UiT
THE ARCTIC UNIVERSITY OF NORWAY

HHT

# Econometrics versus the Bookmakers

*An econometric approach to sports betting*

—

**Sindre Hansen**

*Master thesis in Economics*
*June 2016*

# Forewords

When I asked one of the economy professors at HHT, how I could create the R algorithm used to carry out the forecasts we use in this thesis, he gave me a response that ended with something resembling the following:

"It is just a few lines of code"

While technically, his argument was perfectly sound, I would later find out just how far from the truth it really was.

I did not have experience with the R coding language prior to writing this thesis, nor any other coding language, so even the most basic of operations, such as importing the relevant dataset, presented a challenge initially. The only econometrics related software I had previously used was Stata, which is very easy to grasp compared to R.

My deadline of submitting this thesis was originally 14 December 2015, but due to challenges involving both learning how to use R and later coding the script used for the predictions, I had the deadline postponed to 1 June 2016.

I wish to send thanks to all of the people contributing to R-bloggers, Reddit, Inside-R, Datacamp and YouTube with excellent content for beginners trying to learn R. I am now able to use the software in a productive manner, and with this, I am finally able to conclude six years of studying Economics at the University of Tromsø.

I would also like to thank my supervisor, Sverre Braathen Thyholdt, for his help in coming up with the idea behind this thesis, for providing valuable feedback along the way, and for pointing me in the right direction concerning how I could solve the task.

# Abstract

Econometrics lets us apply our economic knowledge and test it on data samples from the real world. This thesis examines the possibility of using an extended Bradley-Terry model with covariates to predict the outcome of football matches, and use the results as a guideline for placing profitable sports bets. We benchmark performance empirically; using real life data from the 2013/2014 Premier League season, and achieve positive results.

Keywords: Econometrics, Bradley-Terry, profitable sports betting, probability forecast, football, 2013/14 Premier League season

# Disclaimer

The views expressed in this thesis are strictly for research purposes. The author does not advice anyone to engage in gambling activities based on any of the findings in this paper. Gambling addiction is serious; seek help if you suspect addiction. The author holds no responsibility for any losses incurred using any strategies, models or other information from this thesis.

# Table of contents

# 1 Introduction

In the 2010/11 season of the Barclays Premier League, 4.7 billion people watched a game on television at some point. 212 territories around the world saw the Premier League broadcasted. [1]

Sports betting is popular across the globe, and people from a wide range of origins partake in gambling. The reason why they choose to gamble varies from recreation to occupation. As with most markets where there are buyers, there is also sellers. In the sports betting market, the bookmakers are the sellers. How people select which games and results to bet on vary greatly, and there is even a whole range of so-called experts who willingly offer their tips for where people should place their bets.

Using econometrics to create a forecast we can compare with the forecasts of the bookmakers, could in turn automate the whole process of picking bets, and even be profitable if the model and data used are accurate enough.

We can derive the bookmakers' likelihood estimates directly from their valuation of each outcome, also known as the odds they offer on any outcome.

For my thesis, I wish to examine the possibility of creating such a forecast, using an extended version of the Bradley-Terry model (Davidson, 1970), adding covariates, and then use the formula for converting betting odds into probabilities, given by:

$$1.1\ Probability = \frac{1}{Odds}$$

Rewriting this equation yields:

$$1.2\ Betting\ odds = \frac{1}{Probability} = \frac{1}{P}$$

The plan onwards from there will be to make a comparison of our simulated probabilities with the probabilities estimated by the bookmakers, which is easily obtainable by using formula 1.1. In instances with a wide enough positive gap, meaning that our estimate for an outcome is sufficiently higher than that of the bookmakers, I will place fictive bets. For future reference, we refer to this gap as our edge, and we call the minimum edge we require to place a bet our value threshold.

For example, let us look at a hypothetical fixture between Manchester United and Chelsea at Old Trafford. In this example, ignore any costs related to placing a bet and assume perfect free competition between all bookmakers, ensuring that the bookmakers do not subtract an edge on the odds that they offer. The bookmakers give three in odds for a home win; this means that the bookmakers predict that Manchester United has a 33.33% chance to win this particular game. We derive this probability directly from entering the odds into formula 1.1.

If we predict the same likelihood of Manchester United winning to be 45%, we would have an 11.66% edge, according to our own forecast. We would therefore place a bet on this fixture if we had set our value threshold to 11% or below.

The goal is to see if we can beat the bookmakers, using econometrics, over the course of one Premier League season. Granted that we find fixtures fulfilling our value threshold criteria, we can test any model empirically. Should we succeed in creating such an algorithm, we could potentially have a money-generating machine that is able to beat the sports betting market, that anyone could use, without even having any knowledge of football. However, since the betting market is huge, and the bookmakers always try to give themselves an edge, we should expect to lose.

There are a vast number of bookmakers available to choose from, and often the odds vary rather significantly between the providers. There even exists exchange markets, in which bettors are betting against each other, both laying odds as well as placing bets. In the latter case, the bookmaker is other bettors, who in financial terms are shorting a result, providing odds that they deem to be too low. The service provider, for instance Betfair, earns its revenue by charging a small commission from the winner of the bet (typically around 2-5%). For this thesis however, we will not be looking at odds from the exchange market. This seems like a reasonable measure, seeing as the odds there fluctuate a lot. Instead, we will place our fictive bets using the historical odds offered by approximately 50 traditional bookmakers, carefully selecting the best odds for each bet.

# 2 Background

”Soccer clubs need to make fewer transfers. They buy too many Dioufs.” (Szymansky and Kuper, 2014, p.19)

## 2.1 Why is football predictions a relevant topic for an economics thesis?

As mentioned in the introduction, sports betting creates a market. A market in which the bookmakers are the sellers, and the individuals placing their bets are the buyers.

The sports betting market holds some similarity to the financial market, since we can view bets as investments. The outcome of the investments are then given by the results of sports matches, in the case of this thesis it will be soccer, which we will refer to as football throughout the paper.

We use the econometric model as a basis for statistical inference. One of the ways that statistical inference is applied includes predicting economic outcomes. (Hill, Griffiths, & Lim, 2008) Therefore, it follows that predicting the outcomes of football matches is a relevant study within the econometrics discipline by itself, but it becomes an even more relevant topic for economics once we test the predictions empirically on the sports betting market.

## 2.2 Related work

When performing a literature review, searching for related papers published on the topic of predicting football results, I have noticed one clear trend, the fact that computer science students are typically the ones writing master theses on the subject. This is very understandable, as one has to be able to use intricate statistical software in order to carry out modelling of the kind we do in this paper.

Throughout my literature review, I have not been able to find any published papers mentioning football forecasts using an extended Bradley-Terry model including the various wage bills of the relevant clubs.

In the paper “Dynamic Bradley–Terry modelling of sports tournaments” (Cattelan, Varin, & Firth, 2013), the authors examine the possibility of forecasting final league standings in the 2008/09 season of the Italian football league Serie A, using a Bradley-Terry model. They only use information about the final result of previous matches in their forecasts, and suggest that more detailed information about the previous matches could result in better data fitting, and improved forecasts. They mention home advantage as an important covariate. They are not concerned with the betting market, so they do not benchmark their forecasts against the bookmakers.

In the master thesis “Beating the bookie: A look at statistical models for prediction of football matches” (Langseth, 2013), profit is achieved by using a model building on the work of Mike J. Maher and his paper “Modelling association football scores”. (Maher, 1982) Langseth incorporates various explanatory variables to extend Maher’s model. He tests his model empirically on the 2011/2012 and 2012/13 season of the Premier League. Langseth does not mention the effect wages have on the results in the Premier League in his thesis, but instead focuses on in-game events in his modelling approach.

The thesis “What Actually Wins Soccer Matches: Prediction of the

2011-2012 Premier League for Fun and Profit" (Snyder, 2013), investigates various statistical models to carry out a probability forecast of the 2011/12 season in the Premier League. Snyder tests his results empirically against the historical bookmaker odds, and outperforms previous results found for the 2011/12 season. In his work, Snyder employs models that incorporate a wide range of covariates, including wages of the clubs in previous years, transfer budgets, and in-game events, such as yellow cards, shots and successful dribbles. Snyder does not use a Bradley-Terry model, but instead utilizes an R package called glmnet to fit a multinomial logistic regression model with his dataset.

# 3 Theory

"A Liverpool to London return faster than Robbie Keane." –Virgin Trains advertisement [2]

In this thesis, we will try to create a model that can outperform the forecasts of the bookmakers. To achieve this, we will build a model based on the Bradley-Terry probability model extended to accommodate draws, and include our own explanatory variables. (Davidson, 1970). In section 4, we delve deeper into the specifics of our model.

