, for their defensive abilities.

The two models differ on some important features. The models are distinct in the way they assign values to the different actions, which has some implications on the aggregated impact value assigned to each player. This distinction is mainly due to different impact functions and differences in features between the two models. This has especially two large consequences, where the first is with regard to the impact value of shots. Model 1 has a bias towards players with high efficiencies, due to the way the impact function is defined. It seems that most players do not manage to maintain their efficiency rate across seasons, indicated by the low correlation coefficient of G/xG. This is believed to impact the reliability of Model 1, which shows a correlation of 0.1944 between the two seasons. For Model 2, the impact function leads to a favouring of players that regularly attempts high quality shots, because less emphasis is given to the outcome. It seems that players manage to maintain their ability to generate high quality shots across seasons to a larger extent compared with their efficiency. This is believed to positively affect the reliability of Model 2, which shows a correlation coefficient of 0.8988 between the two seasons. Despite the large influence by shots in both models, they also seem to manage to appreciate players that impact the game through other actions. When the correlations are tested on out of sample data, the same result regarding their reliability can be observed. Model 2 shows a correlation of 0.9006, which is almost exactly the same as in sample, while Model 1 shows a mere 0.0932.

The second consequence from the different impact functions is regarding the evaluation of the defensive involvements in football. Model 1 indicates a slightly larger appreciation for the defensive involvements. This is observed from tackles, clearances and ball recoveries being among the top five influential types of actions for one or more positions in Model 1, as seen in Tables 5.10 to 5.12, while only aerial duels is visible among the defensive involvements in Model 2. Since an action is valued on the basis of what is actually lead to in Model 1, it seems that the model manage to occasionally recognise some good defensive involvements, which leads to a more significant impact over the course of a season than for Model 2. For both models, the offensive involvements are larger in magnitude than the defensive ones, which is unsurprising due to them being closer to the objective of the game, namely scoring goals. This leads to a somewhat unfair evaluation of the defensive players. As seen in the player tables in Sections 5.2.3 and 5.3.3, the majority of the impact from the best defenders were from shots and different passing events.

The results from the two models are somewhat different, although a total of 37 out of 60 players in the tables in Sections 5.2.3 and 5.3.3 are the same. An observation that can be made regarding the difference between the two models, is that the impact values are higher for Model 2, which also shows better reliability. A suspicion that might be able to explain this, is that fewer involvements receive a negative impact value in Model 2 than in Model 1. For instance, a successful backward pass might more often be assigned a negative impact value in Model 1. In Model 2, the value of an action is equal to the value of the state, which is defined as the average values of the subsequent states. Thus, even though the pass was played backwards on the field, it can receive a positive impact value if the average impact of the possible events that followed was favourable. This can be seen when taking a closer look at the box plots from each of the models for Mike Jensen, shown in Figure 5.11 and 5.12. When comparing the sample of his passes, it is evident that the median in Model 1 is approximately zero, while it is slightly positive in Model 2. Furthermore, the lower whisker is located lower for Model 1, in addition to several negative outliers being evident. This indicates that the impact values from passes are more evenly distributed around zero in Model 1, while a significantly larger share of the values are positive in Model 2. With this in mind, it is reasonable to believe that the number of actions and the total impact value are positively correlated for Model 2. However, when examining these correlations for both the models, the results show that there is virtually no correlation between the number of actions performed and the total impact value in neither of the models.

This difference between the models is believed to be more evident among the more offensive involvements than the involvements further back on the field. When performing an action on the offensive third of the field, it is likely that the possible subsequent involvements are positive for the team in possession, which leads to a positive impact value for the action in Model 2. In Model 1, the value of the action would depend on the action that follows next, which could be an unfavourable action for the team in possession. Offensive involvements are therefore more likely to be assigned positive values in Model 2, which is believed to result in a high number of offensive involvements being favourable in order to obtain a high impact value. Furthermore, players on the strongest teams are more likely to be in offensive positions more often, hence, Model 2 might be more biased towards the players on the strongest teams. This can be supported by the fact that many of the highest ranked players are from the teams considered the strongest.

In general, both models fulfil their purpose, which is to value each player involvement in Tippeligaen. The results seem to be fairly reasonable, and the players that are considered the best seem to be appreciated and identified as good players by the models. However, the models seem to be very influenced by shots, as well as having a bias towards more offensive involvements. This is therefore considered as potential areas of improvements.

RQ 3 Is it possible to reveal undiscovered talent or identify under- and overvalued players based on the evaluation of individual player involvements?

The results from the three models developed in this thesis show that it might be possible to identify undiscovered talent. The results can be compared to other benchmarks, in order to obtain an objective measure on the performance of a player relative to others.

For the Markov game models, a possible example is the 21 year old central midfielder Giorgi Gorozia, who was rated third and fourth in his position in 2015 by Model 1 and Model 2, respectively. The two subjective player ratings, used as benchmarks, rated him 99th and 101st amongst all players, which indicate that he did not attract much attention from journalists and pundits. He is currently valued at 300 thousand GBP by Transfermarkt. This is only slightly higher than the current average in Tippeligaen of 270 thousand GBP, and substantially lower than the average among the top 10 attacking midfielders of 828 thousand GBP. The players above Gorozia on the list from Model 2, Mike Jensen, Etzaz Hussain and Harmeet Singh, are valued at 1.35 million, 750 thousand and 825 thousand GBP, respectively. He was rated one place above Ole Kristian Selnæs who is valued at 2.25 million GBP by Transfermarkt. All of these players have estimated market values twice that of Gorozia, and could indicate that he is an undervalued player compared to his colleagues.

In addition, although he might not be undiscovered, under- or overvalued, Martin Ødegaard was the highest rated attacking midfielder in both Model 1 and Model 2 in 2014. This shows that both models were able to rate a player that is considered by many as one of the most promising young players in the world as the best.

Only the future can tell how the market value of a player will develop, and it is not possible to conclude from the models in this thesis whether players actually are under- or overvalued.

Nonetheless, as was a big motivation for this thesis, the models can open up new possibilities for clubs both in measuring performance internally and for scouting players. The same detailed data for every on-the-ball contact used in this thesis is collected during every match in 30 leagues and competitions worldwide. Thus, the models developed in this thesis can have several areas of application for clubs. They can use it internally to keep track of performance in own squads throughout a season, to assess what attributes their players possess or can improve, as well as to scout for talent either domestically or internationally.

Furthermore, clubs can use the two Markov game models to quantify what attributes each player is good at compared to others, which can be valuable information when replacing players in the different positions. The values in the state space learnt from data on Tippeligaen can be used to identify players in foreign leagues that are able to obtain high impact values from what is considered good by the model for Norwegian football. On the other hand, the values in the state space can also be learnt on a data set from a foreign league to identify the best players in that country. By doing this, clubs can obtain short lists of players to scout more closely, which can make the process less time consuming and possibly more effective. Scouts can spend more of their time on assessing qualitative considerations, which is still a major part of the process in acquiring players on the transfer market.

For the xG Model, it is possible to follow the same approach. Either by running a binary logistic regression on shots attempted in a foreign league or to use the xG Model built for Tippeligaen, it might be possible to identify the most efficient players in converting shots to goals. However, as argued for in the answer to RQ1, the performance measure G/xG shows limited stability across seasons both in and out of sample, and it may not be wise to make decisions based solely on player efficiency in one season. A possible strategy can be to look for players that obtain a desirable xG/90 min and actually are able to replicate their efficiencies across seasons, and scout them more closely.

Chapter 6

Conclusion

This thesis has described and documented the development of three models for evaluating individual player involvements in association football. To the authors knowledge, the xG Model is the first of its kind for Norwegian football, and the two Markov game models are the first of their kind in the entire football analytics community. The xG Model rated primarily offensive players based on their efficiency in converting attempted shots to goals, while the two Markov game models assigned values to all player involvements during a match, to identify what players had the biggest impact over the course of a season. Answers to the three research questions, introduced in Chapter 1, has been given after the presentation and discussion of the results in Chapter 5.

The results are deemed promising, although some remaining challenges and limitations are identified. The xG model does not account for defender proximity, for which the literature suggests plays an important role in such models. The framework for the Markov models, as adopted from literature on ice hockey and basketball, seems to have a place in football as well. However, the two Markov models show limited abilities in valuing defensive contributions, which is addressed as a subject for further research. In addition, lack of comparable sources make the reliability and validity of the models hard to assess.

Similar data as used in this thesis is available from 30 leagues and competitions worldwide. Thus the three models developed can have several areas of application for clubs. For the xG Model, it is possible to use a similar model building approach to identify the most efficient players in foreign leagues. Clubs can use the two Markov game models to quantify what attributes each player is good at compared to others, which can be valuable information when replacing players in the different positions. All of the three models can also be used to identify players in foreign leagues in order to make short lists of players to scout more closely, which can make the process less time consuming and possibly more effective. By doing this, scouts can spend more of their time focusing on qualitative considerations.

Competition is fierce in the world of football. Only small improvements in scouting, squad management and performance monitoring can have a huge monetary impact considering the money in circulation. Getting it right more often than not can lead to a competitive advantage in the future, and can be the difference between success and failure in the long run. Albeit not flawless, the authors believe the models developed in this thesis can serve as a valuable contribution to Rosenborg Ballklub and the football analytics community, especially with regard to Norwegian football, which has not been studied widely earlier.

Chapter 7

Recommendations for Further Research

The three models developed in this thesis are, to the best of the authors knowledge, the first of their kind for Norwegian football. In the academic community, only one other xG model on football is identified (Lucey et al. (2014)), while three articles were found that applies Markov models to football, all by the same authors (Hirotsu and Wright (2002), Hirotsu and Wright (2003a), Hirotsu and Wright (2003b)). However, no Markov model applied for assessing individual players was found. The novelty of the approaches in this thesis therefore make the recommendations for further research important to address.

As mentioned earlier, the data used in this thesis is not spatiotemporal. Working with event-based data from Opta makes it impossible to account for defender proximity. This is a variable, which in other leagues is found to influence the likelihood of shot outcomes, and incorporating defender proximity would be an important extension to the xG Model developed in this thesis. Applying spatiotemporal data for the Markov models is also an important approach to consider for future research. This would increase the detail level, possibly enabling a more accurate calculation of player impact. Applying spatiotemporal data would therefore be an important step towards developing models that are able to describe more of the complexity of football.

Furthermore, with access to more data it can be possible to include action histories in the definition of a state, similar to the Markov game model for ice hockey presented in Routley (2015). By utilising such an extension, historical effects from the game in a given state would be accounted for, possibly influencing the value of the actions. As described, the two Markov models share a bias towards shots and offensive involvements, although in somewhat different ways. A natural extension, and an important improvement for rating defensive players, is finding a solution for putting less emphasis on offensive events and model involvements further back on the field adequately. Including action histories might be able to appreciate a defensive involvement to a larger extent if the preceding actions were performed by the opposing team.

In addition, a more comprehensive data set would also likely improve the accuracy of the xG model. It would ensure a better basis for regular shot, but especially for static events it would be possible to create models with better discrimination abilities. Situational variables like the home field advantage, team strength and match status are shown to have an impact on match outcome, and are accounted for in this thesis. With more data on hand, it might be possible to incorporate other situational variables, like playing on artificial turf or the fatigue of players. This could be important improvements for both the Markov models and the xG Model.

Another possible extension of the Markov models, is to analyse teams more thoroughly. Especially with more data at hand, it might be possible to examine each team in detail, in order to reveal patterns in their playing style. This could be useful for clubs when scouting their upcoming opponent. Such analyses is believed to especially provide insight if spatiotemporal data is used. Furthermore, if such patterns are possible to reveal, it can be used to simulate matches between the different teams, which could be useful in prediction and for betting companies.

In the two Markov models presented in this thesis, the total player impact is divided into several, equally weighted categories. A possible extension is to create player ratings for specific positions, which can be based on a weighted average of selected actions that are considered important for the different positions. Another interesting approach for utilising different categories of player performance is to find determinants of the market value. Using a set of chosen performance categories, along with other variables like player age, contract duration and nationality, one can specify a regression on market value from Transfermarkt. This might provide insight on which specific capabilities that actually affects the market value of players in different positions, which could be useful for clubs.

Bibliography

Alcock, A., 2010. Analysis of direct free kicks in the women's football World Cup 2007. European Journal of Sport Science 10 (4), 279–284.

Almeida, C. H., Ferreira, A. P., Volossovitch, A., 2014. Effects of match location, match status and quality of opposition on regaining possession in UEFA Champions League. Journal of human kinetics 41 (1), 203–214.

Altomfotball, 2014. `http://www.altomfotball.no/elementsCommonAjax.do?cmd=statistics&subCmd=playerScore&tournamentId=1&seasonId=336&teamId=&useFullUrl=false`, Online; accessed 20-April-2016.

Altomfotball, 2015. `http://www.altomfotball.no/elementsCommonAjax.do?cmd=statistics&subCmd=playerScore&tournamentId=1&seasonId=337&teamId=&useFullUrl=false`, Online; accessed 20-April-2016.

Amir, E., Livne, G., 2005. Accounting, valuation and duration of football player contracts. Journal of Business Finance & Accounting 32 (3-4), 549–586.

Anderson, C., Sally, D., 2013. The Numbers Game: Why Everything You Know About Soccer Is Wrong, 1st Edition. Penguin Books, City of Westminster, London.

Bakkerode, M., Koning, R., 2014. The conversion of penalty kicks in the English Premier League. Master's thesis at Rijksuniversiteit Groningen.

Barreira, D., Garganta, J., Guimarães, P., Machado, J. C., Anguera, M. T., 2014. Ball recovery patterns as a performance indicator in elite soccer. Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology 228 (1), 61–72.

Barreira, D., Garganta, J., Pinto, T., Valente, J., Anguera, M., Nunome, H., Drust, B., Dawson, B., 2013. Do attacking game patterns differ between first and second halves of soccer matches in the 2010 FIFA World Cup? In: Science and football VII: the proceedings of the seventh world congress on science and football. London and New York: Routledge. pp. 193–198.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Matthews, I., 2014a. "win at home and draw away": automatic formation analysis highlighting the differences in home and away team behaviors. Proceedings of the 2014 MIT Sloan Sports Analytics Conference, http://www. sloansportsconference.com.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., Matthews, I., 2014b. Identifying team style in soccer using formations learned from spatiotemporal tracking data. In: Data Mining Workshop (ICDMW), 2014 IEEE International Conference on. IEEE, pp. 9–14.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., Matthews, I., 2014c. Large-scale analysis of soccer matches using spatiotemporal tracking data. In: Data Mining (ICDM), 2014 IEEE International Conference on. IEEE, pp. 725–730.

