Methods for the Identification and Optimal Exploitation of Profitable Betting Scenarios

Tom Flowerdew, M.Math, M.Res

Submitted for the degree of Doctor of Philosophy at

Lancaster University.

October 2015

Abstract

This thesis tackles the issue of how gamblers can profit from betting on the outcome of sporting events. In particular, it addresses issues which have arisen in recent years concerning both the inception of betting exchanges, and the technique of building complex statistical models to accurately predict the sporting outcomes.

This thesis shows that bias in predictive models can be quantified from a collection of model outputs. It is shown that a Bayesian method can be constructed to derive accurate bias estimates, even when the model outputs are merely a collection of independent Bernoulli trials. In addition, the method is expanded, to allow the quantification of a time-varying bias, as long as it changes in a known, deterministic setting. The utility of this method is demonstrated via the correction of a simple football prediction model.

The movements seen in betting markets before the event in question occurs are investigated. It is conjectured that the rate of increase of the amount of capital invested in the betting market is central to understanding other market movements. With this in mind, two approaches are derived, which both use a collection of historic market movements for past events for their predictions. It is shown that in many cases, some mix of the two approaches achieves the most accurate forecasts.

A new gambling strategy, dubbed *consolidated wagering* is introduced. It is demonstrated that consolidated wagering outperforms all other candidate methods when considering string bets (multiple bets on the same event, at different odds). The application of these methods to investing in restricted markets in betting exchanges is demonstrated. Finally, the problem of string wagers under uncertainty is explored.

Acknowledgements

First and foremost, thank you to the EPSRC, and to ATASS Sports, whose funding allowed this research to occur.

Thanks to my supervisors: Jon Tawn, Chris Kirkbride and Kevin Glazebrook who helped and supported me throughout this project. Their time, advice and vast knowledge have been invaluable. Special thanks to Chris Sherlock for the excellent advice which led to a major breakthrough in Chapter 4.

Thank you to the class of '11-'15, in classic alphabetical order: Ben, Dave, Emma, Hugo, Jeddy, Judd, Rhian & Rob. My PhD experience would have not have been nearly as calm, interesting or hilarious without each and every one of you. Thanks and commiserations to others who have shared an office with me over the years: Tim, Shreena, Andy and Jamie. Thanks to Matt Powney for his help in preparing for the viva.

Thanks to my family for their continued support of all of my endeavours. Finally, thanks to Sarah for her love and support, and her genuine and expertly feigned interest in the technicalities of my work frustrations over the years.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Thomas Flowerdew

Contents

Chapter 1

Introduction

> "*Bond didn't defend the practice. He simply maintained that the more effort and ingenuity you put into gambling, the more you took out.*" - Ian Fleming, Casino Royale

Gambling is one of the world's oldest recorded activities, with evidence dating back to around 3000BCE of dice found in Mesopotamia (Schwartz, 2006), and Egyptian gods gambling against the moon for the number of days in a year (Wykes, 1964). In addition, gambling on the outcome of sporting events seems to have been around for as long as the sporting events themselves (see the example of cattle racing in Wykes (1964)).

The history of the mathematical concept of odds, and therefore the history of the mathematical treatment of gambling problems is intrinsically linked to the development of probability theory. Most famously, this is attributed to the 17th Century mathematicians Blaise Pascal and Pierre de Fermat. However, the mathematical treatment of odds can be traced back to Gerolamo Cardano in the 16th Century (Cardano, 1663).

This work shall focus primarily on betting markets in the UK, where gambling on sports is worth around £7.1 billion per year (Wardle and Moody, 2014). The route by which the gambling industry has arrived at this turnover has had many twists and turns, with the industry itself evolving over time as a response to customer preferences, and the proliferation of new technologies.

The betting industry remained mainly unchanged for a number of years, with the large majority of the market being made up of companies offering fixed odds, both 'on-course', by race courses etc, or 'off-course', in betting shops away from sporting venues. The explosion in accessibility to the internet and computing power has irrevocably altered both the way bookmakers operate, and the way investors develop winning strategies.

With the advent of the internet age, a large proportion of betting activity has emigrated online. All major bookmakers now maintain a large internet presence. In 2007, around 20% of UK betting activity was conducted online (Wardle, 2007). In 2014, the online betting industry was valued at £650 million, with 44% of betting activity conducted online Chalabi (2014). In addition, new ways to gamble have been invented, most notably betting exchanges, which allow investors to bet on both an event occurring, and not occurring. In addition, in-play betting allows investors to place bets whilst the event of interest is occurring (a full history of these innovations is available at O'Connor (2015)).

As the betting industry has developed on the back of the explosion in computing power, so has the sophistication of the investor. For most of the 20th century, gaining an edge in placing bets often relied on illegal means, such as the bribery of players (most famously in the 1919 Baseball World Series, see Asinof (2011)), or by learning of the result before others (see Poundstone (2010)). From the late 20th Century onwards, however, statistical methods could use computing power to gain inference about the likelihood of certain events occurring with considerable accuracy (see Section 3.2).

These evolutions have left gaps in understanding in their wake. The advent of betting exchanges has made wagering become more and more akin to playing the stock market, as betting for and against events becomes reminiscent to call and put options (see Benninga (2008)), but with the crucial difference being that the value of positions in betting markets are realised explicitly on the event's conclusion. Similarly, the construction of sophisticated prediction models has raised questions about the best way of exploiting their results.

These gaps in understanding have driven the selection of research topics in this thesis. These topics aim to address questions which has arisen due to the proliferation of betting exchanges, as well as developing methods to complement the use of complex statistical models for the predictions of sports events.

1.1 Thesis Outline

Chapter 2 gives a general introduction to betting markets. This introduction includes a primer on different types of betting markets, as well as methods of setting odds. It also introduces a number of important betting market features, such as overround, efficiency and liquidity.

Chapter 3 discusses different strategies used to profit from sports betting markets. This is divided into two sections. Firstly, techniques which have been developed to predict the outcome of sporting events will be summarised, mostly focussing on football. Secondly, the problem of selecting the correct wager size is considered. This literature is centred on the famed *Kelly Criterion*, which is discussed in depth, along with its extensions in recent years.

Chapter 4 considers the setting of a complex statistical model being used to predict the outcome of a sporting event of interest. The work tackles the problem of quantifying any potential consistent model bias, which may be affecting the probability predictions. The work first applies a Bayesian structure, and demonstrates an MCMC (Monte Carlo Markov Chain) approach to learning about the parameters which form the bias. Later on, the bias is considered to be time-varying, with the MCMC scheme being updated to tackle the additional complexity originating from a more involved parameterisation.

Chapter 5 considers how features of betting markets, such as odds and market size, change in the time before the event occurs. It is demonstrated that the change in market size is not only an important factor for understanding movements in other features, but that its movements demonstrate identifiable properties. This allows a predictive method to be derived, which is successful in simulating occurrences of these properties, allowing predictions to be made. Along with a simpler method, which identifies similar market movements from the past data, a mixed approach is shown to outperform either method in isolation, and to produce accurate forecasts.

Chapter 6 introduces the notion of a string bet. The decisions regarding optimal wager allocation in string bets suffer from downsides, as demonstrated in Proebsting's paradox. The impact of the paradox will be discussed, and as a result, a new type of betting, namely consolidated wagering is developed. Various betting strategies are applied to the related problem of investing in restricted markets in betting exchanges. Finally, these strategies will be re-derived to take into account potential uncertainty around the probability estimate of the event of interest.

Chapter 7 concludes the thesis with a discussion of its contribution, along with ideas regarding potential further development for each of the methods derived.

Chapter 2

Betting Markets

Betting markets can be thought of as places where agreements, or bets can be formed between two parties, which are settled upon the outcome of some future event. Traditionally, one of these parties is a *bookmaker*, who offers the public a range of bets, each with associated *odds*, which indicate what multiple of the stake is returned when a bet is won against the bookmaker. It is understood that an unsuccessful bet results in all wagered money being lost.

In the United Kingdom, odds are most commonly written in their *fractional* form. Fractional odds represent the multiplier given to any bet's winnings, and are usually written as a fraction $o = a/b$. For example, the fractional odds given for Team A to beat Team B could be 3/1; for every £1 wagered, a bettor would win £3 if the event occurs, and would also receive their initial stake back, effectively turning their capital from £1 to £4.

The other commonly-used form are *decimal* odds. Decimal odds give the multiplier given to the initial amount wagered, to give the total amount received after the event.

For the event used as an example before, the decimal odds would be 4.00 (two decimal places are used at all times for decimal odds). Unless otherwise mentioned, this thesis will always use fractional odds as standard.

Of course, odds reflect in some way the probability of an event occurring. Given that an event has a probability, p the 'fair odds', i.e. the odds offered such that the expected return of a wager is the size of the stake is:

$$o = \frac{1-p}{p}. \tag{2.0.1}$$

2.1 Types of Bet

For the remainder of this thesis, wagers will be considered to be placed on sporting events. In particular, association football will be used as a common betting example, and makes up around 15% of the U.K. bookmaker's betting activity (horse racing taking up around 50% of the whole market (Wardle and Moody, 2014)).

Bookmakers now offer a huge selection of sports games, and a large number of events within games to bet upon. The most common sporting outcome is simply the result of a single match, be it a home win, away win, or a draw (in some sports). For football, other possible events for which bets are commonly placed are: the number of goals; the number of corners; the identity of the first goalscorer; the time of the first goal.

Bets do not need to be constrained to events occurring on the sports ground; bookmakers also contrive other more complex events to be wagered on. A selection of these are summarised below:

- **Accumulator Bet**: Involves making a number of selections on non-dependent events, with the bettor only winning if all of the events occur. The decimal odds of all of the events occurring is simply the product of the decimal odds of each of the individual events. Note that accumulator bets can only be placed on independent events.

- **Full Cover Bet**: Involves the betting on an accumulator bet, but also on every possible subset of the accumulator of size 2 or more, of the events which made up the accumulator. For example, if betting on 3 horses in different races, $\{A, B, C\}$, the full cover bet (known as a *Trixie*, for accumulator size 3) is to bet some fixed amount on each of $\{A, B\}, \{A, C\}, \{B, C\}, \{A, B, C\}$. There is also an option to wager on each of the single bets, known as a Full Cover Bet with Singles. In this example, this would result in 7 separate wagers (known as a *Patent* bet).

- **Asian Handicap**: A bet in which the stronger team (or more-likely event) is handicapped against the weaker team (or less-likely event). In the case of a football match, the weaker team would be credited with a number of bonus goals. Say Team A were at home to the much weaker Team B. In an Asian Handicap Market, Team B would be credited with, say, 2.5 goals. As long as Team B loses by 2 goals or fewer, bets on Team B will be denoted as the winning wagers. Note that the possibility of a draw is eliminated, as handicaps are never integers.

2.2 Types of Betting Market

By far the most common type of betting market created by a legally accredited bookmaker. In 2013/2014, bookmakers made up around 90% of the U.K.'s betting industry (based on turnover), with the betting industry itself making up around 65% of the U.K.'s gambling industry (which also includes casinos and lotteries). Of this, betting pools are the second largest betting market, followed by betting exchanges. Within each of these sectors, football consists of around 40% of the non-horse-related betting activity, with tennis the next most popular, at around 20% of the market (all data from (Wardle and Moody, 2014) and (Wardle, 2010), see Figure 2.2.1).

	Betting Activity (%)		
Market	Bookmakers Shop	Online Bookmakers	Betting Exchange
Horse Races	78	7	5
Dog Races	75	6	2
Sports Betting	64	24	10
Non-Sports Betting	73	9	5

Table 2.2.1: The mode of participation in selected betting activities, adapted from (Wardle, 2010).

Figure 2.2.1 shows that betting exchanges only account for a very small proportion of horse racing (5%) and dog racing (2%). However, for other sports (such as football, for which the majority of this thesis is focussed), around 10% of people reported to have used a betting exchange.

Bookmakers

Bookmakers offer odds on a large range of events. Typically, the bookmakers' aim is to set odds such that each possible event outcome results in the bookmaker making a profit, a process known as ‘forming a book’. This process is studied in detail in Boyle (2006). A bookmaker’s profit is achieved primarily through the presence of an *overround*. The overround is also a measure of how much the odds undervalue the probability of events occurring.

As an example, say a bookmaker has offered fractional odds on the result of a football match. The home win is available at 1/2, the draw at 5/2 and the away win at 5/1. From equation (2.0.1), fractional odds are ‘fair’ when $o = p^{-1}(1-p)$, so the implied probability of an event, inferred from the odds are $p = o(1+o)^{-1}$. The implied probability of these events are therefore 0.666, 0.286 and 0.166, respectively. The sum of these probabilities is 1.119, giving an overround for this event as 11.9%. This means that the bookmaker expects to pay out £100 for every £119 taken.

A consideration which should be taken into account for professional investors is the maximum allowable bet size. This can vary from a bet limit of £10,000 (adopted by Ladbrokes, William Hill and others), to a maximum return of around £500,000 (adopted by Totesport and Betfred) (Punter, 2014).

Betting Exchanges

Betting exchanges offer customers the opportunity to gamble on both sides of the market: by placing bets on events occurring; and to offer odds on events occurring (*laying* a bet). Customers can therefore also bet on events *not* happening.

The process of laying an event (creating a betting option) involves both specifying the odds and setting the maximum acceptable loss to be incurred (equivalent to choosing the betting amount). Wagers on this option are only accepted until the potential losses meet the maximum acceptable loss of all the other investors laying at these odds, in which case the option is closed; no further bets are taken. For example, suppose an investor offered odds of 2.00 for Chelsea to win a match at home, and is willing to accept £100 of losses. Once £50 have been matched from people willing to back Chelsea to win at home, the betting opportunity created by the investor is closed, and the investor's maximum losses are set at £100.

If an investor wants to bet at certain odds, then, they would be limited by the total amount of layed capital offered by other exchange users. This is often much smaller than that seen in bookmakers, especially if they are betting a long time before the event occurs, or if the betting event itself is not very popular.

As the betting exchange provider does not take positions in any of the wagers, their profit comes from taking a small commission of any profit made. For the largest betting exchange provider in the UK, BetFair, the commission is around 5%. It is often found that odds found in exchanges are more representative of the true probability of events than those found in bookmakers (Franck et al., 2010). Similarly, betting exchanges tend to offer investors better odds, even after commission has been factored in; see Tsirimpas (2015).

Pools

Pool betting (also called paramutuel betting, or tote betting) is a simpler form of market, most commonly seen in events for which there are a large number of potential outcomes, such as a number of different runners in a horse race. The critical difference between pools and other types of markets is that odds are not offered as the bets are made. Instead, all of the betted money is pooled together, then after the wagering company has taken their cut of the pot, the rest of the pool is divided out amongst the winning parties.

As an example, assume there are 6 horses in a certain race, and the amount of money bet on each of them is shown in Table 2.2.2:

Horse	1	2	3	4	5	6
Staked Money	£5	£10	£8	£20	£15	£5

Table 2.2.2: An example of amounts of money staked on a pools market for a horse race.

In this case, there is a total of £63 in the pot. Say the wagering company takes a cut of around 11%, this would leave a pot for the winners of around £56. Say horse 3 wins, and that the £8 staked was from two bettors, one staking £6 and the other £2. The entire remaining pot is split proportionally between the two, with the first bettor receiving £42 and the second £14, each realising fractional odds of 6/1.

Pool markets do not lend themselves easily to the wider domain of gambling theory, as the odds are not known as bets are placed, rather they are decided by the (mostly) unobserved behaviour of other investors. For this reason, betting in pool markets is not considered further in this work; for a full treatment of the topic, see Stefani (1983).

Spread Betting

Spread betting currently has a small share of the total betting market. Its origins lie in the IG Group's creation of an index for the price of gold in 1974, which allowed investors to speculate on gold's price movements, without requiring actual ownership of the commodity.

The wagering company chooses a 'buy' and 'sell' price for an index, much like the 'bid' and 'ask' prices in financial markets. Investors choose to either buy the index at the buy price, or to sell at the sell price. The eventual result of the event dictates their profit or loss.

As an example, suppose the market was the number of runs the English cricket team score in an innings, with a buy price of 300 runs, and a sell price of 280 runs. If an bettor invests £2 at the buy price, then every run over 300 that England score gives the bettor a profit of £2, so England scoring 325 runs would net the bettor a profit of £50. Conversely, if England scored only 250 runs, the bettor would lose £100, much more than the size of the £2 stake.

On the other hand, if the bettor had chosen to invest £2 at the sell price, then England scoring 325 would lose the bettor £90, and a score of 250 would give a profit of £60.

The first academic treatment of the subject within the domain of sports betting was in Jackson (1994), but it is only with the advent of online betting that interest has significantly grown, with sites such as `sportingindex.com` and `spreadex.com` allowing investors to bet easily on these markets. This work will not consider spread betting markets directly; an exploration of the topic within the context of sports betting can be found in Haigh (2000).

2.2.1 Overround

The use of a bookmaker's overround has already been briefly mentioned in Section 2.2. Put simply, bookmakers set their odds such that the sum of the implied probabilities (see equation (2.0.1)) from each of the possible event outcomes sum to a number larger than 1.

In many situations, (such as comparing a model's prediction of some event probability to that implied by a bookmaker's odds) it is useful to utilise the bookmaker's odds as an indication of the true probabilities of the events. The most obvious method would be to reverse equation (2.0.1) to make the odds the subject. This, however, would necessarily lead to biased probability estimates, as the odds have been transformed away from their true value.

The aim of making the inferred probabilities unbiased, then, is for them to sum to 1. The most obvious way of achieving this objective, and the most common way of removing bookmaker's overround is to scale each probability linearly. Let $p_1^b, \cdots, p_n^b$ be the biased probabilities of n possible event outcomes. The scaling then takes the form

$$p_i^u = \frac{p_i^b}{\sum_{i=1}^{n} p_i^b}$$

where p_i^u is the supposedly unbiased form. This solution to the problem of removing the overround has debatable merit. By using this method, each of the probabilities are shrunk by the same factor; therefore, the change in probability for events with long odds is much smaller than those with short odds. It is shown in Vovk and Zhdanov (2009), that another overround-removal method performs much better (in terms of achieving a superior Brier score, a measure of calibration of probabilities) and is based on predictions taken from the resultant probability estimates. This method gives

$$p_i^u = (p_i^b)^{-\gamma}, \qquad i = 1, \cdots, n \tag{2.2.1}$$

where γ is chosen such that $\sum_{i=1}^{n} p_i^u = 1$.

This chapter has given a general overview of betting markets; how they are made and how investor's bet upon the possible outputs of sporting events. Next, it shall be shown how investors find strategies with the intention to allow their wealth to increase in value over time.

Chapter 3

Strategies

The strategies considered in this chapter have the same central aim: to create profitable strategies for betting on sports markets. These strategies can be split into two themes. Firstly, an investor needs to know what to bet on. This is commonly achieved through statistical means; modelling the sports in question to the extent that accurate forecasts can be made regarding the probability of events occurring.

Secondly, an investor must know how to best utilise the information gleaned from their statistical models. This area of study relating to betting strategies, most commonly revolves around the optimal selection of stake size, given that a particular betting market is deemed to be profitable to the investor.

3.1 Prediction

In general, both the bookmaker and the investor try to accurately predict the probability that the events of interest occur. For the investor, predicting these event probabilities accurately allows them to only wager money on those events for which the bookmakers' offered odds are favourable. The assumption is that some of the bookmakers' odds will be favourable to the investors, which would mean that their odds are not representative of the true probability of events.

3.1.1 Market Efficiency

Efficient markets are those whose prices reflect the information available to all participants. As stated in Fama (1970), an efficient market is "one in which prices always fully reflect available information". The definition of quite what "available information" entails categorises market efficiency into three forms.

- **Weak Efficiency:** Information considered is only the past prices of the market in question.
- **Semi-Weak Efficiency:** Information considered is all publicly available information.
- **Strong Efficiency:** All information is considered, including that held only by small groups of investors, e.g. insider trading.

From this definition, the famed "Efficient Market Hypothesis" can be formed, which states that all financial markets must be efficient, as information is quickly and accurately incorporated into pricing. Applying this notion to betting markets would give the hypothesis that the betting odds always fully reflect the information available to investors.

If this were true, then the creation of complex betting strategies would be pointless. This is because the odds would always be correct, in that they would be accurate at their time of offering, and thus no profit could be made in the long run by betting on such markets. Note that in the context of sports betting, a market could be shown to be inefficient if an investor makes a loss, but still achieves a better return than that implied by the overround.

There have been many papers which have investigated the question "are sports betting markets efficient?". In Kuypers (2000), many of these papers are summarised, focussing on whether sports betting markets can be shown to be efficient in any of the 3 forms stated above. These papers tend to focus on horse racing, most notably Asch et al. (1984) and Snyder (1978) who both show that although US horse racing paramutual markets do not exhibit even weak efficiency, they could not find betting strategies which yielded positive returns. In comparison Ali (1977) and Hausch et al. (1981) both found profitable betting strategies on the same markets (although in the latter case this was a fixed-odds market). The model given in Hausch et al. (1981) was later shown to also produce positive returns in UK horse racing (Ziemba and Hausch, 2008).

Away from horse racing, the investigation of the efficiency of other sports markets has also been an active area of research. In Goddard and Asimakopoulos (2004), it is shown that over a year of football, the market exhibited weak efficiency, which still yielded profitable betting strategies when exogenous variables were added to predictive models. Tennis markets are investigated in Forrest and McHale (2007), primarily in the context of the favourite longshot bias; a common feature of markets where inefficiency seems to affect very likely or unlikely events more often than others and where some investors (usually those without statistical models) tend to undervalue events with short odds and overvalue events with long odds.

The presence of favourite longshot bias has also been noted in baseball (Woodland and Woodland, 1994), football (Cain et al., 2000), horse racing (Sob) and others. In Shin (1991), it is proposed that the favourite longshot bias is a result of bookmakers adjusting their odds to mitigate against insider trading, noting in addition that both this bias and bookmaker profits increase with the number of competitors for each event. These claims are supported empirically in Cain et al. (2003).

In comparison to traditional bookmakers, Smith et al. (2006) claims that betting exchanges exhibit both weak and strong efficiency. Finally, Williams and Paton (1997) propose that the favourite longshot bias was more pronounced in markets with less liquidity, i.e. more bias exists in less popular markets. This particular result is conjectured to be a result of smaller markets having a proportionally larger number of casual bettors (Sob).

In summary, there is evidence to support the claim that sports betting markets do not consistently demonstrate efficiency. With the common presence of favourite longshot bias, the odds offered by bookmakers do not necessarily represent the true probability of the events. These features therefore must point towards the existence of profitable betting strategies for all sports. In the following section, some common sports predictive models will be summarised, each with evidence to attest to their profitable implementation.

3.1.2 Football Prediction

The underlying feature of the vast majority of the academic literature regarding the modelling of football matches is to consider goals scored as being driven by some arrival rate. This was first proposed by Moroney (1956), who drew inspiration from the classic Horse Kicks dataset (von Bortkiewicz, 1898). Moroney, as well as Reep and Benjamin (1968), discusses certain shortcomings of using a homogenous Poisson process to model this arrival stream and also considers a Negative Binomial model. At this time, the arrival rate of goals for each team was judged to be independent of other factors, most notably the strength of the opposing team.

This discrepancy was rectified in Maher (1982), who not only attempted to account for the dependence between the strength of the two opposing teams, but also introduced a parameter used to represent *home advantage*, which accounts for the home team performing more strongly than the away team, on average. Home advantage is a very well-studied phenomenon in football (see Pollard (1986), Boyko et al. (2007) and Pollard (2008)), and is proposed to originate from such sources as: travel fatigue for the away team; home crowds affecting refereeing decisions; and familiarity with the home pitch, amongst others.

The model proposed by Maher (1982), then, can be understood via the two equations:

$$\lambda = \kappa \alpha_i \beta_j,$$

$$\mu = \alpha_j \beta_i.$$

for $\kappa > 0$. Here, λ and μ are the arrival rates for the home and away team, respectively. Each of these is composed of the attack strengths α_i and α_j as well as the defence strengths β_i and β_j for the home and away team respectively. In addition, κ represents the aforementioned home advantage parameter. The arrival rate of goals for the home team, then, is the product of the home team's attack strength, the away team's defensive strength, and the advantage gained by the home team.

Although this model allowed for the strength of the opposing teams to be considered, the direct dependence between the goals scored by the home and away teams is still not accounted for. This was rectified in Dixon and Coles (1997), who used past data to empirically fit a correlation function for low scoring matches. In addition, a new method for estimating the parameters of interest is introduced; a pseudo-likelihood function downweights observations more the further in the past they take place.

Although many years have passed since the publication of Dixon and Coles (1997), the literature regarding the predictions of the outcome of football matches has not progressed to any great degree. It is supposed that advancements and innovations in the field have a substantial commercial value, and thus are not released to the public.

One of the more active areas of research in football prediction, then, concerns in-play betting; see Dixon and Robinson (1998) and Höög (2014). This focusses on the prediction of certain events whilst the match is in progress, such as the next goalscorer or even the time of the next throw-in. For the example of an in-play market for the correct full-time score, the probability of each possible score is updated as game events occur, and as time passes.

In addition, there has been a lot of attention regarding the origins and manifestation of home advantage. The review paper Pollard (2008) gives examples of a vast variety of different factors cited as being in some way beneficial to the home team. These include the impact of the crowd, potential referee bias, travel fatigue for the away team and familiarity with surroundings. Finally, many other factors are hypothesised to affect the outcome of football matches, such as the impact of a red card (Ridder et al., 1994) or yellow card (Titman et al., 2015), and of artificial pitches (Barnett and Hilditch, 1993).

3.2 Betting

This history of betting theory is as varied as it is entertaining. The 2010 book *Fortune's Formula* (Poundstone, 2010) gives an excellent background to all of the material covered in this section, as well as the stories behind their development.

3.2.1 Classic Papers and Results

The earliest major academic paper on the subject of betting can be attributed to Daniel Bernoulli, (reprinted in English in Bernoulli (1954)), drawing on his letters to Nicolas Bernoulli on the subject of the St Petersburg Game (Bernoulli, 1713).

Imagine some casino offered a game where a gambler pays some entrance fee to play. The casino then gives the investor £1, with the promise that every time a fair coin is flipped heads, the investor's current money will be doubled, whereas a tails will end the game, and the investor will keep any money made to that point. As an example, if the coin were to yield three heads followed by a tails, the investor will win £8, whereas an immediate tails would cause the investor to leave the game with only £1.

The St. Petersburg Paradox is a consequence of asking the question “How much should the casino charge an investor to play this game?”. An intuitive way to calculate a sensible answer to this question would be to find the expected return for the gambler:

$$\mathbb{E}(\text{Return}) = \sum_{i=0}^{\infty} 2^i \frac{1}{2^{i+1}} = \sum_{i=1}^{\infty} \frac{1}{2} \to \infty$$

which shows the expected return from the game is infinite. This seems illogical, implying that an investor should play this game, even given an entrance fee of £1000, or even £1 million!

This led Bernoulli to reason that the *utility* of wealth should be an important factor in investment problems. In a hypothetical simple game, such that a gambler won either £10,000 or nothing on a coin flip, a poor person may sell the opportunity to play this game for £3,000, whilst a rich person would happily play if it cost them £4,500. As it is put in Bernoulli (1954),

> “*The determination of the value of an item must not be based on the price, but rather on the utility it yields. There is no doubt that a gain of one thousand ducats is more significant to the pauper than to a rich man though both gain the same amount.*”

Another less satisfying resolution to the paradox is to realise that the casino could not possibly pay-out the winnings when the number of heads flipped increases indefinitely. If the winnings of the player is capped at, say, £10 million, then the expected value of the game decreases dramatically to just £13. This notion, as well as others relating to the use of different increasing and concave utilities are discussed in Samuelson (1977).

Bernoulli's use of a logarithmic utility for wealth when considering betting problems inspired John Kelly's famous 'Kelly criterion' (coined in the blackjack paper, Thorp (1966)) for the selection of bet size in more general investment problems. Kelly worked at Bell Labs at the same time as 'The Father of Information Theory', Claude Shannon, and his methods are strongly influenced by the field.

In Kelly (1956), it is argued that the quantity which investor should seek to maximise is their long-term logarithmic growth rate

$$G = \lim_{n \to \infty} \frac{1}{n} \log \frac{W_n}{W_0} \tag{3.2.1}$$

where W_0 is the initial wealth and W_n is the wealth after n investments have been realised. The key notion is to move away from the idea that investment sizes should be some fixed amount, £20, but instead should be related to the current bankroll of the investor. For this reason, Kelly considers a fractional wager f, of the investor's current wealth.

Imagine, then, some repeated game with probability of occurrence p and fractional odds o. The log of the return of a single wager can be written as

$$p \log(1 + fo) + (1 - p) \log(1 - f)$$

where the investor's current wealth has been normalised to 1. The optimal betting fraction, f^* can be found simply by finding the unique turning point

$$f^* = \frac{p(o+1)-1}{o} = p - \frac{1-p}{o}, \tag{3.2.2}$$

the aforementioned Kelly fraction (or Kelly criterion), which also thus maximised the long-term growth rate shown in equation (3.2.1). Note that the use of the log utility here can be seen as a natural measure, given that repeated fractional bets lead to compounded returns.

In terms of the efficacy of this criterion in real-life use, Kelly crucially demonstrated that not only was this staking strategy optimal asymptotically, but also myopically. This means that if the sequence of opportunities are not identical, investing the Kelly criterion each time still yields the optimal growth for the set of opportunities as a whole.

It should be noted that Kelly's method was discovered independently, and around the same time by Latane (1959) who, unlike Kelly, approached the problem from the viewpoint of an investor. A much larger contribution was made by Louis Breiman, whose papers, Breiman (1960) and Breiman et al. (1961) proved three principles which give strong support for the use of the Kelly criterion for investment decisions.

1. Let two investors face an identical sequence of investment opportunities. Not only will the investor who utilises the Kelly Criterion guarantee a larger return than their competitor as the number of opportunities diverges, but the amount by which they beat their competitor diverges as well. Note that the competitor's

strategy must be *essentially different*; for more details see Breiman et al. (1961).

2. Given some fixed wealth goal, investing using the Kelly criterion asymptotically minimises the time needed for the goal to be reached, as the goal increases.

3. Given a fixed set of opportunities, the strategy which maximises the growth rate of the whole set is independent of the size of the set (a reworking of Kelly's myopic result).

These results, while powerful, are laden with a number of unrealistic assumptions, namely that the returns are independent, identically distributed random vectors and that the investor has an arbitrarily large amount of time to see their investments mature. The first of these assumptions was tackled first by Finkelstein and Whitley (1981) and then by the publication of Algoet and Cover (1988), which extends the three results above such that there is no restriction on the market processes. The short-run properties of the Kelly criterion are exhibited in Bell and Cover (1980), and then with a more general class of utility functions in Bell and Cover (1988).

3.2.2 Practical Use of the Kelly Criterion

A common, general class of utility functions are the power (or isoelastic) utility functions, which take the form

$$u(c) = \begin{cases} \frac{c^{1-\eta}-1}{1-\eta}, & \eta \neq 1 \\ \ln(c), & \eta = 1. \end{cases} \tag{3.2.3}$$

It can be seen, then, that the Kelly criterion would be a specific example of an entire family of utility functions, when the risk aversion parameter, η is set to 1. For convenience, this family of utilities is often written as $u(c) = c^{-\eta}$, a form used in Hakansson (1970) to develop a number of optimal strategies for a variety of investment scenarios.

The use of the power utilities is supported by the fact that they are unique in having a constant relative risk aversion. This means that as wealth increases, the fractional stake in some risky opportunity remains the same, a feature seen in the fractional betting of Kelly. For a more theoretical summary of these features, see Menezes and Hanson (1970), Arrow (1971) and Pratt (1964).

The rationale for choosing a risk aversion greater than 1 is that an investor may wish to bet less than the amount that Kelly recommends. Given the desirable properties of betting using a log utility which have already been mentioned, why would one not bet the Kelly amount? This is the subject of a number of papers, exploring *fractional Kelly* strategies, which promote the use of some scaled-down version of the 'full' Kelly strategy.

In order to understand fractional Kelly's origins, a good place to start is MacLean et al. (1992), which focusses on the minimisation of risk, taking inspiration from Ferguson (1965) who attempts to mitigate the worst-cases. In MacLean et al. (1992), graphs such as that reproduced in Figure 3.2.1 are considered.

The growth rate of a single wager shows its concave, and nearly symmetric shape. The dotted line shows the wager size which maximises this growth rate. The most important feature of this growth rate curve is that there is a clear discordancy between

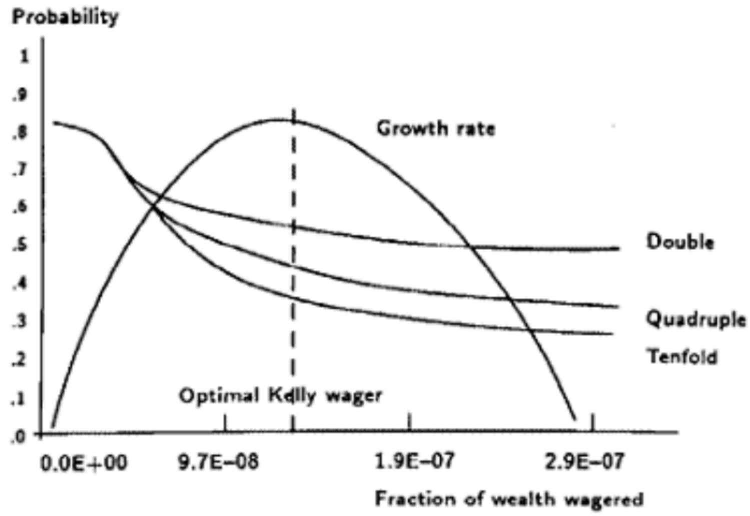


Figure 3.2.1: The probability of doubling, quadrupling and multiplying wealth by ten, before losing half, for a lottery game. Taken from MacLean et al. (1992).

betting more than, or less than this optimal value. By betting a larger fraction of wealth, the investor puts themselves in a position of greater risk, but achieving a lower growth rate. Such a bet is therefore dominated by the optimal bet size. In comparison, investing a smaller fraction of wealth still achieves a smaller growth rate, but now under less risk.

This feature is made clear in Figure 3.2.1, where the probability of achieving certain wealth goals *before wealth drops below some unacceptable value* is shown to be a decreasing function of the fraction wagered. Therefore, the Kelly-optimal strategy can be seen as the most risky of the fractional bets which are not dominated. This feature is explored in depth in MacLean et al. (2010a), which compares investments made using a Kelly strategy against a selection of fractional Kelly methods over realistic investment durations. It shows that the Kelly strategy is indeed very risky, where a sequence of profitable betting opportunities can still lead to large losses. In addition, it is demonstrated in Thorp (2010) that the set of fractional Kelly wagers are equivalent to the efficient frontier of Markowitz-type portfolios (see Markowitz (1952), Markowitz (1968), Thorp (1969), and others). This shows that the choice of a fractional Kelly stake is simply an investor repositioning their attitude towards risk and reward, similar to the choice of η in the power utility family.

In addition to this, Maclean et al. (2010b) gives a useful summary of the good and bad properties of the use of the full Kelly criterion. One interesting feature is that it is possible to actively adjust the choice of fractional wager in order to maximise the probability of achieving a certain growth path (see MacLean et al. (2004) and MacLean et al. (2009)). Possible 'bad points' mostly concern the time the Kelly criterion takes to perform better than other strategies, given that its optimality is proved via asymptotic results. In addition, it is noted that Kelly betting can result in very high stakes, when both the odds and the probability of the event are high (although the upper limit for the size of a stake using the Kelly criterion is the probability of the event itself, see equation (3.2.2)).

Even given the multitude of papers already mentioned concerning the positives for using the Kelly criterion for driving investment decisions, there are those who doubt the wisdom of its use. Much of the discussion regarding the use of growth-maximisation techniques have been between the Nobel Prize winning economist Paul Samuelson and Ed Thorp, who popularised the use of Kelly's methods to a general audience (both inside and out of academia; see Thorp (1969), Thorp (1998) ,Thorp (2010) and Thorp (1966) for his groundbreaking analysis of blackjack and Samuelson (1969), Samuelson and Merton (1974) for Samuelson's contribution to the literature).

Most of the criticisms have been rebuffed, seemingly to the general acceptance of the academic community. However, the unique paper Samuelson (1979) sums up his criticisms with words of only one syllable. His main point is

> "*When you lose - and you sure can lose - with N large, you can lose real big. Q.E.D.*"

which echoes the concerns of MacLean, and others that one stands to potentially lose a significant proportion of wealth before achieving the optimal growth rate 'in the long-run' .