The research problem we try to solve is the following:

**"Can we apply an econometric approach to achieve profitable sports betting?"**

The book Soccernomics by Szymansky and Kuper, in which they analyze the Premier League and Championship teams from the 2003/04-2011/12 season, heavily influences the work in this thesis. (Kuper, 2014) Using statistical methods, they discovered that the wage bills explained over 90 percent of the variation in the participating teams' average league placements over the period, which I must say are truly remarkable findings, given the games complex nature. Moreover, for any one season, they found that clubs' wage spending accounted for approximately 70 percent of the variation in league positions. Szymansky and Kuper make a point to mention that they do not think these findings are because high pay causes good performances. Instead, they are of the belief that the high correlations observed stem from high wages attracting good performers.

From this it follows that wages should be one of the best explanatory variables when predicting the outcome of a football match in the Premier League. Thus, we will want to include this covariate when building our model.

The following is an excerpt from the most recent Premier League Handbook [3]:

"Each Club shall by 1st March in each Season submit to the Secretary a copy of its annual accounts in respect of its most recent financial year or if the Club considers it appropriate or the Secretary so requests the Group Accounts of the Group of which it is a member (in either case such accounts to be prepared and audited in accordance with applicable legal and regulatory requirements) together with a copy of the directors' report for that year and a copy of the auditors' report on those accounts." (Premier League Handbook Season 2015/16 p.104, 2015)

This is good news for us, since it means that the financial data of the clubs becomes accessible after a while. Unfortunately, this also poses a big problem for real-time application of our model. Since the wage bills of the clubs, at any current season, are unknown, one would have to lean on speculative data to incorporate wages as an explanatory variable when predicting results in order to place bets on upcoming fixtures. For research purposes, we will sidestep this problem by using historical data.

What Szymanski and Kuper found to be of not so high significance, was the transfer budgets of the clubs. Somehow, even though clubs spend millions upon millions of pounds in order to acquire the biggest stars, statistics indicate that they are unable to buy success directly from the transfer market. To illustrate this, Szymansky and Kuper use Liverpool from 1998-2010 as a case study. Over the period, the club had two managers, who both kept spending big on transfer fees, yet Liverpool never became one of the biggest title contenders. They go on to mention many purchases that flopped, but the one they deem most strikingly was the acquisition of Robbie Keane. In 2008, Rafael Benitez bought the 28-year old Keane from Tottenham Hotspur for £20 million, only to sell him back to Tottenham six months later at the reduced price of £12 million, netting a deficit of £8 million on one player, over the course of just 6 months. Virgin Trains even ran newspaper advertisements with the slogan "A Liverpool to London return faster than Robbie Keane." Mocking Liverpool for their transfer dealings will not be the theme of this thesis, but we will lean on the findings of Szymansky and Kuper, and omit transfer fees from our model. This is helpful for several reasons, reliable data being the main one. Player transfer fees are often not disclosed, a problem further addressed by Jason Burt in an article for the Telegraph. [4]

After we have built our model, and obtained our forecasts, we will test the results empirically. By doing this we can see how the model would have performed in a real life situation, given that we had known the wage data. In order to maximize expected value, theory must dictate that a profit maximizing individual betting on an outcome at all times will choose to place his wager with the bookmaker offering the best odds. Therefore, we will always assume use of the bookmaker that offered the best odds on the fixtures we predict, when presenting the results. To get the results, we must first compare our forecasts with those of the bookmaker offering the best odds for any one fixture, and look at the difference in our forecasts. We denote this difference our edge, and express it by the following formula:

$$3.1 \; Edge_{A,D,H} = \epsilon_{A,D,H} = \pi_{A,D,H} - \alpha_{A,D,H}$$

Where $\pi_{A,D,H}$ is our forecasted probabilities of an away win, a draw and a home win, respectively, using our extended version of the Bradley-Terry model with explanatory variables.

$\alpha_{A,D,H}$ represents the bookmaker's forecasted probabilities of an away win, a draw and a home win, respectively, given implicitly by the best odds offered for each outcome, using formula 1.1 to convert the probabilities into percentages.

We will then decide how big our edge has to be, in order for us to place a bet. As previously explained, we call this size our value threshold (VT). We denote any predicted outcome satisfying a given value threshold criterion a value bet. By this it follows, that as our value threshold increases, the amount of value bets decreases. In other words, we will be hoping for a large number of value bets that also yield a high return on investment per bet.

The return on investment explains our average return per bet in percentages, we denote this ROI. In theory, our return on investment should increase with our edge, and thus ROI should be an increasing function with respect to the value threshold, since increasing the value threshold means we require a higher edge to place a bet. This is however very hard to test accurately, since samples should be very small at the highest value thresholds, unless our forecasts differ a lot from those of the bookmakers.

Note that as sample size (number of bets in this case) increases, the ROI reported will gradually move closer to our true ROI.

# 4 Method and Data

## 4.1 Data

The data collected for our version of the Bradley-Terry model can be broken down into three parts:

- Historical data on the results of all games played in the Premier League season 2013/2014, including dates, home team, away team and full time results.
- Data on the average weekly wages for the players in all the 20 teams participating in the 2013/2014 season.
- Historical data on the odds offered on all the relevant fixtures.

The website football-data.co.uk [5] was used to obtain the results of all the games played in our sample, since they have data presented as .csv files available on the Premier League, ranging all the way back to the 1993/94 season. I opted to analyze the 2013/14 season, as this was the most recent season with all the financial data required available. Football-data.co.uk contains direct links to a format that is readable by R, thus importing the results data was very convenient.

The downloaded files did also contain historical data on the various odds offered by the bookmakers on all of the fixtures, but football-data.co.uk only include the odds of 10 different bookmakers, whereas oddsportal.com [6] supplies the historical odds from approximately 50 different bookmakers on every fixture. Selecting the latter to collect the historical odds was therefore crucial, as we are trying to maximize the return of all bets; and to do this we must select the best odds offered for any bet we are placing. Adding around 40 additional bookmakers to the pool of odds providers was therefore a most welcome addition.

Gathering the data from oddsportal.com was not as easy as collecting it from the .csv files at football-data.co.uk, but instead it had to be entered manually into excel because their data for the 2013/14 season spans over eight pages. I firmly stand by the fact that this is worth the extra effort, in order to ensure that we obtain the best results possible, when looking at how the model performs later.  Using formula 1.1, I converted all the best odds from all the games of the season into probabilities, in order to compare the bookmakers' forecasted probabilities with those of our own model. Converting the historical odds into probabilities is a very straightforward operation once we have the odds.

It is worth to mention that the historical odds, converted into probabilities, seldom add up to 100%. This is because the bookmakers subtract a small and varying amount, and thus total probabilities of more than 100% can be read as a payback percentage below 100%, whereas total probabilities of less than 100% can be read as a payback percentage of above 100%. This is easy to illustrate:

Say a bookmaker forecasts the probability of a home win, draw and away win to all be equal. This would imply a probability of 33.33% for each outcome, which we can convert into odds using formula 1.2: $\frac{1}{0.3333} = 3$ for each outcome. However, in such a case, the bookmakers would not offer an odds of 3 on all outcomes, seeing as they would only break even (assuming no fees to place a bet, win a bet or deposit or withdraw money from the site in question) if bettors placed bets evenly distributed on all three outcomes. Instead, the bookmakers would offer odds of slightly less than 3, for instance 2.9 on each outcome, thus ensuring that they have an edge over the customers. Had the opposite been the case, they would end up losing money over the long run, not considering other sources of income, and therefore this has to hold true for any profit-maximizing bookmaker. Continuing along this train of thought, if we now convert these slightly lower odds back into probabilities, we would have $\frac{1}{2.9} = 34.48\%$ for each outcome, which in total would add up to $34.48\% * 3 = 103.4\%$.

The average total probability, however, using the best odds offered for each outcome, turned out to be 99.68%. The reason this was possible is that we do not use odds from the same bookmakers on the various outcomes when gathering historical odds on one fixture. For example, Pinnacle could have offered the best home win odds, whereas bet365 had the best draw odds and Nordicbet provided the best away win odds. You could view this as a very mild form of arbitrage, but it is completely possible, since we are free to choose whatever bets we want to place with any given bookmaker. Again, I must stress that selecting the best odds for any given result on any given fixture is crucial in order to maximize profits.