Bjertnes, V., Nørstebø, O., Vabo, E., 2015. Identifying key performance indicators in Norwegian Association Football. Project report at Norwegian University of Science and Technology.

Bojinov, I., Bornn, L., 2016. The Pressing Game: Optimal Defensive Disruption in Soccer. In: Proceedings of the 2016 MIT Sloan Sports Analytics Conference, http://www. sloansportsconference.com.

Bradley, P. S., Lago-Peñas, C., Rey, E., Gomez Diaz, A., 2013. The effect of high and low percentage ball possession on physical and technical profiles in English FA Premier League soccer matches. Journal of Sports Sciences 31 (12), 1261–1270.

Bray, K., Kerwin, D., 2003. Modelling the flight of a soccer ball in a direct free kick. Journal of sports sciences 21 (2), 75–85.

Bray, S. R., Law, J., Foyle, J., 2003. Team quality and game location effects in English professional soccer. Journal of Sport Behavior 26 (4), 319.

Brillinger, D. R., 2007. Modelling some norwegian soccer data. Advances in Statistical Modeling and Inference 3, 3–20.

Brown Jr, T. D., Van Raalte, J. L., Brewer, B. W., Winter, C. R., et al., 2002. World Cup soccer home advantage. Journal of Sport Behavior 25 (2), 134.

Buraimo, B., Forrest, D., Simmons, R., 2010. The 12th man?: refereeing bias in English and German soccer. Journal of the Royal Statistical Society: Series A (Statistics in Society) 173 (2), 431–449.

Buraimo, B., Simmons, R., Maciaszczyk, M., 2012. Favoritism and referee bias in European soccer: evidence from the Spanish league and the UEFA Champions League. Contemporary Economic Policy 30 (3), 329–343.

Carmichael, F., Thomas, D., 2005. Home-field effect and team performance evidence from English Premiership football. Journal of sports economics 6 (3), 264–281.

Castellano, J., Casamichana, D., Lago, C., 2012. The use of match statistics that discriminate between successful and unsuccessful soccer teams. Journal of human kinetics 31, 137–147.

Cervone, D., D'Amour, A., Bornn, L., Goldsberry, K., 2014. Pointwise: Predicting points and valuing decisions in real time with nba optical tracking data. In: Proceedings of the 2014 MIT Sloan Sports Analytics Conference, http://www. sloansportsconference.com. Vol. 28.

Chang, Y.-H., Maheswaran, R., Su, J., Kwok, S., Levy, T., Wexler, A., Squire, K., 2014. Quantifying shot quality in the NBA. In: Proceedings of the 2014 MIT Sloan Sports Analytics Conference, http://www. sloansportsconference.com.

Clarke, S. R., Norman, J. M., 1995. Home ground advantage of individual clubs in English soccer. The Statistician 44 (4), 509–521.

Coleman, B. J., 2012. Identifying the "players" in sports analytics research. Interfaces 42 (2), 109–118.

Collet, C., 2013. The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. Journal of Sports Sciences 31 (2), 123–136.

Constantinou, A. C., Fenton, N. E., Pollock, L. J. H., 2014. Bayesian networks for unbiased assessment of referee bias in association football. Psychology of Sport and Exercise 15 (5), 538–547.

De Baranda, P. S., Lopez-Riquelme, D., 2012. Analysis of corner kicks in relation to match status in the 2006 World Cup. European Journal of Sport Science 12 (2), 121–129.

Duarte, R., Araújo, D., Folgado, H., Esteves, P., Marques, P., Davids, K., 2013. Capturing complex, non-linear team behaviours during competitive football performance. Journal of Systems Science and Complexity 26 (1), 62–72.

Duarte, R., Araújo, D., Freire, L., Folgado, H., Fernandes, O., Davids, K., 2012. Intra-and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. Human Movement Science 31 (6), 1639–1651.

Fearnhead, P., Taylor, B. M., 2010. On estimating the ability of NBA players. Journal of quantitative analysis in sports 7 (3), Article 11.

Fernando, T., Wei, X., Fookes, C., Sridharan, S., Lucey, P., 2015. Discovering methods of scoring in soccer using tracking data. In: Proceedings of 2015 KDD Workshop on Large-Scale Sports Analytics, http://www.large-scale-sports-analytics.org/.

Gazetta, 2014. `http://www.gazzettaworld.com/features/greatest-young-players/5/`, Online; accessed 24-May-2016.

Goumas, C., 2014. Home advantage in Australian soccer. Journal of Science and Medicine in Sport 17 (1), 119–123.

Groebner, D., Shannon, P., Fry, P., Smith, K., 2011. Business Statistics: A Decision-Making Approach, 8th Edition. Prentice Hall/Pearson, Upper Saddle River, N.J.

Gudmundsson, J., Wolle, T., 2014. Football analysis using spatio-temporal tools. Computers, Environment and Urban Systems 47, 16–27.

Gyarmati, L., Hefeeda, M., 2016. Analyzing in-game movements of soccer players at scale abs/1603.05583.
URL `http://arxiv.org/abs/1603.05583`

Gyarmati, L., Kwak, H., Rodriguez, P., 2014. Searching for a unique style in soccer. In: Proceedings of 2014 KDD Workshop on Large-Scale Sports Analytics, http://www.large-scale-sports-analytics.org/.

Halvorsen, P., Sægrov, S., Mortensen, A., Kristensen, D. K., Eichhorn, A., Stenhaug, M., Dahl, S., Stensland, H. K., Gaddam, V. R., Griwodz, C., et al., 2013. Bagadus: an integrated system for arena sports analytics: a soccer case study. In: Proceedings of the 4th ACM Multimedia Systems Conference. ACM, pp. 48–59.

Hirotsu, N., Wright, M., 2002. Using a markov process model of an association football match to determine the optimal timing of substitutions and tactical decisions. Journal of the Operational Research Society 53 (1), 88–96.

Hirotsu, N., Wright, M., 2003a. Determining the best strategy for changing the configuration of a football team. Journal of the Operational Research Society 54 (1), 878–887.

Hirotsu, N., Wright, M., 2003b. An evaluation of characteristics of teams in association football by using a markov process model. Journal of the Royal Statistical Society: Series D (The Statistician) 52 (4), 591–602.

Hosmer, D. W., Lemeshow, S., 2000. Applied Logistic Regression, 2nd Edition. John Wiley & Sons, 605 Third Avenue, New York.

Hughes, M., Franks, I., 2005. Analysis of passing sequences, shots and goals in soccer. Journal of Sports Sciences 23 (5), 509–514.

Hvattum, L. M., Arntzen, H., 2010. Using ELO ratings for match result prediction in association football. International Journal of Forecasting 26 (3), 460–470.

Jankovic, A., Leontijevic, B., Pasic, M., Jelusic, V., 2011. Influence of certain tactical attacking patterns on the result achieved by the teams participants of the 2010 FIFA World Cup in South Africa. Physical Culture 65 (1), 34–45.

Jordet, G., Hartman, E., 2008. Avoidance motivation and choking under pressure in soccer penalty shootouts. Journal of Sports & Exercise Psychology 30, 450–457.

Kerr, M. G. S., 2015. Applying machine learning to event data in soccer. Master's thesis at Massachusets Institute of Technology.

Kim, H.-C., Kwon, O., Li, K.-J., 2011. Spatial and spatiotemporal analysis of soccer. In: Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM, pp. 385–388.

Kumar, V., Kaur, M., Sharma, D. P., 2013. Effect of speed endurance training on crossing accuracy in soccer. International Journal of Movement Education and Sports Sciences 2, 99–104.

Lago, C., Casais, L., Dominguez, E., Sampaio, J., 2010. The effects of situational variables on distance covered at various speeds in elite soccer. European Journal of Sport Science 10 (2), 103–109.

Lago, C., Martín, R., 2007. Determinants of possession of the ball in soccer. Journal of Sports Sciences 25 (9), 969–974.

Lago-Ballesteros, J., Lago-Peñas, C., 2010. Performance in team sports: identifying the keys to success in soccer. Journal of Human Kinetics 25, 85–91.

Lago-Peñas, C., 2012. The role of situational variables in analysing physical performance in soccer. Journal of human kinetics 35 (1), 89–95.

Lago-Peñas, C., Dellal, A., 2010. Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. Journal of Human Kinetics 25, 93–100.

Littman, M. L., 2001. Value-function reinforcement learning in Markov games. Journal of Cognitive Systems Research (2), 55–66.

Lopes, J., Araújo, D., Peres, R., Davids, K., Barreiros, J., 2008. The dynamics of decision making in penalty kick situations in association football. The Open Sports Sciences Journal 1 (1), 24–30.

Lucey, P., Bialkowski, A., Monfort, M., Carr, P., Matthews, I., 2014. quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. MIT Sloan Sports Analytics Conference.

Lucey, P., Oliver, D., Carr, P., Roth, J., Matthews, I., 2013a. Assessing team strategy using spatiotemporal data. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 1366–1374.

Lucey, P., Oliver, D., Carr, P., Roth, J., Matthews, I., 2013b. Assessing team strategy using spatiotemporal data. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '13. ACM, New York, NY, USA, pp. 1366–1374.
URL http://doi.acm.org/10.1145/2487575.2488191

Macdonald, B., 2010. A regression-based adjusted plus-minus statistic for nhl players. Journal of Quantitative Analysis in Sports 7 (3).

Macdonald, B., 2011a. Adjusted plus-minus for nhl players using ridge regression with goals, shots, fenwick, and corsi. arXiv e-print arXiv:1201.317.

Macdonald, B., 2011b. An improved adjusted plus-minus statistic for NHL players. In: Proceedings of the MIT Sloan Sports Analytics Conference, URL http://www. sloansportsconference. com.

Macdonald, B., 2012. An expected goals model for evaluating nhl teams and players. In: Proceedings of the 2012 MIT Sloan Sports Analytics Conference, http://www. sloansportsconference. com.

Macdonald, B., Lennon, C., Sturdivant, R., 2012. Evaluating nhl goalies, skaters, and teams using weighted shots. arXiv preprint arXiv:1205.1746.

Macdonald, B., Weld, C., Arney, D. C., 2013. Quantifying playmaking ability in hockey. arXiv preprint arXiv:1307.6539.

Machado, J. C., Barreira, D., Garganta, J., 2014. The influence of match status on attacking patterns of play in elite soccer teams. Revista Brasileira de Cineantropometria & Desempenho Humano 16 (5), 545–554.

McHale, I., Scarf, P., 2007. Modelling soccer matches using bivariate discrete distributions with general dependence structure. Statistica Neerlandica 61 (4), 432–445.

McHale, I. G., Scarf, P. A., Folker, D. E., 2012. On the development of a soccer player performance rating system for the english premier league. Interfaces 42 (4), 339–351.

Mechtel, M., Bäker, A., Brändle, T., Vetter, K., 2011. Red cards: not such bad news for penalized guest teams. Journal of Sports Economics 12, 621–646.

Mirror, 2015. `http://www.mirror.co.uk/sport/football/news/20-most-promising-youngsters-world-6034191`, Online; accessed 24-May-2016.

Misirlisoy, E., Haggard, P., 2014. Asymmetric predictability and cognitive competition in football penalty shootouts. Current Biology 24 (16), 1918–1922.

Moura, F. A., Martins, L. E. B., Anido, R. D. O., De Barros, R. M. L., Cunha, S. A., 2012. Quantitative analysis of Brazilian football players' organisation on the pitch. Sports Biomechanics 11 (1), 85–96.

Niu, Z., Gao, X., Tian, Q., 2012. Tactic analysis based on real-world ball trajectory in soccer video. Pattern Recognition 45 (5), 1937–1947.

Noël, B., Furley, P., Van der Kamp, J., Dicks, M., Memmert, D., 2015. The development of a method for identifying penalty kick strategies in association football. Journal of sports sciences 33 (1), 1–10.

Opta, 2016. `https://www.optasports.com`, Online; accessed 24-January-2016.

Orth, D., Davids, K., Araújo, D., Renshaw, I., Passos, P., 2014. Effects of a defender on run-up velocity and ball speed when crossing a football. European Journal of Sport Science 14 (1), 316–323.

Pettigrew, S., 2015. Assessing the offensive productivity of nhl players using in-game win probabilities. In: 9th Annual MIT Sloan Sports Analytics Conference. Vol. 2. p. 8.

Pollard, R., Silva, C., Medeiros, N., 2008. Home advantage in football in Brazil: differences between teams and the effects of distance traveled. Revista Brasileira de Futebol (The Brazilian Journal of Soccer Science) 1 (1), 3–10.

Poole, D. L., Mackworth, A. K., 2010. Artificial Intelligence: foundations of computational agents. Cambridge University Press, 32 Avenue of the Americas, New York, NY, USA.

Premier League, 2016. `http://www.premierleague.com/en-gb/players/ea-sports-player-performance-index.html`, Online; accessed 19-May-2016.

Reep, C., Benjamin, B., 1968. Skill and chance in association football. Journal of the Royal Statistical Society. Series A (General) 131 (4), 581–585.

Reep, C., Pollard, R., Benjamin, B., 1971. Skill and chance in ball games. Journal of the Royal Statistical Society. Series A (General) 134 (4), 623–629.

Reilly, B., Witt, R., 2013. Red cards, referee home bias and social pressure: evidence from English Premiership soccer. Applied Economics Letters 20 (7), 710–714.

Ridder, G., Cramer, J., Hopstaken, P., 1994. Down to ten: estimating the effect of a red card in soccer. Journal of the American Statistical Association 89 (427), 1124–1127.

Routley, K. D., 2015. A Markov game model for valuing player actions in ice hockey. Master's thesis at Simon Fraser University.

Ruiz-Ruiz, C., Fradua, L., Fernández-GarcÍa, Á., Zubillaga, A., 2013. Analysis of entries into the penalty area as a performance indicator in soccer. European Journal of Sport Science 13 (3), 241–248.

Sæbø, O. D., 2015. Modelling clubs financial investments in association football players. Master's thesis at Norwegian University of Science and Technology.