Despite this, strong evidence for the use of the Kelly criterion for investment decisions must come from whether or not it is used in real life. In MacLean et al. (2011), it is noted that many billionaire hedge fund and portfolio managers started off as blackjack players; a field which, thanks to Ed Thorp, is in general acceptance of Kelly-type strategies. In addition, Warren Buffet is noted as often being a user of Kelly betting, and is quoted as saying

> "*I have 2 views on diversification. If you are a professional and have confidence, then I would advocate lots of concentration. For everyone else, if it's not your game, participate in total diversification*"

(taken from Thorp (2010)) which further enhances the idea that fractional Kelly strategies and uncertainty go hand-in-hand. In addition, John Maynard Keynes is said to have used a fractional Kelly-type strategy to invest the Kings College Cambridge endowment fund.

3.2.3 Other Innovations

Along with the core of the literature covered in Sections 3.2.1 and 3.2.2, there are a number of other papers which have extended the theory regarding the Kelly Criterion in interesting directions.

There is a need to extend the Kelly criterion away from standard bookmaking and into new betting markets. In terms of betting exchanges, Noon et al. (2013) and Noon (2014) show how bets can be placed on exchanges, and also how markets can be created using similar principles. In addition, Zambrano (2014) discusses how betting in exchange-type markets while using growth-optimal approaches can lead to ruin. This idea is extended further in Chapter 5 in this thesis. Spread betting markets were first discussed in an academic context in Haigh (2000). After the publication of Fitt et al. (2006), which derived a method for valuing the current position of an open spread bet, much like options pricing in finance, Chapman (2006) extended the theory to give not only the optimal fractional wager, but also other familiar results from bookmaking, such as the probability of bankruptcy.

Another area which the core Kelly research did not consider is the notion of simultaneous events. Consider betting on English football matches at 3pm on a Saturday. Bookmakers would potentially offer odds on hundreds of simultaneous events. The first to tackle this problem was Whitrow (2007), who used a fast stochastic approximation technique to find the optimal allocation of fractional wagers to a set of simultaneous events numerically. Interestingly, it is noted that this strategy nearly achieves the optimal growth rate, give a certain set of opportunities, (shown in Edelman (2000)) only if betting on all possible subsets of events was allowed. This was duly done by Grant et al. (2008), who showed that not only does allowing accumulator bets increase the potential growth rate, but it also allows analytical results, regarding optimal stake sizes to be derived. These ideas are demonstrated with a variety of numerical investigations in Grant and Buchen (2012).

A final consideration for the use of any investment strategy is the uncertainty around the parameters of interest, most importantly the probability estimates. Strangely, there have been very few papers published on this issue, perhaps due to the work regarding fractional Kelly giving investors warnings regarding the potential impact of prediction errors on returns. One paper which makes an additional contribution to this area is Browne and Whitt (1996), which brings the basic Kelly work into a Bayesian setting. Given some repeated event, the current estimate of its probability can be updated with observations, along with the uncertainty around this estimate. More generally, and in a likelihood setting Baker and McHale (2013) supposes that the presence of uncertainty regarding the probability estimate should cause the optimal wager size to shrink by some factor. The optimal such shrinking factor is found, both via a numerical method, and analytically given some assumptions regarding the belief distribution for the event probability. These results are then extended for a variety of utility functions.

Chapter 4

Bias Estimation in Sports Predictive Models

4.1 Introduction

Bias in predictive models is notable when the model's predictions are systematically different to observations. Nearly all statistical models will exhibit bias, originating from such sources as the omission of important input variables, selection bias in the training set, and other subject-specific examples. Model validation techniques exist to detect and potentially correct for any bias encountered in model output data.

There are many techniques used to test for bias, which can be generally classified into four categories: subjective assessment, visual techniques, deviance measures and statistical tests (Mayer and Butler, 1993). The state-of-the-art for these techniques tend to be specific to each subject area. For example, climate modelling (Jun et al., 2008), social sciences (Lin et al., 2011) and psychology (Friesen and Weller, 2006), which all use similar deviance measures.

The assessment of bias within sports predictive modelling has been explored for many reasons in the past, mostly to identify whether betting markets exhibit biased behaviour which result in profitable investment strategies. To this end, Gandar et al. (2001) wished to find out whether weaker NFL teams playing at home are underpriced in the betting odds. Woodland and Woodland (1994), find that, against evidence from other sports, baseball teams which are very unlikely to win are priced profitably by bookmakers, in comparison to other teams.

In reality, model outputs do not all occur at one point in time, and instead are generated over some period of interest. For this reason, the model error's distribution will later be considered to be time-varying.

Within the domain of sports, predictive models use the outcomes of past matches as well as other relevant, unscheduled events, such as injuries to important players, to make predictions about future matches. Say a model was fitted to all known data, scheduled or otherwise at the current time. Predictions made at the current time are based on all known information. If the model was not updated, but used as a predictive tool at some future time, the model will not necessarily be based on current information, and the bias may increase.

Conversely, a model which exhibits predictions that are flagged as being potentially biased may result in the user identifying inaccuracies in the model, correcting them accordingly. In this case, the model (given useful interferences) should exhibit less bias over time.

Given an improvement in the performance in some model, it seems natural to think that not only will the 'average' behaviour of the predictive errors observed decrease over time, but so should their variability. Conversely, as the 'average' predictive performance deteriorates over time, the variability in the accuracy of these predictions should also increase.

The purpose of this analysis is two-fold: in the short-term, knowledge of the nature of errors occurring in model predictions would allow the model user to perform an *ad hoc* conversion to the outputs, to force the outputs to become collectively unbiased. Preferably, the information collected from the proposed analysis in this work would allow the model user to infer the cause of their model's erroneous predictions, and correct it accordingly.

This work aims to draw analogies from sports modelling to introduce a Bayesian approach to infer the nature of the error's distribution from the model output. In particular, the output data are assumed to arrive sequentially in pairs, with each pair consisting of an event probability from a model, along with the event's outcome. This work focusses on Bernoulli-type model outcomes as these are commonly observed when betting on sport. For example, in American sports, such as baseball and basketball, the home team will either win or not win. In other sports where draws occur, the popular Asian Handicap markets (see Section 2.1) reduce the outcomes to merely a winner and a loser.

All other output types can be reduced to Bernoulli observations via a loss of information. For example, count data, such as the number of points scored by a basketball team can be reduced to Bernoulli outputs via identity functions conditioned upon the points being greater or equal to some threshold. The use of Bernoulli outcomes is purely a chosen example for this work; the setup of this technique can be easily extended to non-Bernoulli outcomes.

The discussion of this work shall be structured as follows: firstly a simple model with a constant error distribution will be proposed. Its inference will be gained using MCMC (Monte Carlo Markov Chain) techniques, and the approach compared against other methods using simulated data. Afterwards, ways to account for the modelling error being potentially time-varying will be considered. The techniques will be used to analyse how biased bookmakers and betting exchange's predictions are, and how the bias changes between different types of matches, between different leagues, and over time. In addition, the time-varying nature of the errors exhibited by a simple predictive model for football outcomes shall be assessed.

4.2 Simple Model

Let the user of some predictive model receive a set of n estimates of the probability of the event occurring $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_1, \cdots, \hat{\theta}_n\}$, along with a corresponding set of independent Bernoulli outputs $\boldsymbol{y} = \{y_1, \cdots, y_n\}$, such that

$$y_i = \begin{cases} 1 & \text{if event } i \text{ occurs} \\ 0 & \text{if event } i \text{ does not occur} \end{cases}$$

and

$$y_i \overset{\text{indep}}{\sim} \text{Bern}(\theta_i) \quad i = 1, \cdots, n. \tag{4.2.1}$$

The aim is to allow the model user to take the sets $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{y}$ and retrieve some estimate of the bias and variability of the error affecting the model's predictions, assumed to be constant over time. Knowledge of the nature of the error would give vital information to the model user for the purpose of improving the performance of the model in the future.

The main modelling decision regards the way that the error terms should affect outputs. Rather that directly modelling the error of $\hat{\theta}_i$ from θ_i, e.g. $\hat{\theta}_i - \theta_i$, a more natural way to formulate the error is to first transform the probabilities to the log-odds scale. An additive error on the log-odds scale is not constrained by the range of $\theta_i \in [0, 1]$ and hence the error can be modelled as independent of the value of $\hat{\theta}_i$. The odds are defined as the 'fair-odds' that would be offered, such that the expected profit of a wager at these odds is 0. If o_i represent the fair-odds of the i'th event,

$$o_i = \frac{1 - \theta_i}{\theta_i} \tag{4.2.2}$$

and similarly, the estimated fair-odds of the i'th event based on model predictions is:

$$\hat{o}_i = \frac{1 - \hat{\theta}_i}{\hat{\theta}_i}.$$

Given this, and defining $\ell_i = \log o_i$ and $\hat{\ell}_i = \log \hat{o}_i$, the modelling choice for the error can be written as

$$\hat{\ell}_i = \ell_i + \epsilon_i, \quad \epsilon_i \sim \mathrm{N}(\mu, \sigma^2) \tag{4.2.3}$$

for each event $i = 1, \cdots, n$, and for some i.i.d error ϵ_i acting upon the log-odds scale.

From this, define the error's mean term, μ to be the **bias**, and the error's variance, σ^2 as the **model variance**.

As mentioned above, this approach assumes the probability does not affect the bias, i.e. the bias should act identically upon the log-odds of all choices of event probability. This assumption will be explored further in Section 4.2.5.

Apart from σ^2, all of the other parameters of interest, μ and $\boldsymbol{\ell}$ have their support on the whole real line, allowing any exploration of these parameters to be relatively simple to construct. The error's variance is transformed via the function $\xi = \log \sigma^2$, with the support of ξ now the entire real line. Equation (4.2.3) can then be rewritten as

$$\hat{\ell}_i \sim \mathrm{N}(\ell_i + \mu, \exp(\xi)), \quad i = 1, \cdots, n. \tag{4.2.4}$$

The setting, then, is using the $2n$ pieces of outcome data, $(\hat{\boldsymbol{\theta}}, \boldsymbol{y})$ to make inference upon the $n + 2$ parameters of interest (each of the n independent probabilities $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$, as well as the two parameters of the error distribution μ and ξ).

4.2.1 Bias on the Probability Scale

Assume for a moment that inference has been gained about the true value of μ and σ^2, along with some measure of uncertainty. For a Bayesian approach, the belief in these values will be found via the analysis of the values of the chain exploring the joint posterior distribution of the parameters of interest. Let these values be denoted as $\mu_{(1)}, \cdots, \mu_{(j)}, \cdots, \mu_{(m)}$, and $\sigma^2_{(1)}, \cdots, \sigma^2_{(j)}, \cdots, \sigma^2_{(m)}$, where m is the total number of iterations given to the MCMC updates, and where $(\mu_{(j)}, \sigma^2_{(j)})$ are the j'th values of the chain.

Given the relationship $\ell_i = \hat{\ell}_i - \epsilon_i$, taken from equation (4.2.3), then the posterior expected log odds for the i'th event is

$$\mathbb{E}(\ell_i|\hat{\boldsymbol{\ell}}, \boldsymbol{y}) = \hat{\ell}_i - \mathbb{E}(\epsilon_i|\hat{\boldsymbol{\ell}}, \boldsymbol{y}) \approx \ell_i - \frac{1}{m}\sum_{j=1}^{m} \mu_{(j)},$$

with $i > n$. Note that inference is conditioned upon the knowledge of the sets $\hat{\boldsymbol{\ell}}$ and $\boldsymbol{y}$. Therefore after MCMC analysis, the unbiased form of the event probabilities can be recovered. In a betting setting, this allows wagers to be made with a greater confidence in the probability estimates.

In many cases, it would be more useful for the bias to be removed from the probability itself. In motivating this work, it was suggested that any inference upon the nature of the modelling error could be used to adjust the model's outcomes in an *ad hoc* fashion to ensure that the outputs were unbiased. The form of equation (4.2.3) can again be rearranged, this time with the intention of making the true probability, θ_i, the subject:

$$\begin{aligned}
\hat{\ell}_i = \ell_i + \epsilon_i &\Rightarrow \log\left(\frac{1-\hat{\theta}_i}{\hat{\theta}_i}\right) = \log\left(\frac{1-\theta_i}{\theta_i}\right) + \epsilon_i \\
&\Rightarrow \frac{1-\hat{\theta}_i}{\hat{\theta}_i}\exp(-\epsilon_i) = \frac{1-\theta_i}{\theta_i} \Rightarrow \theta_i = \frac{1}{1+\frac{1-\hat{\theta}_i}{\hat{\theta}_i}\exp(-\epsilon_i)}
\end{aligned}$$

and so, for $i > n$

$$\mathbb{E}(\theta_i|\hat{\boldsymbol{\ell}}, \boldsymbol{y}) \approx \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{1}{1+\exp(\hat{\ell}_i - \mu_{(j)} - \sigma^2_{(j)} z_{(j)})}\right\} \tag{4.2.5}$$

where the $z_{(j)}$'s are independent draws from a standard Normal distribution.

By simulating the value of $\mathbb{E}(\theta_i|\hat{\boldsymbol{\ell}}, \boldsymbol{y})$ with $\hat{\theta}_i$ and μ fixed, but with σ^2 varying, the impact of the model error on the unbiased θ_i can be evaluated. Setting $\hat{\theta}_i = 0.5$ and $\mu = 0.1$, gives an unbiased probability of 0.5207 when $\sigma^2 = 0$ and 0.5240 when $\sigma^2 = 1$. This shows that the model error does not have a great impact upon the estimate of the unbiased probability, and therefore, equation (4.2.6) below gives a good approximation to its value when σ^2 is small:

$$\mathbb{E}(\theta_i|\hat{\boldsymbol{\ell}}, \boldsymbol{y}) \approx \frac{1}{1+\frac{1-\hat{\theta}_i}{\hat{\theta}_i}\exp(-\bar{\mu})} \tag{4.2.6}$$

where $\bar{\mu}$ is the mean of the MCMC inference upon the bias. Figure 4.2.1 shows the impact of bias on the log-odds scale on the probability scale.

As would be expected, a fixed bias on the log-odds scale has a greater impact on the probability scale around $\theta_i = 0.5$, and less impact on the chance of the event occurring when the event is near to being certain or impossible.

4.2.2 Prior Choice

In order to perform MCMC analysis on the posterior of the error distribution, given the observations of $\hat{\boldsymbol{\ell}}$ and $\boldsymbol{y}$, prior belief in the error distribution's parameters, as well

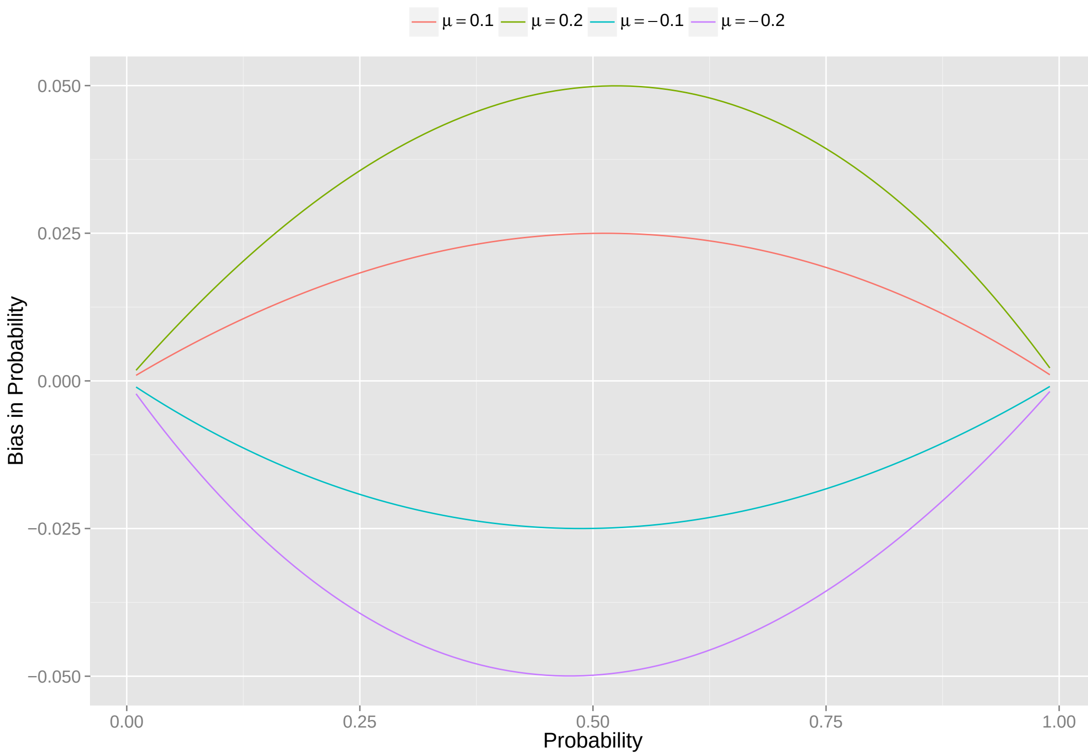


Figure 4.2.1: The equivalent log-odds additive bias acting upon the probabilities, for biases of -0.2, -0.1, 0.1, 0.2.

as the prior belief in the distribution of the independent transformed probabilities must be specified.

From Figure 4.2.2, the three variables requiring the selection of some prior distribution are $\boldsymbol{\ell}$, μ and ξ. The error's parameters, μ and $\xi = \log \sigma^2$ are treated as being independent, both of each other, see Lee (2012), and of the underlying probabilities $\boldsymbol{\ell}$, (as the process which generates the probabilities, and the error terms are completely separate).

Given that μ and σ^2 are the independent mean and variance of some normal distribution, their prior distributions are chosen to be independent and conjugate to the likelihood function. To this end, $\pi(\mu) \sim \mathrm{N}(u, v^2)$ and $\pi(\sigma^2) \sim \text{Inv-Gam}(a, b)$.

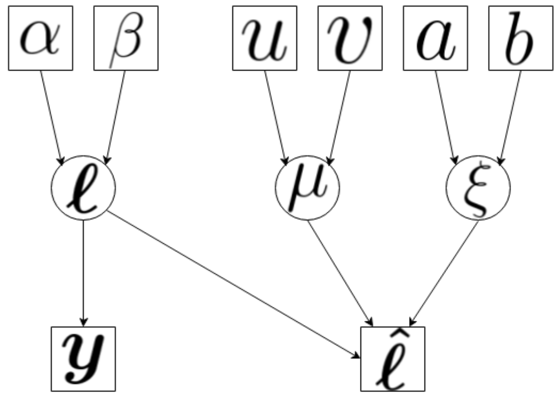


Figure 4.2.2: A DAG representing the model structure of a scheme to gain inference about the nature of some error affecting the Bernoulli outputs of a predictive model.

After transformation, the prior for ξ takes the form:

$$\pi(\xi) = \frac{b^a}{\Gamma(a)} \exp\left\{-a\xi - b\exp(-\xi)\right\}, \quad \xi \in \mathbb{R},\ a > 0,\ b > 0$$

where $\Gamma(a)$ is the Gamma function.

The choice of prior for the true probabilities $\boldsymbol{\theta}$ and thus their transformed counterparts $\boldsymbol{\ell}$ is again influenced by conjugacies, and each of the event probabilities is given an independent $\theta_i \sim \text{Beta}(\alpha, \beta)$ prior, with $\alpha > 0$, $\beta > 0$. After transformation to log-odds, this takes the form:

$$\pi(\boldsymbol{\ell}) = \frac{1}{\mathrm{B}(\alpha, \beta)^n} \prod_{i=1}^{n} \frac{\exp(\ell_i)^\beta}{[1 + \exp(\ell_i)]^{\alpha+\beta}}, \quad \ell_i \in \mathbb{R} \text{ for } i = 1, \cdots, n$$

where $\mathrm{B}(\alpha, \beta)$ is the Beta function.

The key hyperparameter choice constructs the prior for the transformed true probabilities $\boldsymbol{\ell}$. For each ℓ_i, the entire known set of information is the prior knowledge, together with one observation of its outcome. Due to this, the strength of the prior certainty has a large affect on the model's inference upon the nature of the ℓ_i's, ultimately impacting significantly upon the inference of the bias distribution's parameters, μ and σ^2. For more in-depth discussion (see Section 4.2.4).

To investigate this, a dataset containing all of the results of English football matches from the top four divisions (Premier League - League 2) was chosen, spanning the whole of the 2010-2011, 2011-2012 and 2012-2013 seasons, consisting of around 2000 pairs of data, and was taken from www.football-data.co.uk. By considering the home wins, away wins and draws in isolation, any dependence originating from the probabilities coming from the same events is lost, and each set of data can be considered as independent.

These data contain the result of each match in this period, along with a set of bookmakers' odds for each of the outcomes. These odds need to be converted into the probabilities for each of the matches, with the bookmakers' overround removed via the method shown in equation (2.2.1).

This dataset was used to assess whether a fitted Beta distribution can be used to provide a representative prior to describe the distribution of probabilities of certain events occurring in football matches. In this case, the data corresponds to the probability of a football match resulting in a home win.

The shape, α and scale, β, parameters of the Beta distribution were chosen via the method of moments approach, using the mean, m and the variance, s^2 of the observed data, with the fitted values being:

$$\alpha = m\left(\frac{m(1-m)}{s^2} - 1\right), \qquad \beta = (1-m)\left(\frac{m(1-m)}{s^2} - 1\right). \tag{4.2.7}$$

The result of this analysis is shown in Figure 4.2.3, which shows that a Beta distribution provides an adequate fit to the kind of data encountered in sports modelling.

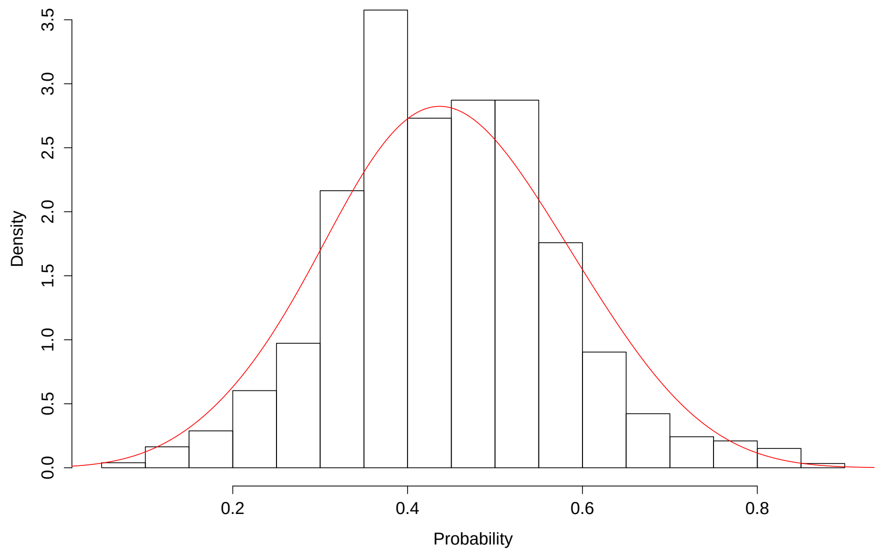


Figure 4.2.3: Histogram showing the implied probabilities of home wins, based on odds offered by bookmakers for English football matches, along with fitted Beta distributions with $\alpha = 6.547$ and $\beta = 8.162$.

In addition, the hyperparameters for the error's parameters, a, b, u and v would also be chosen to demonstrate the prior knowledge, if any, the model user currently has about the nature of the predictive errors.

4.2.3 Posterior

With an assumption of complete independence between the parameters of the error distribution and the event probabilities, the following decomposition of the joint posterior probability can be performed into its interaction terms and the independent prior distributions:

$$\pi(\boldsymbol{y}, \boldsymbol{\ell}, \hat{\boldsymbol{\ell}}, \mu, \xi) = \pi(\boldsymbol{y}|\boldsymbol{\ell})\pi(\hat{\boldsymbol{\ell}}|\boldsymbol{\ell}, \mu, \xi)\pi(\boldsymbol{\ell})\pi(\mu)\pi(\xi). \tag{4.2.8}$$

The first term of the right-hand-side of the joint distribution (4.2.8) is simply a log-odds transformation of the density of $\pi(\boldsymbol{y}|\boldsymbol{\theta})$, and can be written as:

$$\pi(\boldsymbol{y}|\boldsymbol{\ell}) = \prod_{i=1}^{n} \frac{[\exp(\ell_i)]^{1-y_i}}{1+\exp(\ell_i)}.$$

The second term of the joint distribution (4.2.8) can be found simply, given our modelling arrangement (4.2.4), and is simply a reparametrisation of the distribution of a Normal model:

$$\pi(\hat{\boldsymbol{\ell}}|\boldsymbol{\ell}, \mu, \xi) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\xi - \frac{1}{2}\exp(-\xi)(\hat{\ell}_i - \ell_i - \mu)^2\right\}. \tag{4.2.9}$$

The overall joint probability can now be written as:

$$\begin{aligned}
&\pi(\boldsymbol{\ell}, \hat{\boldsymbol{\ell}}, \boldsymbol{y}, \mu, \xi) = \pi(\boldsymbol{y}|\boldsymbol{\ell})\pi(\hat{\boldsymbol{\ell}}|\boldsymbol{\ell}, \mu, \xi)\pi(\boldsymbol{\ell})\pi(\mu)\pi(\xi) \\
&\propto \prod_{i=1}^{n} \frac{[\exp(\ell_i)]^{1-y_i}}{1+\exp(\ell_i)} \times \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\xi - \frac{1}{2}\exp(-\xi)(\hat{\ell}_i - \ell_i - \mu)^2\right\} \times \prod_{i=1}^{n} \frac{\exp(\ell_i)^{\beta}}{[1+\exp(\ell_i)]^{\alpha+\beta}} \\
&\times \exp\left\{-\frac{1}{2v^2}(\mu - u)^2\right\} \times \exp\left\{-a\xi - b\exp(-\xi)\right\} \\
&= \exp\left\{-\frac{1}{2v^2}(\mu - u)^2\right\} \exp\left\{-\left(a + \frac{n}{2}\right)\xi - b\exp(-\xi)\right\} \\
&\qquad \times \prod_{i=1}^{n} \frac{[\exp(\ell_i)]^{1-y_i+\beta}}{[1+\exp(\ell_i)]^{1+\alpha+\beta}} \exp\left\{-\frac{1}{2}\exp(-\xi)(\hat{\ell}_i - \ell_i - \mu)^2\right\}
\end{aligned} \tag{4.2.10}$$

with the model structure being visualised via the DAG in Figure 4.2.2.

This distribution can be explored via a MCMC scheme. Given the perception that there exists at most weak dependence between the parameters, each block update consists of only a single variable at a time. In addition, as the posteriors of each of the parameters was not available in closed form, and due to the domains of the variables being on the real line, Metropolis-Hastings with normal symmetric jumps were proposed, such that the acceptance rate was around 0.23, as recommended by Gilks and Roberts (1996).

4.2.4 Sensitivity to Prior Choice for Probabilities

At this point, analysis has showed that using a Beta distribution as a prior for the underlying event probabilities is justified. What remains to be shown is whether using an uninformed, or weakly informed prior for the event probabilities result in the MCMC analysis on the joint probability shown in equation (4.2.10) providing good estimates for the distribution of μ and σ^2.

Data were simulated from the same Beta distribution as above, i.e, $\theta_i \sim \text{Beta}(6.547, 8.162)$. This was then used to simulate data such that the bias of the data is $\mu = 0.1$, and error variance is $\sigma^2 = 0.05$. A simple Metropolis-Hastings algorithm was run on 2000 pairs of data, with the prior on the θ_i's varying for different runs, with the intention being to observe the sensitivity between the algorithm's outputs and the prior inputs.

An account of this investigation is summarised in Table 4.2.1. It shows that the accuracy of the outputs are highly dependent upon the quality of the prior distribution for the θ_i's. Interestingly, the use of an uninformative prior (in this case,

	Prior Distribution of θ_i's		
Output	Accurate	Inaccurate	Uninformative
μ	Good	Bad	Good
σ^2	Good	Bad	Bad

Table 4.2.1: The performance of a simple MCMC scheme in predicting the size of bias and error variance in simulated data, given different prior choices for the underlying probabilities.

$\theta_i \sim \text{Beta}(1,1)$) gives good posterior estimates for μ, but not for σ^2.

In Tjur (2003), logistic regression problems were considered with binomial outcomes. It is shown that when learning about both the intercept and the coefficients of the exploratory variables, problems with identifiability occur when the number of observed outcomes for each event is set to 1. This identifiability problem emerges as "...we cannot distinguish between the weak influence of the covariate and the high variation between (observations)". Therefore, in order to gain some knowledge of the error variability, some exploratory analysis must be performed beforehand, to give informed prior knowledge regarding the distribution of the underlying probabilities.

The way to solve this problem, and to maintain the ability to make inference upon the error's variance is to consider the modelling problem in three stages:

1. **Exploratory Analysis:** Place an uninformative prior on each of the θ_i, here $\theta_i \sim \text{Beta}(1,1)$. This allows inference upon the bias, whilst giving poor inference upon its variance parameter (the sample correlation between the sample mean and sample variance of a Normal distribution is 0, thus the parameters are orthogonal, so poor learning of the variance does not affect the learning on the mean).

2. **Inference upon α and β:** By fitting a Beta distribution to the observed probabilities,$(\hat{\theta}_1, \cdots, \hat{\theta}_n)$, and by taking the mean of the posterior mean estimate for μ in Step 1, an estimate for the distribution of the underlying ℓ_i's can be found, via transforming the output of equation (4.2.6).

3. **Model Inference:** The MCMC scheme is then run again, with the prior for the ℓ_i's being chosen via the work in Step 2, and the prior for μ also being chosen to match the mean and variance of Step 1's posterior estimate for μ's distribution. The posterior means of this run of the MCMC scheme are taken as being the resultant parameter estimates.

This then, gives a method for inferring information about the true underlying distribution of the the θ_i's, given a prior exploratory analysis which provides an estimate for the bias. Explicitly, the method can be thought of via Algorithm 1.

Algorithm 1 Two-Stage Bias Estimation Approach

Input: Set of n pairs of data, $(\hat{\theta}_i, y_i)$, $i = 1, \cdots, n$.

Perform MCMC analysis with uninformed priors upon the posterior distribution in equation (4.2.10), let $\bar{\mu}$ be the mean of the posterior estimate for the distribution of μ.

for all $\hat{\theta}_i$ **do**

  Find $\mathbb{E}(\theta_i | \hat{\boldsymbol{\ell}}, \boldsymbol{y})$ via equation (4.2.6).

end for

Find the mean and the variance of estimated values $\hat{\boldsymbol{\theta}}$, then fit a Beta distribution via the method of moments in equation (4.2.7). This is an informed prior for the θ_i's.

Perform MCMC analysis of the posterior distribution in equation (4.2.10), outputs taken as being posterior estimate for the distribution of μ and σ^2.

4.2.5 Model Structure Diagnostics

The model structure specified above contains two assumptions which should be explored further in order to ensure confidence that the modelling decisions are appropriate.

Independence of Priors

An assumption highlighted in the main modelling work in Section 4.2 was the independence between the error distribution's parameters (primarily the bias mean, μ) and the underlying probabilities $\theta_1, \cdots, \theta_n$. This states that the errors apply additively and identically to the log-odds of the true probabilities, throughout their whole domain.

To test whether this assumption is realistic, the dataset described in Section 4.2.4 was partitioned according to the ordered observed probabilities $\hat{\theta}_i, \cdots, \hat{\theta}_n$. For each of these partitions, the average log-odds was compared against the proportion of outcomes which did indeed occur. This gives a rough estimate of the bias in the model's predictions; in this case the 'model' is the bookmaker's predictions for the probability of event occurring.

Figure 4.2.4 shows how the average modelling bias, μ changes with the probability estimate of the underlying events, lines describing its variability (expressed as a 95% confidence interval). For data which exhibits bias independently from its underlying probabilities, each partition would exhibit very similar errors on the log-odds scale, with 0 indicating no bias in that particular partition. Figure 4.2.4 shows that in this dataset, at least, the independence assumption seems to hold.

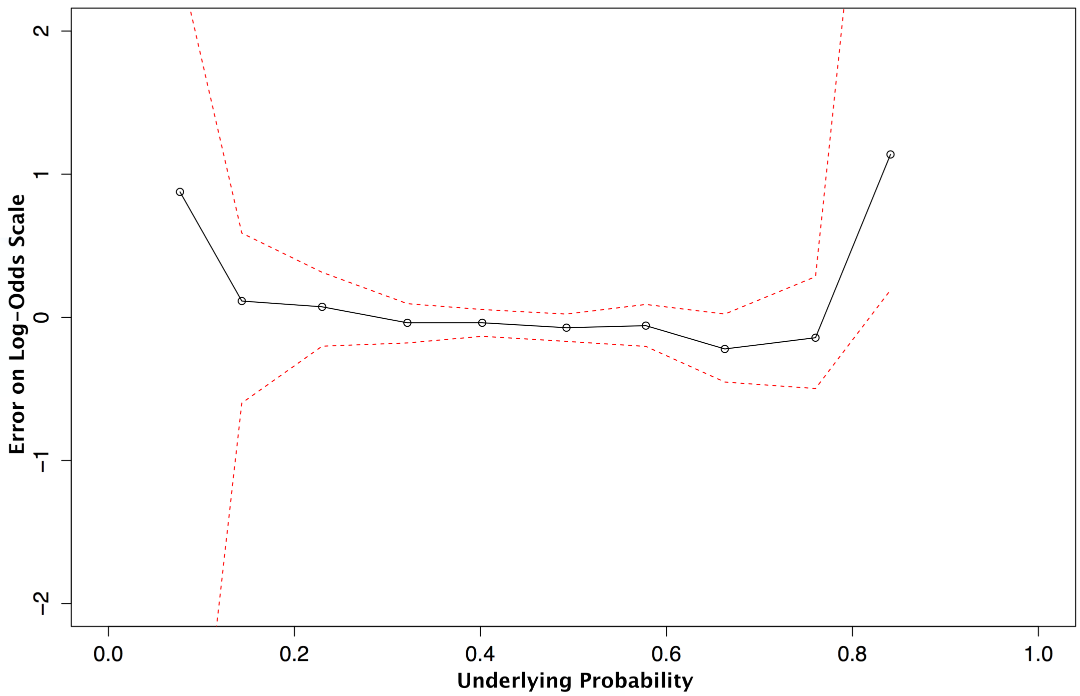


Figure 4.2.4: The average bias in a model's predictions (solid black line), and its variability (dashed red line indicating 95% of the data in each partition), in comparison to the underlying probability of the predicted events occurring.

With one exception, the bias mean of 95% of the errors are distributed constantly around 0. This independence seems to break down, however, once the underlying probabilities are above 0.8. This may conform to the so-called 'longshot-bias', explored in Williams and Paton (1997) and Woodland and Woodland (1994), amongst others. Although this provides evidence that the model-error's distribution may not be independent of the underlying probabilities at high values, this element of the partition only accounts for around 0.6% of the dataset. This feature will be explored in more detail in Section 4.4; for this particular purpose, the independence assumption seems to suffice for datasets of this type.

Dependence upon Priors

Section 4.2.4 explored how there is a strong reliance of the prior distribution upon the underlying probabilities on the quality of the model's outputs. It remains to be seen whether a similar relationship exists for the other two prior choices, namely on the error distribution's parameters μ and σ^2.

Data were generated to assess this assumption. The data were simulated via equation (4.2.3), with the underlying probabilities being random draws from a $\text{Beta}(2, 2)$ distribution, the model error was chosen with a bias of $\mu = 0.1$ and error variability of $\sigma^2 = 0.4$.

The model was initialised with the following priors: the prior for the underlying probabilities was correctly chosen as following a $\text{Beta}(2, 2)$ distribution; the priors for the error distribution's parameters were chosen to be partly-informative, $\mu \sim \text{N}(0, 1)$ and $\sigma^2 \sim \text{Inv-Gam}(2, 0.2)$. The posterior is compared to the prior for μ, σ^2 in Figure 4.2.5.

This clearly shows that the prior choice has very little impact upon the posterior distribution of the model's error parameters. In addition, the posterior distribution shows that there is little correlation between the error's mean and variance.

4.2.6 Comparison

The Bayesian approach is compared against other candidate approaches; stochastic approximation and a very simple diagnostic, which is termed the 'naïve approach', to assess its efficacy. Details of the other two approaches can be found in Appendix A.