Moving on, the final piece of data used in the model was the average weekly salaries for players of all 20 clubs in question. I obtained these by collecting the figures from mirror.co.uk [7] who got their data from sportingintelligence.com [8]. Some level of skepticism is advised for these exact figures, but seeing as all the clubs are obligated to send detailed information regarding their accounts to the government every year, and that this information later is accessible by the public, the older the seasons, the more reliable this data should be (to a certain point, of course). For instance if a paper publishes news on the just finished 2015/2016 season, they are most likely presenting speculative figures, and should be treated as such. Since both the 2013 and 2014 accounts are accessible by anyone, however, I trust the figures reported to be at least close to accurate. I will discuss this topic more extensively in the concluding section.

## 4.2 The Model

The software used to code and obtain the predicted probabilities is R.

The model we use to provide our probability forecasts is an extended Bradley-Terry model with explanatory variables.

The standard Bradley-Terry model is a probability model used to predict the outcome of a comparison. It is useful in situations where individuals from a group repeatedly compete with one another in pairs.

From Hunter's article "MM algorithms for generalized Bradley-Terry models" (Hunter, 2004), we obtain the standard Bradley-Terry model, which we can write as:

$$(4.1)\ P(individual\ i\ beats\ individual\ j) = \frac{\pi_i}{\pi_i+\pi_j}$$

Where $\pi_i$ is a positive parameter associated with individual i, for all the comparisons where individual i faces individual j. Furthermore, we can read $\pi_i, \pi_j$ as sports teams, where $\pi_i$represents the skill level of team i.

If we were to simulate which of two teams were most likely to win, this would be sufficient. We could also link covariates, or explanatory variables to the model. We can think of this the same way we would with a normal regression analysis, in which we would have

$$\ln(\pi_i) = \beta_0 + \beta_1 x_{1,i} + \beta x_{2,i} + \beta_3 x_3, i$$

where $\pi_i$ is the "ability score" or "talent", for team i.

$\beta_o$ is representing some constant, $\beta_i, i = 1,2,3$ would be the coefficients and $x_1, x_2\ and\ x_3$ represent our covariates, namely recent form, salaries and home advantage. This would be the case for a log – linear regression with three explanatory variables.

In order to use these probabilities to place bets on football matches however, we are required to extend the model to include probabilities of a draw being the result.

Thankfully, others have already done this before us.

In Roger R. Davidson's paper (Davidson, 1970), on extending the Bradley-Terry model to accommodate ties in paired comparison experiments, Davidson introduces the Extended Bradley-Terry model.

He bases it on the paper Ties in paired-comparison experiments: a generalization of the Bradley-Terry model (Rao & Kupper, 1967).

Following Davidson's paper, he presents the extended Bradley-Terry model mathematically in the following way:

"2. <u>The mathematical Models</u>.

In paired comparisons one considers a set of t treatments which are presented in pairs. It is assumed that the responses to the treatments may be described in terms of an underlying continuum on which the "worths" of the treatments can be relatively located. Let $\pi_i$ denote the "worth", an index of relative preference, of the $i^{th}$ treatment, $\pi_i \geq 0, \;\; \sum_{i=1}^{t} \pi_i = 1$. The Bradley Terry model postulates that, if $X_i$ $and$ $X_j$ are the responses to treatments I and j respectively, then

$$P(X_i > X_j) = \pi_i / (\pi_i + \pi_j) \qquad (2.1)$$

in the comparison of treatments i and j. One interprets $X_i > X_j$ as indicative of preference for treatment i over treatment j. It is noted by Bradley [1] that in replacing the normal density of the Thurstone-Mosteller model by the logistic (squared hyperbolic secant) density, one obtains the Bradley-Terry model. Specifically, if the difference $Z_{ij} = X_i - X_j$ between the responses is assumed to have logistic distribution with location parameter $(\ln \pi_i - ln\, \pi_j)$

and with distribution function

$$P(Z_{ij} \leq z) = 1/\{1 + \exp[-(z - ln\, \pi_i + ln\, \pi_j)]\}, -\infty < z < \infty, \qquad (2.2)$$

for each treatment pair (i,j), then (2.1) follows by setting z = 0,

and subtracting from unity (cf. Bradley [5]).

With the use of a threshold parameter, $\eta = \ln \Theta$, in conjunction with (2.2), Rao and Kupper [17] obtain

$$p^*(i|i,j) = P(Z_{ij} > \eta) = \frac{\pi_i}{\pi_i + \theta\pi_j}$$

$$p^*(j|i,j) = P(Z_{ij} < -\eta) = \frac{\pi_j}{\theta\pi_i + \pi_j} \qquad (2.3)$$

$$p^*(o|i,j) = P(Z_{ij}| < \eta) = \frac{(\theta^2 - 1)\pi_i\pi_j}{(\pi_i+\theta\pi_j)(\theta\pi_i+\pi_j)}$$

for $i \neq j, i, j = 1, \ldots\ldots, t$. The quantities $p^*(i|i,j)$, $p^*(j|i,j)$ and $p^*(o|i,j)$ represent the Rao-Kupper probabilities of preference for i, preference for j, and no preference respectively, when the treatment pair $(i,j)$ is presented." (Davidson, 1970, p. 2-3).

If we inspect the formulas in 2.3, we notice that if we set the value of $\eta = 0$ or $\theta = 1$, the probabilities estimated are the same, as the ones we would obtain using formula 2.1, the standard Bradley-Terry model. The third of the three formulas is vital here, as it extends the model to include probabilities of a draw as well. Now, we could use this to create a forecast based on the historical results data itself, but the intention of the thesis was always to include some more explanatory variables as well. The explanatory variables used in our version of the extended Bradley-Terry model were recent form, average weekly salaries of team i, where $i = team\ 1,2, \ldots, 20$ and then finally home advantage.

Form is modelled as a changing variable that scores a teams' form based on its performance in the last 5 games, giving 1 point for each victory, 0 points for each draw and -1 points per loss. For example, if we are making a forecast of the probabilities of a game involving Manchester United, and they had won three out of their last five games, drawn one and lost one, their form variable would sum up as follows:

Manchester United's form = 1+1+1+0-1 = 2

Therefore, we give them a form score of two for their upcoming fixture, when doing our analysis.

The average weekly salaries is our key explanatory variable in the model. This is building on the findings of Szymansky and Kuper, in their book Soccernomics, as explained previously in the Theory section.

Lastly, I wanted to include the home advantage variable, based mostly on heuristics, through watching football for many years, but also because of experience with sports betting. Home teams, in my experience, usually get lower odds than in the opposite fixture where they face the same team away. This means that the bookmakers include this factor into their probability forecasts when valuing the three possible outcomes, so I wanted to include it for our forecast as well.

It is possible to expand the model further, but for this thesis, we will settle with the extended Bradley-Terry model estimators, including our own three covariates.

## 4.3 Modelling the data in R

Our model builds on the work of Prof. Øystein Myrland, who performed a simulation of a season in Tippeligaen, using a Bradley-Terry model in R. He was nice enough to give me access to the code he had written in order to carry out his forecast. His model was not concerned with ties, and did not include additional covariates, but instead attempted to predict the final standings of the upcoming league season based on the results of the previous seasons. In other words, it was not possible to use it for our intended betting purposes as it stood.

Having access to his code did however serve as both an introduction to R programming as well as an introduction to practical application of the Bradley-Terry model.

The code of Myrland took advantage of the R package BradleyTerry2, which substantially simplifies matters of simulating a season where each fixture only has two outcomes. However, aforementioned package does not accommodate ties. It follows that the difficult part of writing the code was to incorporate ties in the model, as well as creating and attaching the covariates mentioned in the previous section.

In this section, we will discuss how we created the model we use to predict the outcomes in the 2013/14 season of the Premier League. Starting with our three covariates, we discuss them in order of appearance.

The first covariate introduced is the salaries vector, which attaches all the average weekly salaries per player of each club to its respective clubs. Since we suspect that there should be some form of diminishing returns on the effect of increasing the wages, we investigate the probability outcomes using this covariate on both log and linear form. We can determine the best fit to our model at a later stage, after looking at the betting results empirically.

The second covariate introduced, is the home advantage. Depending on where the fixture we wish to predict is played, this variable takes a form of $[1|0]$. We ensure that this variable is employed correctly by always rewarding the home team with a 1, while the away team gets a 0.