Sæbø, O. D., Hvattum, L. M., 2015. Evaluating the efficiency of the association football transfer market using regression based player ratings. Norsk Informatikkonferanse (NIK).
URL http://ojs.bibsys.no/index.php/NIK/article/view/246

Santín, D., 2014. Measuring the technical efficiency of football legends: who were real madrid's all-time most efficient players? International Transactions in Operational Research 21 (3), 439–452.

Schuckers, M., Curro, J., 2013. Total hockey rating (thor): A comprehensive statistical rating of national hockey league forwards and defensemen based upon all on-ice events. In: 7th Annual MIT Sloan Sports Analytics Conference.

Seçkin, A., 2006. Home advantage in association football: evidence from Turkish Super League. In: ECOMOD Conference in Hong Kong, China, June. pp. 28–30.

Seçkin, A., Pollard, R., 2008. Home advantage in Turkish professional soccer. Perceptual and motor skills 107 (1), 51–54.

Singh, P. K., Ahmad, M., 2015. Performance prediction of players in sports league matches. International Journal of Science and Research (ISJR) 4 (4), 2207–2213.

Taylor, J. B., James, N., Mellalieu, S. D., 2005. Notational analysis of corner kicks in English Premier League soccer. In: Science and Football V: the Proceedings of the Fifth World Congress on Football. pp. 229–234.

Taylor, J. B., Mellalieu, S. D., James, N., Shearer, D. A., 2008. The influence of match location, quality of opposition, and match status on technical performance in professional association football. Journal of Sports Sciences 26 (9), 885–895.

Tenga, A., Holme, I., Ronglan, L., Bahr, R., 2010a. Effects of match location on playing tactics for goal scoring in norwegian professional soccer. Journal of Sport Behavior 33 (1), 89.

Tenga, A., Holme, I., Ronglan, L. T., Bahr, R., 2010b. Effect of playing tactics on achieving score-box possessions in a random series of team possessions from norwegian professional soccer matches. Journal of Sports Sciences 28 (3), 245–255.

Tenga, A., Holme, I., Ronglan, L. T., Bahr, R., 2010c. Effect of playing tactics on goal scoring in norwegian professional soccer. Journal of Sports Sciences 28 (3), 237–244.

Tenga, A., Ronglan, L. T., Bahr, R., 2010d. Measuring the effectiveness of offensive match-play in professional soccer. European Journal of Sport Science 10 (4), 269–277.

Thomas, A., Ventura, S. L., Jensen, S. T., Ma, S., et al., 2013. Competing process hazard function models for player ratings in ice hockey. The Annals of Applied Statistics 7 (3), 1497–1524.

Thomas, A. C., 2006. The impact of puck possession and location on ice hockey strategy. Journal of Quantitative Analysis in Sports 2 (1), 1–16.

Transfermarkt, 2016. www.transfermarkt.com, Online; accessed 20-April-2016.

Tunaru, R., Clark, E., Viney, H., 2005. An option pricing framework for valuation of football players. Review of financial economics 14 (3), 281–295.

Tunaru, R. S., Viney, H. P., 2010. Valuations of soccer players from statistical performance data. Journal of Quantitative Analysis in Sports 6 (2), 1–23.

Unkelbach, C., Memmert, D., et al., 2008. Game management, context effects, and calibration: the case of yellow cards in soccer. Journal of Sport and Exercise Psychology 30 (1), 95.

Vecer, J., 2014. Crossing in soccer has a strong negative impact on scoring: evidence from the English Premier League, the German Bundesliga and the World Cup 2014. Available at SSRN 2225728, http://ssrn.com/abstract=2225728.

Vecer, J., Kopriva, F., Ichiba, T., 2009. Estimating the effect of the red card in soccer: when to commit an offense in exchange for preventing a goal opportunity. Journal of Quantitative Analysis in Sports 5 (1), Article 8.

VG, 2014. http://vglive.no/#85585=&eliteguiden=s-borsen-sid-669, Online; accessed 20-April-2016.

VG, 2015. http://vglive.no/#85585=&eliteguiden=s-borsen-sid-706, Online; accessed 20-April-2016.

Wei, X., Sridharan, S., Lucey, P., Navarathna, R., Morgan, S., 2014. Analyzing and predicting events in soccer and tennis using spatiotemporal data. In: 2014 KDD Workshop on Large-Scale Sports Analytics.
URL http://eprints.qut.edu.au/78738/

Wikipedia, 2016. Logistic Regression — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Logistic_regression, Online; accessed 24-May-2016.

Wright, C., Atkins, S., Polman, R., Jones, B., Sargeson, L., 2011. Factors associated with goals and goal scoring opportunities in professional soccer. International Journal of Performance Analysis in Sport 11 (3), 438–449.

Wright, M., 2015. Operational Research Applied to Sports. Palgrave Macmillan, 175 Fifth Avenue, New Yotk, NY, 10010.

Yamada, H., Hayashi, Y., 2015. Characteristics of goal-scoring crosses in international soccer tournaments. Football Science 12, 24–32.

Appendix A

Tables from Tippeligaen

Appendix A serves as background information on how the 2014 and 2015 seasons of Tippeligaen ended, included for reference to the player ratings from the three developed models. Final league tables, top scorers and players with the most assists are presented in Tables A.1 to A.6.

A.1 2014

Table A.1: League Table Tippeligaen 2014

Position	Team	Games Played	Won	Drawn	Lost	Goals For	Goals Against	Goal Difference	Points
1	Molde	30	22	5	3	62	24	38	71
2	Rosenborg	30	18	6	6	64	43	21	60
3	Odd	30	17	7	6	52	32	20	58
4	Strømsgodset	30	15	5	10	48	42	6	50
5	Lillestrøm	30	13	7	10	49	35	14	46
6	Vålerenga	30	11	9	10	59	53	6	42
7	Aalesund	30	11	8	11	40	39	1	41
8	Sarpsborg 08	30	10	10	10	41	48	-7	40
9	Stabæk	30	11	6	13	44	52	-8	39
10	Viking	30	8	12	10	42	42	0	36
11	Haugesund	30	10	6	14	43	49	-6	36
12	Start	30	10	5	15	47	60	-13	35
13	Bodø/Glimt	30	10	5	15	45	60	-15	35
14	Brann	30	8	5	17	41	54	-13	29
15	Sogndal	30	6	6	18	31	49	-18	24
16	Sandnes Ulf	30	4	10	16	27	53	-26	22

Table A.2: Top 20 players with respect to goals 2014

Position	Player	Team	Goals	Games Played	Average
1	Vidar Örn Kjartansson	Vålerenga	25	29	0.86
2	Christian Gytkjær	Haugesund	15	26	0.58
3	Alexander Søderlund	Rosenborg	13	23	0.57
4	Franck Boli	Stabæk	13	28	0.46
5	Abdurahim Laajab	Bodø/Glimt	13	30	0.43
6	Mohamed Elyounoussi	Molde	13	30	0.43
7	Frode Johnsen	Odd	11	30	0.37
8	Leke James	Aalesund	10	23	0.43
9	Fredrik Gulbrandsen	Molde	10	23	0.43
10	Péter Kovács	Strømsgodset	10	24	0.42
11	Maic Sema	Haugesund	10	26	0.38
12	Daniel Chima Chukwu	Molde	10	27	0.37
13	Fredrik Brustad	Stabæk	10	30	0.33
14	Pálmi Rafn Pálmason	Lillestrøm	9	27	0.33
15	Mahatma Otoo	Sogndal	9	27	0.33
16	Jone Samuelsen	Odd	9	28	0.32
17	Vidar Nisja	Viking	9	28	0.32
18	Jakob Orlov	Brann	9	29	0.31
19	Ghayas Zahid	Vålerenga	9	29	0.31
20	Mike Jensen	Rosenborg	9	29	0.31

Table A.3: Top 20 players with respect to assist 2014

Position	Player	Team	Assists	Games Played	Average
1	Bjørn Helge Riise	Lillestrøm	10	27	0.37
2	Petter Vaagan Moen	Lillestrøm	9	27	0.33
3	Morten Gamst Pedersen	Rosenborg	8	24	0.33
4	Hjörtur Logi Valgardsson	Sogndal	8	26	0.31
5	Zlatko Tripić	Start	8	28	0.29
6	Jone Samuelsen	Odd	8	28	0.29
7	Christian Grindheim	Vålerenga	8	29	0.28
8	Martin Ødegaard	Strømsgodset	7	23	0.30
9	Per-Egil Flo	Molde	7	24	0.29
10	Michael Francis Stephens	Stabæk	7	30	0.23
11	Pål André Helland	Rosenborg	6	21	0.29
12	Lars-Christopher Vilsvik	Strømsgodset	6	25	0.24
13	Håvard Storbæk	Odd	6	27	0.22
14	Mattias Moström	Molde	6	28	0.21
15	Martin Linnes	Molde	6	28	0.21
16	Petter Strand	Sogndal	6	29	0.21
17	Jón Dadi Bödvarsson	Viking	6	29	0.21
18	Papa Alioune Ndiaye	Bodø/Glimt	6	30	0.20
19	Fredrik Brustad	Stabæk	6	30	0.20
20	Magne Hoseth	Stabæk	5	9	0.56

A.2 2015

Table A.4: League Table Tippeligaen 2015

Position	Team	Games Played	Won	Drawn	Lost	Goals For	Goals Against	Goal Difference	Points
1	Rosenborg	30	21	6	3	73	27	46	69
2	Strømsgodset	30	17	6	7	67	44	23	57
3	Stabæk	30	17	5	8	54	43	11	56
4	Odd	30	15	10	5	61	41	20	55
5	Viking	30	17	2	11	53	39	14	53
6	Molde	30	15	7	8	62	31	31	52
7	Vålerenga	30	14	7	9	49	41	8	49
8	Lillestrøm	30	12	9	9	45	43	2	44
9	Bodø/Glimt	30	12	4	14	53	56	-3	40
10	Aalesund	30	11	5	14	42	57	-15	38
11	Sarpsborg 08	30	8	10	12	37	49	-12	34
12	Haugesund	30	8	7	15	33	52	-19	31
13	Tromsø	30	7	8	15	36	50	-14	29
14	Start	30	5	7	18	35	64	-29	22
15	Mjøndalen	30	4	9	17	38	69	-31	21
16	Sandefjord	30	4	4	22	36	68	-32	16

Table A.5: Top 20 players with respect to goals 2015

Position	Player	Team	Goals	Games Played	Average
1.	Alexander Søderlund	Rosenborg	22	27	0.81
2.	Adama Diomandé	Stabæk	17	21	0.81
3.	Olivier Occéan	Odd	15	27	0.56
4.	Ola Kamara	Molde	14	29	0.48
5.	Pål André Helland	Rosenborg	13	21	0.62
6.	Alexander Sørloth	Bodø/Glimt	13	26	0.50
7.	Trond Olsen	Bodø/Glimt	13	29	0.45
8.	Leke James	Aalesund	13	29	0.45
9.	Mohamed Elyounoussi	Molde	12	28	0.43
10.	Marcus Pedersen	Strømsgodset	11	10	1.10
11.	Veton Berisha	Viking	11	14	0.79
12.	Fred Friday	Lillestrøm	11	26	0.42
13.	Iver Fossum	Strømsgodset	11	30	0.37
14.	Erling Knudtzon	Lillestrøm	10	29	0.34
15.	Ernest Asante	Stabæk	10	30	0.33
16.	Christian Gytkjær	Haugesund	10	30	0.33
17.	Tommy Høiland	Molde	9	23	0.39
18.	Fredrik Nordkvelle	Odd	9	24	0.38
19.	Zdenek Ondrásek	Tromsø	9	27	0.33
20.	Matthías Vilhjálmsson	Rosenborg/Start	9	27	0.33

Table A.6: Top 20 players with respect to assist 2015

Position	Player	Team	Assists	Games Played	Average
1.	Mike Jensen	Rosenborg	13	29	0.45
2.	Yassine El Ghanassy	Stabæk	11	26	0.42
3.	Ernest Asante	Stabæk	11	30	0.37
4.	Christian Grindheim	Vålerenga	11	30	0.37
5.	Anders Trondsen	Sarpsborg 08	9	24	0.38
6.	Espen Ruud	Odd	9	28	0.32
7.	Fredrik Midtsjø	Rosenborg	9	29	0.31
8.	Olivier Occéan	Odd	8	27	0.30
9.	Gustav Wikheim	Strømsgodset	8	28	0.29
10.	Pål André Helland	Rosenborg	7	21	0.33
11.	Daniel José Bamberg	Haugesund	7	25	0.28
12.	Per-Egil Flo	Molde	7	25	0.28
13.	Joackim Olsen Solberg	Mjøndalen	7	28	0.25
14.	Bentley	Odd	7	30	0.23
15.	Lars-Christopher Vilsvik	Strømsgodset	6	24	0.25
16.	Øyvind Storflor	Strømsgodset	6	27	0.22
17.	Matthías Vilhjálmsson	Rosenborg	6	27	0.22
18.	Christian Andreas Gauseth	Mjøndalen	6	28	0.21
19.	Trond Olsen	Bodø/Glimt	6	29	0.21
20.	André Danielsen	Viking	6	30	0.20

Appendix B

Results from Expected Goals Model

Appendix B presents additional results from running the xG model. First, the correlation matrix for all variables in the model for regular shots is presented. Thereafter, all players scoring more than 8 goals in Tippeligaen 2014 are presented and rated by G/xG. The results from 2015 are repeated for comparison in Table B.4. In addition, two team tables are presented showing xG difference and goal difference for 2014 and 2015 in Table B.3 and B.5, respectively.