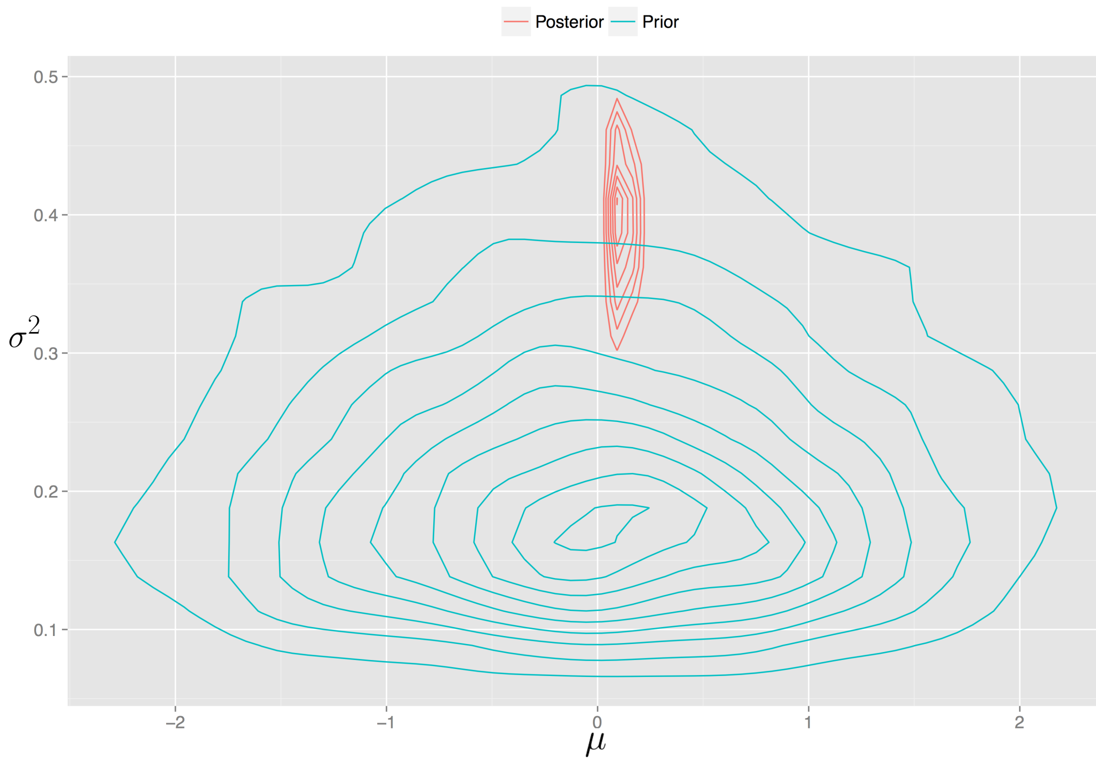


Figure 4.2.5: A comparison of the prior and posterior distribution of a model error with mean of $\mu = 0.1$ and variance $\sigma^2 = 0.4$. The green line represents the prior belief and the red line represents the posterior belief.

The data consists of 100 independent sets of simulated data, each containing 2000 $(\hat{\theta}_i, y_i)$ pairs. This is roughly the number of games played in an English Association Football league season, and would represent the scenario of a bettor using a year's worth of predictions. The data were simulated with a bias of $\mu = 0.1$, and error variability of $\sigma^2 = 0.1$. The underlying probabilities were initially drawn from a Beta$(2, 2)$ distribution, to test the efficacy of the naïve bias estimation technique under idealised circumstances. Later, the probabilities are drawn from a skewed Beta$(5, 2)$ distribution, with mean 0.71 in order to demonstrate a significant drawback of using the naïve technique.

The MCMC scheme had the following starting values. The bias mean was set at an initial value of 0, with prior distribution of $N(0,1)$. The bias variance was set at an initial value of 0.1, with a prior distribution of Inv-Gam$(1,5)$. Each of the individual true probabilities was given an initial value of 0.5, and a prior distribution of Beta$(1,1)$, indicating the assumption of no prior knowledge regarding the distribution of the true probabilities.

Figures 4.2.6 and 4.2.7 show the results of this simulation study, using the posterior means (in the case of the Bayesian method), and a point estimate for the other two methods. The distribution of these means are used to assess and compare the bias-detection methods.

As can be seen in these idealised simulation studies, the Bayesian approach provides a more accurate estimate of the bias more frequently than the other candidate methods. The naïve estimate does surprisingly well in its predictive power, however, as the underlying data were drawn from a symmetric distribution with a mean of 0.5, this was an ideal experimental setup for the naïve method. When the data set is changed to something more skewed, the predictive power of the naïve method is lost entirely (repeating this experiment identically, but with an underlying probability distribution of a Beta$(5,2)$ causes the naïve method to estimate the model bias to be around -0.06). Table 4.2.2 summarises the result.

The Bayesian method provides the best bias estimation of the three measures specified. As expected, the naïve estimate's efficacy is highly dependent upon its major assumption (that being that the distribution of the underlying probabilities is symmetric), with its accuracy in predicting the bias rivalling the other two more sophisticated approaches. When this assumption is violated, the naïve method offers

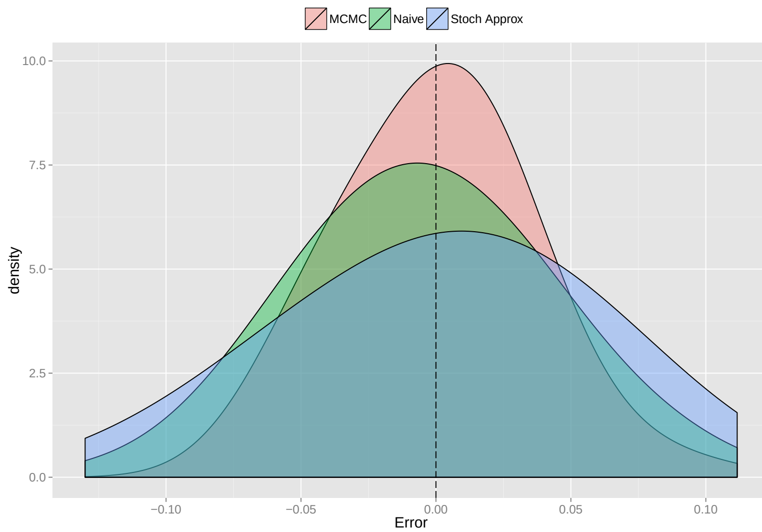


Figure 4.2.6: Density of the difference between model estimates and the truth. The 3 bias-prediction models are approximating the bias' true mean value of $\mu = 0.1$. The underlying probabilities are drawn from a Beta$(2, 2)$ distribution.

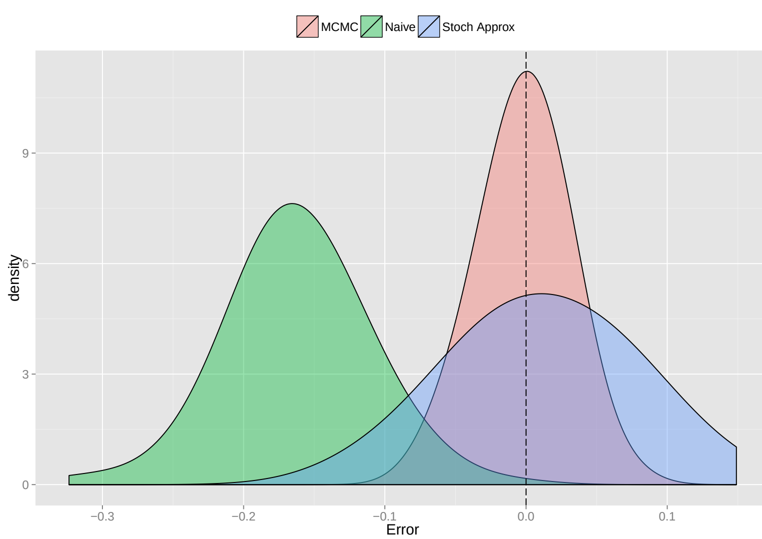


Figure 4.2.7: Density of the difference between model estimates and the truth. The 3 bias-prediction models are approximating the bias' true mean value of $\mu = 0.1$. The underlying probabilities are drawn from a Beta$(5, 2)$ distribution.

	Symmetric Probabilities			Skewed Probabilities		
Technique	Mean Error	RMSE	% close	Mean Error	RMSE	% close
Bayesian	-0.00025	0.3291	25	-0.0019	0.2898	28
Stoch Approx	0.00124	0.5323	15	0.0119	0.6152	15
Naïve	-0.00476	0.4334	10	-0.1625	1.6945	0

Table 4.2.2: A summary of the utility of methods in estimating a bias of 0.1, given data sets containing probabilities drawn from both a symmetric distribution, and a skewed distribution. "% close" indicates the proportion of bias estimates within 10% of the truth.

a user an incorrect insight.

In comparison to the stochastic approximation method, the Bayesian method is also appreciably superior. The difference in efficacy of the methods in terms of the accuracy of predictions is seen most clearly when considering the proportion of inferences which were "close" to the truth, where the Bayesian method's rate of predicting close to the truth is almost twice that of stochastic approximation, for both symmetric and skewed underlying probabilities.

It should be noted that there is a cost involved in this increase in accuracy, namely the amount of time needed to produce the inferences. The stochastic approximation has a computational complexity $\mathcal{O}(n)$, where n is the number of data supplied. In comparison, the Bayesian method requires m updates of $n + 2$ parameter, the n underlying probabilities, as well as the two bias parameters, giving a complexity of $\mathcal{O}(mn)$. In the above example, the Bayesian method was allowed a chain of length 5000 (this was proven to be sufficient to allow the chain to mix well, and to sufficiently explore the posterior distribution), meaning that this method took 5000 times as long as the stochastic approximation. This could potentially provide a problem if inference of the bias was required in a short time-frame. In Section 4.3.1, this underlying

	Symmetric Probabilities			Skewed Probabilities		
Technique	Mean Error	RMSE	% close	Mean Error	RMSE	% close
Bayesian	0.000897	0.0748	51	0.00128	0.0808	48

Table 4.2.3: A summary of the utility of the Bayesian method in estimating the variance term of a bias, given data sets containing probabilities drawn from both a symmetric distribution, and a skewed distribution. "% close" indicates the proportion of bias estimates within 10% of the truth.

modelling structure is adapted to hugely decrease the order of complexity of the MCMC algorithm.

It should also be noted that the Bayesian method produces additional insight, which eludes the other two approaches; namely, estimation of the variance of the model error. Table 4.2.3 shows the performance of the method in this endeavour.

4.3 Time-Varying Bias Parameters

One of the primary assumptions underlying the work of Section 4.2 concerned the constant error parameters. As discussed in Section 4.1, there are many reasons both the error distribution's mean and variance may change over time.

In order to attempt to model the time-varying parameters of the error's distribution, a deterministic structure is imposed. The bias should take some initial value, say μ_0 (which may be 0), which would represent the model's initial state. The bias would then increase, or decrease to 0 in some way over time. The error's variance would also increase in a deteriorating model, and decrease in an improving model. A feature for the variance of any model is that it should not necessarily decrease to 0 when improving, as even an ideal statistical model exhibits some zero-mean noise in its outputs. The deterministic structure of the error's variance should reflect this; what follows is a list of characteristics which the time-varying error parameters μ_t and σ_t^2 should adhere to:

- For an improving model:
  - The bias should decrease to 0 as $t \to \infty$.
  - The model variance should decrease to some lower limit (greater or equal to 0) as $t \to \infty$. This limit, and the rate at which this is achieved are parameters of interest.
- For a deteriorating model:
  - The bias should increase over time. The function describing the change in bias over time should be concave, to prevent the bias 'blowing up' to unrealistic levels.
  - The model variance should increase over time. Again, this increase should be described by some concave function.

The modelling choice for these time-dependent parameters is for the bias at time t, μ_t and the model error at time t, σ_t^2 to be:

$$\mu_t = \mu_0 t^{\gamma}; \quad \mu_0, \gamma > 0, \quad t = 1, 2, \cdots \tag{4.3.1}$$

and

$$\sigma_t^2 = \vartheta + \varphi t^{\nu}; \quad \vartheta, \varphi > 0,\, \nu \in \mathbb{R}, \quad t = 1, 2, \cdots \tag{4.3.2}$$

where μ_0 is the initial bias, γ is a measure of the rate of change of the bias over time. The interpretation of the parameters dictating the change in the error's variance are harder to glean. However, at $t = 1$, $\sigma_1^2 = \vartheta + \varphi$, its initial value.

As stated above, when the model variance decreases, it should have some minimum level. Letting $\sigma_\infty^2 := \lim_{t\to\infty} \sigma_t^2$, if $\nu < 0$ then $\sigma_\infty^2 = \vartheta$. Therefore, for an improving model, the error's time-varying variance will take the initial value of $\vartheta + \varphi$, decreasing to ϑ as $t \to \infty$, with rate determined by ν.

Figures 4.3.1 and 4.3.2 show how the bias and model error are modelled to change over time, given differing parameters. Note that in Figure 4.3.2, the red line exhibits the case where the lower limit is restricted below at $\vartheta = 0.1$.

In comparison to the simple model in Section 4.2, there are now 3 additional parameters to estimate, and 5 in total, written in shorthand as a vector of parameters: $\boldsymbol{\psi} = (\mu_0, \gamma, \vartheta, \varphi, \nu)$. In order to update the MCMC scheme to account for the new parameters of interest, the joint posterior probability must be adapted from equation (4.2.10). The dependence structure between the parameters should also be specified. As before, there should be no dependence between the error distribution's mean and variance time-varying parameters. In comparison, it is envisioned that there may be correlations between the sets of parameters used to construct μ_t and σ_t^2, i.e. within the set $\{\mu_0, \gamma\}$ and the set $\{\vartheta, \varphi, \nu\}$, but not between these sets. This assumption will be analysed in future sections.

Equation (4.2.8) shall be updated in the following form:

$$\pi(\boldsymbol{y}, \boldsymbol{\ell}, \hat{\boldsymbol{\ell}}, \boldsymbol{\psi}) = \pi(\boldsymbol{y}|\boldsymbol{\ell})\pi(\hat{\boldsymbol{\ell}}|\boldsymbol{\ell}, \boldsymbol{\psi})\pi(\boldsymbol{\ell})\pi(\boldsymbol{\psi}). \tag{4.3.3}$$

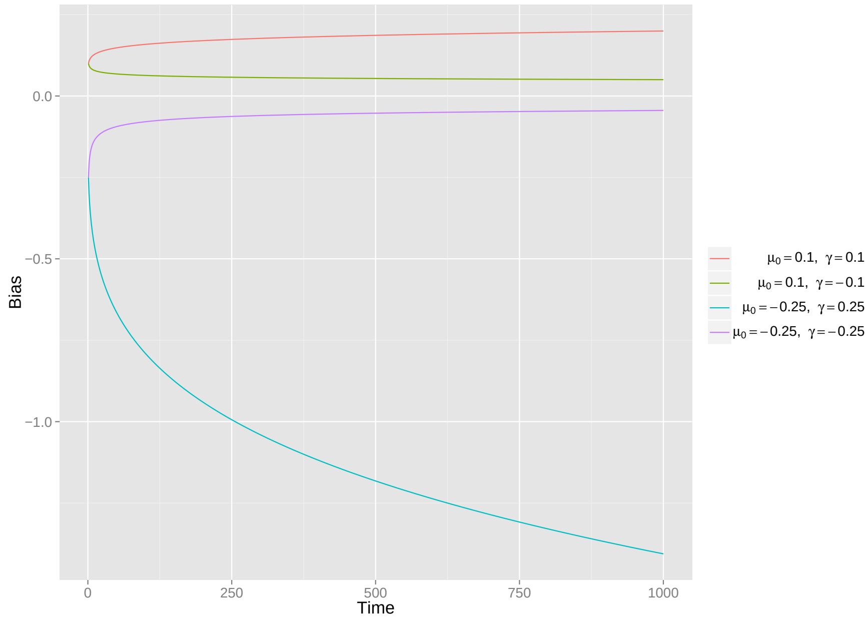


Figure 4.3.1: The change in bias over time, given differing parameters.

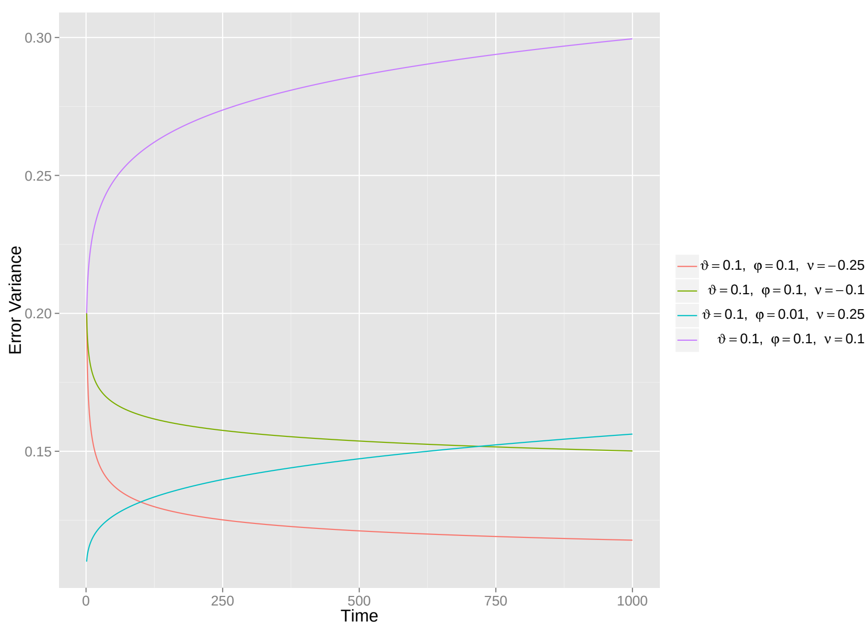


Figure 4.3.2: The change in model variance over time, given differing parameters.

The initial value of the bias, μ_0 takes the same prior form as its constant version: $\pi(\mu_0) \sim \mathrm{N}(u_0, v_0^2)$. The change in μ_t can be both positive and negative, given that the impact of the bias term can be both increasing or decreasing over time. Given this a normal prior is also placed on γ, $\pi(\gamma) \sim \mathrm{N}(u_\gamma, v_\gamma^2)$.

The prior choice for the parameters used to construct σ_t^2 require more thought. As these parameters are components of some time-varying variance of an error distribution, the modelling choice is to place the conjugate prior for a variance on ϑ and φ: $\pi(\vartheta) \sim$ Inv-Gam$(a_\vartheta, b_\vartheta)$ and $\pi(\varphi) \sim$ Inv-Gam(a_φ, b_φ). The variance can both increase and decrease over time, so again a Normal prior is placed on the rate of change parameter: $\pi(\nu) \sim \mathrm{N}(u_\nu, v_\nu^2)$.

An extended version of Figure 4.2.2, to account for the new model structure is shown in Figure 4.3.3.

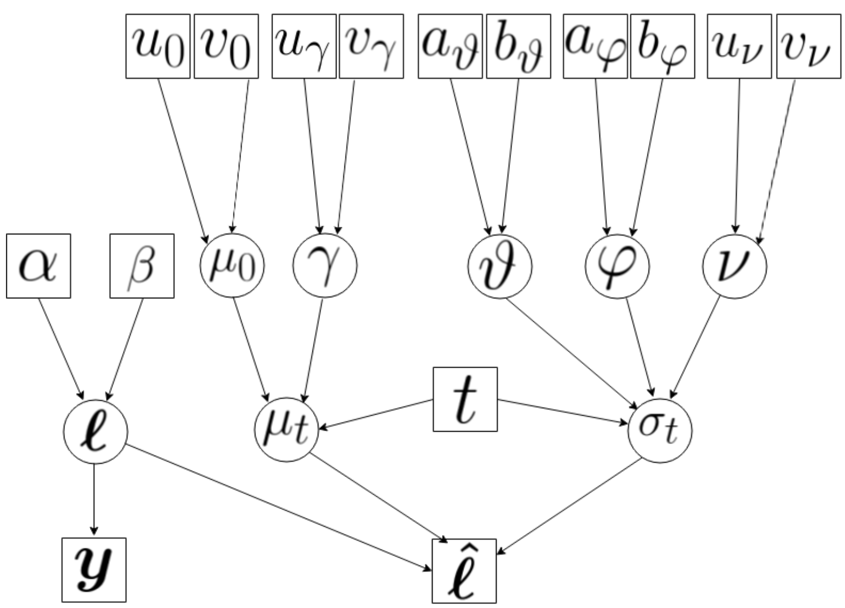


Figure 4.3.3: A DAG to represent the structure of the bias-quantification model shown in equation (4.3.3).

As before, the parameters whose domains are restricted to the positive real line, in this case ϑ and φ are log-transformed, such that the MCMC updates are not restricted. Denote the log-transformed versions of ϑ and φ as ξ_ϑ and ξ_φ, respectively, the notation chosen to signify the link between these parameters and their significance in the modelling of the bias distribution's variance. Redefine $\boldsymbol{\psi} = (\mu_0, \gamma, \xi_\vartheta, \xi_\varphi, \nu)$.

The joint posterior form for static bias parameters, (4.2.10) can now be updated to its full, time-varying form (where the index i has been changed to t to make the time element clearer):

$$\pi(\boldsymbol{\ell}, \hat{\boldsymbol{\ell}}, \boldsymbol{y}, \boldsymbol{\psi}) = \pi(\boldsymbol{y}|\boldsymbol{\ell})\pi(\hat{\boldsymbol{\ell}}|\boldsymbol{\ell}, \boldsymbol{\psi})\pi(\boldsymbol{\ell})\pi(\mu_0)\pi(\gamma)\pi(\xi_x)\pi(\xi_z)\pi(\nu)$$
$$\propto \prod_{t=1}^{n} \frac{[\exp(\ell_t)]^{1-y_t}}{1+\exp(\ell_t)} \times \prod_{t=1}^{n} \exp\left\{-\frac{1}{2}\xi_t - \frac{1}{2}\exp(-\xi_t)(\hat{\ell}_t - \ell_t - \mu_t)^2\right\} \times \prod_{t=1}^{n} \frac{\exp(\ell_t)^\beta}{[1+\exp(\ell_t)]^{\alpha+\beta}}$$
$$\times \exp\left\{-\frac{1}{2v_0^2}(\mu_0 - u_0)^2\right\} \times \exp\left\{-\frac{1}{2v_\gamma^2}(\gamma - u_\gamma)^2\right\} \times \exp\left\{-a_\vartheta\xi_\vartheta - b_\vartheta \exp(-\xi_\vartheta)\right\}$$
$$\times \exp\left\{-a_\varphi\xi_\varphi - b_\varphi \exp(-\xi_\varphi)\right\} \times \exp\left\{-\frac{1}{2v_\nu^2}(\nu - u_\nu)^2\right\} \tag{4.3.4}$$

where ξ_t signifies the log-transform of the time-varying model error, $\xi_t = \log[\vartheta + \varphi t^\nu] = \log[\exp(\xi_\vartheta) + \exp(\xi_\varphi)t^\nu]$.

4.3.1 Improvements to Inference

Improvements via a Probit Link

The comparisons of the simple model against other candidate approaches in Section 4.2.6 showed that, although the Bayesian approach outperformed other methods, it was let down due to its computational speed. If it were possible to re-form the modelling approach to the manifestation of the bias, such that the underlying probabilities $\theta_i, i = 1, \cdots, n$ could be marginalised out of the joint posterior, thus not requiring MCMC updates of each of the probabilities at each iteration, the utility of the method would be much improved.

There will potentially be another benefit for marginalisation of this form. The simple model drew its inference, in part, from a comparison of the both the distribution of the observed, biased probabilities and the distribution of the true probabilities, which changes at every iteration of the MCMC scheme. As this distribution is constantly being updated, the updating of $\boldsymbol{\psi}$ is based upon a different distribution of $\boldsymbol{\theta}$ at each iteration, making the mixing of $\boldsymbol{\psi}$ unstable. Given the more complex structure imposed on the error distribution's parameters introduced in Section 4.3, the challenge of achieving convergence of the chain of updates will be greater, and thus a more stable framework for inference will be essential.

Let the set of parameters representing the distribution of the bias again be written as $\boldsymbol{\psi} = (\mu_0, \gamma, \vartheta, \varphi, \nu)$. The true probabilities $\boldsymbol{\theta}$ are now transformed via a probit link:

$$\eta_t = \Phi^{-1}(\theta_t),\ \eta_t \in \mathbb{R}$$

where $\Phi(.)$ is the cdf of a standard Normal distribution. A latent random variable $Z_t \sim \mathrm{N}(0, 1)$ is introduced, which will make manipulations of the model simpler. Note that $P(Z_t \leq \eta_t) = \Phi(\eta_t) = \theta_t$, so $y_t = 1$ if $Z_t \leq \eta_t$ and $y_t = 0$ otherwise.

Given this, allow the time-varying error parameters to interact with the probit-transformed probabilities in a similar way to the log-odds case from Section 4.2:

$$\hat{\eta}_t - \eta_t \sim \mathrm{N}(\mu_t, \sigma_t^2) \Rightarrow \hat{\eta}_t \sim \mathrm{N}(\eta_t + \mu_t, \sigma_t^2). \tag{4.3.5}$$

A prior for η_t is given explicitly as $\eta_t \sim \mathrm{N}(m, s^2)$, such that the belief in the underlying distribution from which the probit-transformed probabilities are drawn is not time-varying.

Writing the transformed true probabilities in terms of the priors on the observations, along with the error distribution's parameters gives:

$$\begin{aligned}
\pi(\eta_t|\hat{\eta}_t, \boldsymbol{\psi}) &\propto \pi(\hat{\eta}_t|\eta_t, \boldsymbol{\psi})\pi(\eta_t) \\
&= \frac{1}{\sigma_t} \exp\left\{-\frac{1}{2\sigma_t^2}(\hat{\eta}_t - \eta_t - \mu_t)^2\right\} \times \frac{1}{s} \exp\left\{-\frac{1}{2s^2}(\eta_t - m)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{1}{s^2} + \frac{1}{\sigma_t^2}\right]\left[\eta_t - \frac{\frac{m}{s^2} + \frac{\hat{\eta}_t - \mu_t}{\sigma_t^2}}{\frac{1}{s^2} + \frac{1}{\sigma_t^2}}\right]^2\right\}.
\end{aligned}$$

It can thus be written $\eta_t|\hat{\eta}_t, \boldsymbol{\psi} \sim \mathrm{N}(M_t, S_t^2)$ where

$$S_t^2 = \frac{1}{1/s^2 + 1/\sigma_t^2} \quad \& \quad M_t = S_t^2\left(\frac{m}{s^2} + \frac{\hat{\eta}_t - \mu_t}{\sigma_t^2}\right).$$

The separate parts can be brought together, to construct the probability of $y_t = 1$, without reference to the transformed underlying probability η_t.

First, note that $P(Z_t \leq \eta_t|\hat{\eta}_t, \boldsymbol{\psi}) = P(Z_t - \eta_t \leq 0|\hat{\eta}_t, \boldsymbol{\psi})$. However, $Z_t - \eta_t \sim \mathrm{N}(-M_t, S_t^2 + 1)$, so $P(y_t = 1|\hat{\eta}_t, \boldsymbol{\psi}) = \Phi\left(\frac{M_t}{\sqrt{S_t^2+1}}\right)$; note that the η_t's have been marginalised out, and

$$\pi(\boldsymbol{y}|\hat{\boldsymbol{\eta}},\boldsymbol{\psi}) = \prod_{t=1}^{n} \Phi\left(\frac{M_t}{\sqrt{S_t^2+1}}\right)^{y_t} \Phi\left(\frac{-M_t}{\sqrt{S_t^2+1}}\right)^{1-y_t}. \tag{4.3.6}$$

Writing the joint posterior distribution in a similar way to equation (4.3.4) gives:

$$\pi(\hat{\boldsymbol{\eta}},\boldsymbol{\psi}|\boldsymbol{y}) = \pi(\boldsymbol{y}|\hat{\boldsymbol{\eta}},\boldsymbol{\psi})\pi(\hat{\boldsymbol{\eta}}|\boldsymbol{\psi})\pi(\boldsymbol{\psi}). \tag{4.3.7}$$

Given the same prior specifications of the error distribution's parameters as before, only the middle term on the right-hand-side of equation (4.3.7) is left to define. This expression is simple, as equation (4.3.5) can be combined with the prior for η_t to give $\hat{\eta}_t|\boldsymbol{\psi} \sim \mathrm{N}(\mu_t + m, \sigma_t^2 + s^2)$.

The joint posterior distribution can be written, with the individual η_t's marginalised out (in that the η_t's are not required for inference upon the other parameters, however the prior knowledge represented with m and s^2 still has importance), as:

$$\begin{aligned}&\pi(\hat{\boldsymbol{\eta}},\boldsymbol{\psi}|\boldsymbol{y})\\ &\propto \pi(\boldsymbol{\psi}) \prod_{t=1}^{n} \left\{ \Phi\left(\frac{M_t}{\sqrt{S_t^2+1}}\right)^{y_t} \Phi\left(\frac{-M_t}{\sqrt{S_t^2+1}}\right)^{1-y_t} \frac{1}{2\sqrt{(\sigma_t^2+s^2)}} \exp\left[-\frac{1}{2}\frac{(\hat{\eta}_t-\mu_t-m)^2}{\sigma_t^2+s^2}\right] \right\}.\end{aligned} \tag{4.3.8}$$

As stated previously, the main advantage of this form of the joint posterior is that the MCMC scheme does not require the updating of the individual η_t's (or equivalently, the θ_i's), resulting in the mixing of the chains to be both faster and more stable. This means that at every stage of the MCMC updates, there are $|\boldsymbol{\psi}|$ updates, instead of the $n + |\boldsymbol{\psi}|$ from before, speeding up the computational complexity of the algorithm as a whole from $\mathcal{O}(mn)$ to $\mathcal{O}(m)$.

Of course, by setting $\boldsymbol{\psi} = (\mu, \sigma^2)$, the simple model from Section 4.2 can be rewritten too. The ability of the two different methods to estimate the bias and model variance in a simulated dataset, similar to that outlined in Section 4.2.6, will be assessed so that any expected improvement in performance from using the marginalised posterior distribution can be revealed.

A foreseeable issue encountered here is that, as before in the simple model, the prior choice of the underlying distribution of the true probabilities has a large effect on the success of inference. To combat this, a similar technique as before is adopted; treating the problem as taking part in three stages. Note that although the η_t's have been marginalised out, the inference still requires the specification of their prior distribution.

1. **Exploratory Analysis:** Place an uninformative prior upon belief in the underlying distribution for the η_t, here $\eta_t \sim \mathrm{N}(0, 10)$. Gain inference about the bias terms μ_0 and γ.

2. **Inference upon m and s^2:** Let $\hat{m}$ and $\hat{\sigma}^2$ be the mean and variance of the observed $\hat{\boldsymbol{\eta}}$. Then let the estimated 'average' bias size throughout time be

$$\overline{\mu}_t = \frac{1}{n} \sum_{t=1}^{n} \overline{\mu}_0 t^{\overline{\gamma}}. \tag{4.3.9}$$

Note that $\bar{\mu}_0 = \frac{1}{m} \sum_{j=1}^{m} \mu_{0(j)}$, as before. The distribution of the underlying probit-transformed probabilities can then be estimated as

$$\eta_i \sim \mathrm{N}(\hat{m} - \bar{\mu}, \hat{s}^2) \tag{4.3.10}$$

3. **Model Inference:** The MCMC scheme is then run again, with the prior for the η_t's being chosen via the work in Step 2, and the priors for μ_0 and γ also being chosen to match the mean and variance of Step 1's posterior estimate for their distributions. The posterior means of this function of the MCMC scheme are taken as being the resultant parameter estimates.

The equivalent additive bias shown in the previous case as being derived from equation (4.3.10) is written under this new modelling approach as

$$\mathbb{E}(\theta_t|\hat{\boldsymbol{\eta}}, \boldsymbol{y}) \approx \Phi[\Phi^{-1}(\hat{\theta}_t) - \mu_t]. \tag{4.3.11}$$

Due to the change to a probit link, and the non-time-varying parameters, this process changes slightly to that seen before in Algorithm 1, and is outlined below:

Algorithm 2 Two-Stage Bias Estimation Approach for Time-Varying Parameters

Input: Set of n pairs of data, (θ_i, y_i), $i = 1, \cdots, n$.
Perform MCMC analysis with uninformed priors upon the posterior distribution in equation (4.3.8), let $\overline{\mu}$ be that shown in equation (4.3.9)
for all $\hat{\theta}_i$ **do**
Find $\mathbb{E}(\theta_i|\hat{\boldsymbol{\eta}}, \boldsymbol{y})$ via equation (4.3.11).
end for
Find the mean and the variance of $\hat{\boldsymbol{\eta}}$, the fit a Normal distribution. This is an informed prior for the η_i's.
Perform MCMC analysis of the posterior distribution in equation (4.3.8), outputs taken as forming the posterior estimate for the distribution of μ_t and σ_t^2.

4.3.2 Comparison

The case where the error's parameters are time-varying is compared against an approach where the parameters are assumed to be static in a partition of time, but different in different elements of the partition (the "static method"). In the latter case, the data are partitioned into R subsets $\boldsymbol{r} = \{(0, r_1), (r_1, r_2), \cdots, (r_{R-1}, n)\}$, not necessarily of the same size. The static parameter work of Section 4.2 is then used, having been improved via the probit link paradigm to estimate the bias and model variance within each of the data subsets. Given the R parameter estimates for both the bias and model error, curves can be fitted to attempt to match the time-varying structure imposed by equation (4.3.1), with this calculated curve being the basis of inference.

Let the midpoints of the subsets be written as $\tilde{\boldsymbol{r}} = (\tilde{r}_1, \cdots, \tilde{r}_R)$, where $\tilde{r}_i = 0.5(r_{i+1}+r_i)$ etc, and let $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \cdots, \tilde{\mu}_R)$, $\tilde{\boldsymbol{\sigma}}^2 = (\tilde{\sigma}_1^2, \cdots, \tilde{\sigma}_R^2)$ represent the parameter posterior mean estimates from the static model, for each of the R subsets, for the bias and model variance respectively. Given this, a curve is selected as to minimise the Euclidean norm for the distances between some underlying curve of the type defined in equation (4.3.1) and the estimated points $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\sigma}}^2$, weighted by the size of each of the subsets of the data:

$$(\hat{\mu}_0, \hat{\gamma}) = \underset{(\mu,\gamma)}{\arg\min} \sqrt{\sum_{i=1}^{R} w_i \left(\mu \tilde{r}_i^{\gamma} - \tilde{\mu}_i\right)^2},$$

$$(\hat{\vartheta}, \hat{\varphi}, \hat{\nu}) = \underset{(\vartheta,\varphi,\nu)}{\arg\min} \sqrt{\sum_{i=1}^{R} w_i \left(\vartheta + \varphi \tilde{r}_i^{\nu} - \tilde{\sigma}_i^2\right)^2}.$$

where w_i is the length of the i'th subset. The addition of the weightings in the fit give more importance on the fit of the curve being close to observations in the places where more data are found. From the example in Figure 4.3.4, more emphasis is placed on a good fit of the curve from the flatter part of the curve, from the 1000'th data point onwards. This is due to a perceived greater importance of using this work to understand how the model will improve or deteriorate into the future, rather than understanding how the model improved or deteriorated in the past.

In this case, the subsets were split such that they provided inference on the observed data at higher frequency at the beginning of the dataset, where the change in the underlying parameter was most notable, in comparison to the end of the dataset, where the underlying parameter is largely unchanging. In the case where $n = 5,000$, Figure 4.3.4 shows how a particular choice of subsets partitions the dataset.

Clearly, the choice of subsets has a large affect on the efficacy of the static technique. The larger the subset, the more accurate the model inference, however, if the subsets are too large, then the change in the underlying parameter could fail to be captured.

The two techniques were given $10,000$ pairs of data to provide inference. The static method partitions the data such that $\boldsymbol{r} = (250, 500, 1000, 1500, 2000, 3000, 4000, 5000, 7500)$. Both methods were given 100 sets of data, each with the underlying error distribution parameters as those shown in Figure 4.3.4.