Lastly, we introduce the recent form variable. We use a loop to create this variable, collecting data from a maximum of the five last games played per team, starting from the second game. Then we rank the form by summing the scores of the last 5 games, with 5 being the maximum score, and -5 being the minimum score. R-bloggers.com provide an excellent tutorial, written by Martijn Theuwissen, for creating loops. [9]

To provide an explanation of how we later arrive at the model we use for our predictions, first we need to explain the functions needed to create the actual model. We derive our Extended Bradley-Terry model from the formulas listed in section 4.2, in the quoted part of Davidsons' paper:

$$
\begin{aligned}
\text{“}p^*(i|i,j) &= P\big(Z_{ij} > \eta\big) = \frac{\pi_i}{\pi_i + \theta\pi_j} \\
p^*(j|i,j) &= P\big(Z_{ij} < -\eta\big) = \frac{\pi_j}{\theta\pi_i + \pi_j} \qquad (2.3) \\
p^*(o|i,j) &= P\big(Z_{ij}\big| < \eta\big) = \frac{(\theta^2 - 1)\pi_i\pi_j}{(\pi_i+\theta\pi_j)(\theta\pi_i+\pi_j)} \text{“}
\end{aligned}
$$

(Davidson, 1970, p. 3)

We create a function that incorporates these formulas to predict the three possible outcomes (Home win, draw, away win, respectively). Later we modify this to include our own covariates. We then optimize this modified function, using the optim function in R, which uses Nelder and Mead's simplex method for function minimization (Nelder & Mead, 1965), to find the coefficients that best fit our model. We run the optim command on the negative sum of the probabilities, since it minimizes functions by default.

Continuing, we introduce a log likelihood function. On R-bloggers.com, there is a brilliant article by John Myles White, in which he shows examples of how to do a maximum likelihood estimate by hand. The last example he uses proves useful here, as it shows how we can specify a log-likelihood function to do maximum likelihood estimates on any given data set and model. [10]

We let our log-likelihood function be a function of the theta, the response (which is the previous outcomes in the dataset we base our forecast on, so previous results), and the predictors for the home and away team, named x1 and x2. Moreover, since R by default minimizes functions, and we want to maximize our log likelihood estimate, we take the negative sum of the probabilities and store it in a variable, so we can later input it directly when we use the optim function in R.

Moving on, we need to create a function in order to fit the actual data we want to use in order to make our predictions. This function will thus create the modified Bradley Terry model for us. We call this function fitmodel. First, we specify that we want to let fitmodel be a function of the relevant dataset. Then we specify y, which is the full time results for each of the fixtures leading up until the one we want to predict, given as either A, D or H. This way we can make sure that our model does not use data "from the future" that would be unavailable in a scenario where we wanted to forecast probabilities on an upcoming fixture.

Next, we specify x1, which is the relevant explanatory variables that we use for the home team, namely the intercept, the salaries, the home team advantage and the form. Continuing, we specify x2 to be the relevant explanatory variables that we use for the away team, which are the same as for the home team. Lastly, we optimize the parameters ($\beta_0, \beta_1, \beta_2, \beta_3$ $and$ $\theta$) within our model, using R's optim function.

After we have specified a function that lets us fit our own desired covariates to the model, we need a function to let us actually use all that we have gathered so far, to predict the outcome of an upcoming fixture. This function is dependent on the optimal coefficients that we got from the fitmodel function, and can use any input we wish to assign it. We then go on to specify the fixture we want to predict, and get our results based on the maximized parameters that we estimated with the fitmodel function, and the corresponding explanatory variables for both teams.

After we have all the covariates and functions, all that is left is to loop our functions over the entire season, using what we learned when creating the recent form variable. We achieve this by first specifying that we do not want to predict the outcomes of the fixtures played on the first date, and then basing our forecast of the upcoming fixtures on the results leading up to the fixtures on the current date that we are predicting. The loop ensures that we do this until we have estimates for all dates excluding the first one. After we have created this loop, we write out our estimates to a .csv file containing all of our predictions for the season, neatly listed and readable by Excel. Finally, we do this loop one more time, taking the logs of the average weekly salaries per player, in order to be able to compare how our model performs when we take the logs of the wage data, against the results we obtain when we use the wage data on linear form.

# 5 Results

## 5.1 Interpretation of results

In this section, we will discuss the results we obtain by putting our model, explained in the previous section, to use. We compare the predictions of the bookmakers, and the predictions we obtained from our model, in order to find our estimated edge on a bet. Recall section 3, where we defined our edge as given by:

$$3.1\ Edge_{A,D,H} = \epsilon_{A,D,H} = \pi_{A,D,H} - \alpha_{A,D,H}$$

We will look at results for six different value thresholds, the 5%, 8%, 10%, 12%, 15% and 20% levels, respectively. Recall that value thresholds are the minimum edge we require to define a bet as a value bet. We will assume that we bet on all value bets for any particular value threshold. The return on investment, denoted ROI, explains our average return per bet in percentages.

For simplicity and transparency reasons, we will use a fixed bet size of 100 for all of our bets. This makes results easier to read and interpret.

We will run one simulation for all fixtures on a given date, update the relevant variables, run simulations for all fixtures on the next date, and repeat this process until we have a forecast for the entire season.

## 5.2 Results using linear salaries

### Table 5.1

Results of fictive betting, using the 5% value threshold, with salaries on linear form.

| 5% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 48 | 66 | 38 | **152** |
| Bets won | 56 | 23 | 37 | **116** |
| Total bets | 104 | 89 | 75 | **268** |
| Results | -165 | -426 | 692 | **101** |
| Average profit per bet | | | | **0,37686567** |
| ROI per bet | | | | 100,38 % |
| ** 26 cross-bets home win/draw | | | | |
| ** 16 cross-bets away win/draw | | | | |

The table shows the results we would have gotten if we had placed bets of 100 (currency is irrelevant) on all qualifying value bets, using our predictions.

When we examine the 5% VT, we see that we have 268 qualifying value bets, as opposed to the 24 we find when using the 20% VT. This means that the sample size increases more than ten-fold, giving a much more accurate benchmark of our predictions, when using the 5% VT compared to the 20% VT. Below follows, a graphical illustration of the results we would have achieved by using our predicted probabilities to bet on the 2013/14 Premier League season, setting our value threshold to 5%.