Table B.1: xG Model: Correlation matrix

	Length	Angle	Fast break	Take ons	Header	Cross	Through ball	Match status	Home/Away	Elo diff.
Length	1									
Angle	-0.7876	1								
Fast break	-0.0052	-0.0169	1							
Take ons	0.0575	-0.0956	0.0086	1						
Header	-0.4975	0.4598	-0.0493	-0.1043	1					
Cross	-0.5278	0.4501	-0.0432	-0.1142	0.6623	1				
Through ball	-0.0501	-0.0167	0.0371	-0.0093	-0.0535	-0.0655	1			
Match status	-0.0299	0.0244	0.0806	0.0367	-0.0586	-0.0393	0.0486	1		
Home/Away	-0.0301	0.0290	-0.0202	-0.0147	0.0334	0.0224	-0.0066	0.1242	1	
Elo diff.	-0.0492	0.0301	0.0121	0.0057	0.0083	0.0157	-0.0044	0.2273	0.0193	1

B.1 2014

Table B.2: xG Model: All players scoring more than 8 goals in Tippeligaen 2014, ranked by G/xG

Name	Team	Position	Minutes played	Goals	xG	Shots	xG/shots	Goals/xG
Vidar Nisja	Viking	Attacking midfielder	1922	9	4.63	49	0.09	1.94
Jone Samuelsen	Odd	Central midfielder	2498	9	5.57	66	0.08	1.61
Fredrik Gulbrandsen	Molde	Forward	1475	10	7.00	55	0.13	1.43
Papa Alioune Ndiaye*	Bodø/Glimt	Attacking midfielder	2555	9	6.70	98	0.07	1.34
Ernest Asante	Start	Winger	1933	8	5.97	65	0.09	1.34
Tommy Høiland	Molde/Lillestrøm	Forward	672	8	6.20	31	0.20	1.29
Riku Riski	Rosenborg	Winger	2174	8	6.30	61	0.10	1.27
Péter Kovács	Strømsgodset	Forward	1429	10	8.03	45	0.18	1.25
Frode Johnsen	Odd	Forward	2588	11	8.84	67	0.13	1.24
Jakob Orlov	Brann	Forward	2248	9	7.52	66	0.11	1.20
Mike Jensen	Rosenborg	Central midfielder	2590	9	7.53	81	0.09	1.19
Herolind Shala*	Odd	Midfielder	2093	9	7.64	78	0.10	1.18
Franck Boli	Stabæk	Forward	2249	13	11.48	65	0.18	1.13
Fredrik Brustad*	Stabæk	Winger	2261	10	9.18	50	0.18	1.09
Vidar Örn Kjartansson*	Vålerenga	Forward	2608	25	29.99	117	0.20	1.09
Ghayas Zahid	Vålerenga	Attacking midfielder	2216	9	8.34	57	0.15	1.08
Alexander Søderlund*	Rosenborg	Forward	1614	13	12.18	49	0.25	1.07
Christian Gytkjær	Haugesund	Forward	2001	15	14.86	71	0.21	1.01
Mohamed Elyounoussi	Molde	Winger	2533	13	13.14	107	0.12	0.99
Mahatma Otoo	Sogndal	Forward	2077	9	9.15	79	0.12	0.98
Maic Sema*	Haugesund	Attacking midfielder	1803	10	10.29	51	0.20	0.97
Diego Iván Rubio**	Sandnes Ulf	Forward	1794	8	8.32	66	0.13	0.96
Pálmi Rafn Pálmason	Lillestrøm	Attacking midfielder	2079	9	9.55	55	0.17	0.94
Abdurahim Laajab*	Bodø/Glimt	Forward	2468	13	14.43	102	0.14	0.90
Leke James	Aalesund	Forward	2031	10	11.82	76	0.16	0.85
Daniel Chima Chukwu*	Molde	Forward	1998	10	13.66	67	0.20	0.73

* sold to foreign club during or some time after the 2014 season ** on loan from a foreign club in 2014

Table B.3: xG Model: Tippeligaen 2014 teams rated by xG difference

Team	xG	xG received	Real goal difference*	xG difference
Molde	67.80	19.82	38	47.99
Odd	49.21	32.38	20	16.83
Rosenborg	58.17	42.58	21	15.58
Strømsgodset	55.99	42.51	6	13.48
Lillestrøm	47.10	39.42	14	7.67
Vålerenga	54.03	48.20	6	5.82
Viking	50.44	44.96	0	5.48
Haugesund	51.83	47.42	-6	4.41
Aalesund	42.14	44.99	1	-2.85
Stabæk	43.47	52.37	-8	-8.89
Sogndal	31.87	45.83	-18	-13.97
Sandnes Ulf	35.54	50.63	-26	-15.09
Bodø/Glimt	39.98	56.37	-15	-16.40
Sarpsborg 08	33.72	50.08	-7	-17.35
Start	37.88	55.60	-13	-17.73
Brann	36.84	55.25	-13	-18.41

* include own goals

B.2 2015

Table B.4: xG Model: All players scoring more than 8 goals in Tippeligaen 2015, ranked by G/xG

Name	Team	Position	Minutes played	Goals	xG	Shots	xG/S	G/xG
Trond Olsen	Bodø/Glimt	Winger	2510	13	6.55	74	0.09	1.98
Simon Diedhiou*	Haugesund	Forward	1945	9	4.90	52	0.09	1.84
Veton Berisha*	Viking	Forward	1256	11	6.57	47	0.14	1.67
Pål Alexander Kirkevold*	Sandefjord	Forward	1881	8	5.35	66	0.08	1.50
Kristoffer Ajer*	Start	Central Midfielder	2581	8	5.43	43	0.13	1.47
Marcus Pedersen	Strømsgodset	Forward	835	11	8.20	34	0.24	1.34
Alexander Sørloth*	Bodø/Glimt	Forward	1776	13	9.89	58	0.17	1.31
Luc Kassi	Stabæk	Forward	2130	8	6.26	54	0.12	1.28
Alexander Søderlund*	Rosenborg	Forward	2242	22	17.44	87	0.20	1.26
Tobias Mikkelsen*	Rosenborg	Winger	1931	8	6.44	69	0.09	1.24
Ernest Asante	Stabæk	Winger	2643	10	8.06	70	0.12	1.24
Adama Diomandé*	Stabæk	Forward	1850	17	13.92	79	0.18	1.22
Zdeněk Ondrášek*	Tromsø	Forward	2366	9	7.57	74	0.10	1.19
Tommy Høiland	Molde	Forward	1103	9	7.93	36	0.22	1.13
Matthías Vilhjálmsson	Start/Rosenborg	Forward	2087	9	7.99	46	0.17	1.13
Mohamed Elyounoussi	Molde	Winger	2275	12	10.69	84	0.13	1.12
Erling Knudtzon	Lillestrøm	Forward	2595	10	8.91	51	0.17	1.12
Fredrik Nordkvelle	Odd	Attacking Midfielder	1891	9	8.06	48	0.17	1.12
Iver Fossum*	Strømsgodset	Attacking Midfielder	2653	11	10.07	69	0.15	1.09
Pål André Helland	Rosenborg	Winger	1502	13	12.07	89	0.14	1.08
Suleiman Abdullahi	Viking	Forward	1882	8	7.77	70	0.11	1.03
Christian Gytkjær	Haugesund	Forward	2545	10	9.84	57	0.17	1.02
Sander Svendsen	Molde	Forward	1720	8	8.13	60	0.14	0.98
Fred Friday	Lillestrøm	Forward	1770	11	11.22	67	0.17	0.98
Ola Kamara*	Molde	Forward	2384	14	14.69	89	0.17	0.95
Bentley	Odd	Winger	2464	8	8.53	67	0.13	0.94
Leke James*	Aalesund	Forward	2610	13	13.97	91	0.15	0.93
Jón Dadi Bödvarsson*	Viking	Forward	2067	9	10.06	58	0.17	0.89
Gustav Wikheim*	Strømsgodset	Winger	2413	9	11.35	68	0.17	0.79
Olivier Occéan	Odd	Forward	2208	15	20.23	89	0.23	0.74

[*] sold to foreign club during or after the 2015 season [**] on loan from a foreign club in 2015

Table B.5: xG Model: Tippeligaen 2015 teams rated by xG difference

Team	xG	xG received	Real goal difference*	xG difference
Rosenborg	68.66	24.20	46	43.46
Molde	60.32	30.09	31	30.23
Strømsgodset	62.69	33.55	23	29.13
Odd	61.92	34.37	20	27.55
Stabæk	53.67	36.09	11	17.57
Viking	54.56	39.53	14	15.03
Vålerenga	47.56	42.20	8	5.36
Lillestrøm	51.37	49.39	2	1.98
Sarpsborg 08	34.60	41.65	-12	-7.05
Bodø/Glimt	38.86	48.24	-3	-9.38
Tromsø	31.28	46.40	-14	-15.12
Haugesund	30.92	51.44	-19	-20.52
Mjøndalen	38.83	60.22	-31	-21.38
Sandefjord	32.40	59.03	-32	-26.62
Start	32.02	62.42	-29	-30.41
Aalesund	36.43	69.75	-15	-33.32

* include own goals

Appendix C

Results from Markov Game Model 1

Appendix C presents additional results from running Model 1. Top 10 lists for the different playing positions for Tippeligaen 2014 are presented in Tables C.1 and C.2. In Tables C.4 and C.5 are the results for 2015 repeated for comparison. Tables C.3 and C.6 show the league tables that can be obtained for 2014 and 2015, respectively.

C.1 2014

Table C.1: Model 1: Top 10 players in offensive positions, Tippeligaen 2014.
Average value = 0.1118, minimum value = -0.2324

FORWARDS									
Player	Team	Minutes	Total	Shot	Take on	Foul won	Corner won	Aerial	Ball carry
Vidar Örn Kjartansson*	Vålerenga	2608	0.4184	0.3027	0.0020	0.0222	0.0050	-0.0044	0.0696
Fredrik Gulbrandsen	Molde	1475	0.3399	0.1846	0.0528	0.0156	0.0054	0.0004	0.0807
Alexander Søderlund*	Rosenborg	1614	0.3332	0.2981	0.0066	0.0060	0.0018	-0.0009	0.0180
Franck Boli	Stabæk	2249	0.3190	0.1366	0.0509	0.0134	0.0447	0.0037	0.0478
Christian Gytkjær	Haugesund	2001	0.2727	0.2231	-0.0055	0.0066	0.0056	0.0157	0.0181
Abdurahim Laajab*	Bodø/Glimt	2468	0.1979	0.0782	0.0022	0.0177	0.0414	0.0113	0.0426
Daniel Chima Chukwu*	Molde	1998	0.1937	0.0058	0.0271	0.0169	0.0058	0.0022	0.0389
Péter Kovács	Strømsgodset	1429	0.1887	0.1849	-	0.0030	0.0012	0.0200	0.0308
Frode Johnsen	Odd	2588	0.1661	0.0490	-0.0009	0.0123	0.0043	0.0261	0.0100
Diego Iván Rubio**	Sandnes Ulf	1794	0.1604	0.0575	-0.0020	0.0212	0.0061	-0.0106	0.0532
Top 10 average			0.2590	0.1520	0.0133	0.0135	0.0121	0.0064	0.0410
WINGERS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Cross	Corner	Carry
Pål André Helland	Rosenborg	1040	0.3358	0.0298	0.1083	0.0350	0.0878	0.0084	0.0642
Ernest Asante	Start	1933	0.3206	0.1039	0.0731	0.0210	0.0203	-	0.0451
Gustav Wikheim	Strømsgodset	1616	0.2966	0.1089	0.0792	0.0109	0.0506	-0.0017	-0.0009
Zlatko Tripić*	Start	2107	0.2639	0.0144	0.0035	-0.0064	0.0109	0.0484	0.0482
Fredrik Brustad*	Stabæk	2261	0.2471	0.1148	0.0224	0.0212	0.0158	-	0.0520
Øyvind Storflor	Strømsgodset	2029	0.2432	0.0282	0.0841	-0.0039	0.0437	0.0181	0.0211
Riku Riski	Rosenborg	2174	0.2177	0.0276	0.0191	0.0147	0.0271	-0.0005	0.0779
Dane Richards*	Bodø/Glimt	1181	0.2175	0.0615	-0.0004	0.0140	0.0645	0.0266	0.0029
Mohamed Elyounoussi	Molde	2533	0.2132	0.0164	0.0263	0.0043	0.0076	-0.0060	0.0918
Mattias Moström	Molde	2152	0.1875	-0.0387	0.0630	0.0117	0.0109	0.0755	0.0083
Top 10 average			0.2543	0.0467	0.0479	0.0123	0.0339	0.0169	0.0411
ATTACKING MIDFIELDERS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Foul won	Cross	Carry
Martin Ødegaard*	Strømsgodset	1454	0.4030	0.1472	0.1340	0.0038	0.0130	0.0112	0.0154
Maic Sema*	Haugesund	1803	0.2842	0.1967	0.0071	0.0143	0.0025	0.0037	0.0231
Ghayas Zahid	Vålerenga	2216	0.2410	0.0830	0.0625	0.0078	0.0131	-0.0085	0.0485
Petter Vaagan Moen	Lillestrøm	1915	0.2327	0.1107	0.0189	-0.0026	0.0129	0.0164	0.0108
Vidar Nisja	Viking	1922	0.2263	0.1285	0.0073	0.0015	0.0079	-0.0162	0.0601
Herolind Shala*	Odd	2093	0.2163	0.0615	0.0337	0.0062	0.0178	0.0633	0.0259
Håvard Storbæk	Odd	1497	0.2022	0.0618	0.0426	0.0097	0.0086	0.0583	0.0041
Papa Alioune Ndiaye*	Bodø/Glimt	2555	0.1883	0.0039	0.0389	0.0566	0.0092	-0.0036	0.0410
Pálmi Rafn Pálmason	Lillestrøm	2079	0.1801	0.0738	-0.0210	0.0058	0.0565	-0.0034	0.0307
Michael Barrantes*	Aalesund	1608	0.1781	-0.0034	0.0869	0.0269	0.0064	-0.0059	0.0069
Top 10 average			0.2352	0.0864	0.0411	0.0130	0.0148	0.0115	0.0267

* sold to foreign club during or some time after the 2014 season ** on loan from a foreign club in 2014

Table C.2: Model 1: Top 10 players in defensive positions, Tippeligaen 2014.
Average value = 0.1118, minimum value = -0.2324