The static method's prior choices for μ and σ^2 were the same as those seen in the comparison of the simple model, shown in Section 4.2.6, i.e. $\pi(\mu) \sim \mathrm{N}(0, 1)$ and $\pi(\sigma^2) \sim \text{Inv-Gam}(1, 5)$, representing a general lack of knowledge regarding the

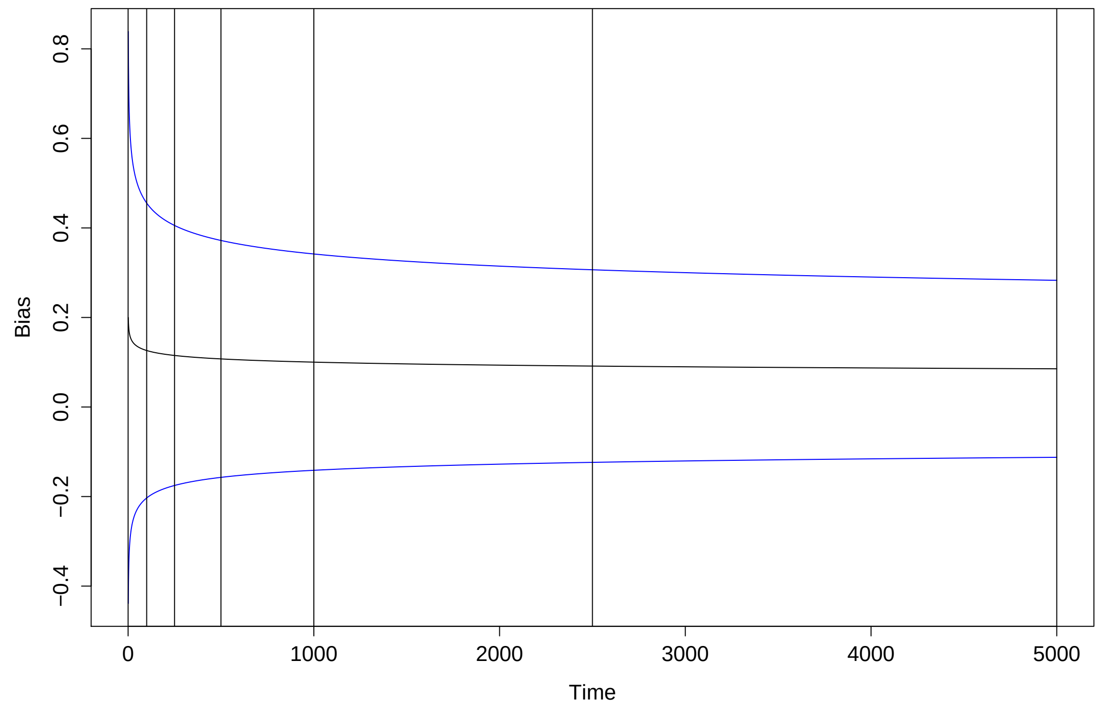


Figure 4.3.4: The change in the bias over time, given a certain set of inputs: $\boldsymbol{\psi} = (0.2, -0.1, 0.002, 0.01, -0.3)$. The black curve shows the change in the bias, flanked by 95% confidence intervals, representing the error over time. The vertical lines partition the data into subsets, where inference is performed separately.

error's behaviour. In the case for the time-varying parameter's prior choice, a similar approach was taken, with μ_0, γ, ν's prior being represented as the party-uninformative Normal distribution, $\mu_0 \sim \mathrm{N}(0, 1)$, and with ϑ, φ's prior choice being the partly-uninformative Inverse-Gamma distribution, $\vartheta, \varphi \sim \text{Inv-Gam}(1, 5)$.

Table 4.3.1 summarises the efficacy of both of these methods with respect to estimating the nature of the underlying curves, describing the error distribution's mean and variance, via calculation of the respective fit's root mean square error (RMSE). The RMSE here is the error in the time-varying parameters, μ_t and σ_t^2, evaluated at each time point $t = 1, \cdots, 10,000$.

	Mean Estimates		Variance Estimates	
Technique	Average RMSE	Max RMSE	Average RMSE	Max RMSE
Static	4.495	6.829	4.578	5.226
Time-Varying	1.818	4.656	2.207	2.797

Table 4.3.1: A comparison of two techniques to estimate the time-varying behaviour of the parameters of the error distribution, for 100 datasets.

As is clear, treating the error distribution's parameters as being time-varying leads to a much better estimate of its behaviour, over the rival technique of assuming that the parameters are not time-varying in a partition of the dataset. This superiority is most evident when the error distribution's variance is estimated, where even the worst fit performed by the time-varying model still performed much better than the best fit achieved by the static approach (with RMSE of 4.256).

4.4 Application Investigation

An interesting application of this work is to consider a betting market to be some form of model, whose output of odds for certain events can be thought of as estimates for the probability of the events occurring. It is therefore interesting to test whether the bookmaker's odds can be used as being predictive of the actual probability of the events occurring. A secondary but complementary study, then, would be to investigate whether there is some range of probabilities for which betting arbitrarily within this range results in profitable wagers.

The dataset is the same as that used in Section 4.2. As was alluded to in Figure 4.2.4, the bias may not be independent of the event probabilities, and it is this feature that the model is attempting to detect.

With this is mind, the data were grouped by the probability implied by the betting odds. The odds were transformed into their implied probabilities, by the method of Khutsishvili (Vovk and Zhdanov, 2009). The probability space $(0, 1)$ was split into equal partitions. Clearly, the amount of data in each interval varies depending on the underlying distribution of the probabilities.

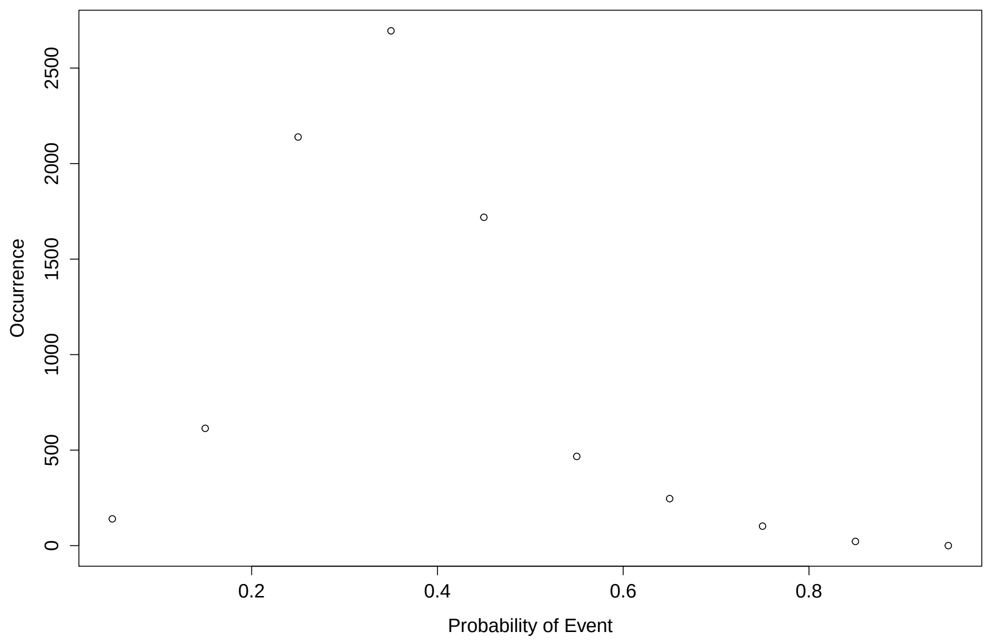


Figure 4.4.1: The number of occurrences of probabilities with subsets of the domain $(0, 1)$. The underlying dataset is the probability of home wins for Premier League football, as defined by the bookmaker, William Hill.

For home wins in isolation, Figure 4.4.1 shows the distribution of probabilities. As can be seen, when the probability of events was close to 0, or above 0.8, there are very few occurrences, with none existing in the interval $(0.9, 1]$. Due to this, the data are partitioned in such a way that each subset of the data contains at least 50 pairs of data. The same is true of all examples shown in this section.

The static model, with probit link, was applied to each subset of the data, with

Figure 4.4.2 showing the estimates for the distribution of the model error for each subset, along with error bars signifying the uncertainty in the MCMC posterior estimates. The black line represents the mean of the posterior estimate for the bias for each of the subsets, with the error bars giving a representation of the uncertainty in this posterior estimate. Similarly, the red lines show a 95% confidence interval representing the mean for the posterior of the model variance. Again, the error bars around the model variance lines show the uncertainty regarding the model variance's posterior estimates.

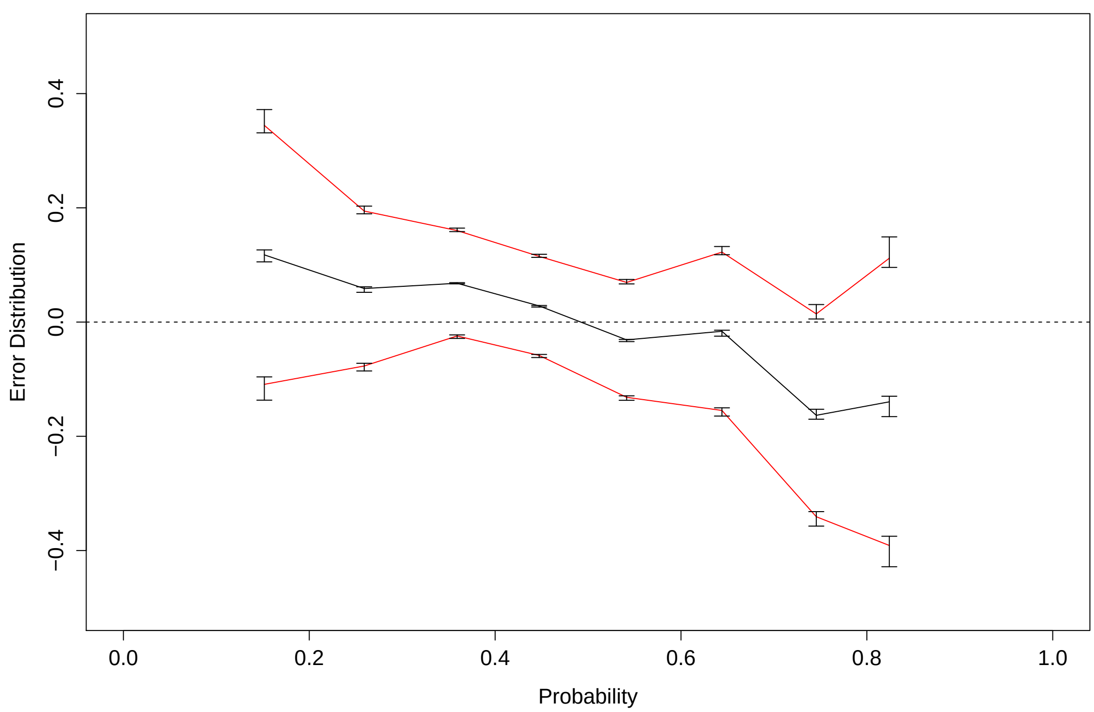


Figure 4.4.2: The modelling error (as defined by equation (4.3.5)) in bookmaker's odds of home wins, broken down into subsets, dependent upon the estimated probability for each event. The black line indicates the bias, with the error bars indicating the uncertainty in the error. The red lines represents the size of the model error, via a 95% confidence interval.

When considering the bookmakers' markets as being in some way predictive of the true probability of events occurring, the location of the black line in comparison to the horizontal dotted line in Figure 4.4.2 gives an idea of whether the probabilities are over or under-estimated. An estimated bias being greater or less than zero implies that the probability inferred from the odds are less than or greater than the true probabilities, respectively.

To make this relationship clearer, equation (4.3.11) is used to transform the results shown in Figure 4.4.2 onto the probability scale, and is shown in Figure 4.4.3.

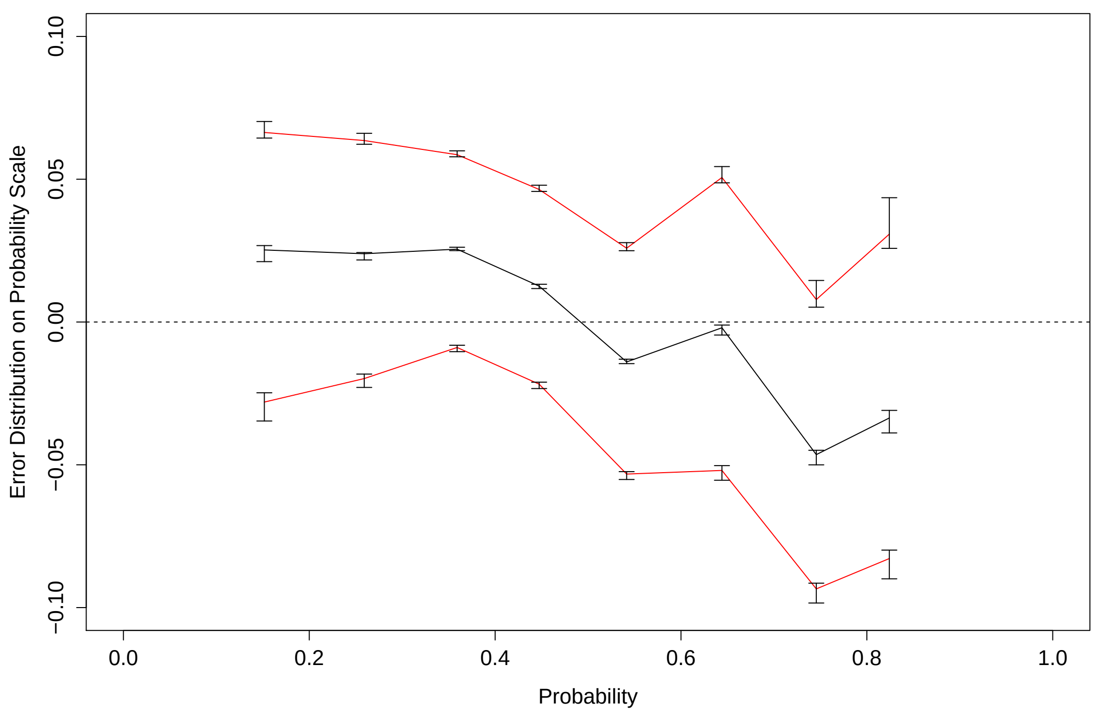


Figure 4.4.3: The error in probability (as defined by equation (4.3.11)) in bookmaker's odds of home wins, broken down into subsets, dependent upon the estimated probability for each event. The black line indicates the bias, with the error bars indicating the uncertainty in the estimates. The red lines represents the size of the model error, via a 95% confidence interval.

Both Figures 4.4.2 and 4.4.3 show the same general relationship, when the underlying probability of the event is small, the probabilities tend to be underestimated, while the probabilities are overestimated when the underlying probabilities are high. When the probability is in the range $(0, 0.5)$ (which accounts for the majority of all offered odds, see Figure 4.4.1), the probabilities from the bookmakers are overestimating the truth. This shows that if the bookmakers' odds were being used as a predictive tool for estimating the probability of events occurring, the majority of the time the probabilities will be overestimated.

Similar analysis can also be used to answer the related question regarding the profitability of betting on events with different underlying probabilities. The main difference in the analysis is that the process to remove the overround from the markets should be omitted, as any investor would have no choice but to bet at the offered odds. When this is taken into account, the detected error over each of the probabilities can be seen in Figure 4.4.4.

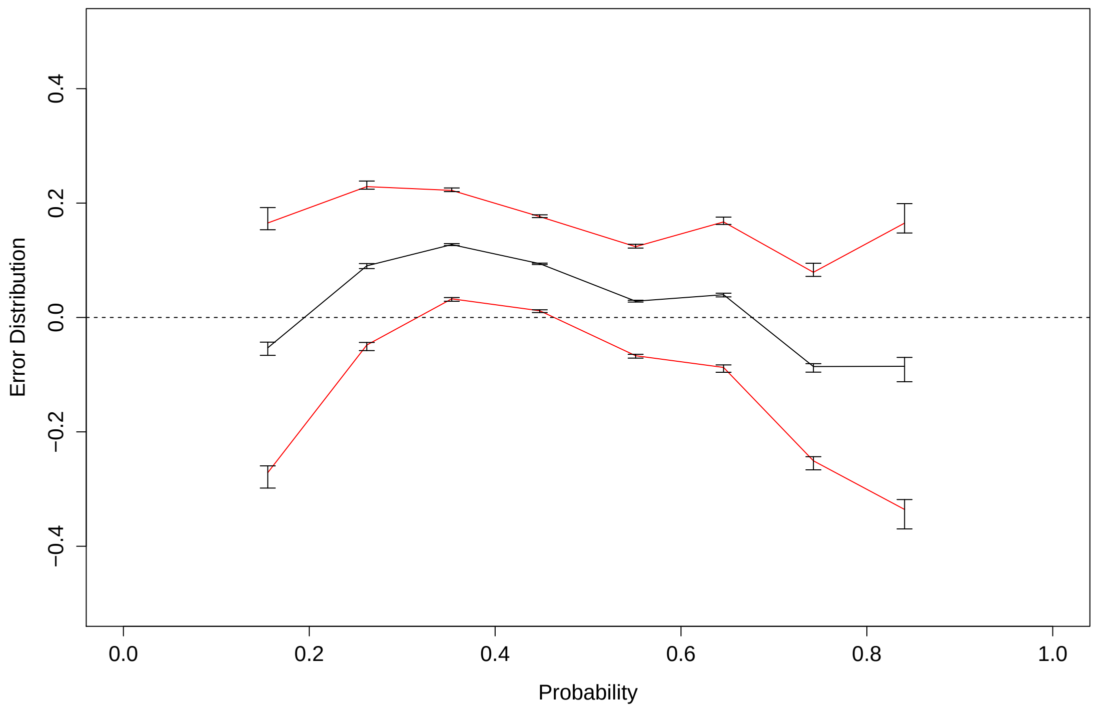


Figure 4.4.4: The modelling error (as defined by equation (4.3.5)) in bookmaker's odds of home wins, broken down into subsets, dependent upon the estimated probability for each event without considering overround. The black line indicates the bias, with the error bars indicating the uncertainty in the error. The red lines represents the size of the model error, via a 95% confidence interval.

Again the log-odds are transformed onto the probability scale, with the results shown in Figure 4.4.5.

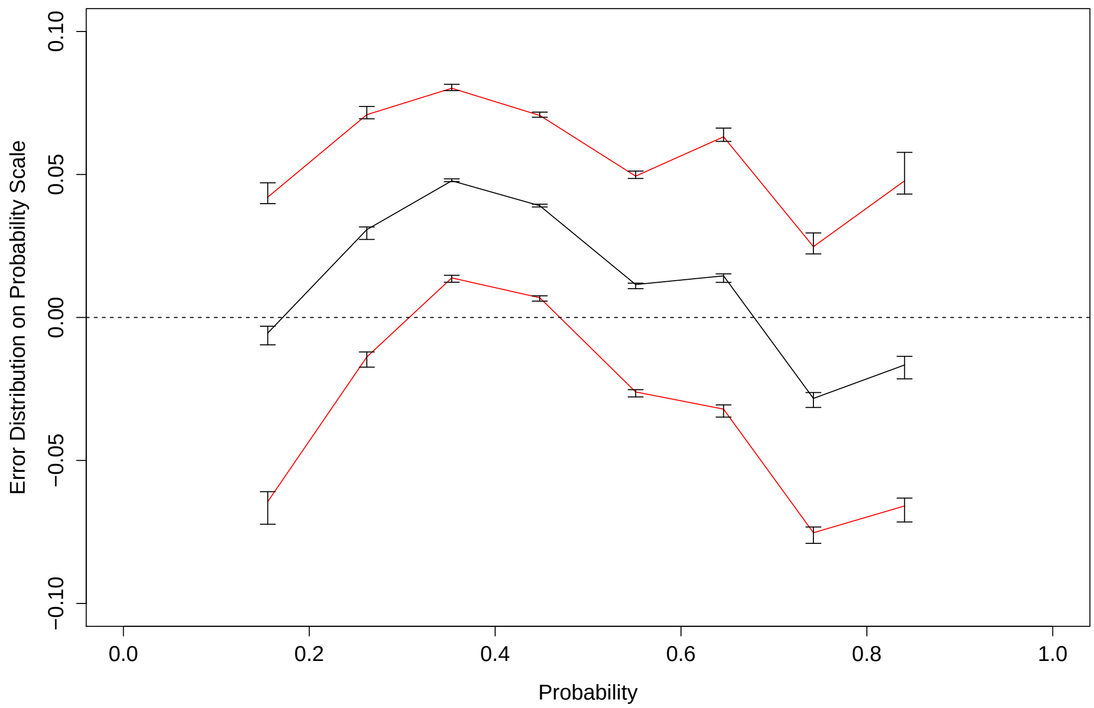


Figure 4.4.5: The error in probability (as defined by equation (4.3.11)) in bookmaker's odds of home wins, broken down into subsets, dependent upon the estimated probability for each event, without considering overround. The black line indicates the bias, with the error bars indicating the uncertainty in the error. The red lines represents the size of the model error, via a 95% confidence interval.

The profitability of betting on this particular bookmaker market is seen in Figure 4.4.5. For the large bulk of the data, (for underlying probabilities in the interval $(0.15, 0.65)$, around 76% of the dataset), the probabilities derived by the odds were higher than the truth. This means that the odds offered were lower than they should have been in a "fair" market, and bets are not profitable. Interestingly, probabilities outside of this region seem to give profitable outcomes, as the odds offered by the bookmakers are generous in comparison to the truth. This potential truth can be cross-referenced against Appendix B, in which this kind of analysis can be shown for different bookmakers and for different football leagues.

Appendix B shows that betting where the underlying probability of the event in question is very high seems to remain profitable when different bookmakers are considered. This becomes even more noticeable when the investor is considered to shop around the available bookmakers to find the highest offered odds. When this is done, see Figure B.0.2, the profitability of these bets reveals itself more sharply. In comparison, Figure B.0.1 seems to show that this relationship doesn't necessarily map over to different UK football leagues. Indeed, for the Championship and for League 2, the opposite seems to be true; betting on events with a very high probability of success seems to offer worse value than other bets.

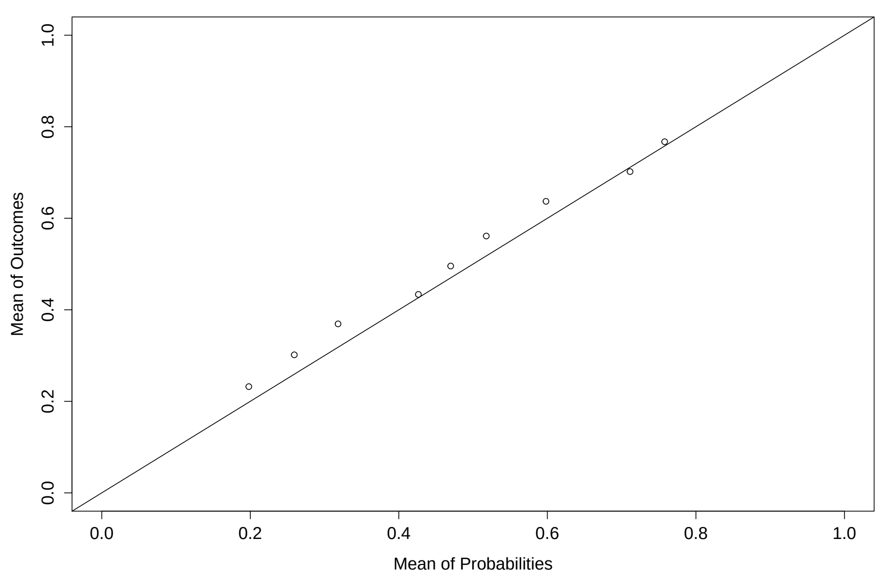


Figure 4.4.6: The proportion of outcomes resulting in a home wins, compared to the mean of their quoted probabilities, for each partition of the data.

This result is checked via a simple qq-style plot, shown in Figure 4.4.6, where the intervals are chosen to match those used in Figure 4.4.5, and others. It affirms the notion that certain subsets of the data may actually present profitable betting opportunities for an investor (although, it must be noted that this feature may merely be a quirk of this dataset).

Figures 4.4.7 and 4.4.8 describe the same comparison of bias against the event's underlying probabilities but for the match outcomes being an away win or a draw, respectively.

The clear relationship observed for the home wins is not so evident for the other two possible match outcomes. For away wins, there is no obvious trend linking the probability of events and the bias detected. The upper and lower confidence bands lie on either side of the line representing no modelling bias, the line of "fair prices".

In the case of the match outcomes being draws, there is one clear message, the odds always overestimate their probability. This also gives rather clear evidence that betting on draws on most events does not give rise to profitable wagers. This work is also applied to comparing the bias in bookmaker's odds across football leagues, and across different bookmakers in Appendix B. Figures B.0.1 and B.0.2 compare the model bias across football leagues and across bookmakers.

Figure B.0.1 shows the bias across the main 4 English football leagues. One of the most obvious differences between the leagues is that the Premier League exhibits more extreme odds, both high and low, than the other leagues. This is probably due to the fact that the Champions League teams, (usually the top 4 teams from the previous season) have a much higher ability to create money, and therefore invest in their teams. This means that when the top teams play the bottom teams, there is a higher difference in the quality of the teams than in the lower leagues.

The general relationship noted before seems to hold, albeit weaker in the lower

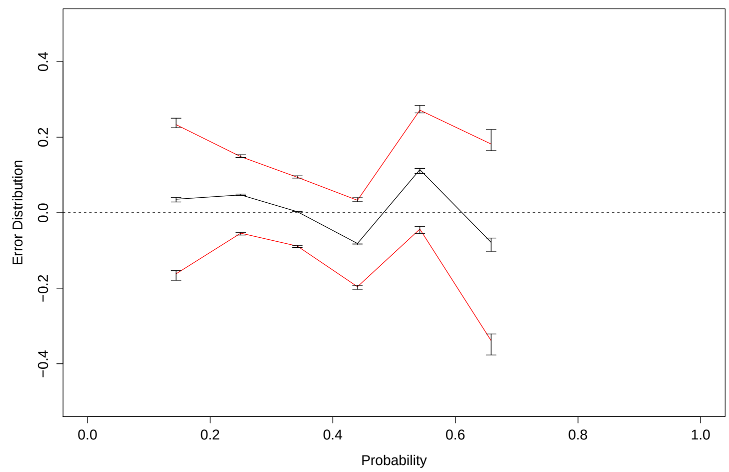


Figure 4.4.7: The modelling error in bookmaker's odds of away wins, broken down into subsets, dependent upon the estimated probability for each event. The black line indicates the bias, with the error bars indicating the uncertainty in the error. The red lines represents the size of the model error.

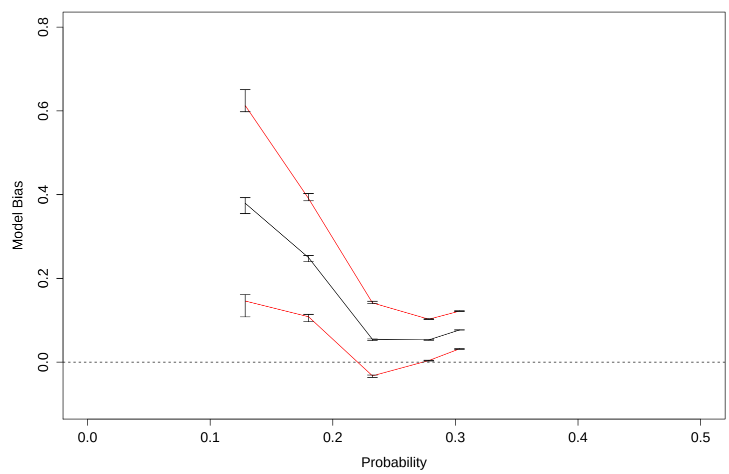


Figure 4.4.8: The modelling error in bookmaker's odds of draws, broken down into subsets, dependent upon the estimated probability for each event. The black line indicates the bias, with the error bars indicating the uncertainty in the error. The red lines represents the size of the model error.

leagues; that wagering on events with higher probabilities, and thus lower odds, results in more profitable outcomes than bets placed at lower probabilities. However, in both the Championship and League 2, events with the highest probabilities, being the events with the greatest difference in perceived quality between the two teams, the odds are unprofitable, relative to the events with slightly lower probabilities. This feature is also noted in the Premier League, but strangely, the reverse seems to occur in League 1.

Figure B.0.2 shows the bias across different bookmakers (William Hill and Bet365), along with the average odds of all of the bookmaker and the maximum odds offered by any of the bookmakers for each event (data collected from BetBrain). On the whole, the relationship between the bias and the underlying probabilities seems to be relatively unchanging among the different bookmakers. This is to be expected, as there tends to be not much difference between the offered odds of the different bookmakers, due to: their methods of setting odds being very similar; the makeup of their customer base being similar; and them being able to use each other's odds to set their own.

4.4.1 Time-Varying Bias

The time-varying model is demonstrated in two different contexts. Firstly, the bookmaker's data from Section 4.4 is tested to see if its bias is in any way time-dependent. Secondly, a simple predictive model for football is fit to past data, with the change in bias being detected as the model makes its inferences from a dataset which increases with size over time.

Bookmaker's Model

Given that there have been no great changes in how bookmaker's odds are formed in recent years, there is likely to be no systematic increase or decrease in the bias in bookmaker's odds over time. The model used was of the probit-link-type, with vague priors placed on each of the parameters ($\pi(\mu_0) = \pi(\gamma) = \pi(\nu) \sim \text{N}(0, 1)$, $\pi(\varphi) = \pi(\vartheta) \sim \text{Inv-Gam}(2, 0.2)$). The results of this analysis for home wins are shown in Figures 4.4.9 and 4.4.10.

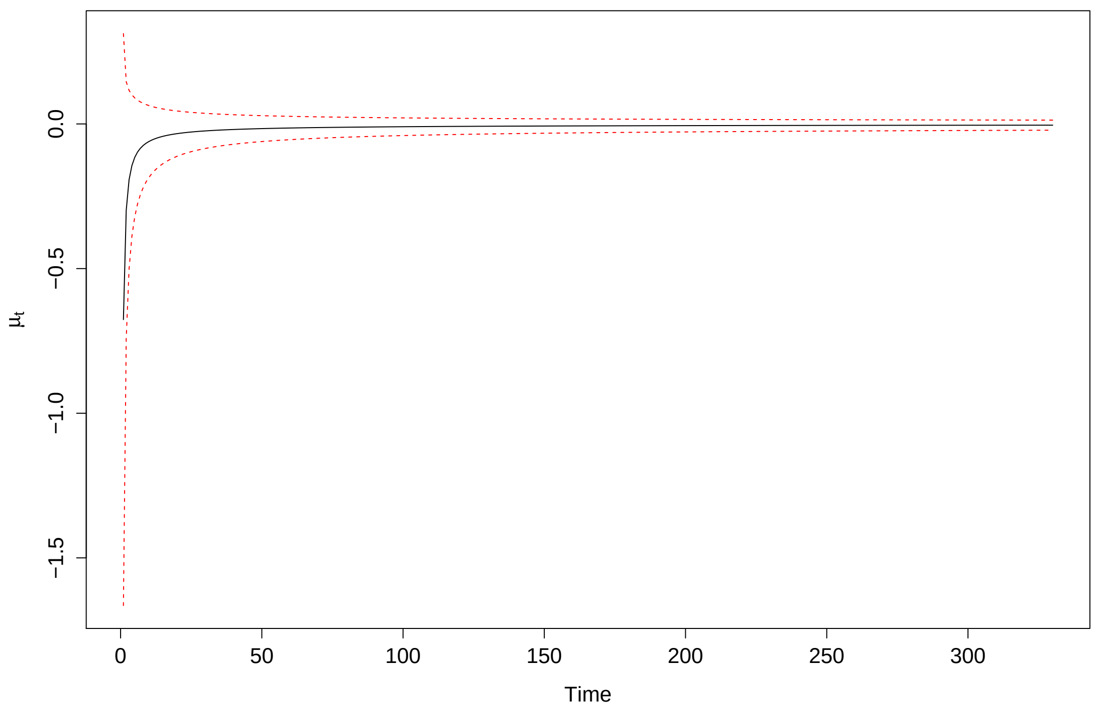


Figure 4.4.9: The posterior mean and 95% credibility region for the time-varying model bias μ_t for bookmaker's odds of home wins.

The model bias is initially detected to be small and negative at time 0, increasing and converging very quickly to 0. In addition, its 95% posterior credibility interval includes 0 for all time, indicating that there is no evidence of bias at any time. This

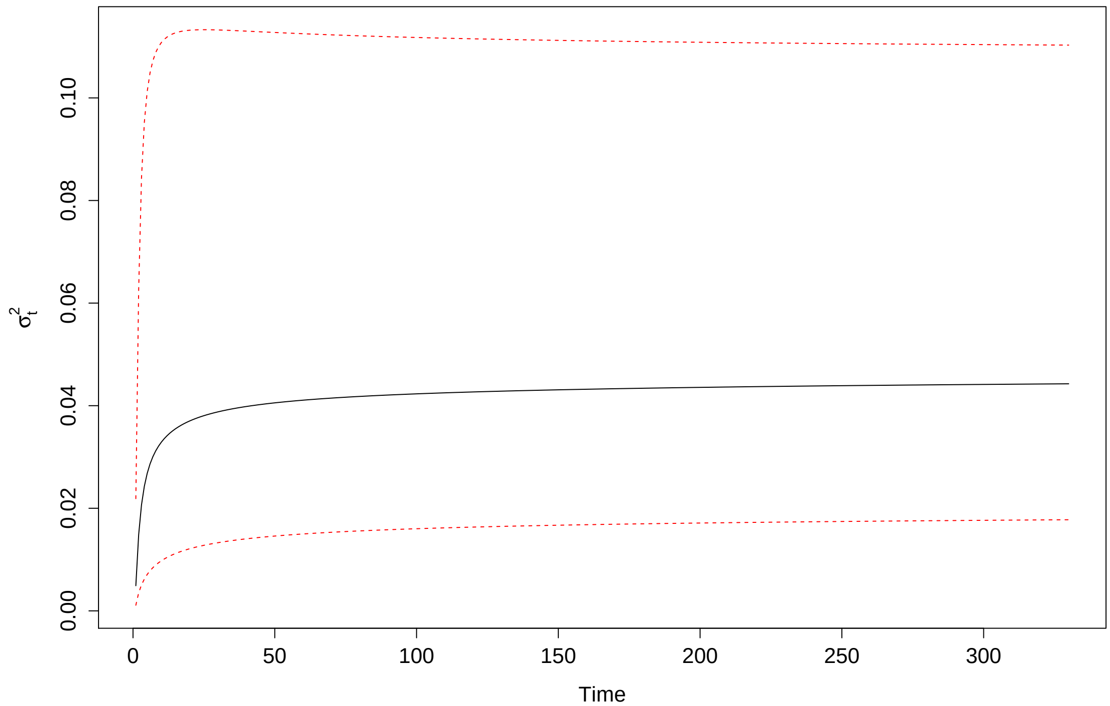


Figure 4.4.10: The posterior mean and 95% credibility region for the time-varying model variance σ_t^2 for bookmaker's odds of home wins.

credibility region shrinks over time, showing that more certainty regarding the bias size is being acheived over time.

The model variance increases quickly from 0.01 then plateaus to a fixed level of around 0.04. In this case, the posterior credibility region does not shrink over time to the extent seen for the model bias.

This demonstration has given useful insight into the nature of inference when the true bias is static, and not time-varying. Although the mean estimates for the model bias and error variance imply some striking time-varying behaviour, the 95% posterior credibility regions show that there is not enough evidence to support this conclusion.

Football Modelling

In Spiegelhalter and Ng (2009), a basic approach to predict the probability of the outcomes of football matches is presented. The approach is based on the idea that matches at the beginning of the season can be used to make inference upon the likelihood of match outcomes at the end of the season. The model is basic, providing an approximate form of the model given in Maher (1982), covered in Section 3.1.2. The average number of goals conceded by both home and away teams are calculated for the season as a whole, and for all teams. These values are multiplied by factors quantifying the relative attacking and defensive skills of the individual teams (note that home advantage is not considered). These values are then considered as arrival rates, and the probabilities of events are found via simulating from the Poisson distribution.

Data were again taken from `www.football-data.co.uk`, this time only consisting of the 2012-13 Premier League season. The odds data were not used; instead, the goals for and against teams were used to fit the model specified in Spiegelhalter and Ng (2009). As predictions are made on the basis of previous observations, predictions were only calculated after each team had played 5 games.