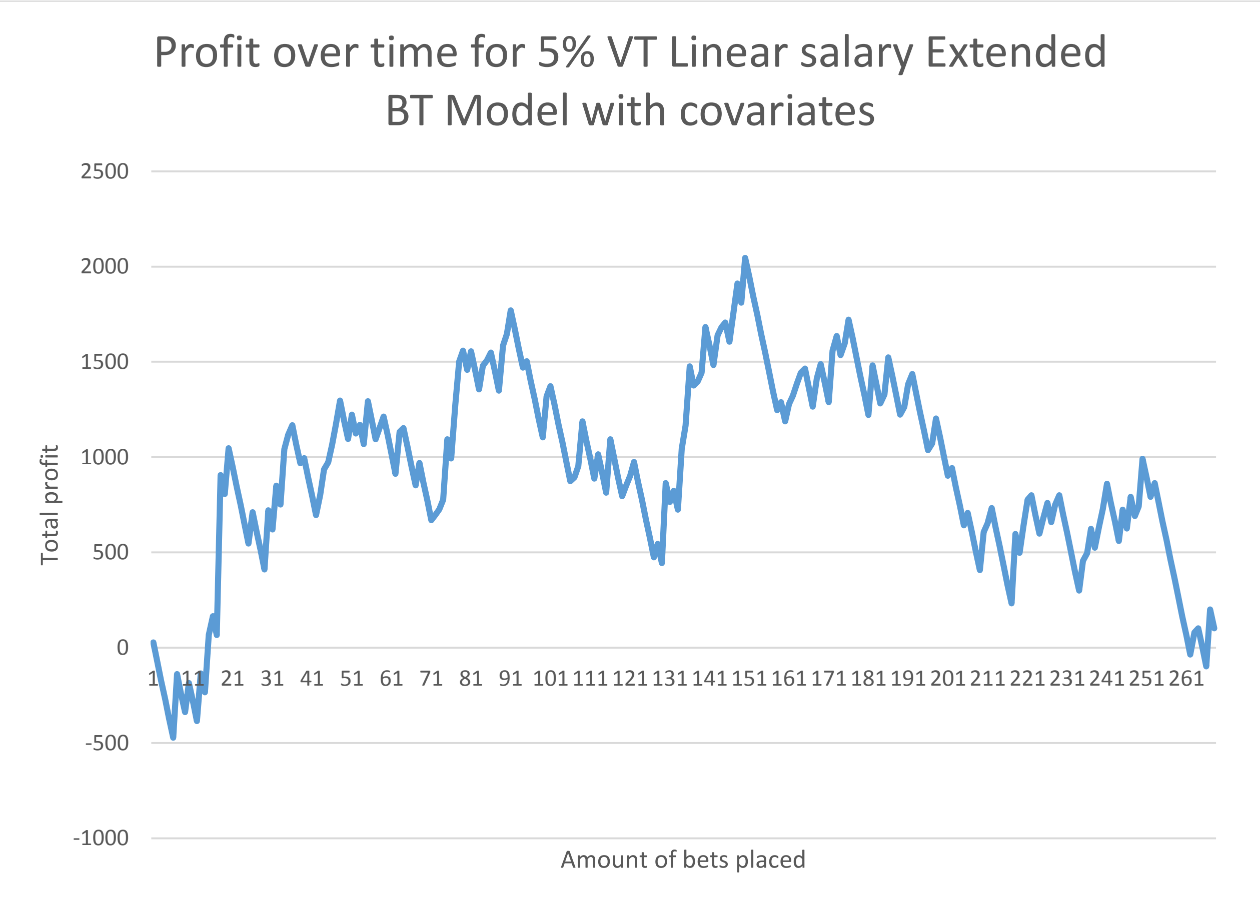


Figure 5.1

The graph in figure 5.1 illustrates the profit we would have gotten over the 2013/14 Premier League season, if we had used our model with salaries on linear form, setting a 5% value threshold, carefully selecting the bookmaker that offers the best odds for every bet placed, and using a fixed bet size of 100.

This result is obviously not great. When looking at figure 5.1, we see no immediate trend in the graph. Furthermore, the total profit reported is only 101 over the 268 bets. Out of the six value thresholds used, this one yielded the worst result.

Table 5.2

Results of fictive betting, using the 8% value threshold, with salaries on linear form.

| 8% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 32 | 18 | 17 | **67** |
| Bets won | 45 | 10 | 28 | **83** |
| Total bets | 77 | 28 | 45 | **150** |
| Results | 759 | 942 | 1781 | **3482** |
| Average profit per bet | | | | **23,2133333** |
| ROI per bet | | | | 123,21 % |
| ** 12 cross-bets home win/draw | | | | |
| ** 2 cross-bets away win/draw | | | | |

Comparing the various value thresholds, we see that we would have achieved the highest total profit and the highest ROI per bet by using the 8% value threshold criterion. Over the season, using an 8% VT, we would have achieved a remarkable 23.21 average profit per 100 bet, which translates to a ROI of 123.21%. Our total profit would have been 3482, over 150 qualifying value bets.

The 5% value threshold gives us the biggest sample size for our benchmark, and the 8% value threshold yields the best results. Another point that stands out is that we have a positive return on investment, using any of the value thresholds listed, when using our forecasts to bet on the Premier League 2013/14 season. For further comparison of the various value thresholds, see tables in the appendix.

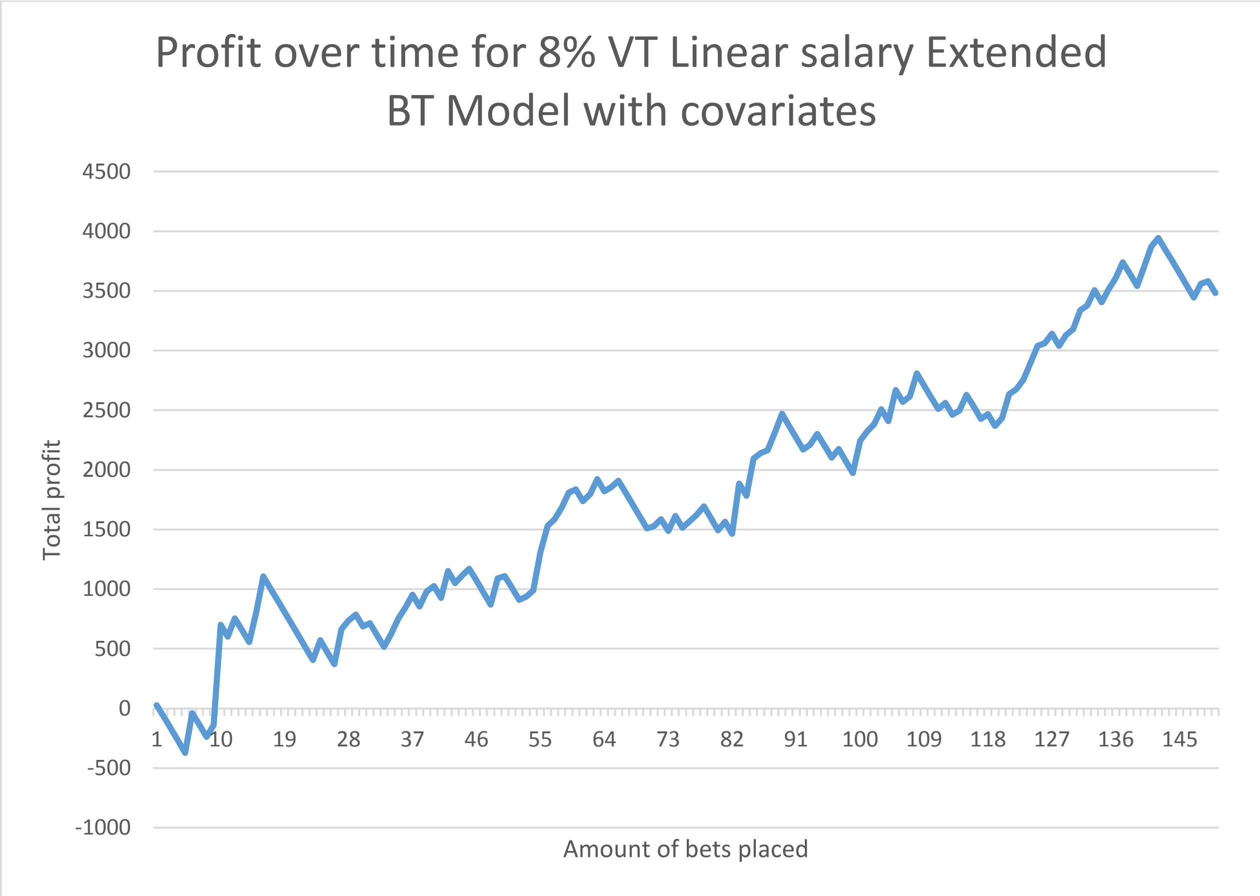


Figure 5.2

The graph in figure 5.2 illustrates the profit we would have achieved over the 2013/14 Premier League season, if we had used our model with salaries on linear form, setting a 8% value threshold, carefully selecting the bookmaker that offers the best odds for every bet placed, and using a fixed bet size of 100.

Notice the positive trend of the graph; it seems to be moving steadily upwards, suggesting that we do indeed have a profitable model for picking bets.

We can also note that we would win 83 out of the 150 bets placed, for a win percentage of 55.33%, but this statistic is not a very interesting one, since it can easily be misleading. If our goal were simply to maximize win percentage, we would just always bet on the favorite, in other words the outcome with the lowest odds.

However, for comparisons sake, if we were to bet on the favorite in all the same games that we bet on using the 8 percent VT, our total profit would be 1950 instead of 3482. Instead of 150 bets, we would only place 136, since we only bet on one of the three possible outcomes in the 14 cross-bets (fixtures where we bet on more than one outcome, when following our predictions). The ROI over the same sample would then be lower, more specifically 114.33%, but the win percentage would be higher as we would have won 84 out of the 136 bets, yielding a win percentage of 61.7%. For the remainder of this paper, we will not compare this approach of picking bets to our results, since the relevant question is whether we can beat the bookmakers using econometrics to pick bets. Whether we could make money by simply choosing the bookmaker that offers the best odds on the most probable outcome, and always betting on the favorite, is another discussion that we will not be concerned with for this paper.

On that note, a much more interesting comparison is to see how the model performs when we make a change to the most important explanatory variable, namely the average weekly salaries that the clubs pay their players.

## 5.3 Results using log of salaries

A quick and effective way of incorporating diminishing returns for increasing the wage bills is to take the logs of the salaries used in our data. Below follows a table containing the results we would have gotten, using our model with salaries on log form, using the 5% value threshold.