CENTRAL MIDFIELDERS									
Player	Team	Minutes	Total	Shot	Pass	Tackle	Ball rec.	Cross	Corner
Jone Samuelsen	Odd	2498	0.3338	0.0948	0.0741	0.0136	0.0098	0.0857	-0.0017
Sakari Mattila*	Aalesund	1688	0.3123	0.2082	0.0178	0.0141	0.0086	0.0011	0.0282
Fredrik Midtsjø	Rosenborg	1754	0.2535	0.0700	0.0379	0.0099	0.0059	0.0292	0.0024
Christian Grindheim	Vålerenga	2519	0.2429	0.0598	0.1136	0.0147	0.0102	0.0065	0.0199
Mike Jensen	Rosenborg	2590	0.2299	0.0636	0.0466	0.0102	0.0087	0.0261	0.0046
Makhtar Thioune*	Viking	1115	0.2198	0.0326	0.1161	0.0159	0.0041	0.0002	0.0405
Björn Daníel Sverrisson	Viking	2275	0.2036	0.0618	0.0810	0.0090	0.0059	0.0050	-0.0024
Daniel Berg Hestad	Molde	1255	0.1549	-0.0178	0.1549	0.0108	0.0097	0.0175	-0.0005
Peter Orry Larsen	Aalesund	1887	0.1402	0.0741	0.0112	0.0031	0.0030	0.0049	-0.0008
Ole Kristian Selnæs	Rosenborg	1770	0.1310	0.0120	0.0359	0.0207	0.0127	0.0146	0.0556
Top 10 average			0.2222	0.0659	0.0689	0.0122	0.0079	0.0191	0.0146

FULL BACKS									
Player	Team	Minutes	Total	Shot	Pass	Tackle	Clearance	Ball rec.	Cross
Per-Egil Flo	Molde	1959	0.2597	-0.0166	0.0774	0.0083	0.0127	0.0102	0.0629
Jarkko Hurme	Odd	1018	0.2543	0.0458	0.0093	0.0093	0.0295	0.0063	0.1244
Claes Phillip Kronberg	Sarpsborg 08	2501	0.2242	0.0861	0.0484	0.0159	0.0147	0.0142	0.0196
Mikael Dorsin	Rosenborg	2072	0.2154	0.0662	0.0616	0.0131	0.0133	0.0091	0.0177
Martin Linnes*	Molde	2520	0.1991	0.0221	0.0552	0.0130	0.0166	0.0167	0.0629
Andreas Vindheim	Brann	1874	0.1952	0.0038	0.0027	0.0279	0.0095	0.0108	0.0571
Amin Nouri	Start	1482	0.1898	0.0388	0.0404	0.0214	0.0100	0.0132	0.0406
Kristoffer Haugen	Viking	1717	0.1752	0.0273	0.0384	0.0244	0.0085	0.0146	0.0292
Jonas Svensson	Rosenborg	1934	0.1728	0.0419	0.0765	0.0228	0.0098	0.0097	0.0009
André Danielsen	Viking	2474	0.1643	-0.0129	0.0392	0.0179	0.0190	0.0134	0.0471
Top 10 average			0.2050	0.0303	0.0449	0.0174	0.0144	0.0118	0.0462

CENTRE BACKS									
Player	Team	Minutes	Total	Shot	Pass	Foul won	Tackle	Clearance	Ball rec.
Vegard Forren	Molde	2467	0.2217	0.0344	0.0902	0.0055	0.0118	0.0425	0.0147
Jonas Grønner	Brann	1711	0.1853	0.0501	0.0486	0.0067	0.0042	0.0089	-0.0005
Brede Moe	Bodø/Glimt	2250	0.1636	0.0559	0.0543	0.0052	0.0113	0.0296	0.0020
Jon Inge Høiland	Stabæk	1873	0.1613	0.0967	0.0528	0.0026	0.0048	0.0044	0.0104
Indridi Sigurdsson	Viking	2326	0.1565	0.0491	0.0537	0.0018	0.0067	0.0134	0.0109
Martin Jensen	Sarpsborg 08	1874	0.1489	0.0514	0.0672	0.0083	0.0101	0.0143	0.0076
Jonatan Tollås Nation	Aalesund	1036	0.1383	0.0482	0.0649	0.0078	0.0135	0.0299	0.0052
Simon Andreas Larsen	Vålerenga	2691	0.1352	0.0503	0.0590	0.0025	0.0051	0.0071	0.0079
Nikolaj Høgh	Vålerenga	1192	0.1324	0.0067	0.0658	0.0033	0.0155	0.0166	0.0076
Marius Amundsen	Lillestrøm	1216	0.1323	0.0521	0.0484	0.0034	0.0062	0.01986	0.0036
Top 10 average			0.1575	0.0495	0.0605	0.0047	0.0089	0.0186	0.0069

* sold to foreign club during or some time after the 2014 season ** on loan from a foreign club in 2014

Table C.3: Model 1: League table 2014 based on the model compared to the real table

Team	Value/match	Table	Real table	Diff
Molde	1.1825	1	1	0
Rosenborg	0.5863	2	2	0
Odd	0.4795	3	3	0
Lillestrøm	0.3836	4	5	-1
Vålerenga	0.2055	5	6	-1
Strømsgodset	0.0502	6	4	2
Stabæk	0.0367	7	9	-2
Aalesund	0.0340	8	7	1
Viking	-0.0016	9	10	-1
Sarpsborg 08	-0.1370	10	8	2
Haugesund	-0.1618	11	11	0
Start	-0.3338	12	12	0
Brann	-0.4886	13	14	-1
Bodø/Glimt	-0.5749	14	13	1
Sandnes Ulf	-0.6263	15	16	-1
Sogndal	-0.6343	16	15	1

C.2 2015

Table C.4: Model 1: Top 10 players in offensive positions, Tippeligaen 2015. Average value = 0.1124, minimum value = -0.1248

FORWARDS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Foul won	Cross	Carry
Adama Diomandé*	Stabæk	1850	0.5448	0.2736	0.0252	0.0287	0.0317	0.0454	0.0519
Veton Berisha*	Viking	1256	0.4721	0.3702	0.0475	-0.0167	0.0063	0.0097	0.0562
Alexander Søderlund*	Rosenborg	2242	0.3896	0.3674	-0.0244	0.0003	0.0094	0.0047	0.0154
Tommy Høiland	Molde	1103	0.3771	0.2945	0.0077	0.0175	0.0045	0.0090	0.0099
Fred Friday	Lillestrøm	1770	0.3130	0.1145	-0.0011	0.0574	0.0207	0.0034	0.0954
Alexander Sørloth*	Bodø/Glimt	1776	0.3108	0.2304	-0.0481	0.0258	0.0154	0.0091	0.0250
Luc Kassi	Stabæk	2130	0.2759	0.1064	0.1279	0.0156	0.0140	0.0051	-0.0036
Erling Knudtzon	Lilestrøm	2595	0.2481	0.1051	0.0665	0.0139	0.0118	0.0030	0.0374
Simon Diédhiou*	Haugesund	1945	0.2327	0.1287	0.0097	0.0285	0.0129	-0.0224	0.0431
Matthías Vilhjálmsson	Start/Rosenborg	2087	0.1961	0.0941	0.0450	0.0047	0.0060	0.0146	0.0317
Top 10 average			0.3360	0.2085	0.0256	0.0175	0.0133	0.0082	0.0362

WINGERS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Foul won	Cross	Carry
Pål André Helland	Rosenborg	1502	0.5553	0.2431	0.0381	0.0055	0.0357	0.0021	0.1153
Trond Olsen	Bodø/Glimt	2510	0.3749	0.1932	0.0374	0.0244	0.0110	0.0503	0.0503
Moryké Fofana*	Lillestrøm	1300	0.3683	0.1777	0.0728	0.0400	0.0109	-0.0013	0.0098
Zymer Bytyqi	Viking	1258	0.3111	0.0527	0.0236	0.0130	0.0043	0.1469	0.0315
Ernest Asante	Stabæk	2643	0.2961	0.0470	0.0850	0.0226	0.0172	0.0206	0.0741
Gustav Wikheim*	Strømsgodset	2413	0.2935	-0.0084	0.1016	0.0708	0.0037	0.0528	0.0439
Espen Børufsen	Start	2154	0.2209	0.0815	-0.0075	0.0107	0.0084	0.0619	0.0145
Ole Jørgen Halvorsen	Odd	1393	0.2160	0.0144	0.0047	-0.0057	0.0038	0.1064	0.0295
Magnus Andersen	Tromsø	2663	0.1950	0.0428	0.0321	-0.0030	0.0036	0.0343	0.0374
Mohamed Elyounoussi	Molde	2275	0.1855	0.0892	0.0513	-0.0051	0.0185	-0.0146	0.0165
Top 10 average			0.3017	0.0933	0.0439	0.0173	0.0117	0.0459	0.0423

ATTACKING MIDFIELDERS									
Player	Team	Minutes	Total	Shot	Pass	Cross	Corner	Carry	FK
Fredrik Nordkvelle	Odd	1891	0.2719	0.1954	0.0041	0.0174	-0.0011	-0.0069	-0.0006
Daniel Fredheim Holm	Vålerenga	1974	0.2687	0.1749	0.0404	0.0097	0.0026	-0.0116	0.0021
Michael Barrantes*	Aalesund	1000	0.2261	0.0503	0.0358	-0.0020	0.0122	0.0397	0.0297
Eirik Hestad	Molde	941	0.2161	0.0232	0.0963	0.0041	0.0195	-0.0072	0.0381
Papa Alioune Ndiaye*	Bodø/Glimt	1286	0.2087	-0.0279	0.0383	0.0189	0.0098	0.0321	-0.0010
Ghayas Zahid	Vålerenga	2383	0.2055	0.0082	0.0851	-0.0040	-0.0007	0.0479	-
Iver Fossum*	Strømsgodset	2653	0.2045	0.0744	0.0371	0.0174	-0.0003	0.0226	-0.0030
Gjermund Åsen	Tromsø	1709	0.2026	-0.0156	0.0095	0.0334	0.0814	0.0355	0.0277
Henrik Furebotn	Bodø/Glimt	2157	0.1731	0.0769	-0.0008	0.0395	0.0030	0.0047	0.0241
Thomas Kind Bendiksen*	Molde	931	0.1407	-0.0503	0.0233	0.0156	0.0900	0.0084	0.0332
Top 10 average			0.2118	0.0509	0.0369	0.0150	0.0216	0.0165	0.0150

* sold to foreign club during or some time after the 2015 season ** on loan from a foreign club in 2015

Table C.5: Model 1: Top 10 players in defensive positions, Tippeligaen 2015.
Average value = 0.1124, minimum value = -0.1248

CENTRAL MIDFIELDERS									
Player	Team	Minutes	Total	Shot	Pass	Tackle	Ball rec	Cross	Corner
Christian Grindheim	Vålerenga	2675	0.3087	0.0985	0.1239	0.0069	0.0105	0.0040	0.0125
Malaury Martin	Lillestrøm	975	0.2947	0.1446	0.0453	0.0045	0.0088	0.0103	0.0336
Giorgi Gorozia	Stabæk	1467	0.2326	-0.0578	0.1369	0.0065	0.0241	0.0063	0.0858
Kristoffer Ajer*	Start	2581	0.2187	0.1016	0.0243	0.0077	0.0122	0.0066	-
Bismark Adjei-Boateng**	Strømsgodset	1293	0.1847	0.0590	0.0192	0.0214	0.0026	0.0440	-0.0002
Morten Konradsen	Bodø/Glimt	1362	0.1740	0.1439	-0.0058	0.0044	0.0071	0.0202	-0.0009
Kamal Issah	Stabæk	1846	0.1733	0.0274	0.0838	0.0143	0.0163	0.0082	-
Fredrik Midtsjø	Rosenborg	2396	0.1718	0.0227	0.0833	0.0152	0.0118	0.0010	-
Johan Andersson	Lillestrøm	973	0.1664	0.1054	0.0429	0.0046	0.0117	-0.0056	-0.0014
Ole Kristian Selnæs*	Rosenborg	1951	0.1553	0.0157	0.0782	0.0133	0.0217	0.0226	-0.0025
Top 10 average			0.2080	0.0661	0.0632	0.0099	0.0127	0.0118	0.0127

FULL BACKS									
Player	Team	Minutes	Total	Shot	Pass	Cross	Ball rec	Corner	FK
Per-Egil Flo	Molde	1998	0.3290	0.0396	0.1003	0.0327	0.0092	0.0700	0.0087
Espen Ruud	Odd	2476	0.2953	0.0457	0.0925	0.0831	0.0070	0.0087	0.0204
Lars-Christopher Vilsvik	Strømsgodset	2126	0.2692	0.0542	0.0197	0.1462	0.0134	0.0153	-0.0021
Jo Nymo Matland*	Aalesund	1425	0.2429	0.0605	-0.0015	0.0426	0.0049	0.0580	0.0501
André Danielsen	Viking	2700	0.2329	0.0995	0.0428	0.0459	0.0155	0.0179	0.0050
Kent-Are Antonsen	Tromsø	2248	0.2275	0.0557	0.0029	0.0277	0.0160	0.0102	0.0261
Joachim Olsen Solberg	Mjøndalen	2486	0.2087	-0.0588	0.0230	0.0192	0.0143	0.0508	0.1031
Zarek Chase Valentin*	Bodø/Glimt	2074	0.2006	0.0056	0.0112	0.0498	0.0114	-0.0025	0.0014
Jørgen Skjelvik	Rosenborg	1812	0.1855	0.0154	0.0801	0.0449	0.0120	0.0070	0.0006
Mikael Dorsin	Rosenborg	1582	0.1492	0.0365	0.0294	0.0210	0.0106	-0.0011	-0.0011
Top 10 average			0.2341	0.0354	0.0400	0.0513	0.0114	0.0234	0.0212

CENTRE BACKS									
Player	Team	Minutes	Total	Shot	Pass	Clearance	Ball rec	Cross	Long pass
Johan Bjørdal	Rosenborg	1093	0.2398	0.0461	0.1655	-0.0001	0.0107	-0.0026	0.0012
Rhett Bernstein*	Mjøndalen	1234	0.1927	0.1214	0.0301	-0.0068	0.0077	0.0002	-0.0038
Joona Toivio	Molde	1706	0.1781	-0.0190	0.0665	0.0274	0.0093	0.0650	0.0125
Andreas Nordvik*	Sarpsborg 08	1846	0.1667	0.0532	0.0586	0.0168	0.0065	0.0036	0.0403
Brede Moe	Bodø/Glimt	2408	0.1657	0.0544	0.0545	0.0304	0.0072	-0.0016	-0.0082
Lars-Kristian Eriksen	Odd	2340	0.1456	0.0235	0.0746	0.0132	0.0084	0.0003	0.0110
Ole Christoffer Heieren Hansen	Sarpsborg 08	2059	0.1383	0.0634	0.0578	0.0337	0.0051	0.0000	-0.0025
Vegard Forren	Molde	2416	0.1196	-0.0375	0.0902	0.0199	0.0069	0.0052	0.0256
Morten Sundli	Mjøndalen	2018	0.1050	0.0314	0.0381	-0.0137	0.0092	0.0029	-0.0034
Steffen Hagen	Odd	2610	0.1000	0.0174	0.0813	-0.0033	0.0089	-0.0002	0.0079
Top 10 average			0.1551	0.0354	0.0717	0.0117	0.0080	0.0073	0.0081

* sold to foreign club during or some time after the 2015 season ** on loan from a foreign club in 2015

Table C.6: Model 1: League table 2015 based on the model compared to the real table

Team	Value/match	Table	Real table	Diff
Rosenborg	1.3148	1	1	0
Molde	0.8748	2	6	-4
Strømsgodset	0.7249	3	2	1
Stabæk	0.4698	4	3	1
Odd	0.4641	5	4	1
Viking	0.4397	6	5	1
Vålerenga	0.3452	7	7	0
Lillestrøm	0.1848	8	8	0
Bodø/Glimt	-0.0747	9	9	0
Tromsø	-0.3564	10	13	-3
Sarpsborg 08	-0.4349	11	11	0
Aalesund	-0.4681	12	10	2
Haugesund	-0.7172	13	12	1
Start	-0.8898	14	14	0
Mjøndalen	-0.9365	15	15	0
Sandefjord	-0.9400	16	16	0

Appendix D

Results from Markov Game Model 2

Appendix D presents additional results from running Model 2. Top 10 lists for the different playing positions for Tippeligaen 2014 are presented in Tables D.1 and D.2. In Tables D.4 and D.5 are the results for 2015 repeated for comparison. Tables D.3 and D.6 show the league tables that can be obtained for 2014 and 2015, respectively.