The probit-linked bias estimation method was used to estimate the time-varying nature of the model's bias. This estimate was then used to adjust, using equation (4.3.11), the predictions made for the next season in order to create unbiased predictions for the new season of matches. It is assumed, then, that the time-varying bias for the 2012-13 season is going to take a similar form to that in the 2013-14 season.

The data were also split, so that the probabilities of home wins, away wins and draws were considered separately. This should highlight if certain results' predictions are different to others. Figure 4.4.11 shows the results of this analysis.

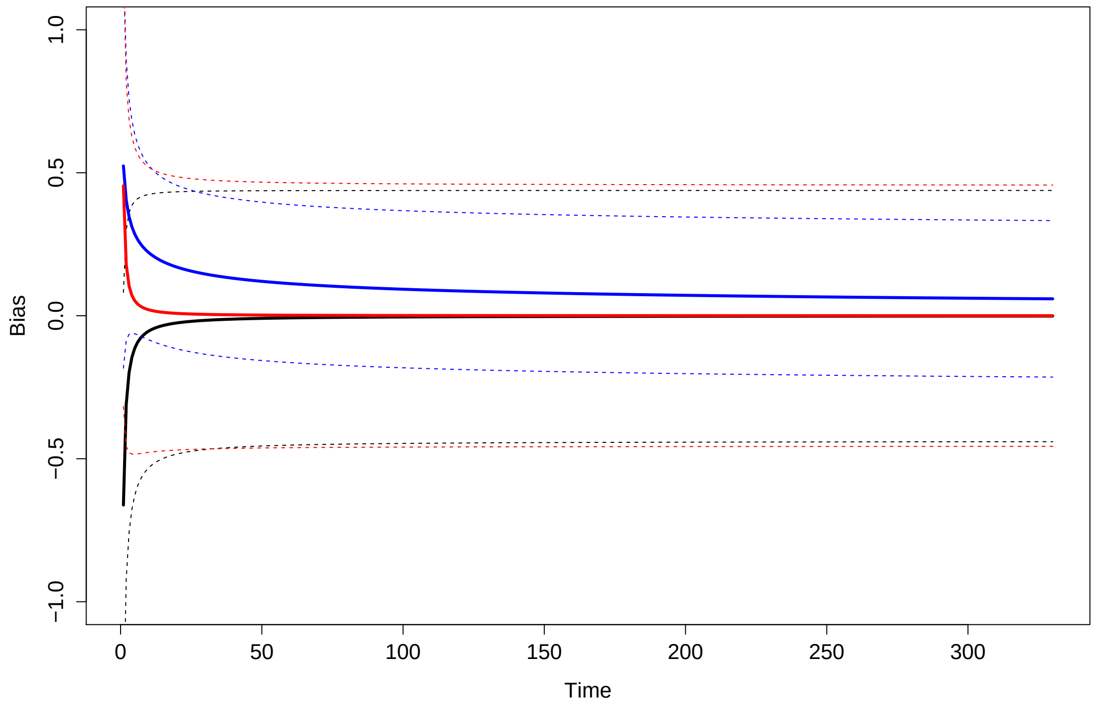


Figure 4.4.11: Model bias exhibited by the football prediction model described in (Spiegelhalter and Ng, 2009). The thick black line signifies the changing bias in the prediction of home wins, with the black dashed lines showing confidence intervals for the bias size. Similarly, the blue lines represent the probability of draws, and the red lines represent the probability of away wins.

At the initiation of the model, draw and away win predictions exhibited a positive bias, whilst home win predictions showed a negative bias. Whilst the home and away win predictions quickly improve as more results are observed, there remains an indication that the probability of a draw occurring might be being over-estimated for the duration of the prediction window, as seen by the solid blue curve in Figure 4.4.12.

On the probability scale, the additive error can be calculated via the transformation described in equation (4.3.11). For the draw probabilities, the model bias has a value of around 0.175 after two weeks of predictions with a 95% posterior credibility interval of $(-0.13, 0.4)$, whilst at the end of the season this has decreased to around 0.05, with a 95% posterior credibility interval of $(-0.26, 0.31)$.

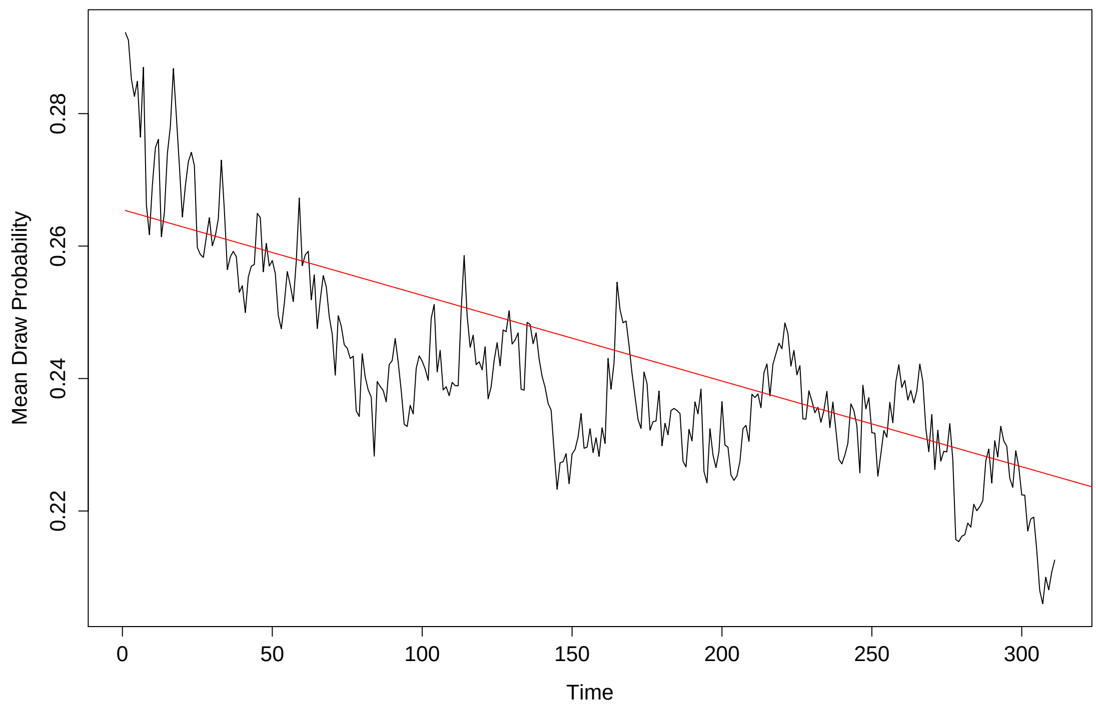


Figure 4.4.12: A moving average, of window size 20, of the probabilities of draws predicted by (Spiegelhalter and Ng, 2009) through the 2013-14 season. The red line is a fitted linear relationship, with equation $p = 0.2656 - 0.00014t$, where p is the probability of a draw and t is the time index for a match.

Inserting the values $\hat{\theta}_i = 0.26$, $\mu = 0.175$ into equation (4.3.11) gives a true probability of 0.2066, corresponding to the true probability of a draw at the start of the season. The true probability of a draw at the end of the season is found in the same way, this time with $\hat{\theta}_i = 0.22$, $\mu = 0.05$, giving a probability of 0.2055. This indicates that in reality, the probability of a draw does not change significantly throughout the season, and that the change observed by the model outputs are down to its bias, which slowly corrects as more data are collected.

In comparison, the probability of draws derived from the Spiegelhalter and Ng model seems to exhibit a downward trend, represented by the simple linear fit, shown as the red line in Figure 4.4.12, which tallies with the bias estimation, shown in Figure 4.4.11.

In order the assess the efficacy of using the bias estimation outputs in order to improve the model performance in the future, the Spiegelhalter and Ng (2009) model was run again, this time using data from the 2012-2013 Premier League season. The aim of this work is to use the estimates for the bias movements over time from the analysis on the 2013-2014 season to provide better estimates for the new data.

Given that the evidence suggested that the home and away probability estimates were, on the whole, accurate, whilst the draw probability estimates improved over time, bias correction is attempted only on the draw probabilities. Figure 4.4.13 gives the evolution of the bias for the 2013-2014 season, after the draw probabilities have been 'corrected' given the previous year's bias estimate.

Clearly, the bias exhibited for draw probability predictions has decreased significantly (shown by the thick blue line). Figure 4.4.14 gives a moving average relating how the draw probability estimates change over time, using the corrected model.

Figure 4.4.14 clearly shows that that the draw probability estimates created by the corrected model do not change significantly over time, like they did in the uncorrected model. Indeed, aside from the first 30 estimates, the draw probability estimates seem to not drift, as shown by the linear fit's gradient being very close to 0. The low initial values are another indication that the time-varying bias structure, imposed in

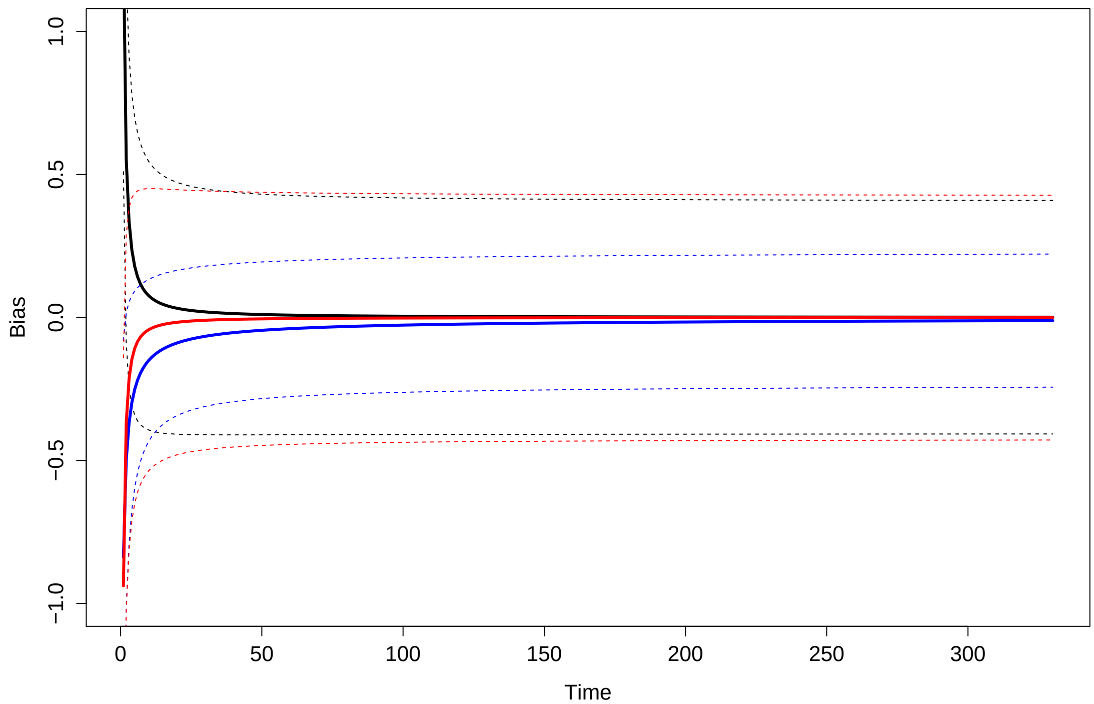


Figure 4.4.13: Model bias exhibited by the football prediction model described in (Spiegelhalter and Ng, 2009). The thick black line signifies the changing bias in the prediction of home wins, with the black dashed lines showing confidence intervals for the bias size. Similarly, the blue lines represent the probability of draws, and the red lines represent the probability of away wins.

equation (4.3.1) are not appropriate, as the large initial bias deviates the corrected model away from the truth. This indicates that, although correcting models in this way gives improved model performance, giving an accurate form for the time-varying bias to take is also essential.

4.5 Conclusion

This work has demonstrated how Bayesian methods can be used to assess how bias is manifesting itself in a predictive model which outputs Bernoulli outcomes. This

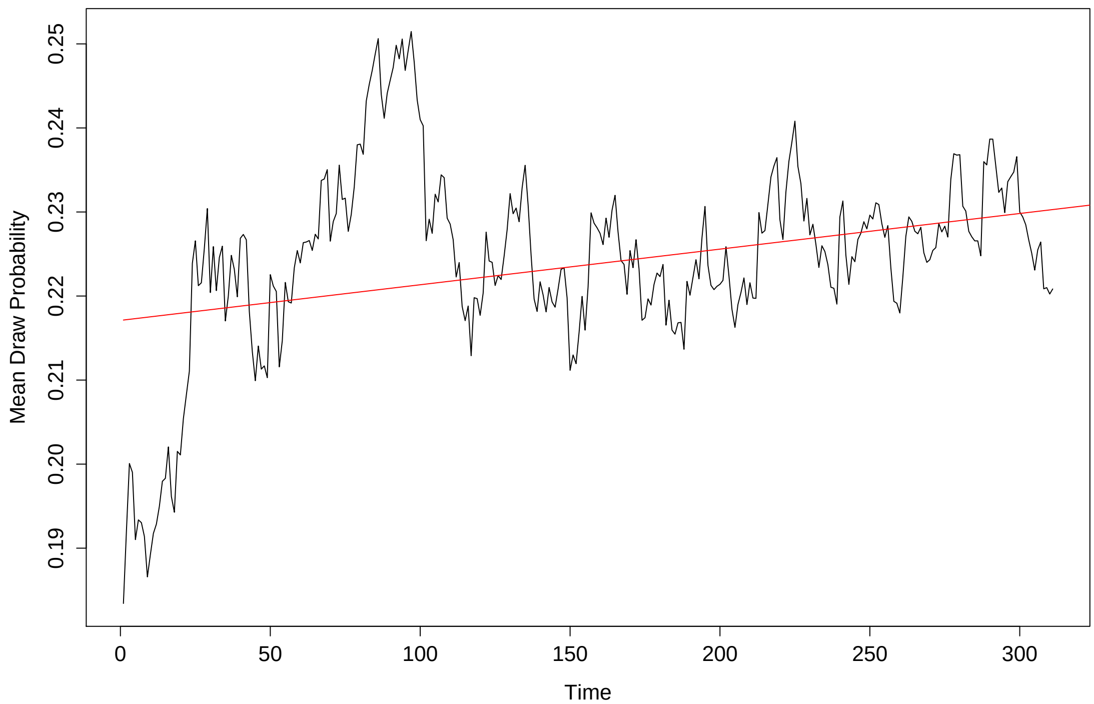


Figure 4.4.14: A moving average, of window size 20, of the probabilities of draws predicted by (Spiegelhalter and Ng, 2009) through the 2012-13 season. The red line is a fitted linear relationship, with equation $p = 0.217 + 0.00004t$, where p is the probability of a draw and t is the time index for a match.

inference is extended to include both static and time-varying forms of both bias and model variance.

It is demonstrated that, by using a probit-link on the model probability estimates, faster and more accurate bias evaluation can be carried out by the use of Bayesian methods. It is also demonstrated (in Section 4.3.2) that by treating time-varying bias as a smooth function over time, then inference gained is more accurate than if the bias is treated to be piecewise static.

Analysis on betting markets has shown that for the large majority of potential bets, bookmakers gain a significant advantage, as their odds predictions are biased in their favour. There exists, however, a subset of bets (for the Premier League, this seems to be for probabilities higher than 0.75) for which the advantage is with the bettor.

An investigation into a simple sports modelling method showed the utility of the bias detection method in estimating a time-varying bias, as well as a potential downfall of the method. In reality, the form of the time-varying bias and model variance, shown in equations (4.3.1) and (4.3.2) can be adjusted without the form of the joint posterior, (4.3.8) changing in any problematic way. It can be foreseen that the structure of the time-varying parameters can be chosen, given some idea about bias evolution in the particular dataset of interest.

Another potential area for future study would be to investigate how different forms of model output affect the ability of this technique to assess the bias. As an example, if the Bernoulli outputs are replaced by Binomial outputs, each observation (now of the number of trials resulting in a success for each probability estimate) gives much more information than before, and would therefore foreseeably give more accurate estimates of the bias.

Chapter 5

Pre-Match Market Movements

5.1 Introduction

Sports betting markets evolve between their inception and the time when the event in question is completed, and all bets are settled. These evolutions can be broadly split into two major time-periods:

1. **Pre-Match Market:** The section of the betting market which covers the time when the market is created up to the time when the event of interest commences. This period of time could last a few days, but for events organised long in advance (like a qualifying match for a major football competition), this period can cover weeks, or even months.

2. **In-Play Market:** The section that covers the duration of the event of interest. This period of time contains the playing time of the event, as well as any breaks in play, such as half-time (football), or the changing of batting team (baseball

or cricket), etc.

Recently, many papers have been published analysing how betting odds are formed, how fair they are, and how they evolve for the duration of a sporting event, see Section 3.2, and Williams (1999), ? for further information. In comparison there has been no work investigating how pre-match markets evolve over time. This is surprising, as not only does most of the betting activity occur pre-match (note that intra-market betting is a very recent invention), but also there is some clear structure to the movements in particular features of pre-match betting markets. In order to investigate this structure, a dataset was gathered, consisting of the pre-match movements of a number of betting market features. A summary of the dataset, along with a description of some pre-analysis preprocessing is described at length in Section 5.1.1

As motivation, consider a summary of the movements in a market on a betting exchange. Figure 5.1.1 shows the pre-match movements of four market features for a particular football match, over a period of 9 days. Note that each of these series can be thought of as time-series, and will be referred to as such for the rest of this work.

The overround quoted does not, as is usually the case refer to the sum of the implied probabilities for each of the possible event outcomes (here, home win, away win, draw). Instead, it refers to the overround between the back and lay prices. Recall from Section 2.2 that the set of possible outcomes, $\{H, D, A\}$ have the property that betting on one of them is equivalent to laying the other two options, and vice versa. If only the odds for backing and laying a particular event are quoted, the overround can be found via:

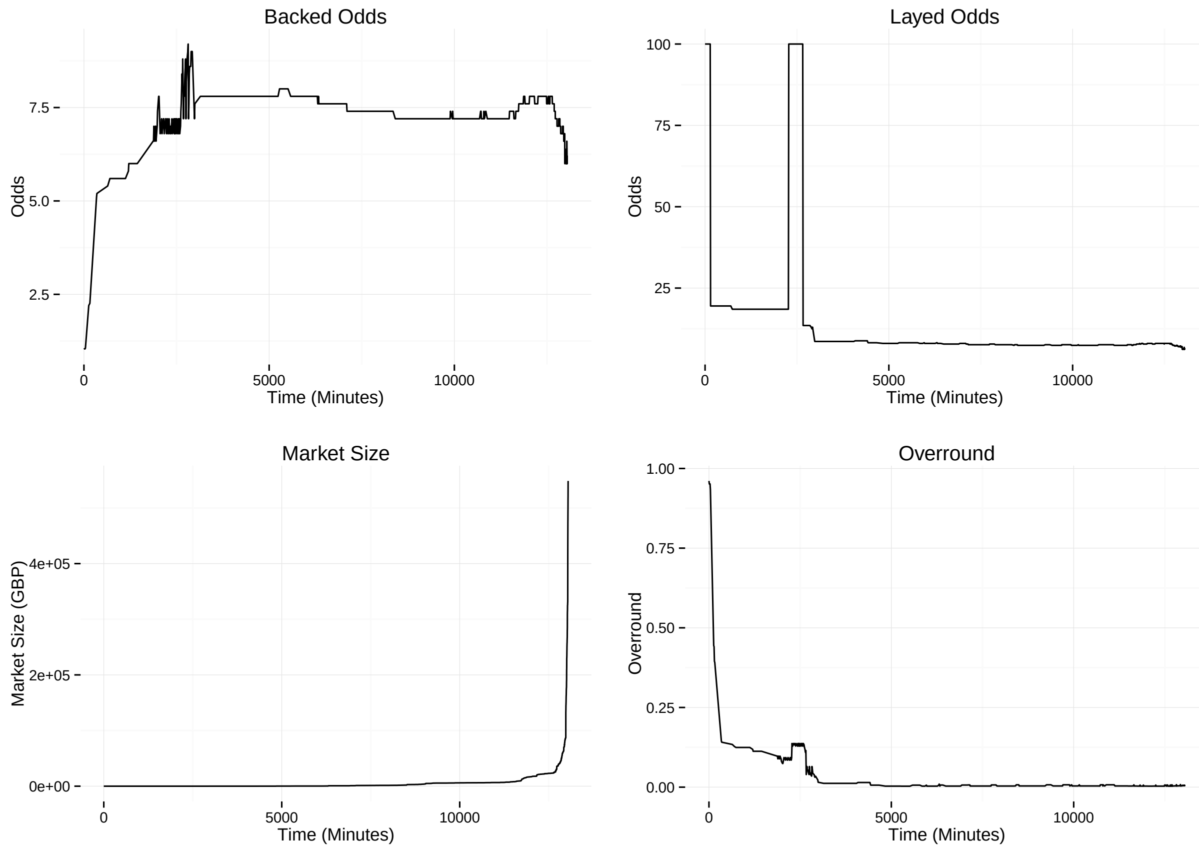


Figure 5.1.1: The change in backed odds, layed odds, market size and overround during the pre-match period of the betting market for Sunderland beating Manchester City on the 1st January 2012.

$$\kappa_t = \frac{1}{b_{H,t}} + \frac{l_{H,t} - 1}{l_{H,t}} - 1$$

where $b_{H,t}$ signifies the odds available for backing the home win, at time t, similarly $l_{H,t}$ signifies the odds available for laying the home win at time t. Note that a simpler, alternative measure could be achieved by simply taking the difference between the backing and laying odds; however this measure is sensitive to the scale of the odds.

Given this definition, many features of the evolution of the markets contained within Figure 5.1.1 can be noted.

- The change in the odds for both backing and laying are highly volatile as the betting market begins. This volatility seems to settle down after a few days, before increasing again for the hours before the event commences.

- The size of the market's overround decreases quickly as the market evolves from its commencement. This is an indication that the odds become more fair and closer to their fair values as the market evolves.

- The amount of money invested in the market increases at a faster rate immediately before the event starts. Therefore, nearly all of the capital bet on the event is committed in the last few hours and minutes before the event starts, when all the information which is pertinent to the event is available (such as news concerning the teams, weather, etc.).

At the beginning of the market, the volatility shown is due to the market in question not being *well-formed*. Due to betting exchanges being created entirely by investors backing and laying events, the initial market can be altered significantly by very little activity. At the very beginning of the market, the backing and laying odds offered are very poor, as there is little competition amongst investors. The convergence of the backing and laying odds can be seen as being the result of increasing competition between the opposing investors (see Williams and Paton (1997), etc.).

The formation of the market can therefore be identified via two potential factors, the market size and the overround. Figure 5.1.2 shows the market represented in Figure 5.1.1, but with burn-in removed, defined as being the period where both the overround is under 5%, and the market size is over £1000. In addition, the market size has been replotted on a log scale, such that the movements can be more easily seen. This now represents the period where the market is well-formed.

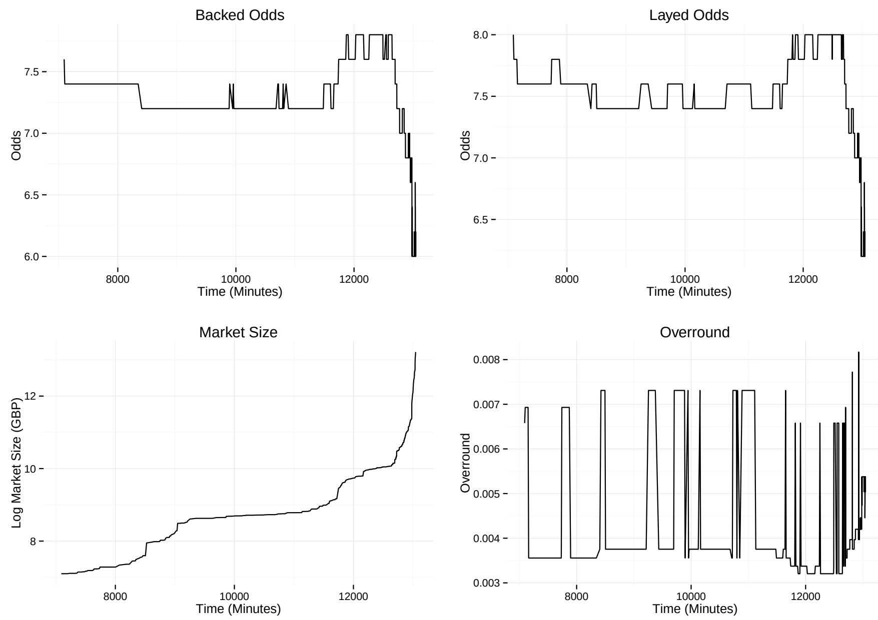


Figure 5.1.2: The change in backed odds, layed odds, log market size and overround during the pre-match period of the betting market for Sunderland beating Manchester City on the 1st January 2012, after removal of a suitable burn-in period.

Removing the burn-in leaves around 4 days of data leading up to the event. Given that the noisy initial set of market movements has been removed, some of the finer structure can be observed.

- The market size seems to increase roughly linearly on the log scale, up to a point very close to the event's commencement, when the rate of money being invested into the market increases sharply.
- The sharp increase in money invested in the market has a direct effect on the odds movements, which exhibit a concurrent and significant change.
- The overround remains at some very low value, once the market has been formed.

Both the market liquidity and the rate at which capital enters the betting market have an effect on how other important market features evolve. When the market is illiquid, small stakes by investors can cause the odds to move a large amount. Liquidity implies both a fair price, and a greater resilience of the market to change, given a constant stake size. A sudden increase in market size implies a release of information into the market with the odds reacting concurrently to these changes. For this reason, this work will concentrate primarily on gaining insight into how the market size increases over time. The aim of gaining this insight is to use predictions of the growth of market size to inform other predictions, such as the movement of the backing odds over time.

5.1.1 Summary of the Dataset

In total, the dataset covers 1244 football matches, covering a two year period from January 2012 to March 2014, and relates to both the English Premier League and the Spanish Primera Division. The duration of the market data range from 5 to 40 days before the event, but the vast majority covered the period of 5 days preceding the time of the matches. The pertinent features of betting markets which are covered by the dataset are not only the backing and laying odds, but also the amount of money in the market (the *market size*), and the amount of money available for exchange at the best current odds for both backing and laying (for a full treatment, see Chapter 6). All data was taken from BetFair, a popular betting exchange (see Section 2.2 for more information).

The most popular bet seen in the data set was for Arsenal to beat Newcastle, at home, which took over £4.5 million in bets. The least popular market was for Almeria to beat Athletico Madrid away from home, which only matched £850.

The intention is that inference regarding the evolution of the market size will be made from a set of historical observations of market movements, whose structures are assumed to be similar to current observations. Recall that rapid market changes occur during the last few hours before the start of the event (see Figure 5.1.2). Within the second tranche of the data, only a few data points are collected for the whole of this very interesting period of time, which does not allow the structure of the market movements to be captured. For this reason, the second tranche is discarded, leaving only the finer-detailed first tranche to be used for analysis. Unfortunately, this decreases the number of matches available in the dataset from 1244 to only 242.

The data is also simplified by restricting the time-period of the pre-match market movements to 5 days. This is important as it forces each of the time-series in the historical set to commence at the same point in time, making analysis of the dataset much easier. At this point, the matches which show the most, and least interest are quite different. The biggest interest in a market 5 days before the event itself was for Manchester City to beat QPR at home, in which £125,000 had been matched. The least popular market was for Espanyol to beat Seville away from home, for which £0 has been matched 5 days before kick-off. Interestingly, this shows that by far the most popular markets are those for which the home team is the heavy favourite, and the least popular markets are those for which the away team is the overwhelming underdog.

The dataset is now somewhat limited, consisting of 242 time series, with an average length of around 1800 data points, representing the 7200 minutes covering the last 5 days of the betting market. The 242 time series are further split into two sets, 200 time series becoming the training data, with the remaining 42 becoming the prediction set used to assess the efficacy of any derived methods. Taking stock, the challenge is to use the 200 past time-series (which are assumed to be independent a priori) of length 1800 to predict the future movements of a time series which has been observed up to the current time. For these reasons, heuristic approaches will be used in order to gain insight into the movements of a particular market. However, it will be shown that statistical approaches can be used to classify certain features of the market size movements.

This work will thus attempt to answer the following question: *"Given there are τ minutes until the event occurs, and the market currently has £W_τ invested in it, how will the market evolve in the future, and what will its final size be?"*. It will be shown that in answering this question, other information is required, such as the time of day, and whether the event of interest is a home win, away win, or draw.

Consider again an example of the evolution of the size of a betting market, leading up to the event commencement.

Say that an investor wished to predict the future evolution of the market, 2000 minutes before the commencement of the event (the dashed red line in Figure 5.1.3). Two sets of information would be very useful for such an individual:

1. Given the behaviour witnessed already by the market, how do similar markets behave, in general, as they near the event commencement?

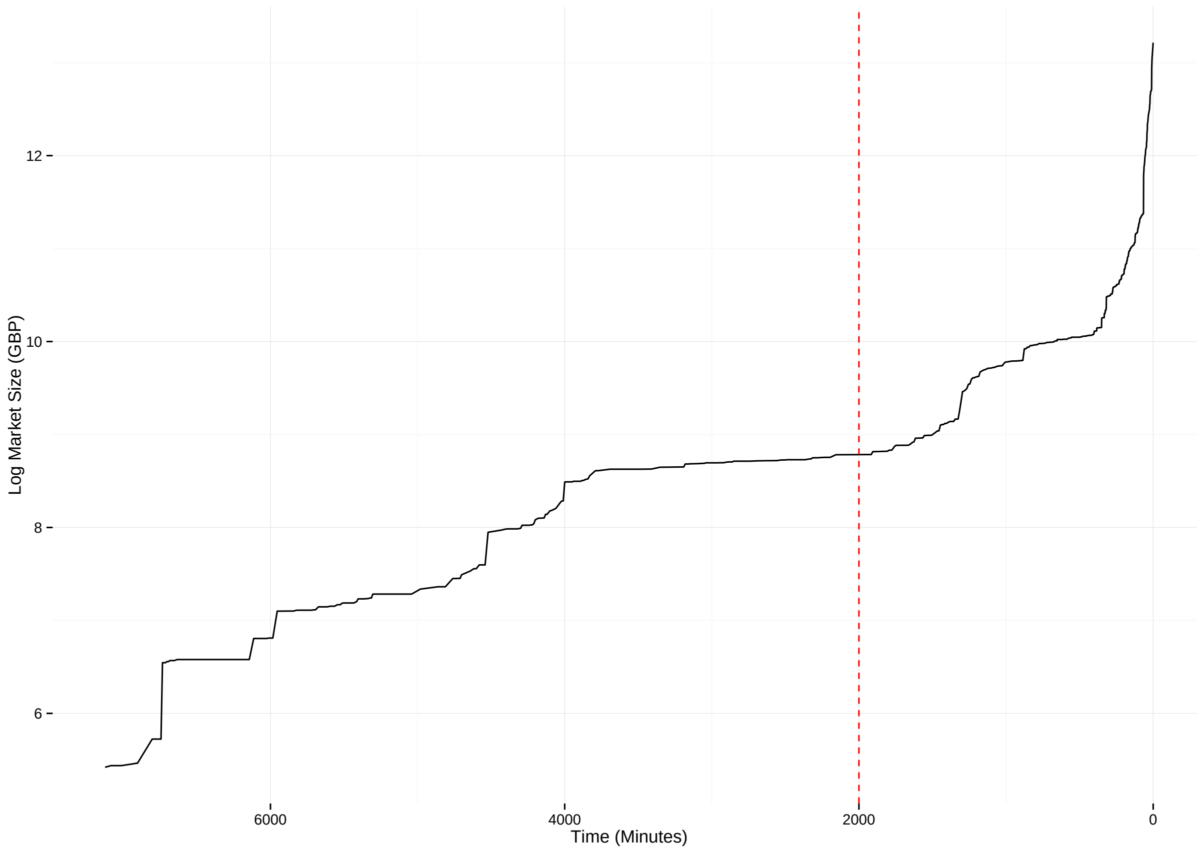


Figure 5.1.3: The change in market size during the pre-match period of the betting market for Sunderland beating Manchester City on the 1st January 2012, with an appropriate burn-in period removed.

2. Can different types of market movement be appraised, such that extrapolation from the current point can be simulated?

The first question above can be approached via data-mining, that is, looking at a collection of past market size evolution, finding the 'most similar', then using these to predict future movements. This data-mining approach will be developed in Section 5.2. The second question can be tackled by simulation, once the features of the market movements have been identified and quantified in some way. This simulation approach will be developed in Section 5.3. It is the intention of this work that these two methods should complement each other to provide accurate forecasts.

Given that the structure of the evolution of the market size is most evident on the log scale, all of the following work will assume that this scaling is being used. Clearly, the actual market size can be recovered from such analysis via a simple transformation.

In the spirit of clarity, in all notation the convention will be that time will run from 1 to 7200, representing the physical time passed since the beginning of the market (which has been truncated at this point). However, in graphics it is clearer to represent time as the number of minutes until the event's commencement.

5.2 Data-Mining Approach

The data-mining approach manipulates the data into a useful structure, then selects the m closest time series from the historical collection of past time-series observed on past pre-market movements. The m closest time series are then used to extrapolate the market size forwards from the current point. The method will be broken down into its individual components, which are investigated individually in the following subsections

5.2.1 Interpolation

The interpolation stage is necessary in this case, as the fidelity of the data in the training set is not equal through the time series. The aim of the interpolation stage is therefore to ensure that the fidelity of the all time series in the training set is equal and constant. It should be noted that techniques do exist which allow for time series to have different frequencies, most notably Dynamic-Time Warping, or

DTW (introduced in Bellman and Kalaba (1959) and reviewed in Müller (2007)). Importantly, DTW is relatively slow, requiring $\mathcal{O}(n^2)$ operations for data of length n. It is therefore faster to first enforce equal frequencies, by means of interpolation, then using one of the other methods explored, which require only $\mathcal{O}(n)$ operations.

The interpolation method is simple; the data is extended by forcing a record at every minute throughout the whole time series. This is achieved by making the assumption that the market does not change between successive observations. So, given a set of market size observations, $y_{a_{1:a_r}}$, where $r < 7200$, and $\boldsymbol{a}_{1:r} = (a_1, \cdots a_i, \cdots, a_r)$ represent the time indices where data is known, then for any $j \notin \boldsymbol{a}_{1:r}$, $y_j = y_{\max(a_i<j)}$.

5.2.2 Select Closest Time-Series

There are two choices considered as a good metric to judge the 'closeness' of two time series, namely the Minkowski distance and a distance based on the correlation between the two time series.

The Minkowski distance (also known as the L_q distance), of order q, is defined as

$$d_{L_q}(y_{1,1:n}, y_{2,1:n}) = \left(\sum_{i=1}^{n} |y_{1,i} - y_{2,i}|^q \right)^{\frac{1}{q}}. \tag{5.2.1}$$

for two time series $y_{1,1:n} = (y_{1,1}, \cdots, y_{1,n})$ and $y_{2,1:n} = (y_{2,1}, \cdots, y_{2,n})$ whose time indices are identical. Most commonly, q is chosen to be equal to 2, known as the Euclidean distance. This distance measure is very sensitive to transformation, most notably by scaling; if two time series were identical, apart from a scaling factor, the Minkowski distance could still be large. In addition, if there are few but significant outliers in one of the time series, the Minkowski distance measure would still report a poor likeness.

Alternatively, a measure is available which is based on the correlation between the two series $y_{1,1:n}$ and $y_{2,1:n}$. Let ρ be the Pearson's correlation between the two time series then the correlation distance is defined as

$$d_{\rho}(y_{1,1:n}, y_{2,1:n}) = \sqrt{2(1-\rho)} \tag{5.2.2}$$

which was first introduced in Golay et al. (1998). This measure is clearly only based on the correlation between the two time series, which extracts the extent of linear dependence between them. The clear advantage of using the correlation distance over the Minkowski distance, is that the measurement remains invariant under changes in the location and scale.

The choice of a suitable distance metric depends on the structure of the training dataset. If the magnitude of the time series are similar, then the focus of finding close time series is for the movement to be similar over time. In this case, the correlation distance would be a good choice. In comparison, if there is a large disparity between the magnitude of the time series (like in this case), then "close" time series should have a similar size too. In the rest of the work, then, it is assumed that the Euclidean distance, as described in equation (5.2.2), with $q = 2$ will be used. This is due to exploratory analysis on a small subset of the dataset showing that the Euclidean distance giving more robust predictive estimates.

This stage of the process considers all of the training set, and returns the k time series which are judged as being the closest to the current observations.