Table 5.3

Results of fictive betting, using the 5% value threshold, taking logs of salaries.

| 5% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 41 | 60 | 49 | **150** |
| Bets won | 52 | 29 | 39 | **120** |
| Total bets | 93 | 89 | 88 | **270** |
| Results | 744 | 1679 | -98 | **2325** |
| Average profit per bet | | | | **8,61111111** |
| ROI per bet | | | | 108,61 % |
| ** 23 cross-bets home win/draw | | | | |
| ** 31 cross-bets away win/draw | | | | |

Looking at the results for the 5% VT, observe that we would have a ROI per bet of 108.61% over 270 bets. When looking at the results for the same VT using linear salaries, we have a ROI of 100.38% over 268 bets. One could argue that 5% is the most important threshold, since it contains the most value bets, and therefore the ROI reported is statistically more likely to be closer to the true ROI, than for any of the other thresholds.

It is very interesting that we are able to measure such a spread in performance over the two models on the 5% VT level, with salaries on linear and log form, respectively. It suggests that taking the logs of salaries, when predicting soccer results, gives us estimates that are more reliable than the estimates of the linear counterpart.

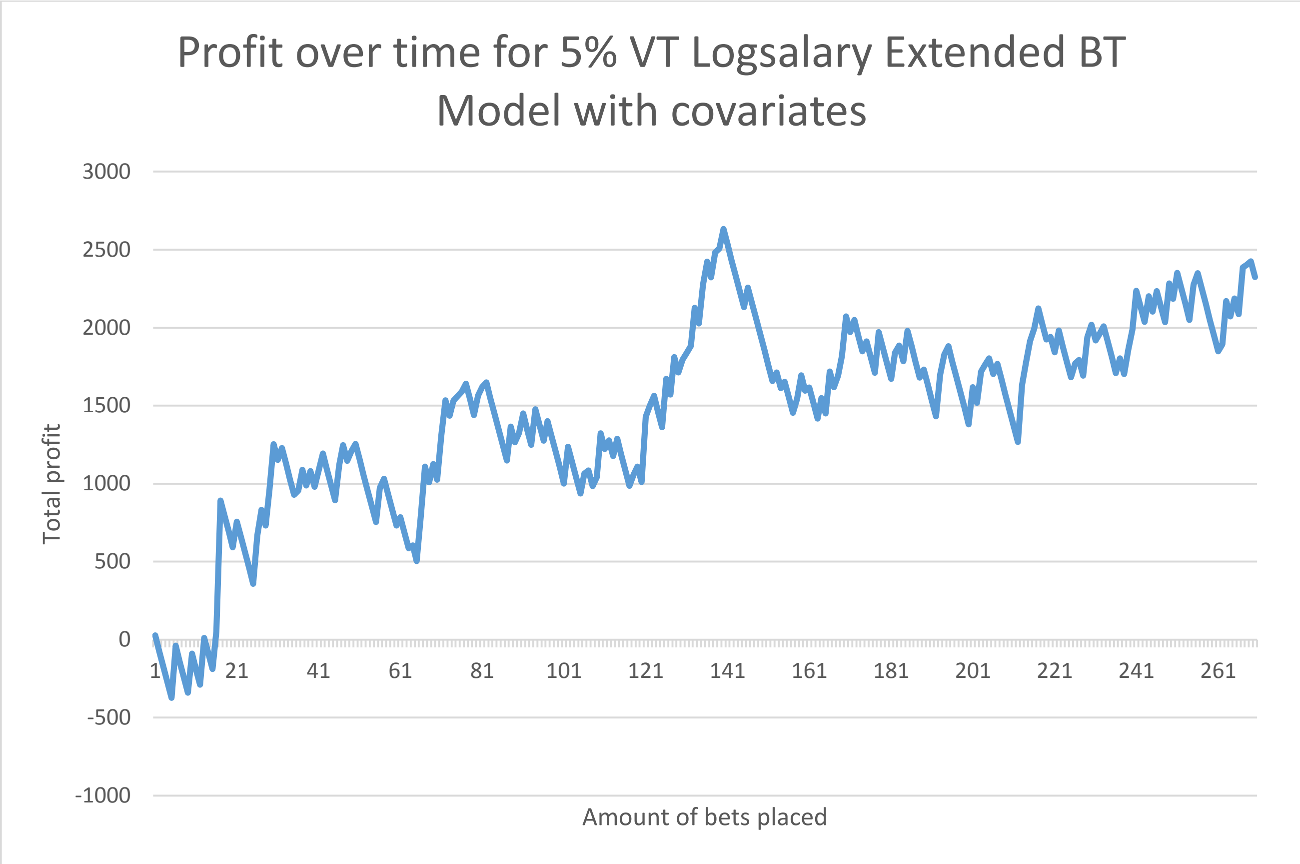


Figure 5.3

The graph in figure 5.3 illustrates the profit we would have gotten over the 2013/14 Premier League season, using our model with salaries on log form, setting a 5% value threshold, carefully selecting the bookmaker that offers the best odds for every bet placed, and using a fixed bet size of 100.

We can observe a positive trend, but there also seems to be more variance in the results, than in the graphs depicted in figure 5.2 and 5.4. A possible explanation could be that the average estimated edge is higher in those two graphs, and therefore the variance is lower. A more likely explanation for this graphs' more swingy nature is that it could be down to the variation in the sample, or it possibly be a combination of the two.

Table 5.4

Results of fictive betting, using the 8% value threshold, taking logs of salaries.

| 8% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 30 | 17 | 24 | **71** |
| Bets won | 40 | 9 | 28 | **77** |
| Total bets | 70 | 26 | 52 | **148** |
| Results | 995 | 772 | 1368 | **3135** |
| Average profit per bet | | | | **21,182432** |
| ROI per bet | | | | 121,18 % |
| ** 6 cross-bets home win/draw | | | | |
| ** 6 cross-bets away win/draw | | | | |

Again, as we had in the case of linear salaries, we see that 8% is the VT that performs best out of the six thresholds inspected.

From table 5.4, we see that the ROI measures to 121.18% per bet, over 148 bets. The total profit is 3135.

When comparing the results for the 8% VT, using salaries on both linear and log form, we can see that the difference in total profit is only 347. The difference in ROI is also small, with the linear version of the model performing 2.03% better per bet. The total amount of value bets estimated is almost identical as well, with 150 and 148 for the linear and log version, respectively.