D.1 2014

Table D.1: Model 2: Top 10 players in offensive positions, Tippeligaen 2014. Average value = 0.4668, minimum value = 0.0118

FORWARDS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Foul won	Cross	Ball carry
Aaron Samuel*	Sarpsborg 08	1032	0.9668	0.4466	0.1850	0.0576	0.0528	0.0098	0.1126
Jón Dadi Bödvarsson*	Viking	2129	0.9221	0.4332	0.1714	0.0322	0.0165	0.0280	0.1341
Daniel Chima Chukwu*	Molde	1998	0.9213	0.4031	0.2486	0.0489	0.0204	0.0102	0.1069
Leke James*	Aalesund	2031	0.9086	0.4005	0.2146	0.0436	0.0282	0.0058	0.1326
Vidar Örn Kjartansson	Vålerenga	2608	0.8462	0.5047	0.1383	0.0152	0.0404	0.0070	0.0915
Fredrik Gulbrandsen	Molde	1475	0.8138	0.3853	0.1360	0.0572	0.0221	0.0187	0.1460
Diego Ivan Rubio**	Sandnes Ulf	1794	0.7807	0.3213	0.1904	0.0080	0.0314	0.0460	0.1077
Matthías Vilhjálmsson	Start	1681	0.7746	0.3531	0.2165	0.0056	0.0191	0.0163	0.0725
Erik André Huseklepp	Brann	1989	0.7592	0.1992	0.1729	0.0351	0.0158	0.1086	0.0994
Franck Boli	Stabæk	2249	0.7547	0.3451	0.1687	0.0680	0.0225	0.0027	0.1029
Top 10 average			0.8448	0.3792	0.1842	0.0371	0.0269	0.0253	0.1106
WINGERS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Cross	Corner	Ball carry
Trond Olsen	Bodø/Glimt	1443	1.1593	0.3857	0.1951	0.0563	0.1644	0.0852	0.1448
Mohamed Elyounoussi	Molde	2533	1.0804	0.4095	0.3019	0.0435	0.0320	0.0252	0.1798
Pål André Helland	Rosenborg	1040	1.0453	0.3037	0.2246	0.0758	0.1571	0.0307	0.1401
Elbasan Rashani*	Odd	1229	0.9340	0.3730	0.1500	0.0597	0.1047	0.0563	0.1098
Zlatko Tripić*	Start	2107	0.9238	0.2287	0.1725	0.0258	0.0777	0.0980	0.1182
Moryké Fofana*	Lillestrøm	2051	0.9187	0.2673	0.2737	0.0826	0.0335	0.0014	0.1785
Mattias Moström	Molde	2152	0.8384	0.1138	0.3682	0.0349	0.0350	0.0578	0.1489
Øyvind Storflor	Strømsgodset	2029	0.8279	0.1402	0.3235	0.0107	0.0919	0.0668	0.0854
Gustav Wikheim*	Strømsgodset	1616	0.8206	0.1549	0.3301	0.0636	0.0844	0.0005	0.1218
Morten Gamst Pedersen	Rosenborg	1724	0.7943	0.1197	0.2108	0.0147	0.0656	0.1647	0.0388
Top 10 average			0.9343	0.2496	0.2551	0.0468	0.0846	0.0587	0.1266
ATTACKING MIDFIELDERS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Cross	Corner	Ball carry
Martin Ødegaard*	Strømsgodset	1454	1.1504	0.1496	0.5469	0.0594	0.0117	0.0552	0.2207
Michael Barrantes*	Aalesund	1608	0.9481	0.3274	0.2714	0.0199	0.0509	0.0366	0.1092
Ghayas Zahid	Vålerenga	2216	0.8955	0.2570	0.3054	0.0347	0.0521	0.0111	0.1518
Herolind Shala*	Odd	2093	0.8760	0.3077	0.2135	0.0172	0.1131	0.0298	0.0967
Papa Alioune Ndiaye*	Bodø/Glimt	2555	0.7923	0.2930	0.1647	0.0720	0.0377	0.0496	0.0817
Bojan Zajić	Sarpsborg 08	1826	0.7890	0.2704	0.2190	0.0265	0.0261	0.0572	0.1052
Petter Vaagan Moen	Lillestrøm	1915	0.7322	0.1986	0.2174	0.0062	0.0451	0.1187	0.0534
Maic Sema*	Haugesund	1803	0.7183	0.2742	0.2175	0.0195	0.0412	0.0012	0.1120
Fredrik Nordkvelle	Odd	1110	0.6898	0.2094	0.2382	0.0312	0.0546	0.0012	0.0984
Vidar Nisja	Viking	1922	0.6572	0.2714	0.1900	0.0178	0.0222	0.0004	0.0914
Top 10 average			0.8249	0.2559	0.2584	0.0304	0.0455	0.0361	0.1121

* sold to foreign club during or some time after the 2014 season ** on loan from a foreign club in 2014

Table D.2: Model 2: Top 10 players in defensive positions, Tippeligaen 2014.
Average value = 0.4668, minimum value = 0.0118

CENTRAL MIDFIELDERS

Player	Team	Minutes	Total	Shot	Pass	Cross	Long	Corner	Ball carry
Harmeet Singh	Molde	2424	0.8135	0.1562	0.3984	0.0183	0.0699	-	0.1238
Jone Samuelsen	Odd	2498	0.7631	0.2123	0.2439	0.0718	0.0182	0.0009	0.1285
Etzaz Hussain*	Molde	1279	0.7284	0.1048	0.3832	0.0073	0.0582	0.0286	0.0939
Mike Jensen	Rosenborg	2590	0.6937	0.2325	0.2375	0.0340	0.0280	0.0088	0.0737
Fredrik Midtsjø	Rosenborg	1754	0.6386	0.1738	0.1936	0.0341	0.0146	0.0162	0.0932
Daniel Berg Hestad	Molde	1255	0.6357	0.0838	0.3833	0.0244	0.0266	0.0013	0.0990
Gudmundur Thórarinsson*	Sarpsborg 08	2392	0.6334	0.1153	0.2450	0.0265	0.0348	0.0849	0.0837
Makhtar Thioune*	Viking	1115	0.6315	0.0498	0.2933	0.0190	0.0568	0.0661	0.0747
Michael Francis Stephens*	Stabæk	2551	0.6275	0.0895	0.3010	0.0377	0.0475	0.0136	0.0968
Mix Diskerud*	Rosenborg	1762	0.6187	0.1420	0.2917	0.0086	0.0271	0.0077	0.0631
Top 10 average			0.6784	0.1360	0.2971	0.0282	0.0382	0.0228	0.0930

FULL BACKS

Player	Team	Minutes	Total	Shot	Pass	Cross	Corner	Throw	Ball carry
Lars-Christopher Vilsvik	Strømsgodset	2152	0.6473	0.1104	0.2500	0.0835	0.0557	0.0356	0.0654
Per-Egil Flo	Molde	1959	0.6468	0.0583	0.2645	0.0753	0.1014	0.0385	0.0714
Martin Linnes*	Molde	2520	0.6171	0.1086	0.2250	0.1004	-	0.0361	0.0734
André Danielsen	Viking	2474	0.4822	0.0479	0.1403	0.0899	0.0524	0.0276	0.0643
Joakim Våge Nilsen	Haugesund	2625	0.4590	0.0501	0.2143	0.0402	0.0016	0.0235	0.0553
Thomas Grøgaard	Odd	2700	0.4452	0.0131	0.2188	0.0849	0.0228	0.0449	0.0430
Andreas Vindheim*	Brann	1874	0.4445	0.0844	0.1032	0.1441	0.0069	0.0322	0.0667
Birkir Már Sævarsson	Brann	1231	0.4408	0.0991	0.1564	0.0714	-	0.0297	0.0782
Kristoffer Haugen	Viking	1717	0.4406	0.0237	0.1009	0.1267	0.0828	0.0232	0.0364
Jo Nymo Matland*	Aalesund	1067	0.4390	0.0710	0.1101	0.0892	0.0528	0.0295	0.0349
Top 10 average			0.5062	0.0667	0.1784	0.0905	0.0376	0.0321	0.0589

CENTRE BACKS

Player	Team	Minutes	Total	Shot	Pass	Aerial	Cross	Long	Ball carry
Ruben Gabrielsen	Molde	1954	0.3575	0.0989	0.1387	0.0409	0.0299	0.0264	0.0273
Vegard Forren	Molde	2467	0.2867	0.0706	0.1422	0.0142	0.0058	0.0496	0.0456
Jarl André Storbæk	Strømsgodset	1783	0.2495	0.0395	0.1260	-0.0002	0.0339	0.0235	0.0439
Lars-Kristian Eriksen	Odd	2230	0.2476	0.0411	0.1049	-0.001	0.0353	0.0233	0.0349
Azar Karadaş	Brann	1640	0.2437	0.0873	0.1134	0.0210	0.0045	0.0174	0.0224
Tor Arne Andreassen	Haugesund	2462	0.2384	0.0911	0.0975	-0.0011	0.0215	0.0189	0.0302
Stefan Strandberg*	Rosenborg	1643	0.2343	0.0957	0.1088	0.0122	0.0036	0.0406	0.0291
Marius Christopher Høibråten	Strømsgodset	1372	0.2279	0.0631	0.1386	-0.0026	0.0019	0.0244	0.0328
Daniel Mojsov*	Brann	1393	0.2092	0.1065	0.0882	0.006	0.0064	0.0295	0.0220
Tore Reginiussen	Rosenborg	2065	0.2024	0.1201	0.0904	-0.0044	0.0009	0.0164	0.0286
Top 10 average			0.2491	0.0814	0.1149	0.0085	0.0144	0.0270	0.0317

* sold to foreign club during or some time after the 2014 season ** on loan from a foreign club in 2014

Table D.3: Model 2: League table 2014 based on the model compared to the real table

Team	Value/match	Table	Real Table	Diff
Molde	3.6921	1	1	0
Strømsgodset	2.0150	2	4	-2
Odd	1.7909	3	3	0
Rosenborg	1.4025	4	2	2
Viking	0.5882	5	10	-5
Vålerenga	0.5035	6	6	0
Lillestrøm	0.3532	7	5	2
Brann	-0.3729	8	14	-6
Haugesund	-0.5862	9	11	-2
Sogndal	-0.7550	10	15	-5
Aalesund	-0.8503	11	7	4
Sarpsborg 08	-0.9429	12	8	4
Start	-1.3133	13	12	1
Stabæk	-1.3528	14	9	5
Bodø/Glimt	-1.7765	15	13	2
Sandnes Ulf	-2.3953	16	16	0

D.2 2015

Table D.4: Model 2: Top 10 players in offensive positions, Tippeligaen 2015. Average value = 0.4608, minimum value = 0.0059