5.2.3 Extrapolate

The extrapolation step simply uses the k time series from the training set which are deemed to have been the closest to the observed series, and uses them to extrapolate an estimate of the current series for the remaining time. In order to bring the selected series together, two questions must be answered: firstly, how many past time series should be used to predict the future movement? Secondly, how should the k series be weighted to yield the predictions, based on their respective distances from the series of interest.

The weighting of the k series is achieved via an exponentially weighted formula, shown in equation (5.2.3), where $(d_1, \cdots, d_k)$ are the distance measures from the k selected series. The weights should be higher for the time series with smaller distance metrics. The weighting formula is shown in equation (5.2.3) (more penalisation for poorer-fitting series are controlled by larger values of κ). The weight for data set i is chosen to be

$$w_i = \frac{\exp(-\kappa d_i)}{\sum_{i=1}^{k} \exp(-\kappa d_i)} \tag{5.2.3}$$

Clearly, a good choice of k depends on the choice for κ. If κ is large, then the poorly-fitting series from the training set will be downweighted to such an extent that their contribution to the extrapolated predictions would be minimal. Such a choice for κ would also heavily downweight any time series other than that of the smallest distance measure, this would result in predictions where the best-fitting series dominates all others, essentially giving $k = 1$. It is therefore recommended that the value of k is fixed for all predictions, and κ is chosen via some knowledge of the dataset.

The choice of κ for a particular prediction is recommended to be chosen according to two criteria:

1. What is the size of the dataset? The larger the dataset, the larger the probability (in general), that a number of close time series will exist, and therefore the larger the value of κ, as the rest of the time series need not be considered.

2. How unusual is this event? If the event in question is special in some way, e.g. a domestic cup final, then the market movements may well be different to the training set. In this case, a small choice of κ would be more suitable, and the prediction is made by taking the average of a large number of potential time series, with none of them weighted much higher than any other.

For the dataset used in this work, its size is small (only 242 matches), but all of the events come from the same two major leagues, so there should not be any obvious outliers. The actual choice of k and κ will be revisited in Section 5.5.

The prediction, then, is based on the distance measures of the k-closest time series, along with the adjustable parameter κ, all used to calculate the set of weights $w_{1:k} = (w_1, \cdots, w_k)$. The estimated wealth over the rest of the time series, $\tilde{y}_{(\tau+1):7200}$ is

$$\tilde{y}_{\tau+1:7200} = \sum_{i=1}^{k} w_i y^*_{i,\tau+1:7200} \tag{5.2.4}$$

where $y^*_{1:k}$ are the k-closest time series selected by the above process.

In summary, the data-mining approach can be visualised via Figure 5.2.1.

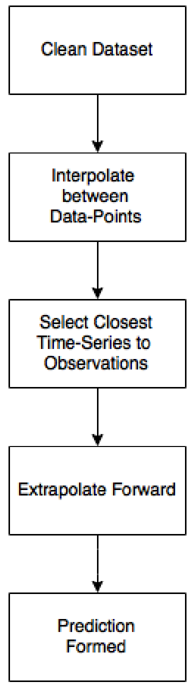


Figure 5.2.1: A diagram depicting the data-mining approach to forecasting changes in market size.

5.3 Simulation Approach

The simulation approach proposes a very different method of prediction to the data-mining approach. The data-mining approach uses a “top-down” paradigm, focussing on the time series as a whole in order to detect similarities between what is being observed and what has been observed in the past. In contrast, the simulation approach breaks down the market movements into quantifiable pieces, allowing the future of a particular set of observations to be inferred from the ground up.

Crucially, the simulation method relies on the detection of changepoints, which can be thought of as the locations in the time series where the statistical properties change. It is intended that the changepoints witnessed in the data correspond to a surge of interest in the market at that time. This could be due to a number of phenomena, most notably the introduction of information to the market, such as injury or team news, which could trigger investment in the market. This is a well-known market phenomenon, see Jiang et al. (2011). Another possible reason for a changepoint to occur is much more particular to betting markets. As the betting market nears its end, more and more of the information regarding the event, such as the weather or pre-match injuries are known, and thus investors tend to choose this time to make their decisions. A change in the rate of investment is also detected as a changepoint (see the last few hundred minutes in Figure 5.3.1).

The modelling approach is to simply conjecture that the log market movements are linear in between the occurrence of changepoints, which cause the market size to be boosted to some higher value (note that the market size is guaranteed to not decrease over time).

This approach is visualised in Figure 5.3.1. The black dashed lines represent the detected changepoints, with the blue line giving a linear fit to the data between the changepoints. As can be seen, in many instances the trend of the data between changepoints does not seem to be linear in nature, most notably in the interval from 4500-4000 minutes before the event. However, the linear fit between changepoints is a convenient structure to use, which evidently capture the movement of the market well.

What is noticeable about Figure 5.3.1 is that the frequency of the changepoints

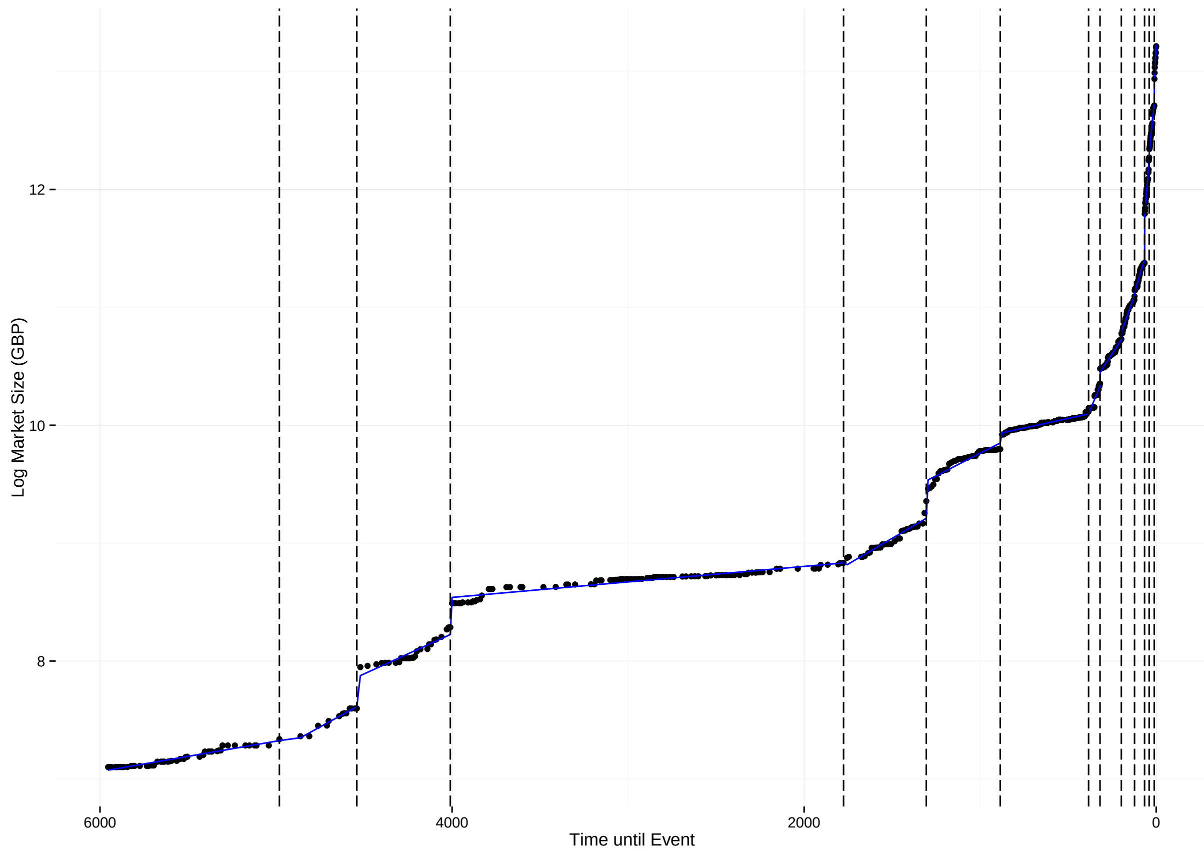


Figure 5.3.1: The change in market size during the pre-match period of the betting market for Sunderland beating Man City on the 1st January 2012, overlaid with a changepoint-based model-fitting approach.

increases noticeably as the time approaches the commencement of the event. In particular, when the market is a few hours from the event's commencement, the frequency of changepoints increases, and the market size increases markedly. This feature occurs in every time series in the training set, and will have to be accounted for in the model fitting procedure.

Like the data-mining method, the simulation approach contains many stages, which will be explored separately, before being brought together as a full method in Section 5.3.6

5.3.1 Detecting Changepoints

PELT is chosen as the method used to detect changepoints in the time series, which is implemented via the R package "changepoints" (Killick et al., 2014). Details regarding why PELT is chosen, as well as a details of its derivation are given in Appendix C.

The implementation of the PELT algorithm first requires the setting of a cost function, $\mathcal{C}(.)$, which represents the measure of closeness-of-fit, and is chosen within the package to be twice the negative log-likelihood of the data. The choice of the penalty against overfitting remains that of the user. Most of the popular choices for the penalty are linear with the number of changepoints. For example, the AIC uses twice the number of changepoints. Experimentation with these penalties, however, results in a poor characterisation of the changepoints observed in the training set, with the AIC's use in changepoint selection being show in Figure 5.3.2.

For this reason, a manual penalty is chosen for the PELT method. That is, the penalty function is a constant, chosen by the user. It is found that choosing a manual penalty in the region of 0.01 to 0.05 produces classifications of changepoints which look consistent to the data. The plot shown in Figure 5.3.1 was found by using a manual penalty of 0.03, a penalty found to detect the occurrence of market movements well for the training set of data. This penalty is used throughout the following sections.

5.3.2 Number of Changepoints

Given that the training set is to be analysed with changepoint methods, the first feature to model is the number of changepoints seen in the time series. It is expected

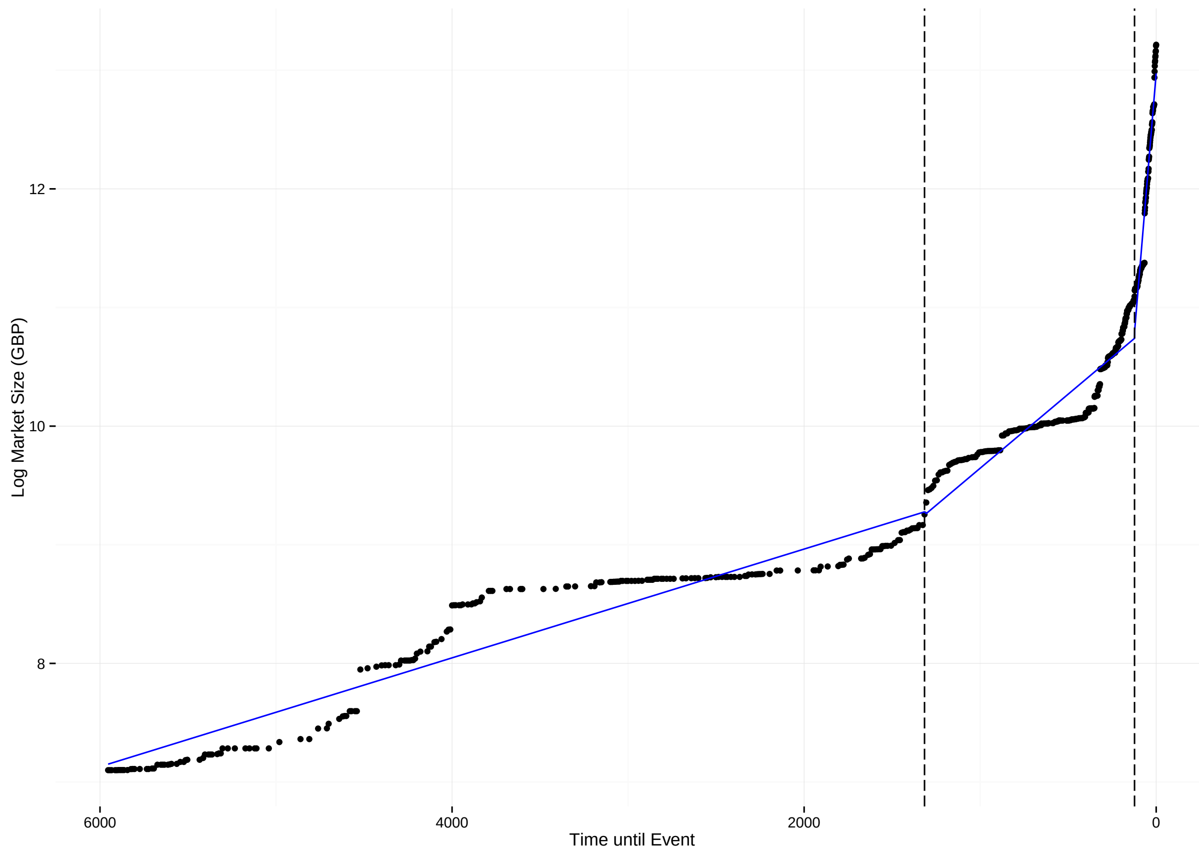


Figure 5.3.2: The change in market size during the pre-match period of the betting market for Sunderland beating Man City on the 1st January 2012, overlaid with a changepoint-based model-fitting approach with an AIC penalty function.

in this and other modelling problems for there to be subsections of the training data which behave differently, i.e. the data can be partitioned into different categories, such that there is a difference in the structure of the data in each category.

The most obvious potential categorisation is in the time of the event. Betting activity most commonly occurs during the daytime, as this is the time where most of the potential bettors are awake. In addition, most of the betting activity occurs in the hours leading up to the event. Therefore, it is expected for the rate of capital being invested in events taking place late in the evening to be different to when the event is in the morning. In this spirit, the data is categorised into the events taking place early in the day (before 15:00), middle of the day (15:00 up to 18:00) and late in the day (after 18:00). Of the 242 matches, 52 were categorised as being early, 80 as middle, 110 as late.

Figure 5.3.3 shows the difference in the distribution in the number of changepoints, as the time of the event changes. There is a clear observable difference between the distributions: most notably, the matches with late starts exhibit a multimodal shape, which the other start times do not demonstrate. Again, it is presumed that this feature is merely a quirk of the data, and does not demonstrate some wider truth.

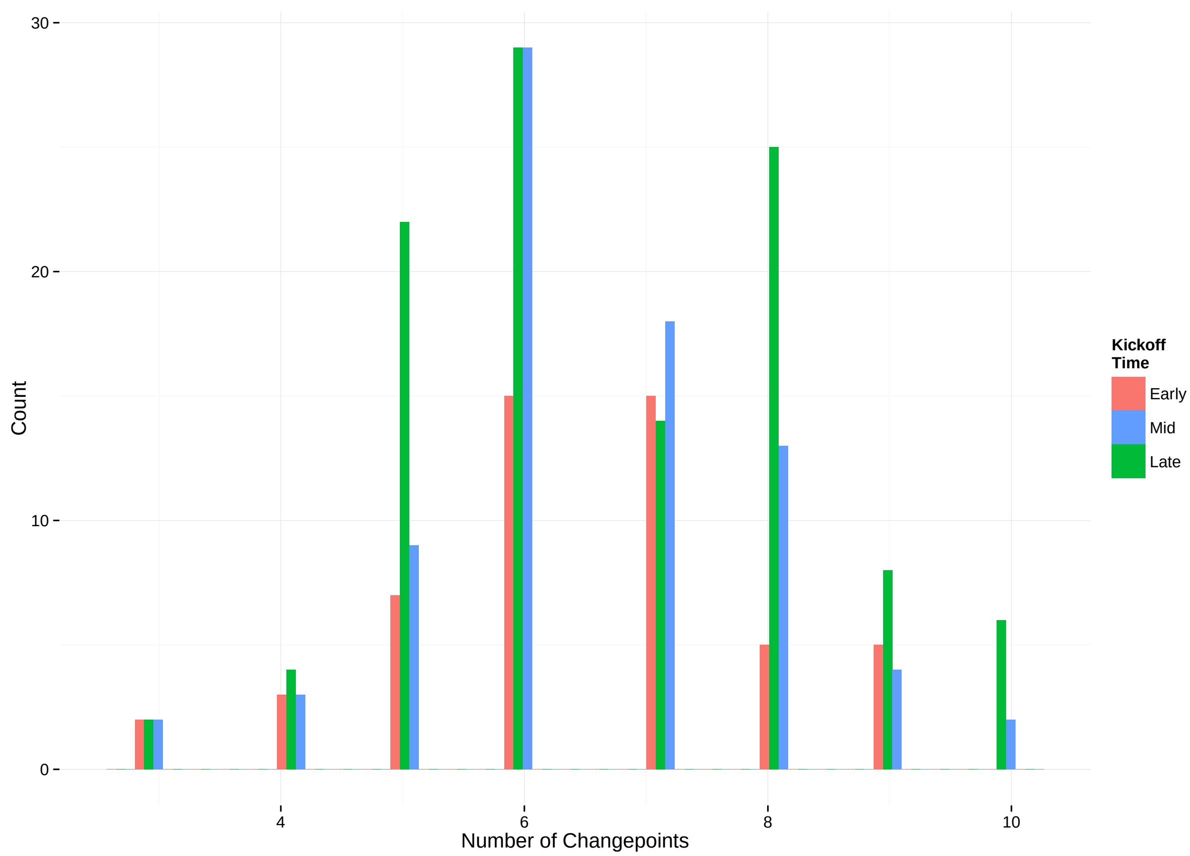


Figure 5.3.3: Bar chart showing the number of changepoints detected in the log market size, for football matches commencing at different times in the day.

Recall in Section 5.3.1, it is conjectured that changepoints arise both due to the release of new information into the market, and also from the inevitable ramping-up of betting activity towards the commencement of the market. The different starting times of event would change the time at which changepoint occurred, but would not change their total number.

The different event start times are therefore considered separately, both here, and in all of the subsequent modelling sections. Therefore, in the use of this model, the time of the match becomes a requisite input, in order for effective inference to be conducted.

The output from this section is that the number of changepoints in each series in the training set is stored, along with the time series' time location, i.e. early, mid or late. Let this stored data be written as $\mathrm{CP}_{\mathrm{num},T}$, where T takes the values E, M or L depending on the time location.

5.3.3 Location of Changepoints

Another important feature needed in order to simulate the occurrence of changepoints in the future is how they are distributed along the duration of the pre-match market. As noted earlier, both the release of information into the market, and the time at which investors are more likely to be awake affect the timing of changepoints. Figure 5.3.4 shows the distribution of changepoints over time for both the whole data, and also for the categorised event times.

The effect of the time of the day on the market is very strong. Consider the density representing all of the data in Figure 5.3.4; there is a clear multi-modal structure, with the modes occurring roughly once a day (1440 minutes). In addition, anti-modes occur in the evenings, representing periods where low betting activity is likely. These modes and anti-modes shift in time according to the start-time of the event, as expected, which gives strong evidence supporting the treatment of events with different start-times separately. The times indices are therefore recorded separately for the different start times, as in Section 5.3.2.

For use in future work, the following notation will be adopted: $\zeta_{1:n} = (\zeta_1, \cdots, \zeta_n)$ signifies the time-index of n detected changepoints in a particular time series. The time-index is defined to be the time until the event's occurrence, in minutes. In addition $y_{\zeta_{1:n}} = (y_{\zeta_1}, \cdots, y_{\zeta_n})$ will then signify the size of the market at the time of the n changepoints.

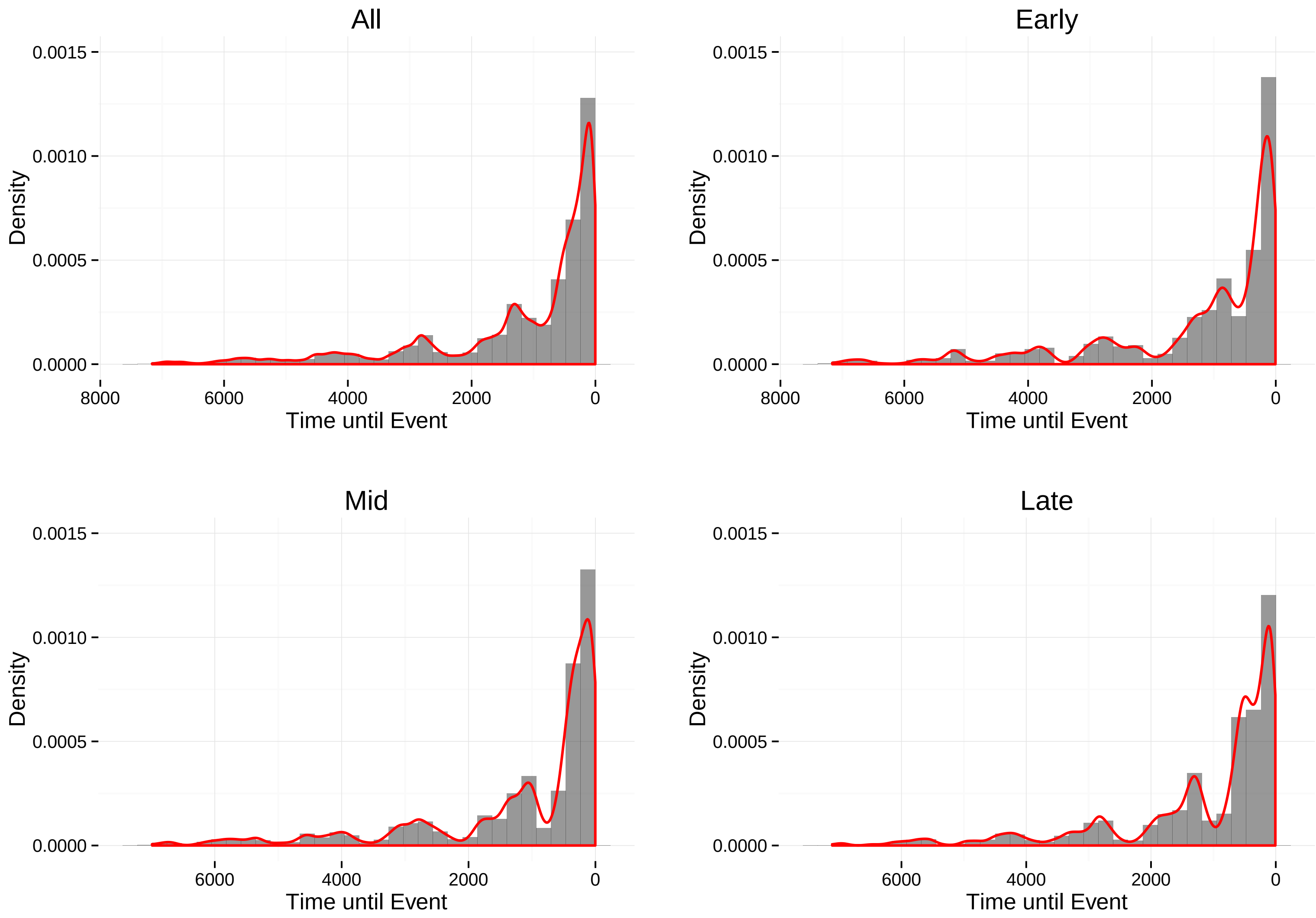


Figure 5.3.4: Density fits for the location of changepoints detected in the log market size, for football matches commencing at different times in the day.

The output of this section of work is to store the time location of the detected changepoints in the training set separately for each of the event's time locations. Let this stored data be written as $\mathrm{CP}_{\mathrm{loc},T}$, as before.

5.3.4 Size of Changepoints

So far, knowledge has been gained about the likely number of changepoints to be observed in a time-series representing the growth in the log market size, as well as where they are likely to occur. What remains to be grasped is the magnitude of the effect of a changepoint upon the size of the market. As changepoints occur, the market jumps up from one value to another.The difference in the values can be calculated as

$$z_i = \frac{y_{\zeta_i+1}}{y_{\zeta_i}}. \tag{5.3.1}$$

The size of the changepoint, then, is defined as being the multiplicative factor linking the (log) size of the market before and after the occurrence of the changepoint. In this case, not only is the size of the changepoints potentially dependent upon the time of the event, but it is proposed that there may be dependence of the time of the changepoint on its size. The size of the changepoints is therefore plotted against their time of occurrence, and is shown in Figure 5.3.5.

This relationship has some interesting structure. Firstly, the day and night structure from Section 5.3.3 is seen again, with the changepoints occurring in 'waves' through time. In addition, the distribution of changepoint sizes seems be segmented into two sets: the time immediately prior to the event; and the rest of the time series.

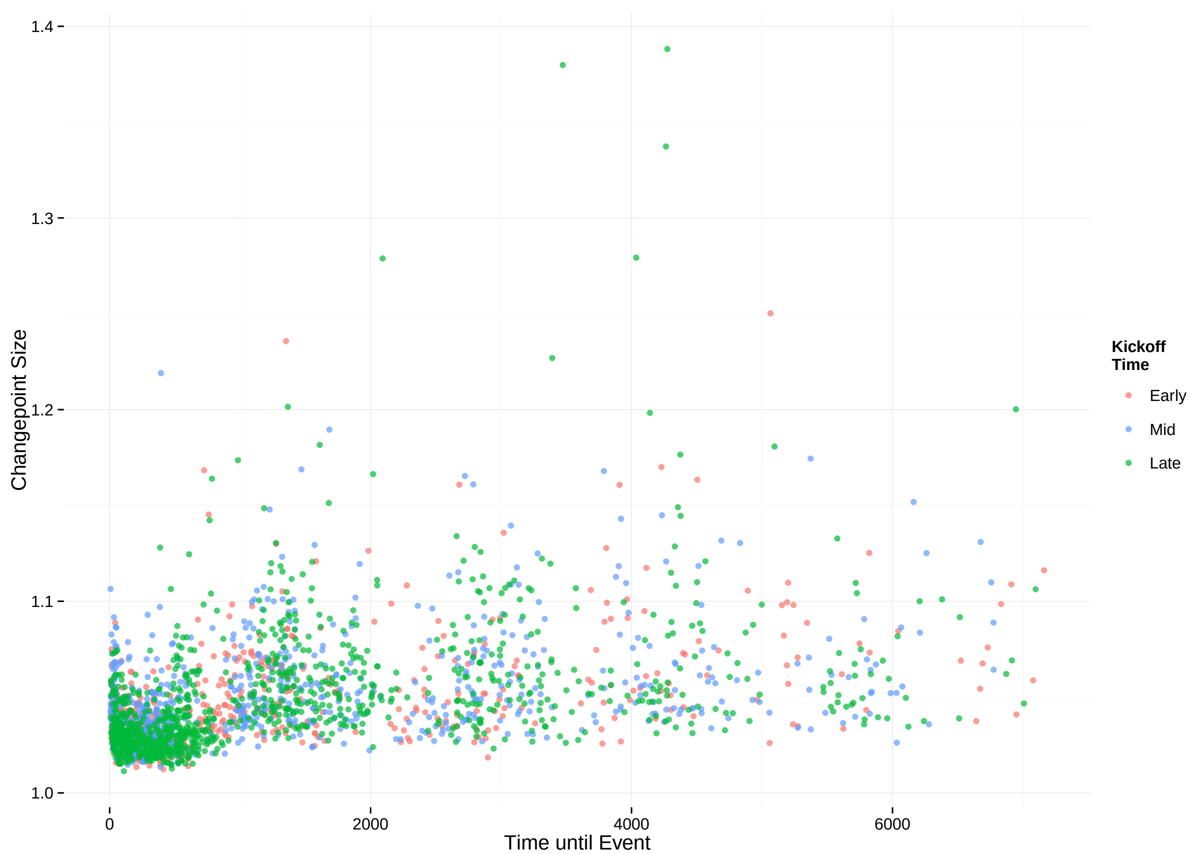


Figure 5.3.5: The relationship between the size of changepoints and how long before an event they occur.

On the day of the event, the changepoint size is small, rarely forcing a jump in the size of the market by more than 10%. In comparison, in the days beforehand, the size of the changepoints seems to be not only on average higher, but with more variability, with jumps regularly being over 10% and increasing up to around 40%. These differences are summarised in Table 5.3.1.

	Summary Statistics	
Section	Mean	Variance
Day of Event	1.0421	4.5×10^{-4}
Days before Event	1.063	1.17×10^{-3}

Table 5.3.1: The difference in mean and variance between the size of changepoints on the day of the event, and in the days before the event of interest.

Due to this clear disparity in the structure of the data as the time until the event changes, merely treating the size of the changepoints as i.i.d. for simulation purposes is not sufficient. Instead, inference from this data is made only in the time-point's neighbourhood via sampling only from the neighbourhood of the current point. This method is described in Algorithm 3

Algorithm 3 Neighbourhood Sampling

Input: Data containing a set of changepoint sizes $z_{1:n} = (z_1, \cdots, z_n)$, along with their times $\zeta_{1:n} = (\zeta_1, \cdots, \zeta_n)$

A number, s representing the size of the neighbourhood.

A number, m representing the required size of sample.

The current time, t.

1: Define $\delta_{1:n} = |t - \zeta_{1:n}|$.

2: Let $a_{1:m} = (a_1, \cdots, a_m)$ be the indices of the m smallest elements of $\delta_{1:m}$.

3: Let $z^*_{1:m} = z_{a_{1:m}}$

4: Sample s elements from $z^*_{1:m}$, with replacement.

Output: A set of m size samples from the neighbourhood of the time point.

The logic behind neighbourhood sampling is that samples should only be drawn from the training data which are similar to the current state of the market, in this case, the current time. The algorithm only samples the changepoint sizes from the m changepoints with time closest to the current time.

As a result of this section, three sets of data are stored, representing the size of the changepoints and their times, for all of the early, mid and late matches. Let these samples be written as $\mathrm{CP}_{\mathrm{size},T}$.

5.3.5 Linear Fit between Changepoints

The last feature of the market movements to model is how it increases between the changepoints. As previously stated, a modelling decision is to let the market increase linearly over time between sucessive shocks. There is presumably some dependence between the gradient of the linear fit and other factors, such as the time remaining until the event occurs (inferred from the mid-point of the intervals), or the length of the intervals themselves. These relationships are shown in Figures 5.3.6 and 5.3.7, respectively.

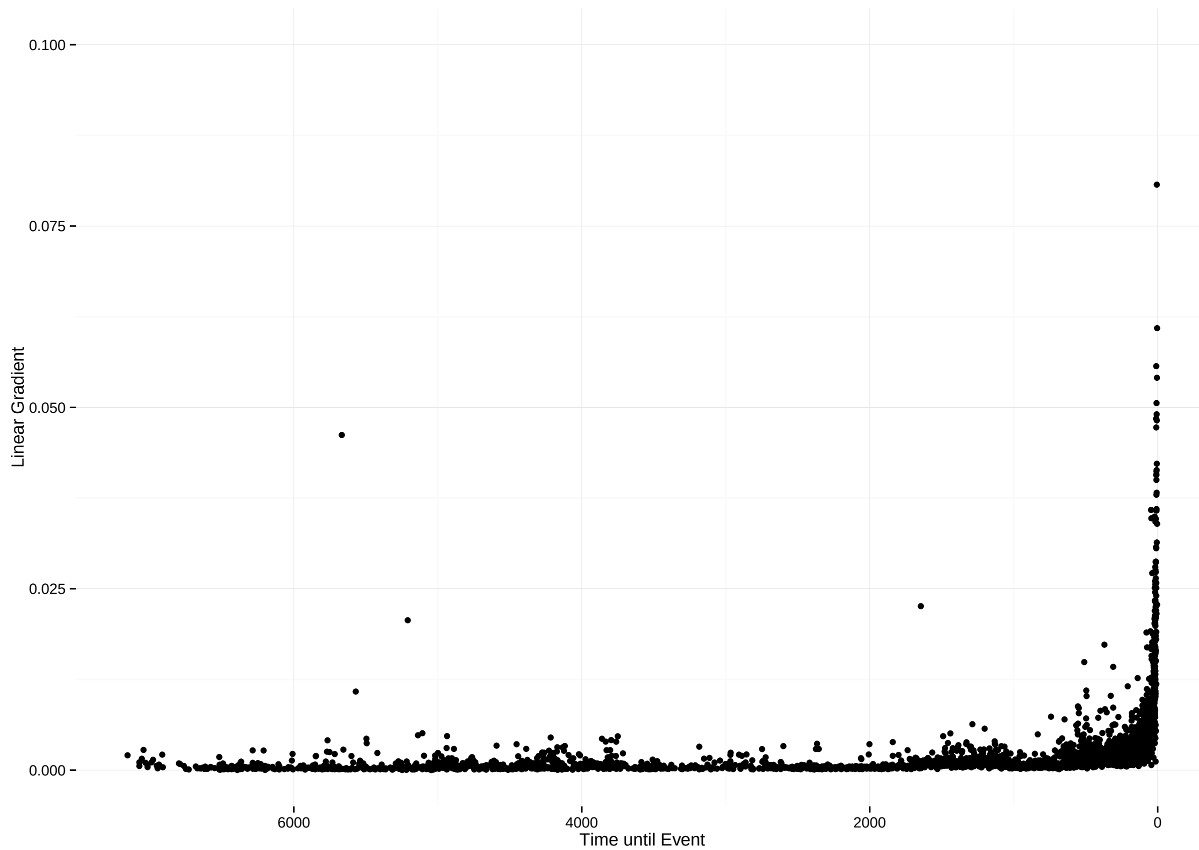


Figure 5.3.6: The relationship between the gradient of the linear fits, and the time until the event occurs.

Both of these relationships are strong. In Figure 5.3.6, there is a strong dependence between the time until the event and the gradient of the linear fit. As the time for the event approaches, the average size of the linear gradients increases markedly. Figure 5.3.7 shows a similar pattern. As the size of the interval becomes very short, the

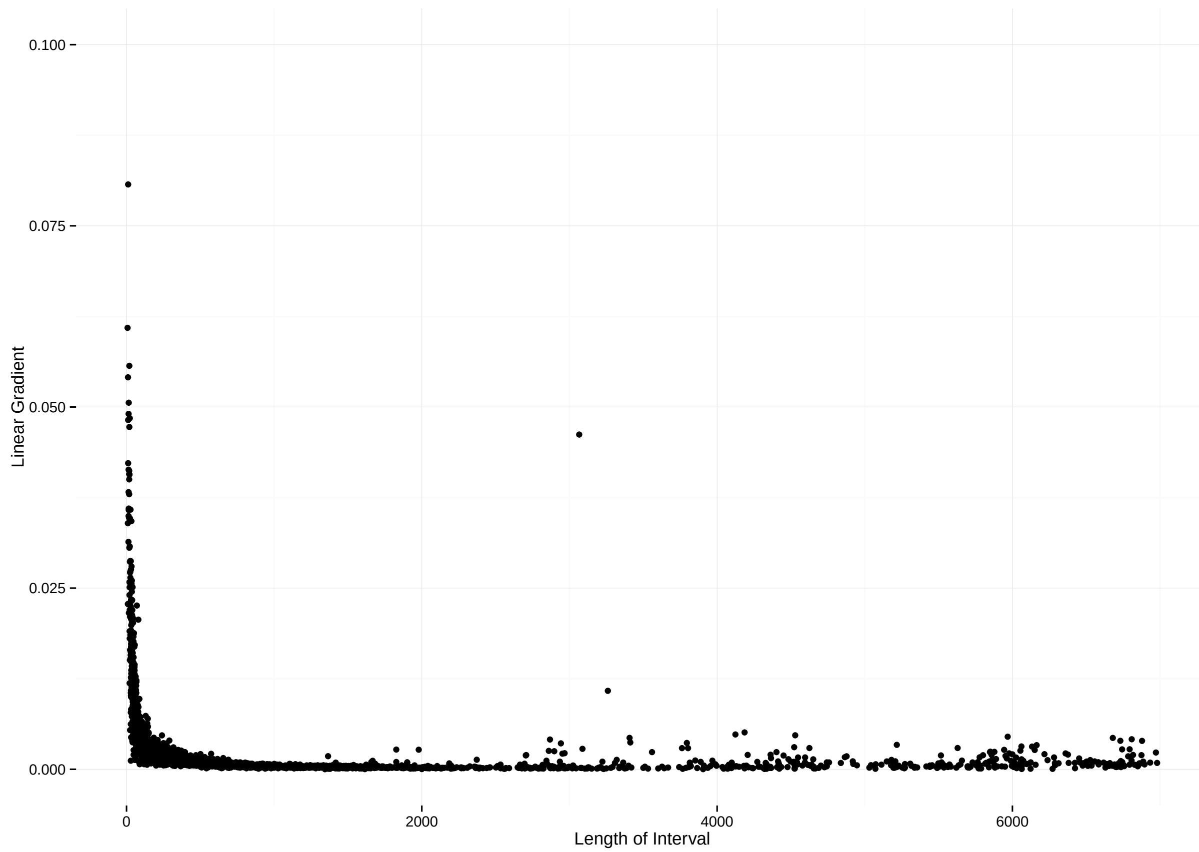


Figure 5.3.7: The relationship between the gradient of the linear fits, and the length of their intervals.

gradient of the linear fit increases. These relationships go hand-in-hand. As the market reaches its climax, the market activity increases, and many changepoints are detected in a small amount of time, inevitably causing the time between changepoints to be short.