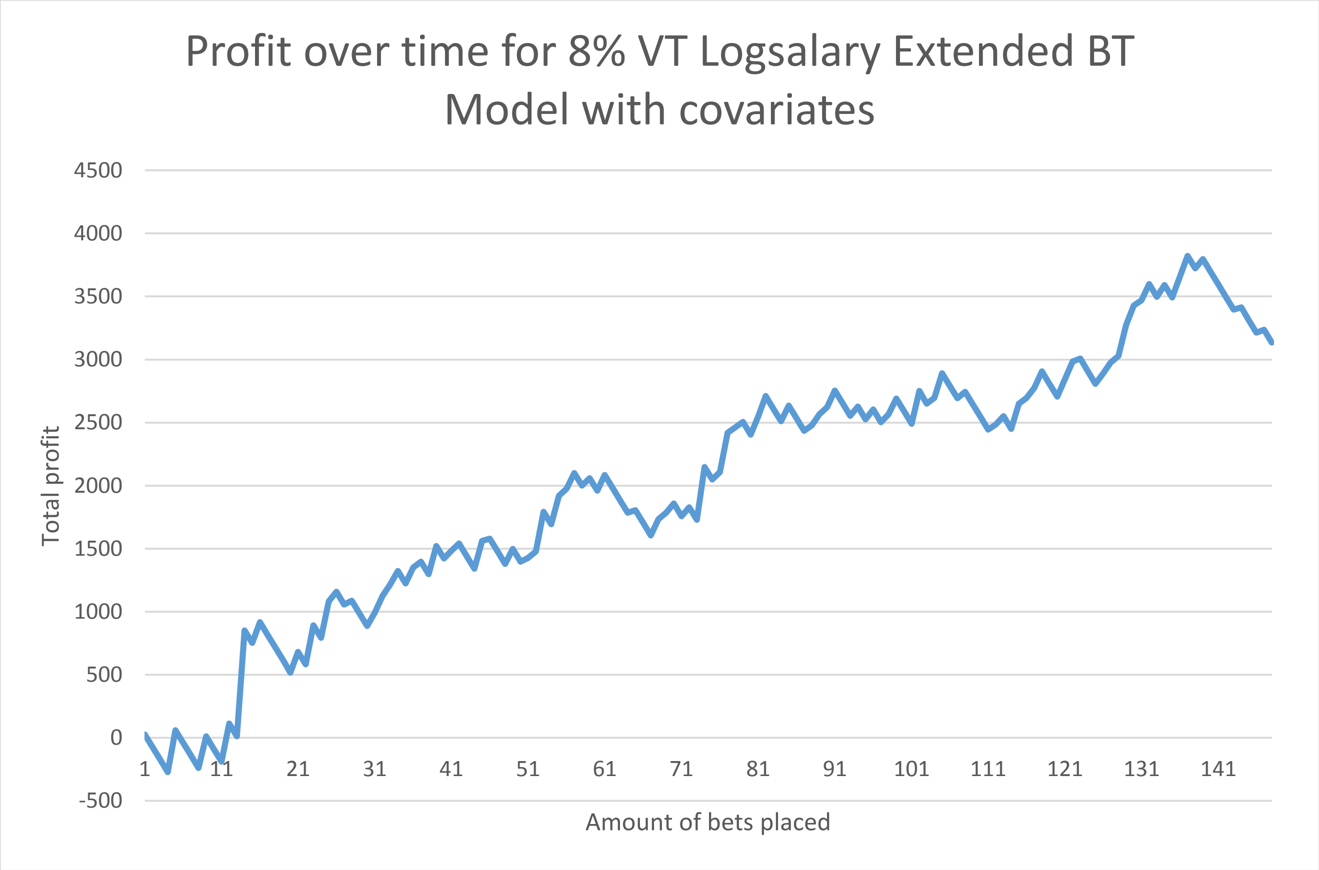


Figure 5.4

The graph in figure 5.4 illustrates the profit we would have achieved over the 2013/2014 Premier League season, using our model with salaries on log form, setting a 8% value threshold, carefully selecting the bookmaker that offers the best odds for every bet placed, and using a fixed bet size of 100.

Upon inspection of the graph in figure 5.4, we again observe a positive trend, very similar to the corresponding 8% VT results we got from using the model with salaries on linear form. This is a very uplifting result, since it represents one more indicator that we have indeed met our goal, which was to provide profitable forecasts.

When comparing all of the thresholds, we can see that placing bets using any of the VT's except for 20% would yield a positive result. Furthermore, and more importantly, we see that the 5% value threshold performs much better than in the model with linear salaries. This is particularly important because the 5% VT represents the biggest sample size out of the six thresholds we have chosen.

We can also see that the model where we take the logs of salaries delivers a more stable result across the various value thresholds. It should be noted that the 20% VT is slightly losing in this version of the model, but with a sample of only 18 bets, this result can easily be written off as unfortunate variance.

The 10, 12 and 15 percent value thresholds perform consistently in the 17-20% ROI range, but the samples observed are decreasing for each increase in VT. The total profit is also lower for each increase in VT, which is logical, since the amount of bets decrease significantly for each increase in VT, but the ROI only varies with 1-4%. For further comparison of the different VT's using the log of salaries model, see the appendix.

Benchmarking the results of our model against other similar models used on the same season would provide valuable insight into how our model performs, but unfortunately, we have not been able to find such results anywhere.

# 6 Conclusion

## 6.1 Salaries – Weakness and strength of the model

The fact that we would not know what salaries the various clubs are paying their players at any current date means that we probably will be unable to reproduce the results shown in this thesis in real time. However, we could make some educated guesses as to what the current salaries would be, taking various approaches. We could look at the historical wages paid over several seasons, and try to linearize the development in order to achieve an estimate. Another approach could be to adjust last year's salaries for inflation, and employ these estimated figures when predicting results. Another approach could be to take the speculative figures reported by various newspapers. The pitfalls are many, so it is probably best to avoid this approach, as it could potentially lead to us basing our economic decisions on the guesswork of some journalists who are under pressure to deliver stories for their respective newspaper.

Salaries appear able to explain a huge part of the variance in football results. It therefore seems best to include them in one way or another, when trying to predict the outcome of football matches. The average salaries of the clubs is the only explanatory variable that we would need in order to be able to achieve forecasts of the same quality as the ones in this thesis. If we find a solution to this, we can potentially carry out forecasts in real time and provide profitable betting tips based on our model.

## 6.2 Econometrics 1 - 0 bookmakers

There are very few bets that fulfill the 20 percent value threshold criterion, regardless of which of the two models we used to predict outcomes. This means that the biggest edges are rare, but more importantly, it means that our predictions match those of the bookmakers to some extent. This result is inspiring, since it should mean that our model is able to predict the probabilities of the various outcomes quite accurately. Even more uplifting is the fact that out of six different value thresholds, and two different versions of our model, only one of the combinations yielded a negative return. This was also the combination with the smallest sample, namely the 20% value threshold applied to the predictions from the log of salaries version of the model.

In this thesis, we have demonstrated that it is possible to beat the bookmakers over a Premier League season by taking an econometric approach, assuming that we had known what wages to put into our model. The half time score is therefore Econometrics 1 – 0 bookmakers. We will need bigger samples to verify the results further, but even though it is still early days, the positive trends of figures 5.2-5.4 holds promise of a bright future.

Our thesis does not incorporate commissions and fees into the results, something that is often present at online bookmakers. However, the effect these fees would have on the results would be small and negligible. Taking advantage of sign-up bonuses and other promotions offered by the betting companies would probably offset this effect for beginning sports bettors, and the commissions charged are typically only around the 2% range for winning bets. Even after employing the fees and commissions, we would still have a ROI of over 120% per bet at the 8% value threshold, irrespective of which version of our model we used, for the 2013/14 Premier League season. Why the 8% VT is performing best in both versions of the model could be down to several reasons, but it is likely that variance in the sample is playing a big part in this.

In the future, we can extend our model even further. One explanatory variable that we could add in, would be the bookmakers own odds, which in effect would mean that we would be using their own forecasts against them, by improving our own predictions. This would likely mean smaller, but more accurate edge estimates. The coefficients of this variable would definitely be negative, as increasing odds means decreasing probability. We could also use other statistics to extend our model, and incorporate in-game events such as shots on target, dribbles and tackles, among others.

We have not considered bet sizing strategy for this thesis, as the main goal is not about bankroll management and how to get rich, but whether or not we can apply an econometric approach to essentially make better predictions than those of the bookmakers.

Another point to consider, when we think about real life application of the model, is that we should not just blindly follow our own predictions. The bookmakers often have good reason to devaluate an outcome, as well as increasing another. This could be down to reasons that are hard to incorporate into an econometrics model, such as an unsettled squad, illness in the team, suspensions and other non-quantifiable variables. Hypothetically, consider a case where Jose Mourinho took the Manchester United first team squad out to dinner at one of the fancy restaurants in Manchester, and the whole team got food poisoning. Mourinho would have to field a team of youngsters from the reserve team in the upcoming fixture, but our model would predict the likelihoods of each outcome assuming nothing had happened, while in reality Anthony Martial would be lying sick at home in bed. Naturally, the chances of United winning would have decreased significantly, and the bookmakers would quickly adjust the odds accordingly. Taking no precautions, we would just happily bet on Manchester United to win, if we had estimated an edge on them winning that satisfied our value threshold, and we would likely end up being punished for it.

This brings us to the next point. Return on investment does not increase with the value threshold, contradicting theory. In theory, we should have gotten a higher ROI as we demanded a higher edge per bet, thus increasing the average edge. When we studied our empirical results, this did not hold true. On the contrary, the biggest value thresholds also have the smallest samples, so it is hard to draw accurate conclusions. One possible explanation for the biggest edges estimated performing poorly, could be variables not recognized by our model, such as in the aforementioned example.

The results of any one season are going to vary greatly, and there are definitely complex details in football, that no existing econometrics model is able to predict accurately. The 2015/16 season just finished, seeing Leicester finishing in first place with 81 points! Leicester, who had finished the 2014/15 season in 14$^{th}$ place, with a total of 41 points. In the 2013/14 season, their wage bill was £36.3m, which seems small when you compare it to Manchester United's £215.8m in the same season. [11] The difference becomes even bigger when you consider the fact that Leicester was playing in the Championship. Following a successful 2013/14 season, Leicester were promoted to the Premier League for the 2014/15 season. Following their promotion, their wage bill increased to £57m. Manchester United had reduced their wage bill down to £203m this season. [12] The wages for the 2015/16 season have not yet been published, but it is probably safe to assume that Manchester United, who finished fifth this year, were still paying their players far more on average than Leicester who won. Therefore, had we applied our model to this year's season, it is likely that we would have lost on a lot of Leicester's fixtures, due to basing our predictions on the salaries, rather than in-game frequencies, and therefore underestimating Leicester.