FORWARDS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Foul won	Aerial	Ball carry
Adama Diomandé*	Stabæk	1850	1.0543	0.5001	0.2130	0.0656	0.0333	0.0312	0.1160
Olivier Occéan	Odd	2208	0.9867	0.5049	0.2432	0.0128	0.0407	0.0888	0.0616
Fred Friday	Lillestrøm	1770	0.9133	0.4026	0.1939	0.0740	0.0303	0.0115	0.1419
Ola Kamara*	Molde	2384	0.9085	0.4298	0.2208	0.0143	0.0246	0.0210	0.1069
Veton Berisha*	Viking	1256	0.8831	0.3770	0.2317	0.0165	0.0232	0.0258	0.1247
Sander Svendsen	Molde	1720	0.8309	0.3897	0.1975	0.0471	0.0101	0.0047	0.1174
Jón Dadi Bödvarsson*	Viking	2067	0.8080	0.3123	0.1932	0.0326	0.0273	0.0365	0.1328
Leke James*	Aalesund	2610	0.8047	0.3676	0.1920	0.0487	0.0212	0.0609	0.0723
Alexander Søderlund*	Rosenborg	2242	0.7993	0.4564	0.1677	0.0131	0.0279	0.0491	0.0550
Alexander Sørloth*	Bodø/Glimt	1776	0.7916	0.3829	0.1769	0.0233	0.0265	0.0719	0.0762
Top 10 average			0.8781	0.4123	0.2030	0.0348	0.0265	0.0401	0.1005
WINGERS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Corner won	Cross	Ball carry
Pål André Helland	Rosenborg	1502	1.4047	0.4942	0.2508	0.1021	0.0379	0.0729	0.1821
Gustav Wikheim*	Strømsgodset	2413	1.1020	0.3125	0.3692	0.1477	0.0246	0.0381	0.1562
Yassine El Ghanassy*	Stabæk	1879	0.9840	0.2875	0.2355	0.1028	0.0242	0.0531	0.1261
Mohamed Elyounoussi	Molde	2275	0.9445	0.3526	0.2887	0.0509	0.0213	0.0255	0.1259
Tobias Mikkelsen*	Rosenborg	1931	0.9015	0.3444	0.2291	0.0446	0.0197	0.0630	0.1587
Moryké Fofana*	Lillestrøm	1300	0.8512	0.2796	0.2384	0.1325	0.0241	0.0176	0.0852
Ernest Asante	Stabæk	2643	0.8326	0.2662	0.2453	0.0436	0.0195	0.0683	0.1435
Samuel Adegbenro	Viking	1778	0.8245	0.3356	0.1451	0.0648	0.0212	0.0554	0.1234
Bentley	Odd	2464	0.8027	0.2561	0.1901	0.0436	0.0294	0.1292	0.0829
Ole Jørgen Halvorsen	Odd	1393	0.7688	0.2770	0.1771	0.0272	0.0271	0.1195	0.0823
Top 10 average			0.9416	0.3206	0.2369	0.0760	0.0249	0.0643	0.1266
ATTACKING MIDFIELDERS									
Player	Team	Minutes	Total	Shot	Pass	Take on	Long pass	Corner	Ball carry
Iver Fossum*	Strømsgodset	2653	0.8587	0.2702	0.3297	0.0242	0.0253	0.0004	0.1227
Papa Alioune Ndiaye*	Bodø/Glimt	1286	0.7967	0.2872	0.1726	0.0670	0.0585	0.0119	0.0929
Ghayas Zahid	Vålerenga	2383	0.7618	0.2373	0.2650	0.0449	0.0048	0.0036	0.1352
Michael Barrantes*	Aalesund	1000	0.7167	0.2097	0.1888	0.0051	0.0689	0.0703	0.0817
Fredrik Nordkvelle	Odd	1891	0.6806	0.2128	0.2292	0.0198	0.0071	0.0009	0.1100
Gjermund Åsen	Tromsø	1709	0.6745	0.1602	0.1151	0.0125	0.0021	0.1190	0.0678
Bojan Zajić	Sarpsborg 08	1760	0.6574	0.2360	0.2166	0.0304	0.0140	0.0221	0.0550
Aron Elís Thrándarson	Aalesund	1152	0.6464	0.2980	0.1567	0.0286	0.0163	0.0095	0.0726
Daniel Fredheim Holm	Vålerenga	1974	0.6449	0.1309	0.2683	0.0439	0.0124	0.0110	0.0978
Eirik Hestad	Molde	941	0.6258	0.0661	0.2771	0.0212	0.0201	0.0846	0.0704
Top 10 average			0.7064	0.2108	0.2219	0.0298	0.0229	0.0333	0.0906

* sold to foreign club during or some time after the 2015 season   ** on loan from a foreign club in 2015

Table D.5: Model 2: Top 10 players in defensive positions, Tippeligaen 2015. Average value = 0.4608, minimum value = 0.0059

CENTRAL MIDFIELDERS

Player	Team	Minutes	Total	Shot	Pass	Cross	Long pass	Corner	Ball carry
Mike Jensen	Rosenborg	2466	1.0463	0.2562	0.3341	0.0737	0.0285	0.0929	0.1235
Etzaz Hussain*	Molde	1956	0.7513	0.1343	0.3162	0.0099	0.0458	0.0275	0.1208
Harmeet Singh	Molde	2212	0.7429	0.0985	0.3671	0.0203	0.0847	0.0128	0.0946
Giorgi Gorozia	Stabæk	1467	0.6988	0.1110	0.2833	0.0304	0.0205	0.0891	0.0612
Ole Kristian Selnæs*	Rosenborg	1951	0.6865	0.0757	0.3080	0.0319	0.0675	0.0511	0.0875
Fredrik Midtsjø	Rosenborg	2396	0.6609	0.1205	0.3048	0.0143	0.0246	-	0.0885
Jone Samuelsen	Odd	2292	0.6473	0.1435	0.2440	0.0476	0.0114	0.0024	0.0960
Malaury Martin	Lillestrøm	975	0.6427	0.1252	0.2152	0.0290	0.0442	0.1375	0.0372
Christian Grindheim	Vålerenga	2675	0.6139	0.0649	0.2832	0.0138	0.0285	0.0600	0.0567
Bismark Adjei-Boateng**	Strømsgodset	1293	0.5535	0.1999	0.2271	0.0348	0.0149	0.0007	0.0433
Top 10 average			0.7044	0.1330	0.2883	0.0306	0.0371	0.0474	0.0809

FULL BACKS

Player	Team	Minutes	Total	Shot	Pass	Cross	Corner	Throw	Ball carry
Per-Egil Flo	Molde	1998	0.7505	0.0849	0.2954	0.1053	0.1187	0.0333	0.0470
Jonas Svensson	Rosenborg	2610	0.7108	0.1126	0.3389	0.0730	-	0.0412	0.0709
Espen Ruud	Odd	2476	0.6584	0.0881	0.1954	0.1233	0.0518	0.0496	0.0471
Lars-Christopher Vilsvik	Strømsgodset	2126	0.6179	0.1047	0.2076	0.1542	0.0237	0.0371	0.0593
Martin Linnes*	Molde	2439	0.5079	0.0545	0.1931	0.0908	0.0071	0.0459	0.0430
Joachim Olsen Solberg	Mjøndalen	2486	0.4978	0.0895	0.0798	0.0517	0.1017	0.0222	0.0208
André Danielsen	Viking	2700	0.4634	0.0612	0.1425	0.0892	0.0351	0.0383	0.0523
Birger Meling	Stabæk	2263	0.4592	0.1000	0.1665	0.0279	0.0484	0.0301	0.0585
Jørgen Skjelvik	Rosenborg	1812	0.4468	0.0762	0.1873	0.0587	0.0030	0.0256	0.0620
Akeem Latifu*	Aalesund	2564	0.4454	0.0657	0.1206	0.1373	-	0.0363	0.0558
Top 10 average			0.5558	0.0837	0.1927	0.0911	0.0389	0.0360	0.0517

CENTRE BACKS

Player	Team	Minutes	Total	Shot	Pass	Aerial	Cross	Long pass	Ball carry
Johan Bjørdal	Rosenborg	1093	0.4473	0.1057	0.2274	0.0002	0.0018	0.0490	0.0803
Stefan Strandberg*	Rosenborg	1151	0.4149	0.1064	0.1797	0.0215	0.0057	0.0613	0.0883
Hólmar Örn Eyjólfsson	Rosenborg	2357	0.3758	0.0974	0.1835	0.0124	0.0005	0.0260	0.0742
Rhett Bernstein*	Mjøndalen	1234	0.3516	0.2193	0.1001	0.0650	0.0034	0.0046	0.0067
Joona Toivio	Molde	1706	0.3293	0.1184	0.1363	0.0107	0.0193	0.0259	0.0290
Jørgen Horn*	Strømsgodset	1260	0.3004	0.0737	0.1623	-0.0008	0.0125	0.0335	0.0423
Oddbjørn Lie	Aalesund	1549	0.2715	0.0495	0.1161	-0.0024	0.0381	0.0198	0.0311
Morten Sundli	Mjøndalen	2018	0.2690	0.1332	0.0907	0.0136	0.0106	0.0180	0.0205
Vegard Forren	Molde	2416	0.2614	0.0402	0.1300	0.0067	0.0123	0.0700	0.0353
Andreas Nordvik*	Sarpsborg 08	1846	0.2366	0.0830	0.0950	0.0037	0.0072	0.0630	0.0229
Top 10 average			0.3258	0.1027	0.1421	0.0131	0.0111	0.0317	0.0430

* sold to foreign club during or some time after the 2015 season ** on loan from a foreign club in 2015

Table D.6: Model 2: League table 2015 based on the model compared to the real table

Team	Value/match	Table	Real table	Diff
Rosenborg	3.9116	1	1	0
Molde	2.8190	2	6	-4
Strømsgodset	1.7488	3	2	1
Odd	1.5592	4	4	0
Stabæk	1.2503	5	3	2
Viking	0.3503	6	5	1
Vålerenga	0.1922	7	7	0
Lillestrøm	-0.9348	8	8	0
Sarpsborg 08	-0.9524	9	11	-2
Aalesund	-0.9742	10	10	0
Tromsø	-1.0293	11	13	-2
Mjøndalen	-1.1722	12	15	-3
Haugesund	-1.2649	13	12	1
Bodø/Glimt	-1.2931	14	9	5
Sandefjord	-2.0329	15	16	-1
Start	-2.1776	16	14	2

Appendix E

Markov Game Models: Validation from Team Evaluations

As mentioned in Sections 4.3.4 and 4.4.5 the values of the individual involvements can be aggregated for each team each match. From this, the difference between the two opponents in a match, in order to evaluate which team that performed the best. These values can again be aggregated for each team over a season, from which a league table can be generated. This can also be done after each matchday, which leads to a kind of time series for each team. The results from such an experiment is shown in Figure E.1, where the table position of each team is compared each match day across the two different models and with the real league table.

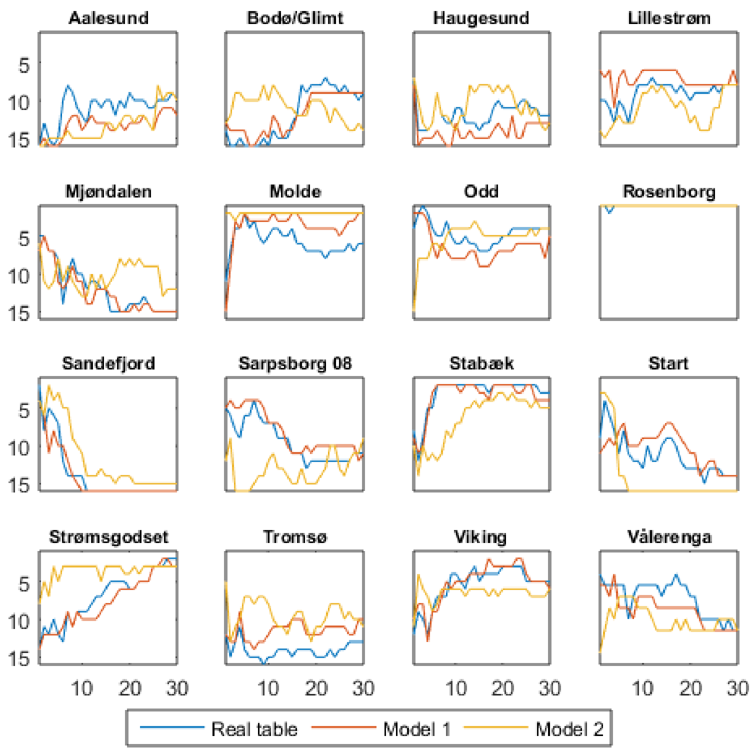


Figure E.1: Model 1 and Model 2: Real table position compared to estimated each matchday

The results, as can be seen from the figure, is somewhat differing amongst the teams. In order to evaluate how accurate each of the models assigns positions to each team, an Augmented Dickey Fuller (ADF) test is conducted to test the cointegration between the time series from each model with the real league positions. In the ADF two lags are included in order to remove any autocorrelation in the residuals. With two or more lags the test conclusions are the same, that is if the series are cointegrated or not. In addition, both the Kendall rank correlation coefficient (Kendall's tau) and Spearman's rank correlation coefficent (Spearman's rho) are calculated for each team and each model. The results from the ADF test with two lags and the correlation coefficients are shown in Table E.1.

Table E.1: Model 1 and Model 2: p-values from the ADF test, Kendall's tau and Spearman's rho in 2015

	Model 1			Model 2		
Team	p-value	Kendall's tau	Spearman's rho	p-value	Kendall's tau	Spearman's rho
Aalesund	0.00	0.5872	0.6788	0.00	0.4815	0.5736
Bodø/Glimt	0.02	0.7873	0.8904	0.09	-0.3181	-0.4569
Haugesund	0.01	0.5155	0.6212	0.01	0.2012	0.2346
Lillestrøm	0.12	0.2633	0.3089	0.10	0.6907	0.8283
Mjøndalen	0.01	0.8093	0.8834	0.06	-0.0270	-0.0322
Molde	0.16	0.5842	0.6962	0.03	-0.1974	-0.2206
Odd	0.06	0.7093	0.7989	0.02	-0.3636	-0.4390
Rosenborg	0.00	N/A	N/A	0.00	N/A	N/A
Sandefjord	0.00	0.8903	0.9457	0.01	0.7849	0.8614
Sarpsborg 08	0.02	0.6002	0.7494	0.25	-0.0848	-0.1149
Stabæk	0.07	0.7717	0.8171	0.00	0.4255	0.4982
Start	0.00	0.6212	0.7446	0.16	0.5587	0.6455
Strømsgodset	0.08	0.8597	0.9554	0.45	0.3536	0.4081
Tromsø	0.02	0.3029	0.3524	0.06	-0.1284	-0.1506
Viking	0.00	0.7632	0.8657	0.01	0.3542	0.3884
Vålerenga	0.14	0.7354	0.8190	0.39	0.0635	0.0912

From the table it can be seen that the test statistic for the ADF test is significant at 5% for 10 teams with Model 1, and for 8 teams with Model 2. Furthermore, the two correlation coefficients are positive for all teams in the first model. This indicates conformity for all the teams regarding the real trend and the estimated development in table position. The coefficients for Rosenborg are not applicable due to the almost absence of variations in position throughout the season. These quite good results for Model 1 were expected, due to the heavy impact from the scored goals.

For the second model, the two correlation coefficients are negative for six teams. This indicates that the model estimates a somewhat opposite trend in league position than reality. A good example of such an opposite trend can be seen through the yellow line for Bodø/Glimt in Figure E.1. These results are a bit surprising, but very interesting since Model 2 does not account for the outcome of each shot, and therefore is less impacted by the goals scored.

Appendix F

Case Studies

In Appendix F box plots of actions performed either by Mike Jensen or Martin Ødegaard are presented. For Mike Jensen, three plots from the 2014 season are presented, in addition to the plot of his Model 1 values from 2015 including the shots. For Martin Ødegaard, only the plot of his Model 1 values from 2014 is presented.