If the correlation between the length of the interval and the time until the event were very strong, then considering both of the factors would not be necessary, observing the gradient of the linear fit for a certain interval length could be used to infer the time of the event, and vice versa. In order to see whether this is true, the interval times are compared against their lengths, in Figure 5.3.8. The correlation exhibited

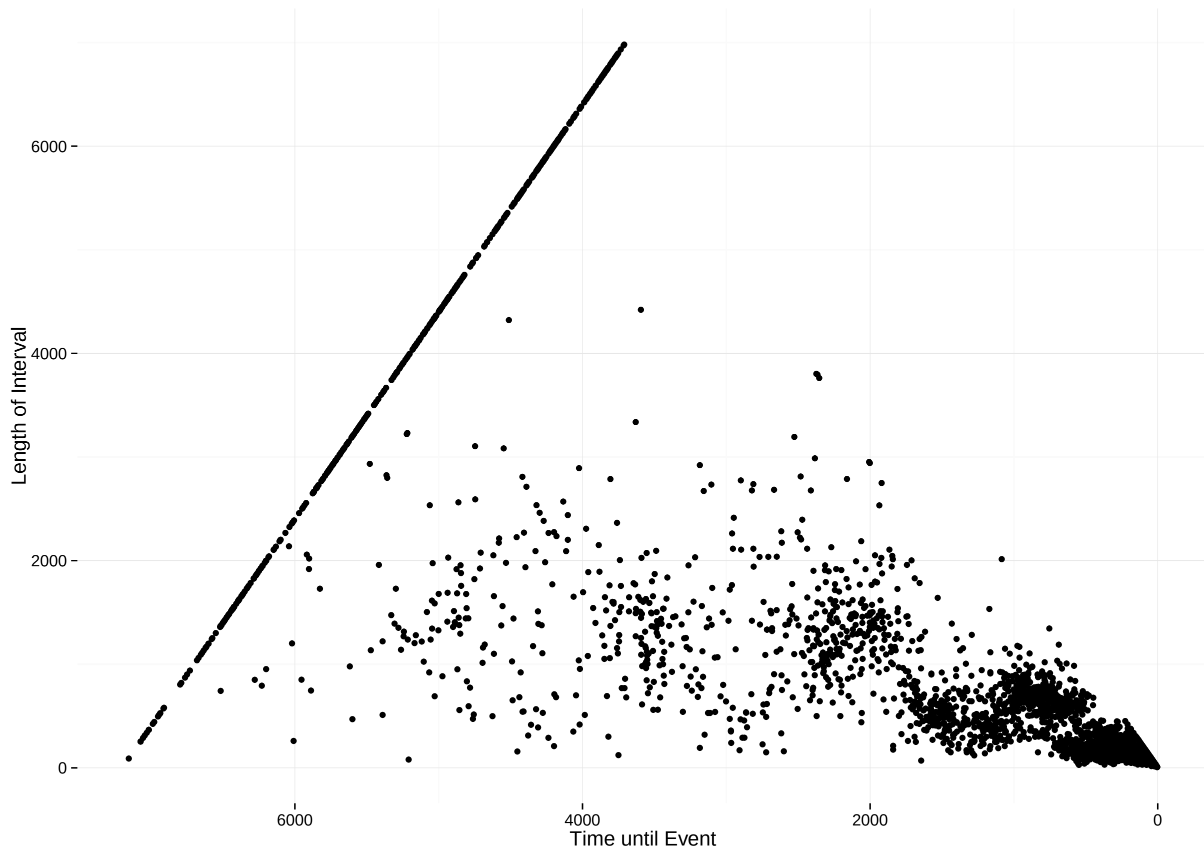


Figure 5.3.8: The relationship between the time of the event of changepoints, and the length of the intervals between them.

between the variables in Figure 5.3.8 is clearly not strong enough to discount one of the variables as being uninformative. In particular, the clear straight line on the left hand side of the graph shows the deterministic relationship between the variables for intervals at the beginning of the time series (the intervals which occur between the market's commencement and the first changepoint detected) whereas the remainder have a more random relationship.

In this light, when attempting to sample values of the gradient of the linear fit between changepoints, both the time of the interval, along with the length of the interval should be taken into account. In order for this to be achieved, the neighbourhood sampling approach shown in Algorithm 3 needs to be adapted to take into account more than one variable of interest, and is described in Algorithm 4.

Let the gradient of the linear fits for all n instances in the training set be written as $\tilde{y}_{1:n} = (\tilde{y}_1, \cdots, \tilde{y}_n)$. In addition, let the other π features of these linear fits be written as a set of vectors $\tilde{x}_{1:\pi,1:n}$ (in this case, $\pi = 2$ and $\tilde{x}$ would contain the length and time to the event for the linear fits). The weights in Algorithm 4 refer to how the different features are weighted towards the overall distance measure δ.

Algorithm 4 π-dimensional Neighbourhood Sampling

Input: A vector of output data $\tilde{y}_{1:n} = (\tilde{y}_1, \cdots, \tilde{y}_n)$, along with π vectors of explanatory data $\tilde{x}_{1:\pi,1:n}$, matching the output data.
A number, s representing the size of the neighbourhood.
A number, m representing the required size of sample.
The current state, $x_{1:\pi}$.
A vector of weights $\omega_{1:\pi}$.

1: Define $\delta_{1:n} = \sum_{i=1}^{\pi} \omega_i |x_i - \tilde{x}_{i,1:n}|$.
2: Let $a_{1:m} = (a_1, \cdots, a_m)$ be the indices of the m smallest elements of $\delta_{1:m}$.
3: Let $\tilde{y}^*_{1:m} = \tilde{y}_{a_{1:m}}$
4: Sample s elements from $\tilde{y}^*_{1:m}$, with replacement.

Output: A set of m size samples from the neighbourhood of the time point.

Therefore, the gradients of the linear fits are stored, along with their corresponding time and length, to be used to sample linear gradients in the full simulation. As before, the time of the event impacts significantly upon the distributions, and therefore this stored data will be separated for each of the different classifications of start time.

This section of the work now allows sampling the gradients for the linear fits between changepoints, whilst taking into account the dependence of the time and interval length on the gradient. Let these samples be written as $\mathrm{CP}_{\mathrm{lg},T}$.

5.3.6 Simulation Formation

Now that the nature of the occurrence and impact of changepoints, as well as the change in market size between changepoints has been modelled, a scheme to simulate forward from a particular market state can be formalised. Recall that the sets of data collected in the previous sections are written in shorthand as the number of changepoints, $\mathrm{CP}_{\mathrm{num},T}$, the location of the changepoints $\mathrm{CP}_{\mathrm{loc},T}$, the length of the changepoints $\mathrm{CP}_{\mathrm{size},T}$ and the size of the gradient of the linear fit $\mathrm{CP}_{\mathrm{lg},T}$.

The full algorithm for the simulation technique is shown in Algorithm 5 and can be visualised via Figure 5.3.9. The logic behind the process is the changepoint locations, along with their sizes are simulated first, as only then will the location of the linear intervals be revealed. After all of these elements have been simulated, predictions are made sequentially through the remaining time-steps, with the movements being defined either by the impact of a changepoint, or the underlying linear movement of the market in between the changepoints. The current state of the market $x_{y,\tau,T}$ is characterised by the current market size, y, the time until the event τ and the event time T.

Note that in Algorithm 5, the mean of the N simulations is used as the overall prediction for each time step. In some instances, exceptional simulation outcomes could result in extremely high or low market size estimations, in this case, the median is recommended instead.

Algorithm 5 Simulation Approach

Input: The current state of the market, $x_{y,\tau,T}$
N, the number of simulations

for i=1 **to** N **do**
  Sample once the number of changepoints from $\text{CP}_{\text{num},T}$ and denote this as γ.
  Sample γ times from $\text{CP}_{\text{loc},T}$ and denote these as $L_{1:\gamma} = (L_1, \cdots, L_\gamma)$.
  Remove all elements such that $L_{1:\gamma} < \tau$.
  Let $\tilde{\tau} = 7200 - \tau$
  For each $L_{1:\gamma}$, sample one changepoint size from $\text{CP}_{\text{size},T}$ from their neighbourhoods. Denote these $z_{1:\gamma} = (z_1, \cdots, z_\gamma)$.
  Define $\tilde{L}$ as $(0, L_{1:\gamma}, 7200)$.
  Define the $\gamma + 1$ intervals in $\tilde{L}$ as $I_{1:(\gamma+1)}$.
  for $k = 1$ **to** $\gamma + 1$ **do**
    Calculate the length of I_k.
    Calculate the time of the mid-point of I_k.
    Sample one linear gradient size from the neighbourhood of the interval's length and mid-point, using $\text{CP}_{\text{lg},T}$. Label these as $\alpha_{1:(\gamma+1)}$
  end for
  Define a $N \times (\tilde{\tau} + 1)$ matrix $\Theta_{1:N,1:\tilde{\tau}}$, with elements in the first column all equal to y, the current market size.
  Set count=1.
  for $j = 1$ **to** $\tilde{\tau}$ **do**
    if $\tau + j \in L_{1:\gamma}$ **then**
      $\Theta_{i,j+1} = \Theta_{i,j} z_{\text{count}}$.
      $\text{count} += 1$.
    else
      $\Theta_{i,j+1} = \Theta_{i,j} + \alpha_{\text{count}}$
    end if
  end for
end for

Output: The mean estimate for the market at time j, $\hat{y}_j = \frac{1}{N} \sum_{i=1}^{N} \Theta_{i,j}$

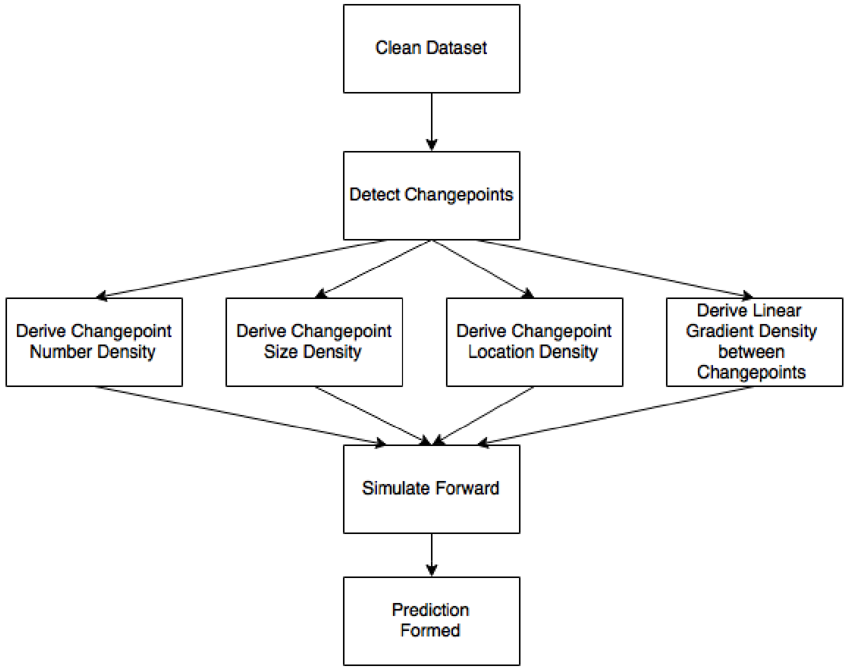


Figure 5.3.9: A diagram depicting the simulation approach to forecasting changes in market size.

5.4 Full Approach

The full approach takes advantage of blending both the ‘top down’ paradigm of the data-mining method and the ‘bottom up’ paradigm of the simulation method. This is achieved by forming some weighted average between the predictions of both approaches, with the entire process being shown in Figure 5.4.1.

Let the predictions from the data-mining approach be written as $y^d_{1:n}$, whilst the predictions from the simulation approach be written as $y^s_{1:n}$, then the full prediction $y^*_{1:n}$ is found via

$$y^*_{1:n} = \frac{y^s_{1:n} + wy^d_{1:n}}{1+w}. \tag{5.4.1}$$

for some $w \geq 0$.

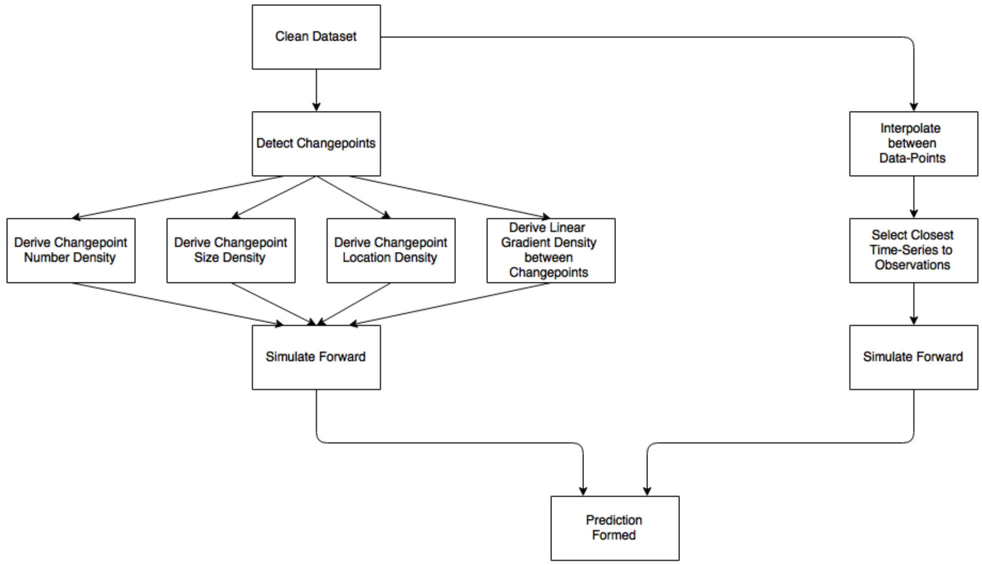


Figure 5.4.1: A diagram depicting the full approach to forecasting changes in market size.

If the full prediction is formed on the basis of this weighted average of the two approaches, then the individual predictions can be recovered by choosing $w = 0$ in the case for simulated prediction, and some very large value of w for the data-mined prediction.

5.5 Assessing Predictive Performance

In order to assess the efficacy of both the individual approaches, and the full, mixed approach, a suitable objective must be defined. This objective will represent the intentions of the user. The two most common applications are as follows:

1. If the use of this method is to help inform an investor of the best time to bet, then the performance of the predictive method throughout the whole duration of the time series is important. In this case, some simple measure of fit of the whole series, such as the root mean squared error, (RMSE) is suitable.

2. If instead the use of the predictive methods is to inform an investor how large the market might be at a pre-defined point in time, most likely just prior to the commencement of the event, then only the prediction of the last point will be of interest. In this case, the simple percentage difference between the true terminal market size and the predicted terminal market size will suffice.

Given some chosen objective, then, two separate assessments should be made, which will provide valuable insight into the performance of the two techniques, both in terms of a single time series, and an average performance over a collection of time series.

Recall that the dataset consisted of a training set of 200 time series, along with a prediction set of 42 time series. The 42 series will be used in both of the performance assessments. Initially, the two separate method will be analysed separately, with the introduction of the 'full method' coming after some inference has been made about suitable choices for the weighting value w between the two methods.

The efficacy of the methods are investigated for the 42 matches, with predictions taking place 5 days, 1 day, and 6 hours before the event's commencement. As a modelling decision in this case, the simulation approach will use 50 iterations (i.e. $N = 50$ in Algorithm 5), whilst the small training sample implies that the data-mining approach uses a small number of time series for its inference, in this case 10, with the tuning parameter chosen as $\kappa = 0.2$, (see equation (5.2.3) (this setup rewards historical time series with close fits to the observations).

A summary of the results of this study is shown in Figure 5.5.1, where the densities represent the overall performance for each of the methods. Note that the end error signifies the difference between the predicted log market size and the true log market size.

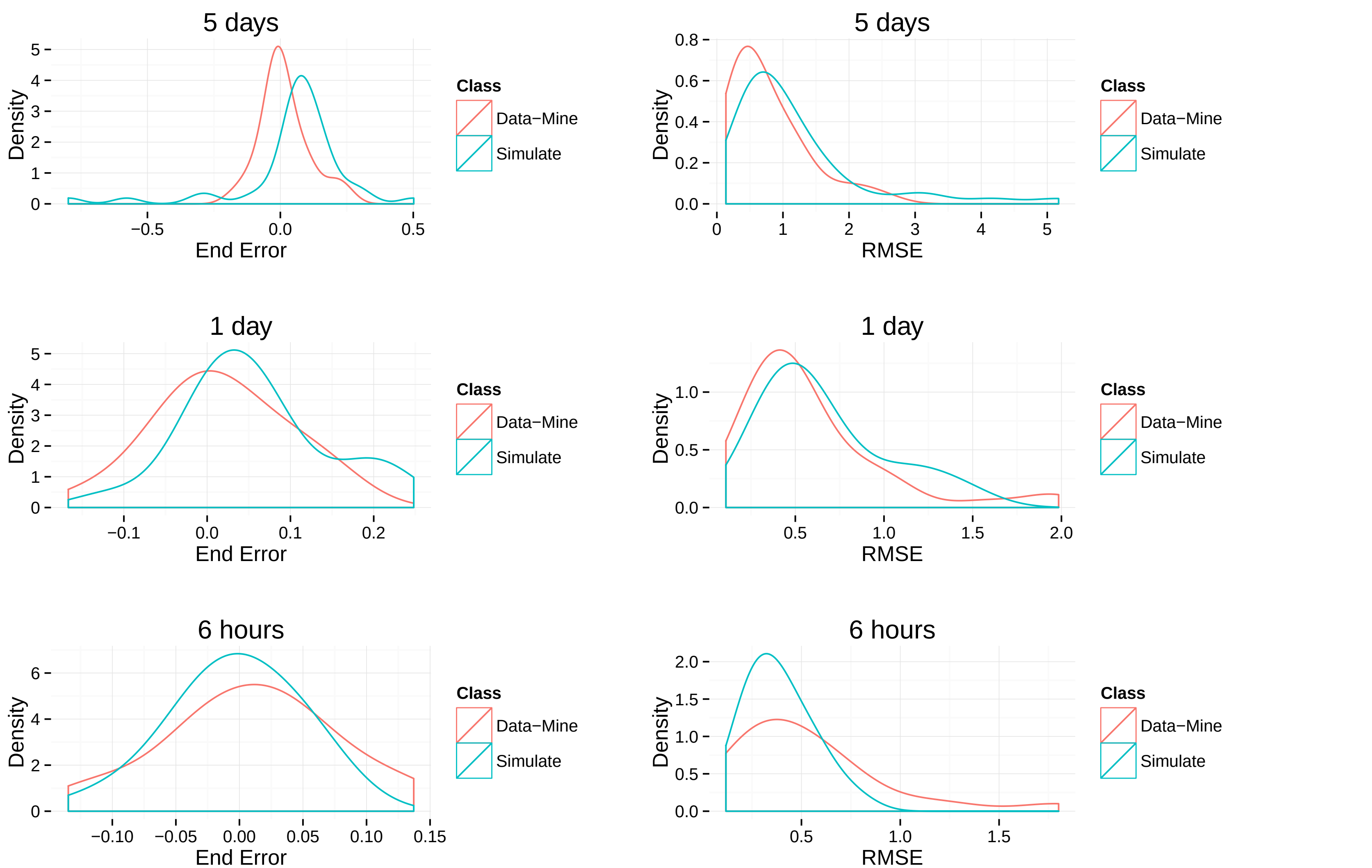


Figure 5.5.1: A summary of the performance of both the simulation-based and data-mining-based methods for predicting market sizes, for a range of objectives and based on a range of prediction points.

First consider the end error of the time series, seen in the left-hand column of Figure 5.5.1. With a large amount of time remaining, the data-mining approach clearly has the edge, with the distribution of its estimates clustered more tightly around a 0% error. It can be noted that when predictions are made a long time before the event's commencement, there are occasions where extremely poor predictions are formed, observable in the left tail of the simulation method's density. Taking a closer look at the individual series, in these cases (where the final market value is underestimated by over 70%) it is apparent that the predictions are being attempted from unformed markets, emphasising the importance of a correct classification of market formation.

This feature is, naturally, not observed when the predictions are made from a later point, as all markets are formed at this point. It can be seen that the simulation method performs better, and is comparable to the data-mining method in the case where predictions are made 1 day from the event's commencement. The mean and variance of the end error of the simulation method and data mining method are shown in Table 5.5.1.

	End Error	
Method	Mean	Variance
Data-Mine	0.0177	0.0062
Simulation	0.0576	0.0075

Table 5.5.1: The performance of two methods to predict the terminal market size 1 day in the future.

This shows that although the simulation method still underestimates the terminal size of the market, on average, its performance in comparison to the data-mining approach has improved markedly.

Looking finally at the attempt to predict just 6 hours into the future, it is no surprise that both methods improve their predictive accuracy. The mean and variance of their errors are shown in Table 5.5.2.

	End Error	
Method	Mean	Variance
Data-Mine	9.68×10^{-3}	4.48×10^{-3}
Simulation	-3.32×10^{-3}	2.47×10^{-3}

Table 5.5.2: The performance of two methods to predict the terminal market size 6 hours in the future.

In this case, the simulation method performs better than the data-mining method, both in the average prediction of the market's end value, but also in the consistency of these estimates. It is conjectured that this points towards a limitation of the data-mining method; a lack of richness in the historical dataset.

If the movements of the current market is not mirrored closely by one or more time series from the training set, it is unlikely that the predictions will be accurate. In comparison, the simulation method does not rely on any individual historic time series, and therefore this limitation does not apply, giving the method no upper-bound on its accuracy, due to this feature.

Consider the other objective of the predictions, to minimise the RMSE of the predictions. Initially, the data-mining method performs much more accurately that the simulation method, whilst both methods contribute a small number of very poor predictions. As the length of the predictions shortens, the performance of the simulation method improves, both in absolute terms and in comparison to the data-mining method. The mean and variance of the RMSE for both of the methods, with 6 hours until the event's commencement are shown in Table 5.5.3.

	RMSE	
Method	Mean	Variance
Data-Mine	0.539	0.152
Simulation	0.388	0.027

Table 5.5.3: The performance of two methods to predict the evolution in market size 6 hours in the future.

As the general performance of the methods, then, is dependent upon the amount of time remaining in the time series, the 'full method' may have value in providing an approach which performs at a consistent level. Applying the weighting of $w = 1$, i.e. equally weighting both approaches gives the performance measures summarised with the underlying performances of the supporting parts, in Figure 5.5.2.

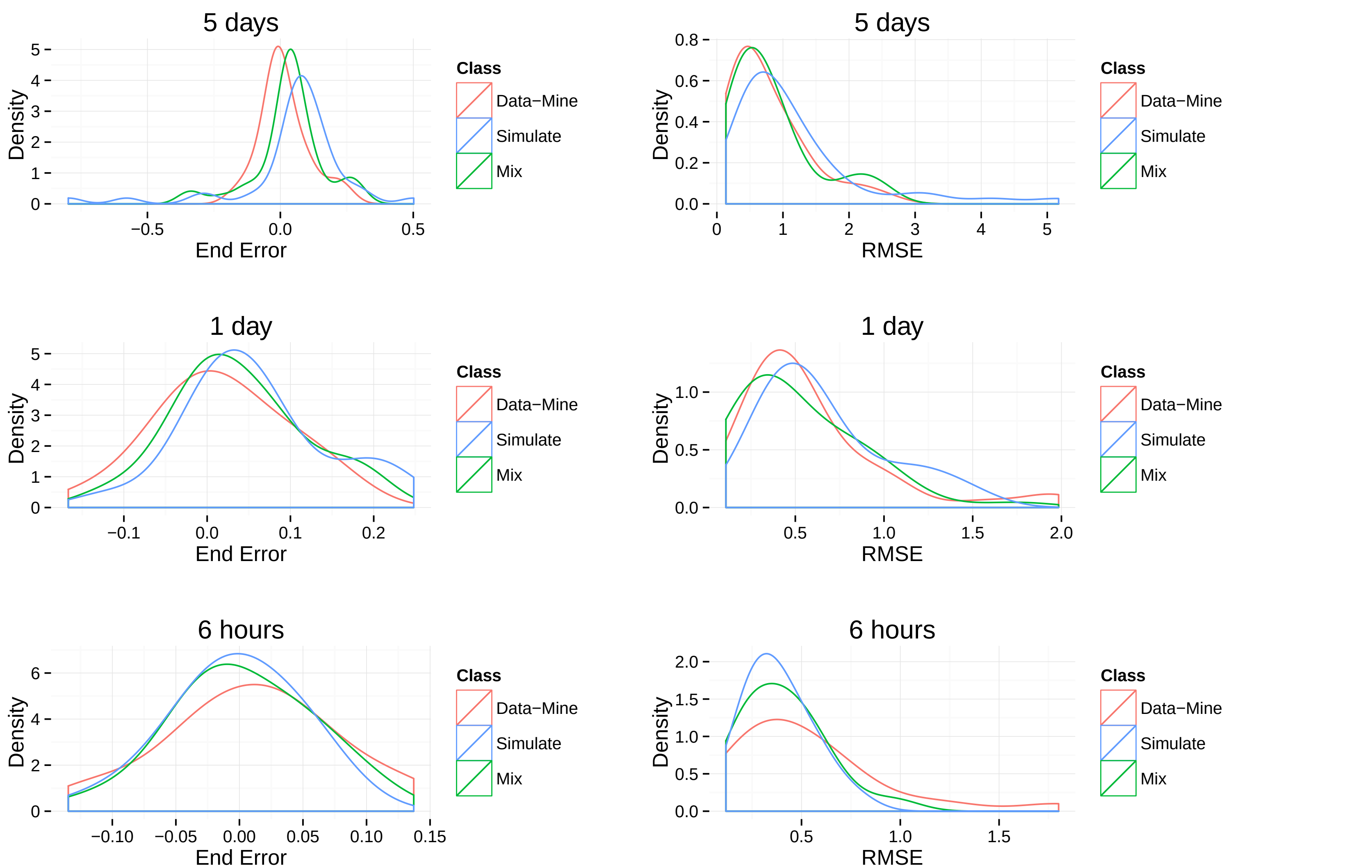


Figure 5.5.2: A summary of the performance of: the simulation-based method; the data-mining-based method; and a mixed method for predicting market sizes, for a range of objectives and based on a range of prediction points.

In each of the cases shown in Figure 5.5.2, using a mix of the two underlying methods gives a result which is never the worst performing of the three, and therefore may be recommended as a suitable robust approach for some general predictive challenge.

5.6 Conclusion

Two modelling paradigms have been introduced and explored for the intention of predicting the increase in size of betting markets. These methods both rely on the user having access to a collection of historical time series of similar markets.

The data-mining approach is shown to provide the better predictions when the market of interest still has a number of days left to run, whilst the accuracy of its predictions are, in part, determined by the quality of the underlying historical dataset.

In comparison, the simulation approach is less sensitive to the individual time series in the historical dataset. This method is much more complicated to set up, given that each time series is broken down into a number of smaller components. However, this allows the method to provide very accurate predictions of market movements in the final hours before the market terminates.

It is also shown that mixing the two methods via a weighting gives a consistent estimate of the future market movements, and is recommended for most applications.

Is is foreseen that this work can form the basis of the larger question of predicting the movement of other features of betting markets, most notably the betting odds.

Chapter 6

Optimal Wager Allocation for String Bets

6.1 Introduction

The idealised betting paradigm takes the following form: an investor calculates the true probability of some event occurring; if they can find offered odds from a bookmaker (or other source) that are favourable compared to the event's probability, then they wager capital on that event, with the betted amount being proportional to their edge.

There are many reasons why this idealised setting may not be true-to-life. Whilst some of these reasons have been explored already (see Section 3.2.3 for a full treatment), a major assumption, not yet tackled by the literature, is that of constant odds.

Consider the setup from Chapter 5. An investor selects a market of interest, then attempts to predict the market's movements such that bets can be placed at some favourable time, such as when the odds are at their highest value. Let o_t be the odds at time t, and let τ be the time point selected to make a wager. If $o_t > o_\tau$ for any $t > \tau$, this not only means that the investor failed to find the 'top-of-the-market' for their bet, but also means that the investor is presented with another opportunity to place a wager, at greater odds than their previous bet.

Given the dangers of overbetting, (see Figure 3.2.1 and the subsequent discussion), it is not immediately clear whether betting again would improve the investor's log growth rate. However, if at some time t, $o_t >> o_\tau$, then intuitively, betting again must become profitable. Given this situation, how does the investor go about calculating how much to wager on the second opportunity, and potentially multiple others?

This scenario has become more relevant since the advent of betting exchanges (for an in-depth description, see Section 2.2). In the past, most betting activity would be made through bookmakers, whose maximum allowable stake at some odds would be beyond most gamblers' bankroll. For betting exchanges, however, the maximum allowable wager at some odds is merely the sum of other gamblers' laid bets, and can be any value. If an investor wishes to wager more than this amount, then they must move onto other, less attractive odds.

Figure 6.1.1 gives an example of an exchange market. The best possible decimal odds for backing Leeds to win are 7.6, with a maximum allowable wager size of £100. If an investor wished to wager more than this 'price limit' on the event, then they would have to move to the less-favourable odds of 7.4, for which the pot limit is £110. How much capital an investor should stake in the inferior betting option (and any

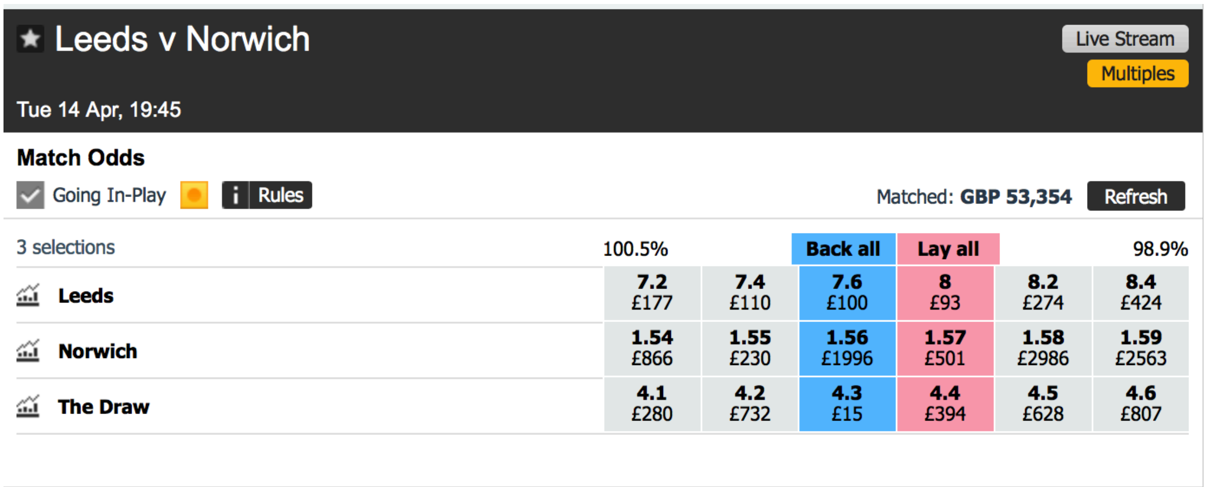


Figure 6.1.1: A Betfair market for the match outcome of a football game in the English Championship.

subsequent betting options) is one of the problems tackled in this work.

A 'string bet' is therefore be defined as: "*a bet such that more than one wager is made at different odds, but on the same exact event*". String bets will be considered in two separate scenarios: betting at markets with pre-event odds movements, "*pre-event betting*"; and betting on exchanges, "*exchange betting*".

This chapter will be organised as follows: some general results regarding string wagers is initially explored, especially Proebsting's Paradox, which brings to the fore questions regarding how string bets should be approached. This analysis shall then be applied to a number of problems: firstly, how should wagers be placed on a general string wager (most commonly faced when odds increase after a bet has been made). Secondly, how should an investor place a bet when their preferred bet size is greater than the event's price limit?

6.2 String Bets

Recall from Chapter 3 that the objective of many betting strategies is to maximise the long-term growth rate of wealth, under the assumption that the game of interest is repeated endlessly. This idea is extended in Breiman et al. (1961), which ensures that maximising the growth rate of a single wager myopically gives the optimal growth rate for a sequence of non-identical opportunities. If there are a number of odds presented for the same event, regardless of the reasons, the growth rate can again be written with the assumption of an endlessly repeated game, and is equal to

$$G_n = p\log(1 + f_1 o_1 + \cdots + f_n o_n) + (1-p)\log(1 - f_1 - \cdots - f_n), \tag{6.2.1}$$

where the notation G_n signifies that this is the overall growth-rate from n wagers at the same event, $f_1, \cdots, f_n$ represent the fractions of the *initial wealth* wagered at (not necessarily ordered) odds $o_1, \cdots, o_n$. Throughout this work it is understood that upper case notation with subscript n refers to a collection of n elements, (for example G_n being the growth rate for a collection of n betting fractions). In addition lower case notation with subscript n refers to the n'th element of such a collection.

The aim, then, is to optimise over $f_i, i = 1, \cdots, n$, a *string bet*. Note that the optimisation challenge depends on the scenario. For the problem of pre-event betting, f_1 will have already been chosen, thus the problem at hand simplifies to merely selecting f_2. In comparison, given exchange betting, the whole collection of fractional wagers must be specified.

6.2.1 Proebsting's Paradox

Consider the situation previously alluded to: a bettor is initially offered fractional odds of o_1 on an event for which they know that the probability of it occurring is p. Assuming that the bettor believes that the odds are fixed, or perhaps even may decrease, they would bet the Kelly fraction of $K(p, o_1) = p - (1-p)/o_1$ (assuming, of course that their aim was to maximise the growth rate of their capital). Before the event occurs, but after they have placed their non-returnable wager, the bettor is offered new odds, $o_2 > o_1$, and must make a decision regarding how much additional capital they wager at the new and improved odds, if any.

The intuitive approach to tackle this problem is to imitate the method from the original Kelly problem; maximise the long-term growth rate of capital via maximising the expected log-utility of the gambling decision step-by-step. In this case, the aim would be to maximise the expression:

$$G_2 = p\log[1 + f_1 o_1 + f_2 o_2] + (1-p)\log[1 - f_1 - f_2] \tag{6.2.2}$$

The growth rate G_2 is then simple to maximise, with the optimal wager on the second option, f_2^* being:

$$f_2^* = \frac{p o_2(1-f_1) - (1-p)(1+f_1 o_1)}{o_2}. \tag{6.2.3}$$

Note that if the first wager is assumed to be the Kelly amount, equation (6.2.3) becomes:

$$f_2^* = \frac{p(1-p)(o_1+1)(o_2-o_1)}{o_1 o_2}. \tag{6.2.4}$$

In addition, if there were n such betting opportunities, and the investor has previously bet the fractional stakes $f_1, \cdots, f_n$ at odds $o_1, \cdots, o_n$, then their log growth rate, given the $n+1$ opportunities is

$$G_{n+1} = p \log\left[1 + \sum_{i=1}^{n} f_i o_i + f_{n+1} o_{n+1}\right] + (1-p)\log\left[1 - \sum_{i=1}^{n} f_i - f_{n+1}\right].$$

which gives the optimal wager size on the $n+1$'st opportunity as

$$f_{n+1}^* = \frac{p o_{n+1}\left(1 - \sum_{i=1}^{n} f_i\right) - (1-p)\left(1 + \sum_{i=1}^{n} f_i o_i\right)}{o_{n+1}}.$$

Any string-betting strategy which calculates a betting amount via maximising the growth rate of the bet as a whole will be referred to as a '*Kelly system*'. By substituting a set of values into equation (6.2.4), the supposed paradox can be explored.