In this thesis, we have managed to show that any profit maximizing individual, regardless of football knowledge or tips from experts, using only an econometric model, and having access to the wage bills of the relevant clubs, could have been a profitable sports bettor in the 2013/14 season of the Premier League by following our predictions. Once the 2016/17 season starts, we will be able to test our model empirically in real time, completely risk free, as we do not necessarily have to place any bets in order to verify profitability.

# 7 References

## 7.1 Academic books and articles


Cattelan, M., Varin, C., & Firth, D. (2013). Dynamic Bradley–Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 62*(1), 135-150. doi:10.1111/j.1467-9876.2012.01046.x

Davidson, R. R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association, 65*(329), 317-328.

Hill, R. C., Griffiths, W. E., & Lim, G. C. (2008). *Principles of econometrics* (Vol. 5): Wiley Hoboken, NJ.

Hunter, D. R. (2004). MM algorithms for generalized Bradley-Terry models. *Annals of Statistics*, 384-406.

Kuper, S. (2014). *Soccernomics: Why England Loses, Why Spain, Germany, and Brazil Win, and Why the US, Japan, AustraliaÑand Even IraqÑAre Destined to Become the Kings of the World's Most Popular Sport*: Nation Books.

Langseth, H. (2013). *Beating the bookie: A look at statistical models for prediction of football matches.* Paper presented at the SCAI.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica, 36*(3), 109-118.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal, 7*(4), 308-313.

Rao, P., & Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. *Journal of the American Statistical Association, 62*(317), 194-204.

Snyder, J. A. L. (2013). What Actually Wins Soccer Matches: Prediction of the 2011-2012 Premier League for Fun and Profit.


## 7.2 Online references


[1] Premierleague.com, “The world’s most watched league”, 2016. Website:

http://www.premierleague.com/en-gb/about/the-worlds-most-watched-league.html Date: 30.05.2016

[2] Dailymail.co.uk, "Tottenham star Keane is ridiculed for his quick Liverpool exit in cheeky train advert, Sportsmail Reporter, 2009. Website:

http://www.dailymail.co.uk/sport/football/article-1141982/Tottenham-star-Keane-ridiculed-Liverpool-exit-train-advert.html Date: 30.05.2016

[3] Premier League Handbook Season 2015/16, 2015. Website:

http://www.premierleague.com/content/dam/premierleague/site-content/News/publications/handbooks/premier-league-handbook-2015-16.pdf Date: 30.05.2016

[4] "Not revealing how much football clubs pay in their transfer dealings is an insult to fans", telegraph.co.uk, Burt, Jason, 2015. Website:

http://www.telegraph.co.uk/sport/football/11378383/Not-revealing-how-much-football-clubs-pay-in-their-transfer-dealings-is-an-insult-to-fans.html Date: 30.05.2016

[5] Football-data.co.uk, 2016. Website:

http://www.football-data.co.uk/englandm.php Date: 27.11.2015

[6] Oddsportal.com, 2016. Website:

http://www.oddsportal.com/soccer/england/premier-league-2013-2014/results/#/page/8/ Date: 27.11.2015

[7] Premier League Wages: where does YOUR club rank in sport's salary table, mirror.co.uk, Richards, Alex, 2015. Website:

http://www.mirror.co.uk/sport/football/news/premier-league-wages-your-club-5729730 Date: 20.02.2016

[8] Sportingintelligence.com, 2016. Website:

http://www.sportingintelligence.com/ Date: 30.05.2016

[9] "How to write the first for loop in R", Theuwissen, Martijn. Website:

http://www.r-bloggers.com/how-to-write-the-first-for-loop-in-r/ Date: 20.02.2016

[10] "Doing Maximum Likelihood Estimation by Hand in R, White, J.M., 2010. Website:

http://www.r-bloggers.com/doing-maximum-likelihood-estimation-by-hand-in-r/ Date: 20.02.2016

[11] "Manchester United wage bill higher than Manchester City and Chelsea as what every Premier League club pays their players is revealed", Rice, S., 2015. Website:

http://www.independent.co.uk/sport/football/premier-league/revealed-manchester-united-wage-bill-is-10-times-more-than-burnley-and-higher-than-manchester-city-10193532.html Date: 30.05.2016

[12] "Premier League finances: the full club-by-club breakdown and verdict, Conn, D., 2016. Website:

https://www.theguardian.com/football/2016/may/25/premier-league-finances-club-by-club-breakdown-david-conn Date: 30.05.2016

# 8 Appendix

## Table 8.1

Results of fictive betting, using the 10% value threshold with salaries on linear form.

| 10% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 23 | 10 | 13 | **46** |
| Bets won | 39 | 5 | 21 | **65** |
| Total bets | 62 | 15 | 34 | **111** |
| Results | 466 | 383 | 1251 | **2100** |
| Average profit per bet | | | | **18,9189189** |
| ROI per bet | | | | 118,92 % |
| ** 7 cross-bets home win/draw | | | | |
| ** 2 cross-bets away win/draw | | | | |

## Table 8.2

Results of fictive betting, using the 12% value threshold with salaries on linear form.

| 12% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 19 | 8 | 7 | **34** |
| Bets won | 30 | 3 | 18 | **51** |
| Total bets | 49 | 11 | 25 | **85** |
| Results | 319 | 57 | 1191 | **1567** |
| Average profit per bet | | | | **18,4352941** |
| ROI per bet | | | | 118,44 % |
| ** 5 cross-bets home win/draw | | | | |
| ** 1 cross-bets away win/draw | | | | |

Table 8.3

Results of fictive betting, using the 15% value threshold with salaries on linear form.

| 15% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 13 | 5 | 5 | **23** |
| Bets won | 24 | 0 | 13 | **37** |
| Total bets | 37 | 5 | 18 | **60** |
| Results | 383 | -500 | 794 | **677** |
| Average profit per bet | | | | **11,2833333** |
| ROI per bet | | | | 111,28 % |
| ** 3 cross-bets home win/draw | | | | |
| ** 1 cross-bets away win/draw | | | | |

Table 8.3

Results of fictive betting, using the 15% value threshold with salaries on linear form.

| 20% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 6 | 1 | 3 | **10** |
| Bets won | 9 | 0 | 5 | **14** |
| Total bets | 15 | 1 | 8 | **24** |
| Results | 41 | -100 | 154 | **95** |
| Average profit per bet | | | | **3,95833333** |
| ROI per bet | | | | 103,96 % |

Table 8.4

Results of fictive betting, using the 10% value threshold, taking the logs of salaries.

| 10% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 24 | 12 | 16 | **52** |
| Bets won | 33 | 6 | 24 | **63** |
| Total bets | 57 | 18 | 40 | **115** |
| Results | 92 | 511 | 1393 | **1996** |
| Average profit per bet | | | | **17,3565217** |
| ROI per bet | | | | 117,36 % |
| ** 4 cross-bets home win/draw | | | | |
| ** 4 cross-bets away win/draw | | | | |

Table 8.5

Results of fictive betting, using the 12% value threshold, taking the logs of salaries.

| 12% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 15 | 7 | 11 | **33** |
| Bets won | 27 | 3 | 19 | **49** |
| Total bets | 42 | 10 | 30 | **82** |
| Results | 535 | 167 | 886 | **1588** |
| Average profit per bet | | | | **19,3658537** |
| ROI per bet | | | | 119,37 % |
| ** 2 cross-bets home win/draw | | | | |
| ** 1 cross-bets away win/draw | | | | |

Table 8.6

Results of fictive betting, using the 15% value threshold, taking the logs of salaries.

| 15% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 9 | 3 | 6 | **18** |
| Bets won | 17 | 2 | 10 | **29** |
| Total bets | 26 | 5 | 16 | **47** |
| Results | 402 | 251 | 313 | **966** |
| Average profit per bet | | | | **20,553191** |
| ROI per bet | | | | 120,55 % |
| ** 2 cross-bets home win/draw | | | | |
| ** 1 cross-bets away win/draw | | | | |

Table 8.5

Results of fictive betting, using the 20% value threshold, taking the logs of salaries.

| 20% VT | Results | | | |
|---|---|---|---|---|
| | Home | Draw | Away | Totals |
| Bets lost | 4 | 1 | 3 | **8** |
| Bets won | 6 | 0 | 4 | **10** |
| Total bets | 10 | 1 | 7 | **18** |
| Results | 8 | -100 | 60 | **-32** |
| Average profit per bet | | | | **-1,77777778** |
| ROI per bet | | | | 98,22 % |