F.1 Mike Jensen

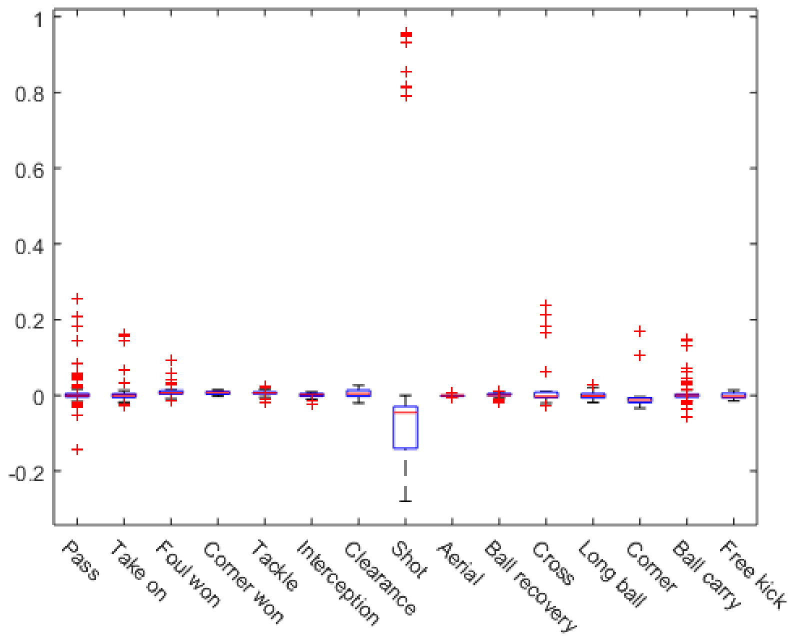


Figure F.1: Model 1: Box plot for Mike Jensen 2014 including shots

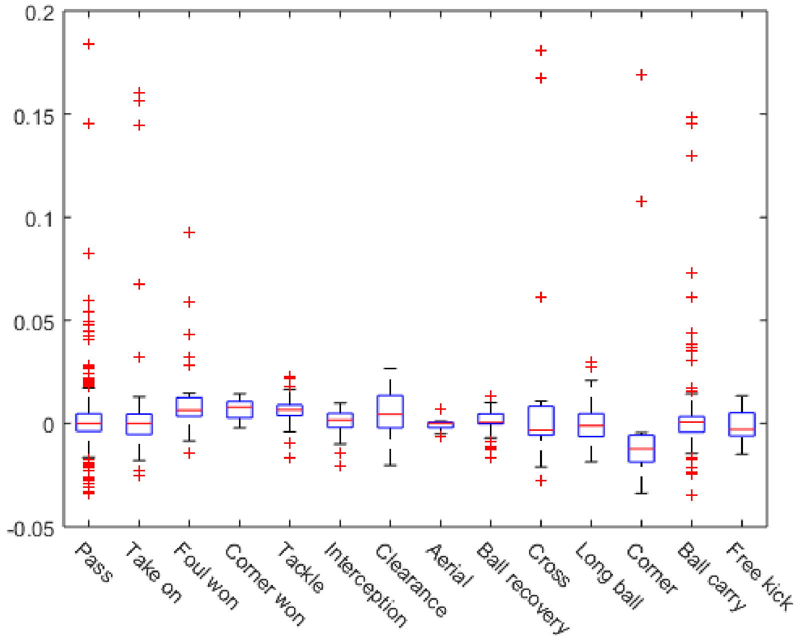


Figure F.2: Model 1: Box plot for Mike Jensen 2014 excluding shots

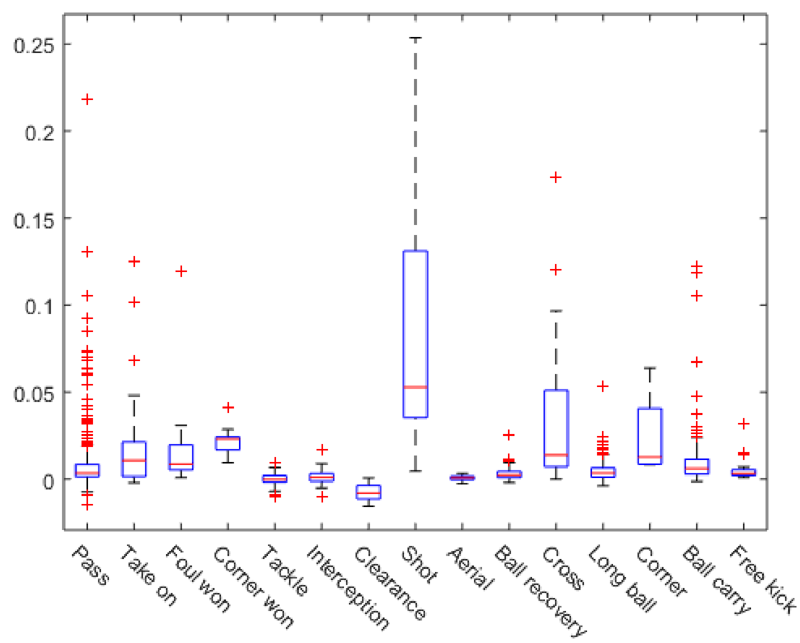


Figure F.3: Model 2: Box plot for Mike Jensen 2014

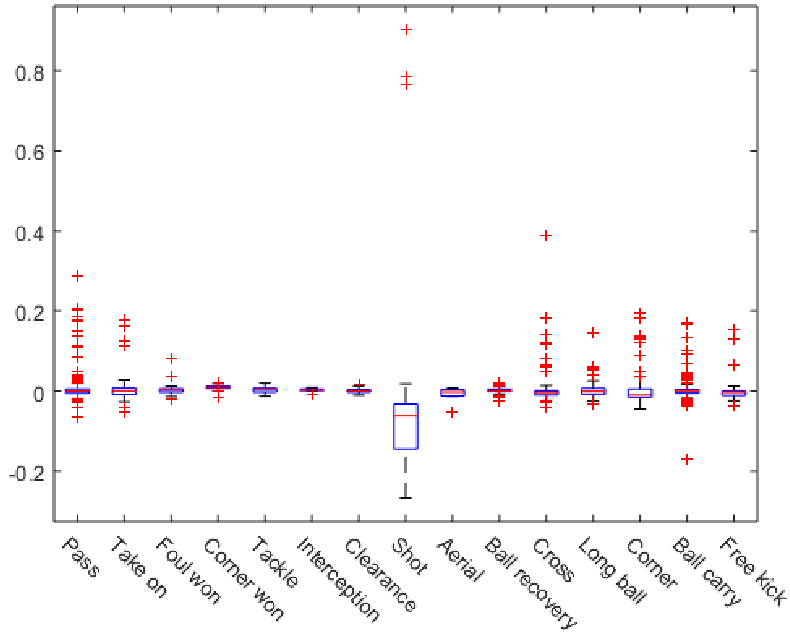


Figure F.4: Model 1: Box plot for Mike Jensen in 2015 including shots

F.2 Martin Ødegaard

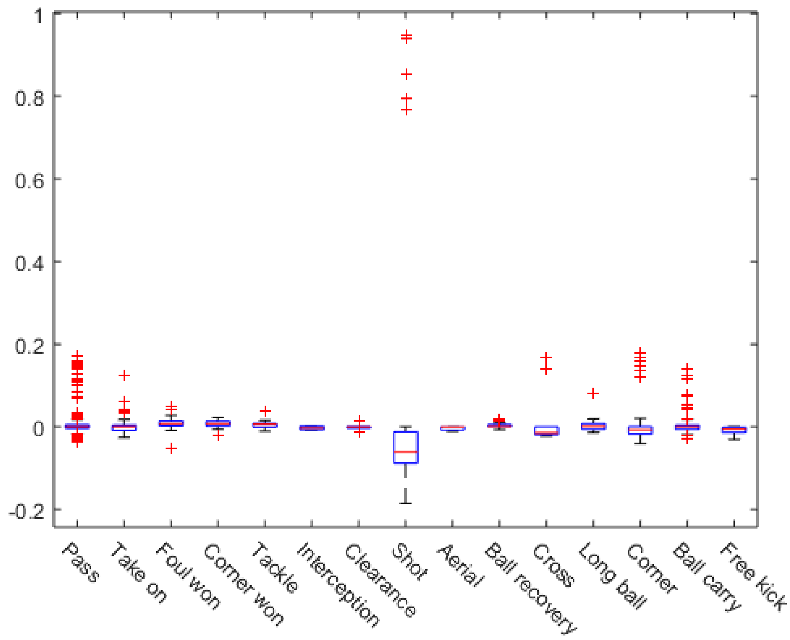


Figure F.5: Model 1: Box plot for Martin Ødegaard 2014 including shots

Appendix G

Results for 2016

Appendix G illustrates results from running the three models on the first 13 match days of the 2016 Tippeligaen season. The results are presented as top 20 lists, including players in all outfield positions. A minimum of 390 minutes on the field is demanded. In the xG Model, only players with three or more goals are considered.

G.1 Expected Goals Model

Table G.1: xG Model: Top 20 players 2016 with respect to G/xG

Name	Team	Position	Minutes Played	Goals	xG	Shots	xG/S	G/xG
Milan Jevtović	Bodø/Glimt	Winger	949	6	1.79	21	0.09	3.35
Luc Kassi	Stabæk	Forward	1050	3	0.90	22	0.04	3.33
Fredrik Semb Berge	Odd	Centre Back	932	3	1.02	4	0.25	2.95
Petter Strand	Molde	Winger	940	3	1.13	20	0.06	2.66
Mathias Bringaker	Viking	Forward	565	3	1.61	17	0.09	1.87
Øyvind Storflor	Strømsgodset	Winger	859	3	1.66	15	0.11	1.81
Fredrik Nordkvelle	Odd	Attacking Midfielder	1056	4	2.23	24	0.09	1.79
Bassel Jradi	Strømsgodset	Attacking Midfielder	1049	3	1.71	26	0.07	1.75
Fredrik Gulbrandsen	Molde	Forward	515	4	2.29	17	0.13	1.75
Bentley	Odd	Winger	994	3	1.77	20	0.09	1.69
Torbjørn Agdestein	Haugesund	Forward	944	6	3.55	24	0.15	1.69
Malaury Martin	Lillestrøm	Central Midfielder	1097	4	2.43	34	0.07	1.65
Fred Friday	Lillestrøm	Forward	1080	8	4.87	50	0.10	1.64
Steffen Ernemann	Sarpsborg 08	Central Midfielder	565	3	1.91	9	0.21	1.57
Marcus Pedersen	Strømsgodset	Forward	670	5	3.29	24	0.14	1.52
Thomas Lehne Olsen	Tromsø	Forward	712	3	2.01	16	0.13	1.49
Matthías Vilhjálmsson	Rosenborg	Forward	726	3	2.02	20	0.10	1.49
Tokmac Nguen	Strømsgodset	Attacking Midfielder	912	3	2.09	24	0.09	1.43
Kristoffer Tokstad	Sarpsborg 08	Winger	1013	3	2.32	21	0.11	1.29
Deshorn Dwayne Brown	Vålerenga	Forward	991	6	4.95	45	0.11	1.21

G.2 Markov Game Model 1

Table G.2: Model 1: Top 20 players in all positions 2016.
Average tota value = 0.0752, minimum total value = -0.4143

Name	Team	Position	Minutes Played	Total value
Milan Jevtović	Bodø/Glimt	Winger	949	0.4873
Fredrik Gulbrandsen	Molde	Forward	515	0.4612
Haris Hajradinović	Haugesund	Winger	670	0.4176
Steffen Ernemann	Sarspborg 08	Central midfielder	565	0.4174
Fredrik Midtsjø	Rosenborg	Central midfielder	1006	0.4167
Fredrik Semb Berge	Odd	Centre back	932	0.3566
Fred Friday	Lillestrøm	Forward	1080	0.3492
Torbjørn Agdestein	Haugesund	Forward	944	0.3266
Simen Kind Mikalsen	Lillestrøm	Full back	900	0.3172
Moussa Nije	Stabæk	Winger	547	0.3097
Mattias Moström	Molde	Winger	421	0.2862
Alexander Gersbach	Rosenborg	Full back	440	0.2807
Petter Strand	Molde	Winger	940	0.2685
Malaury Martin	Lillestrøm	Central midfielder	1097	0.2682
Øyvind Storflor	Strømsgodset	Winger	859	0.2659
Per-Egil Flo	Molde	Full back	990	0.2657
Filip Kiss	Haugesund	Central midfielder	1010	0.2616
Mathias Bringaker	Viking	Forward	565	0.2490
Mos	Aalesund	Forward	656	0.2462
Fredrik Nordkvelle	Odd	Attacking midfielder	1056	0.2408

G.3 Markov Game Model 2

Table G.3: Model 2: Top 20 players in all positions 2016.
Average total value = 0.4541, minimum total value = 0.0356

Name	Team	Position	Minutes Played	Total value
Pål André Helland	Rosenborg	Winger	505	1.1727
Ghayas Zahid	Vålerenga	Attacking midfielder	949	0.9958
Mike Jensen	Rosenborg	Central midfielder	1080	0.9790
Mohamed Elyounoussi	Molde	Winger	971	0.9497
Mahatma Otoo	Sogndal	Forward	624	0.9487
Gilbert Koomson	Sogndal	Winger	948	0.9293
Fred Friday	Lillestrøm	Forward	1080	0.8992
Samuel Adegbenro	Viking	Winger	1079	0.8878
Yann-Erik de Lanlay	Rosenborg	Winger	967	0.8802
Olivier Occéan	Odd	Forward	1080	0.8486
Patrick Mortensen	Sarpsborg 08	Forward	860	0.8295
Malaury Martin	Lillestrøm	Central midfielder	1097	0.8152
Øyvind Storflor	Strømsgodset	Winger	859	0.8076
Anders Konradsen	Rosenborg	Central midfielder	1168	0.7993
Ole Jørgen Halvorsen	Odd	Winger	418	0.7957
Espen Ruud	Odd	Full back	1080	0.7938
Fredrik Midtsjø	Rosenborg	Central midfielder	1006	0.7918
Sofien Moussa	Tromsø	Forward	617	0.7888
Christian Gytkjær	Rosenborg	Forward	840	0.7834
Daniel Fredheim Holm	Vålerenga	Attacking midfielder	738	0.7508