Let $p = 0.5$; a generous party first offers a game where wagered money is tripled if a fair coin lands heads, and is lost if the coin lands tails (giving $o_1 = 2$). The Kelly fraction from these offered odds is $f_G^* = K(0.5, 2) = 0.25$, so the investor will wager 25% of their current wealth on this opportunity, and is labelled Situation G (for 'Good', the naming convention being taken from Zambrano (2014)). A very similar situation would be if the investor was offered the better odds of $o_2 = 5$ on the outcome of a coin-flip; again, wagering the Kelly fraction of $f_B^* = K(0.5, 5) = 0.4$ of their current wealth on opportunity B (for 'Better').

Now consider a third situation, named opportunity M (for 'Mixed'). After the wager, given the original odds of o_1 has been placed, but before they flip the coin, the generous party offers the better odds of $o_2 = 5$. By equation (6.2.4), the optimal fraction to wager (in terms of the initial wealth) on the second option becomes 0.225. Note that by using the Kelly system, the total wager size is now 47.5% of their initial wealth, i.e. $f^*_M = 0.475$. Note that $f^*_M > f^*_B$ even though, on average, the odds in opportunity M are worse than those in opportunity B.

Using a Kelly system therefore seems to lead to illogical results. This scenario is the subject of what is dubbed '*Proebsting's Paradox*', after Todd Proebsting who first noted this phenomenon and corresponded with Ed Thorp; an exchange which is recounted in Thorp (2010). The demonstrated betting scenario seems to show that, by structuring a sequence of betting odds in a particular way, the investor who attempts to allocate their wealth 'optimally' becomes over-invested in the opportunity. To make matters worse, Thorp (2010) then went on to show that if the odds offered were structured such that $o_i = 2^i$, then the bettor using a Kelly strategy, as above, would asymptotically invest their entire wealth on this event. Clearly this feature is detrimental to the case of using the Kelly Criterion to drive investment decisions.

6.2.2 Analysis of Proebsting's Paradox

A comprehensive discussion of the origins of this 'paradox' is given in Zambrano (2014), which reproduces unpublished correspondence between the financier Aaron Brown and Ed Thorp. The key contribution form this work is to introduce the 'cash-equivalent wealth' of a bettor, W^C. This is the amount of cash the bettor possesses, given a valuation of their current assets (here, the pending wagers), and given the current state of information.

Consider the opportunities G and B from before, and consider how much the investor would have to invest in opportunity B to achieve the same growth rate as that given in opportunity G. At the lower odds, the investor achieved a growth rate of $G_1 = 0.5 \log(1 + 0.25 \times 2) + 0.5 \log(1 - 0.25)$, so they stand to make twice as much in the event of a winning bet in comparison to a losing bet. How much cash would the investor have to sacrifice, in the event of betting at odds of 5, to make the wager as profitable as it was before, and no more? i.e. what is the value of lost capital L, such that the growth rate in this new case, $\tilde{G}_1 = p \log(1 + 5f_2 - L) + (1 - p) \log(1 - f_2 - L)$ is the same as before? Solve the following simultaneous relationship:

$$\begin{cases} 5f_2 - L = 0.5 \\ f_2 + L = 0.25 \end{cases} \tag{6.2.5}$$

which has the unique solution of $f_2 = L = 0.125$.

This solution gives two pieces of information. Firstly, $L = 0.125$ implies that 12.5% of the initial wealth is lost as a result of the odds changing from 2 to 5. Secondly, f_2 shows that under these conditions, a wager of 12.5% of the current wealth is equivalent to the Kelly wager of 25% on the initial odds.

The cash-equivalent wealth of the investor changes, then, merely from the changing odds. In this case the investor has, in effect, lost 12.5% of their wealth as a result of the odds movement. The new wealth, known as the marked-to-market wealth, W^C, can be shown, in general, to be:

$$W^C = \frac{1 + f_1 o_1 + (1 - f_1) o_2}{1 + o_2} W. \tag{6.2.6}$$

where W is the investor's initial wealth.

Why is the second bet in the paradox equal to the seemingly illogical value of 0.225, then? The intuitive explanation is that the second wager should be the Kelly fraction, in the scenario explored above, given that the investor has already made a small wager at these odds before, and given that they have already lost some wealth as the odds change. The optimal wager size should be the naïve Kelly wager on the second odds, scaled by the wealth lost, minus the equivalent amount already wagered at these odds, i.e.:

$$\begin{aligned} f_2^* &= (1 - L) K(p, o_2) - f_2 \qquad (6.2.7) \\ &= 0.875 \times 0.4 - 0.125 = 0.225 \end{aligned}$$

where f_2 and L relate to the simultaneous equations (6.2.5).

Additional explanations to convince the reader of the logic behind the optimal wagers in opportunity M are made in Zambrano (2014). The work could suggest that the Kelly betting system is fundamentally flawed, however this is countered by Thorp (2010), which states

"*In contrast to Proebsting's example, the property that betting Kelly or any fixed fraction thereof less than one leads to exponential growth is typically derived by assuming* *a series of independent bets or, more generally, with limitations on the degree of dependence between successive bets*".

So the explored phenomena are not a consequence of the betting system somehow not performing as expected, but are rather due to the Kelly betting system being derived on the key assumption of the independence between wagers, which is explicitly broken in the explored scenarios.

The conclusion of Zambrano (2014) is not that the outcome of the paradox is somehow erroneous, but instead is a vulnerability of the betting system. As previously stated, a bookmaker can derive a series of structured wagers such that a bettor whose objective is the maximisation of some expected utility can be drawn into wagering their entire capital (known as *skimming*). It is therefore of use to derive new betting strategies such that there is no possibility of being skimmed, i.e. as the sequentially offered odds diverge, the overall fractional wager converges to some value smaller than 1.

6.2.3 Alternative Betting Approaches

One reason why bettors are able to be skimmed is that as the odds rise, the bettor remains optimistic that their evaluation of the odds is correct. The solution offered in Zambrano (2014) is to update the bettor's belief in the true probability of the event, after each observation of offered odds, under the assumption that the odds offered are predictive of the event's probability.

Doubly-Conservative Wagering

This new betting system, dubbed 'doubly-conservative', in which the movement of the odds inform the bettor about the true probability of the event, is contrived purposefully to avoid the possibility of being skimmed, with the bettor becoming ever-more pessimistic about the true probability of the event occurring as the odds increase. This paradigm is somewhat analogous to Bayesian updating; this analogy is explored further later.

Doubly-conservative wagering asserts that the belief in the probability of the event deteriorates at least as fast as $\ln(o)^{-1}$. According to Zambrano (2014), this guarantees that this betting scheme avoids the possibility of being skimmed. One example presented is 'logarithmic fractional Kelly betting':

$$\log(\bar{p}) = c\log(p) + (1-c)\log\left(\frac{1}{1+o}\right) \tag{6.2.8}$$

where $\bar{p}$ can be thought of as the posterior belief in the probability of the event, given some prior belief p, and an observation of fractional odds o. The rate at which previous beliefs are discarded in favour of what is implied by observations is described by some 'remembering factor' $c \in (0,1)$.

By choosing $c = 0.5$, Proebsting's betting scenario unravels as follows. Let the belief in the probability of the event be summarised by a single value $p = 0.5$. After the observation of odds $o_1 = 2$, the belief in the probability is $\bar{p} = 0.408$, resulting in a wager of size $K(0.408, 2) \approx 0.112$. After the observation of odds $o_2 = 5$, the probability is again, updated this time to 0.261, for which the maximisation of G_2

from equation (6.2.1) gives the second wager as $f_2 \approx 0.506$, giving a total stake of $F = f_1 + f_2 = 0.163$, far below the value required for the paradox to be observed.

It is shown in Zambrano (2014) that by placing beliefs in this way (namely putting the prior belief in the probability at 0.5, and $c = 0.5$), the maximum total wager, no matter the structured offered odds, is $F \approx 0.192$.

A Bayesian Alternative

The premise behind doubly-conservative wagering is that the belief in the probability of the event should be updated as observations are made. The example of logarithmic fractional Kelly betting gives an updating scheme such that the possibility of being skimmed is nullified. The aim of this alternative betting system is to give this form of update a more familiar statistically setting.

Assume that some previous process (perhaps a predictive model) gives an estimate for the probability of the event of interest occurring. Let this information be represented via the prior distribution $p \sim \text{Beta}(\alpha_m, \beta_m)$.

As discussed, the investor then observes a sequence of odds offered on the outcome of the event of interest. The assumption is that the investor can use these observations as being in some way informative of the underlying event probability. If this were true, the observation of betting odds are also observations of estimates of the probability of the event, via $\hat{p} = (1 + o)^{-1}$, where $\hat{p}$ can be thought of as the probability estimate inferred from the odds.

In order to justify some updating system, given sequential observations of odds, let an inferred probability be the expected value of some distribution $p \sim \text{Beta}(\alpha_0, \beta_0)$

with $\mathbb{E}(p) = \hat{p}$ and $\mathbb{V}(p) = \sigma^2$. The parameters α_0 and β_0 can then be fitted via the method of moments (as seen before in equation (4.2.7)), such that

$$\alpha_0 = \hat{p}\left(\frac{\hat{p}(1-\hat{p})}{\sigma^2} - 1\right), \qquad \beta_0 = (1-\hat{p})\left(\frac{\hat{p}(1-\hat{p})}{\sigma^2} - 1\right). \tag{6.2.9}$$

By combining the two pieces of information (the distribution of the prior belief and the observation of the inferred odds), the posterior belief can be found via

$$\begin{aligned}\pi(p|\hat{p}) &\propto p^{\alpha_0-1}(1-p)^{\beta_0-1}p^{\alpha_m-1}(1-p)^{\beta_m-1} = p^{(\alpha_0+\alpha_m-1)-1}(1-p)^{(\beta_0+\beta_m-1)-1} \\ &\sim \text{Beta}(\alpha_0+\alpha_m-1, \beta_0+\beta_m-1).\end{aligned}$$

So the expectation of the posterior belief in the probability is

$$\begin{aligned}\mathbb{E}(p|\hat{p}) &= \frac{\alpha_0+\alpha_m-1}{\alpha_0+\beta_0+\alpha_m+\beta_m-2} \\ &= \left(\frac{\alpha_m+\beta_m}{\alpha_0+\beta_0+\alpha_m+\beta_m-2}\right)\frac{\alpha_m}{\alpha_m+\beta_m} + \left(\frac{\alpha_0+\beta_0-2}{\alpha_0+\beta_0+\alpha_m+\beta_m-2}\right)\frac{\alpha_0-1}{\alpha_0+\beta_0-2} \\ &= c\mathbb{E}_{\text{prior}}(p) + (1-c)\frac{\alpha_0-1}{\alpha_0+\beta_0-2}\end{aligned}$$

where

$$c = \frac{\alpha_m+\beta_m}{\alpha_0+\beta_0+\alpha_m+\beta_m-2}.$$

It follows that

$$\mathbb{E}(p|\hat{p}) \approx c\mathbb{E}_{\text{prior}}(p) + (1-c)\hat{p} = c\mathbb{E}_{\text{prior}}(p) + \frac{1-c}{1+o}$$

where the approximation holds whenever α_0 and β_0 are large, i.e. whenever σ^2 is small. This is simply the weighted sum of the prior belief and the observation, and is therefore very similar to the form of logarithmic fractional Kelly betting in equation (6.2.8). Note that given $c \neq 1$, the probability deteriorates at the rate of $1/c$, so not at a rate deemed quick enough to entirely avoid the risk of skimming in the worst case, but still gives protection against normal market movements leading to unprofitable betting situations.

As an illustration of this, consider again the classic Proebsting scenario. Let the prior belief in the probability of the event be 0.5. Given that $c = 0.5$, the posterior belief in the probability becomes 0.375 in situation G, leading to a wager size of $K(0.375, 2) \approx 0.625$ and 0.333, leading to a wager size of $K(0.333, 2) = 0$ in situation B. Interestingly, the odds in situation B are so profitable, that the investor using the Bayesian Alternative method adjusts their estimate of the probability to such an extent that wagering on the event becomes unprofitable. Trivially from this point, the fractional wager on the string bet in situation M, is the same as that in situation B (as the investor would turn down the second betting opportunity), and thus the paradox does not appear.

Thresholded Wagering

Thresholded wagering gives a cruder solution, which can be thought of as the 'naïve' betting strategy in this scenario. When entering into a string wager, a maximum total wager is specified before the first bet is made. When making a decision regarding the size of the i'th wager at a single event, then, the fraction is chosen such that:

$$f_i = \min \{\arg\max[G_i | f_1, \cdots, f_{i-1}], \Theta - (f_1 + \cdots + f_{i-1})\}$$

where Θ is the gambling threshold.

For Proebsting's scenario, then, if $\Theta = 0.33$, $f_1 = 0.25$ as before, but $f_2 = \min\{0.225, 0.33 - 0.25\} = 0.07$. Every other odd offered from this point onward would be ignored.

Both of the resolutions to the possibility of skimming offered thus far require some form of tuning; for doubly-conservative betting the remembering factor must be specified, whilst for thresholded wagering, the threshold must be specified. A new approach will now be introduced, which doesn't require the choice of a tuning factor, and which brings the problem back to the familiar territory of simple utility-maximisation.

This approach, named consolidated betting is a novel technique for deciding upon multiple wagers. Consolidated wagering will be shown to have many positive attributes, and is one of the contributions of this chapter.

6.2.4 Consolidated Betting

Consider the log-growth rate of a series of n fractional wagers, $f_1, \cdots, f_n$, given a sequence of fractional odds $o_1, \cdots, o_n$ and with the investor's wealth being normalised to $W = 1$:

$$G_n = p \log(1 + f_1 o_1 + \cdots + f_n o_n) + (1 - p) \log(1 - f_1 - \cdots - f_n).$$

The growth rate can be rewritten such that it can be interpreted as being the result of a single wager, where the size of the wager is the sum of the fractional wagers placed on all of the odds individually:

$$G_n = p \log \left(1 + \sum_{i=1}^{n} f_i \frac{\sum_{i=1}^{n} f_i o_i}{\sum_{i=1}^{n} f_i} \right) + (1-p) \log \left(1 - \sum_{i=1}^{n} f_i \right). \tag{6.2.10}$$

Denote the total wager size, given n individual wagers, as $F_n = \sum_{i=1}^{n} f_i$, and the *consolidated* odds as

$$O_n = \frac{\sum_{i=1}^{n} f_i o_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i o_i}{F_n}.$$

The consolidated odds can be seen as the average odds encountered, weighted by the relative amounts wagered. Given this notation, equation (6.2.10) can be rewritten as

$$G_n = p \log(1 + F_n O_n) + (1-p) \log(1 - F_n). \tag{6.2.11}$$

Note that the growth-rate is now being viewed as the result of a single wager at a single odd.

The simplified growth rate G_n from n wagers can then be maximised by inputting F_n and O_n into the Kelly criterion formula, to find the optimal total wager size as $F_n^* = K(p, O_n)$. Note that as the decision regarding how much to wager at each odd changes, the consolidated odds change, thus calculating F_n^* requires the balancing of the two.

Given a set of n previous wagers and new odds o_{n+1}, finding the optimal current wager, f^*_{n+1} requires optimising the growth rate in equation (6.2.11), solving

$$F_n + f^*_{n+1} = \frac{p(F_n O_n + f^*_{n+1} o_{n+1}) - (1-p)(F_n + f^*_{n+1})}{F_n O_n + f^*_{n+1} o_{n+1}}.$$

So the total optimal wager, given the sum of previous wagers on this opportunity is the Kelly criterion, given the new total bet size and updated consolidated odds.

Finding the optimal stake of the $n+1$'st opportunity is equivalent to solving the quadratic equation

$$f^{*2}_{n+1} o_{n+1} + f^*_{n+1}[F_n(O_n + o_{n+1}) - p(o_{n+1}+1) + 1] + [F_n^2 O_n - F_n(p(O_{n+1}+1) - 1)] = 0. \tag{6.2.12}$$

The term in equation (6.2.12) which is constant in f^*_{n+1} is zero either if $F_n = 0$ or $F_n = K(p, O_n)$. In this case, the equation is easily solvable, with either the risk-free case, $f^*_{n+1} = 0$, or

$$f^*_{n+1} = \frac{p(o_{n+1}+1) - 1 - F_n(O_n + o_{n+1})}{o_{n+1}} = K(p, o_{n+1}) - K(p, O_n)\frac{O_n + o_{n+1}}{o_{n+1}}.$$

So the optimal wager size is the Kelly fraction given the current odds, minus some expression, which takes into account the amount already staked, and the marginal improvement to the odds of the current opportunity.

When the first wager isn't either zero, or the Kelly wager, however, such a simplification is not possible and the optimal wager size is found by completing the square in equation (6.2.12):

$$f_{n+1}^* = \frac{1}{2}K(p, o_{n+1}) - \frac{1}{2}F_n\frac{O_n + o_{n+1}}{o_{n+1}} \pm \sqrt{\frac{1}{4}\left[F_n - \frac{O_n + o_{n+1}}{o_{n+1}} - K(p, o_{n+1})\right]^2 - \frac{F_n}{O_n o_{n+1}}[F_n - K(p, O_n)]} \tag{6.2.13}$$

Given $n = 2$, and under the assumption that at the first available odds, the investor bets the Kelly fraction, the optimality equation can be rearranged to give the somewhat more palatable relationship

$$f_2^* = \frac{(1-p)o_2 - po_1^2}{o_1 o_2} \tag{6.2.14}$$

and with $o_1 = 2$, $o_2 = 5$, and $p = 0.5$, this gives $f_2^* = 0.05$, thus $F_2 = 0.3$. In general, this procedure reframes the betting decision such that it is being made in relation to a single wager at a single odd. Consolidated betting therefore forces the betting decision to be made on the surface of outcomes encountered with a single wager. An example of a surface of this type is visualised in Figure 6.2.1.

The important feature of consolidated wagering is that it is immune to skimming. Recall that the Kelly fraction can be written as $K(p, o) = p - (1 - p)/o$. As $o \to \infty$, the Kelly fraction tends to p. Therefore, there cannot exist a systematic series of odds such that the investor invests more than the maximum wager encountered when there is only one betting option, and thus the investor can not be skimmed when using consolidated odds to drive investment decisions.

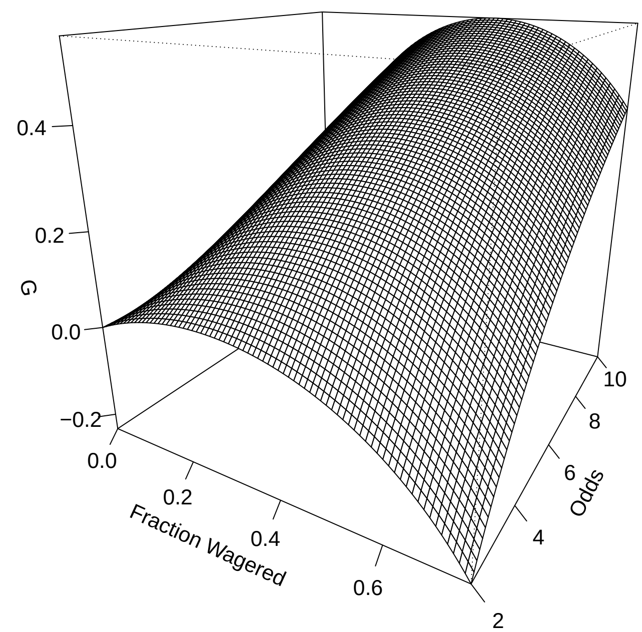


Figure 6.2.1: The log growth-rate (G) achieved by wagering varying fractions of wealth and varying odds, given that the true probability of the event, $p = 0.5$.

6.2.5 Comparison & Evaluation

The three possible betting strategies designed to avoid the possibility of being skimmed are compared, given different scenarios. These strategies will be assessed by their growth rates, and will be compared against the growth rate achieved by the straight utility-maximisation approach, which falls foul of skimming.

In the use of doubly-conservative wagering, the remembering factor must be chosen. One approach to choose c would be for it to be selected such that the resultant growth-rate is maximised. This approach would be unadvised, however, as this results in a remembering rate of around 1 to be chosen in many scenarios, such as Proebsting's example, for which the choice of c is represented in Figure 6.2.2.

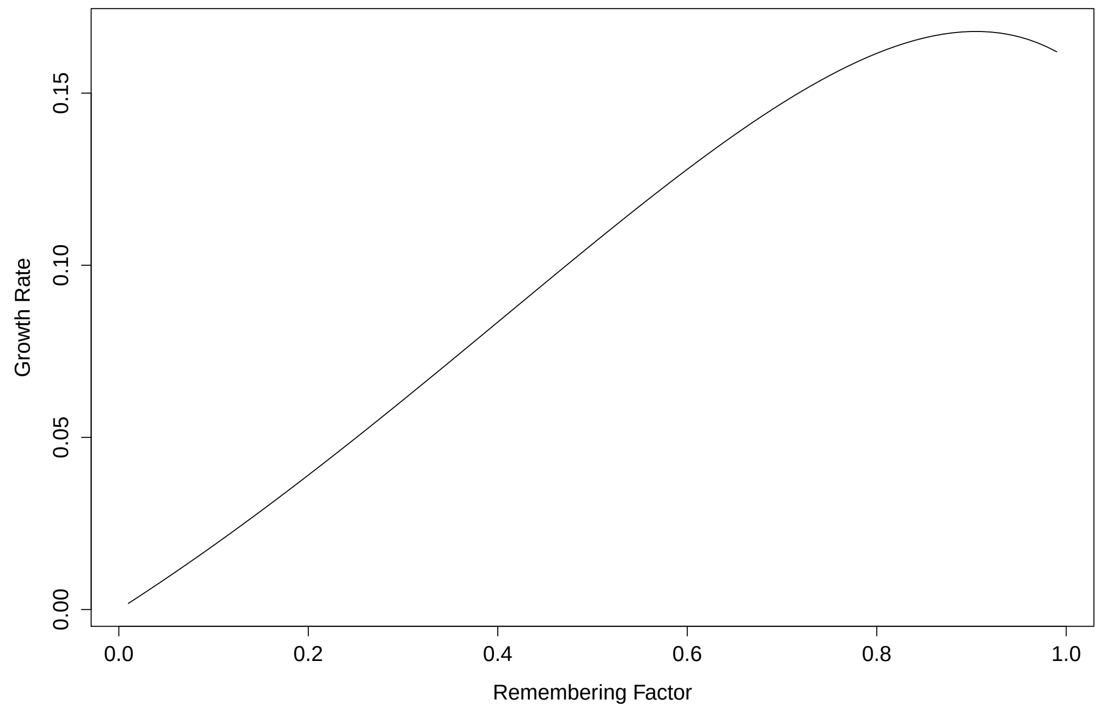


Figure 6.2.2: The growth rate achieved by doubly-conservative betting, with a range of remembering factors, given that the bettor is offered odds of $o_1 = 2$ and $o_2 = 5$, with the $p = 0.5$.

In Figure 6.2.2, the optimal choice of the remembering factor is around $c = 0.9$. It should be noted, however, that this method of wagering avoids only the possibility of the investor being skimmed, i.e. that they wager their entire bankroll on the outcome of a single event. It does not mitigate against the bettor investing a very large stake in the outcome of the event. For that reason, simple maximisation of the resultant growth-rate would not be a good criteria for a cautious gambler. Instead, the method will be assessed with a range of remembering factors, designed to represent a range of attitudes towards the risk of being skimmed. This range of values for c shall also be applied to the Bayesian alternative betting system.

Similarly, the thresholded bettor must choose their maximum overall stake, Θ.

Again, the choice of the threshold will be driven by the attitude of the investor with regards to the risk of being skimmed, and therefore a range of thresholds will be analysed.

Each strategy will be examined given a sequence of odds, with betting decisions being made sequentially, with no knowledge how the odds will appear in the future. A range of different odds movements will be trialled, and will represent markets that are both increasing and decreasing over time, as well as the case where the odds drift randomly.

Five sets of odds will be considered, with the event probability fixed at 0.5 in each case. In the first set of the odds, the initial price is 1.2, rising by 5% each time to a maximum of 1.86. In the second set, the initial price is 2, reducing as by 5% each time to a minimum of 1.26. The third set of odds represents the situation where the odds rise and fall over time, and was generated via a simple random walk. A set of 100 random odds movements were generated via the random walk formula

$$o_i = o_{i-1} + \delta_i, \quad \delta_i \sim \mathrm{N}(0, 0.1^2)$$

with $o_1 = 1.2$, where each wagering strategy was applied to each of the sets of odds individually, and inference was made about the mean of the resultant growth rates. The fourth set of odds represent the unrealistic situation where the odds rise at such a rate that optimising the growth-rate of wealth at each step causes a vast proportion of the bankroll to be risked. These odds simply take the form $o_i = 2^i$, $i = 1, \cdots, 10$.

Finally, the fifth and final set of odds is created from the betting exchange data used and explored in Chapter 5. For 100 random matches from the data, the last 20 odds movements were extracted, for which each of the potential betting strategies were applied.

For doubly-conservative wagering and its Bayesian alternative, three remembering factors are considered, representing a range of beliefs regarding the importance of rescaling the belief in the event probability, given observations of successive odds. Additionally, for thresholded betting, three separate thresholds were considered, again representing a range of beliefs, this time towards the maximum allowable wager size.

From this point on, shorthands are adopted for each of the potential betting strategies:

- GRM: the "growth-rate maximisation" approach, seen in Section 6.2.1.
- DCW: the "doubly-conservative" wager.
- BA: the Bayesian alternative to the doubly-conservative wager.
- TW: the thresholded wagering approach, relating to the method which adopted the GRM method with some upper limit on the stake size.
- CW: the consolidated wagering approach, outlined in Section 6.2.4.

	GRM	DCW			BA			TW			CW
		$c = 0.25$	$c = 0.5$	$c = 0.75$	$c = 0.25$	$c = 0.5$	$c = 0.75$	$\Theta = 0.2$	$\Theta = 0.3$	$\Theta = 0.4$	
5% Increase	0.0068	0.0018	0.0031	0.0042	0.0018	0.0031	0.0045	0.0059	0.0068	0.0068	0.0132
	0.2621	0.02	0.0407	0.064	0.0208	0.0417	0.0672	0.2	0.2621	0.2621	0.1392
5% Decrease	0.0589	0.0233	0.0418	0.0542	0.0266	0.0448	0.0554	0.0584	0.0589	0.0589	0.0589
	0.25	0.0533	0.1124	0.1778	0.0625	0.125	0.1875	0.2	0.25	0.25	0.25
Random Walk	0.0062	0.0041	0.0066	0.0074	0.0061	0.0061	0.0061	0.006	0.0061	0.0061	0.0058
	0.1378	0.0322	0.0658	0.101	0.0225	0.0454	0.0740	0.1287	0.1327	0.1327	0.0921
Exponential	0.5889	0.0486	0.1634	0.512	0.0881	0.5444	1.235	0.0196	0.087	0.1156	1.0852
Increase	0.9437	0.0646	0.1657	0.3563	0.1087	0.24	0.535	0.2	0.3	0.4	0.4848
Exchange	0.0058	0.0058	0.0101	0.0126	0.0059	0.0097	0.0122	0.0124	0.0129	0.0130	0.0127
Odds	0.0821	0.0195	0.0405	0.0633	0.0206	0.0407	0.0610	0.0811	0.0865	0.0874	0.0790

Table 6.2.1: A comparison of betting strategies, given five scenarios describing the potential evolution of the odds. GRM related to the strategy which blindly maximises the growth rate at each stage. For each result two figures are given: the upper cell shows the growth rate for the chosen set of wagers; the lower cell shows the proportion of the initial wealth invested to achieve this growth rate.

The results shown in Table 6.2.1 give many interesting insights. The growth rates are calculated via repeatedly applying the investment rules to the set of odds sequentially. In the case of the random walk, the results reported represent the median of the 100 simulated odds sets.

When the odds are increasing over time, the highest growth rate is achieved by CW, which recommends investing a total of 13.9% of initial wealth to achieve a growth rate of 0.0132. In comparison, the GRM approach achieves roughly half the growth rate whilst investing roughly double the wealth. The reason for this is clear; CW is naturally more cautious with string bets, so as the odds increase, the consolidated wagerer has more capital to invest in the better opportunities. Note that DCW and BA seem to have the opposite problem; for all three choices of risk c, the bettor who updates their belief in the event probability seems to become pessimistic about the probability of the event quickly, meaning that even as the odds improve, the event doesn't seem profitable, and thus little capital is invested in the better odds.

When the odds are decreasing, it is conjectured that the optimal strategy should be to bet the Kelly amount at the first odds observed, then invest nothing in any future opportunity (considered more formally and proven in Lemma 6.4.1). For the odds given, this meant that the optimal strategy was to wager 25% of total wealth at the first offering, then stopping. This gives a growth rate of 0.0589. Both GRM and CW do just this, as does TW, whenever the threshold is set to be equal to, or larger than 0.25. Again, all choices of risk result in the investor under-utilising the best odds. In this case, it is due to them re-evaluating the probability of the event immediately, decreasing their belief in the value of the event's probability.

In the instance of the odds evolving as a realisation of a random walk, CW does not do as well as other candidate methods, in terms of growth-rate achieved. As an example, the GRM achieves a growth rate of 0.00619, compared to CW's 0.00576 a 6.5% decrease. CW creates this growth rate while committing around 67% of GRM's investment to wagers. In addition, here DCW seems to perform the best, in terms of its growth rate when $c = 0.75$, whilst investing only 10% of the investor's wealth.

When considering the betting exchange's odds movements, the performance of consolidated wagering seems to echo what is seen in the other cases. In short, although it does not always achieve the greatest growth rate amongst the candidate methods, it tends to achieve a competitive growth rate whilst committing a lower total stake.

On the whole, CW seems to be a good choice when wagering with uncertainty about future odds movements. When odds increase, or decrease steadily over time, CW achieves better than or equal growth rates relative to the candidate methods, with less risk. When the odds move randomly over time, there is evidence to say that CW does *nearly* as well as the best result achieved by other approaches, but again does so with less risk.

In terms of the use of these techniques in real betting scenarios, it is likely that an investor would invest at every opportunity in a market with random movements. More likely would be the situations previously alluded to which the odds strictly increase over time when the investor places a bet at a non-optimal time. Another likely situation is an investor betting on a set of odds which are strictly decreasing, a situation that occurs when betting on exchanges.

6.3 Betting under Uncertainty

So far, the betting discussion in this chapter has, made the assumption that the probability of the event occurring is some known quantity p. No matter what method is being used to estimate the value of p, be it a complex statistical model, or a complete guess, there must inevitably be some uncertainty around the value of these estimates. If the value of wagers is claimed to be optimal, then, the uncertainty around the estimate of the event probability must be considered.

There have been a number of papers in the literature which have attempted just that. In Medo et al. (2008), the authors consider bets where the probability is known to be one of two discrete values, with the conclusion being that an investor who knows which of the two values is correct benefits from a larger growth rate of wealth than an investor who does not have this information.

In Sinclair (2014), confidence intervals are derived for the Kelly criterion, given uncertainty about the event probability. The intention of this work is not to reoptimise the Kelly fraction, but instead inform what the wager size such that the probability of overbetting is bounded above by some fixed amount.

The most comprehensive approach is given in Baker and McHale (2013). The original work contained in this section follows closely from this paper, which will be summarised here. Baker and McHale (2013) asserts that the investor's belief of the probability can be described via some density function $g(q)$, *with mean of the true probability*, p and some variance σ^2. This can be thought of as the probability estimate being drawn from some random variable, Q, whose probability density function is $g(q)$.

Say an investor samples a probability estimate q, and they wager the Kelly fraction, $K(q, o)$, then the expected (maximised) growth rate is given as:

$$\mathbb{E}(u^*) = \int_0^1 g(q)\{p \log[1 + oK(q, o)] + (1 - p) \log[1 - K(q, o)]\}dq$$

where u^* (a function of the betting fraction, f) represents the maximised growth.

Introduce a scaling factor to the fraction wagered, $\lambda > 0$. If $\lambda \in (0, 1)$, this indicates that the optimal wager under uncertainty is less than the naïve estimate (which is the Kelly criterion in this example).

In order to find the optimal value of this scaling factor, the expected utility can be considered:

$$\mathbb{E}(u^*) = \int_0^1 g(q)\{p \log[1 + o\lambda K(q, o)] + (1 - p) \log[1 - \lambda K(q, o)]\}dq. \tag{6.3.1}$$

The idea is to optimise equation (6.3.1) with respect to the scaling factor λ. If the optimal scaling factor is λ^*, with $\lambda^* < 1$, then the introduction of uncertainty has indeed altered the optimal betting decision.

If only the mean and variance of the sampling distribution of the predicted probability are specified, then it may be reasonable for g to be approximated by a Beta distribution with parameters determined by the equations (6.2.9). Equation (6.3.1) can be solved exactly using an iterative method, such as Newton-Raphson, which would use the procedure:

$$\lambda_{n+1} = \lambda_n - \frac{d\mathbb{E}(u^*)/d\lambda}{d^2\mathbb{E}(u^*)/d\lambda^2}.$$

We term this the “Beta method”.

Alternatively, let some general function describing the optimal betting fraction, given some event probability p, be written as $s^*(p)$. By Taylor series expanding equation (6.3.1) around $s^*(p)$ an approximation for the optimal expected is

$$\mathbb{E}(u^*) \approx \mathbb{E}[u(s^*(p))] + \frac{1}{2}\frac{\partial \mathbb{E}[u(f)]}{\partial f}\bigg|_{f=s^*(p)} \int_0^1 [\lambda s^*(q) - s^*(p)]^2 g(q)dq. \qquad (6.3.2)$$

The maximisation of the growth rate in equation (6.3.2) requires differentiation with respect to λ, and then setting the whole expression equal to 0 to solve for λ. The optimisation of (6.3.2) over λ reduces to merely solving

$$\frac{\partial}{\partial \lambda}\int_0^1 [\lambda s^*(q) - s^*(p)]^2 g(q)dq = 0.$$

This gives

$$\Rightarrow \lambda^* = \frac{s^*(p)\int_0^1 s^*(q)g(q)dq}{\int_0^1 s^*(q)^2 g(q)dq}. \qquad (6.3.3)$$

Now suppose that the form of $s^*(q)$ is linear in q, i.e. $s^*(p) = ap+b$ for some constants a and b. Then the expression for the optimal wager size reduces further to

$$\lambda^* = \frac{s^*(p)^2}{\int_0^1 s^*(q)^2 g(q)dq},$$

which utilises the fact that $\mathbb{E}(Q) = p$. Further, noting that $\mathbb{E}(Q^2) = p^2 + \sigma^2$, then

$$\lambda^* = \frac{a^2p^2 + 2abp + b^2}{\int_0^1 [a^2q^2 + 2abq + b^2]g(q)dq} = \frac{a^2p^2 + 2abp + b^2}{a^2(p^2 + \sigma^2) + 2abp + b^2} = \frac{s^*(p)^2}{s^*(p)^2 + a^2\sigma^2} \quad (6.3.4)$$

The scaling, then depends entirely on the uncertainty of the event probability estimate and the coefficient of the p term of $s^*(p)$. As the uncertainty tends to zero, the optimal scaling factor approaches 1, and thus no scaling takes place. On the contrary, as the variance diverges, the optimal scaling tends to 0, and the bet size is shrunk to 0.

As an example of this method, Baker and McHale (2013) calculate the optimal scaling for the simple Kelly wager, $s^*(p) = K(p, o_1)$, given a measure of uncertainty in the event probability. The coefficient of p in the Kelly fraction is $(1 + o_1)/o_1$, so the optimal scaling for the Kelly wager is

$$\lambda^* = \frac{K(p, o_1)^2}{K(p, o_1)^2 + \sigma^2 \left(\frac{1+o_1}{o_1}\right)^2} \quad (6.3.5)$$

which yields an optimal wager size as

$$s^*(p, \sigma^2) = \lambda^* K(p, o_1) = \frac{K(p, o_1)^3}{K(p, o_1)^2 + \sigma^2 \left(\frac{1+o_1}{o_1}\right)^2}.$$

For forms of $s^*(p)$ for which higher orders of p exist, a closed form simplification for λ^* of this type is not found.

The severity of the scaling depends primarily upon the amount of uncertainty, σ^2 in the prediction for p. When the probability is certain, no scaling takes place. As the uncertainty increases, the scaling tends towards zero, and the optimal wager similarly tends to zero, see Figure 6.3.1. This relationship is concave, thus the rate of decrease in the scaling of the optimal wager is most pronounced when the variance is small. As the size of the probability estimate's variance increases from 0 to 0.05, the scaling of the wager decreases from 1 to 0.357, whereas as the variance increases from 0.05 to 0.1, the scaling decreases only by a further 0.14 to 0.217.

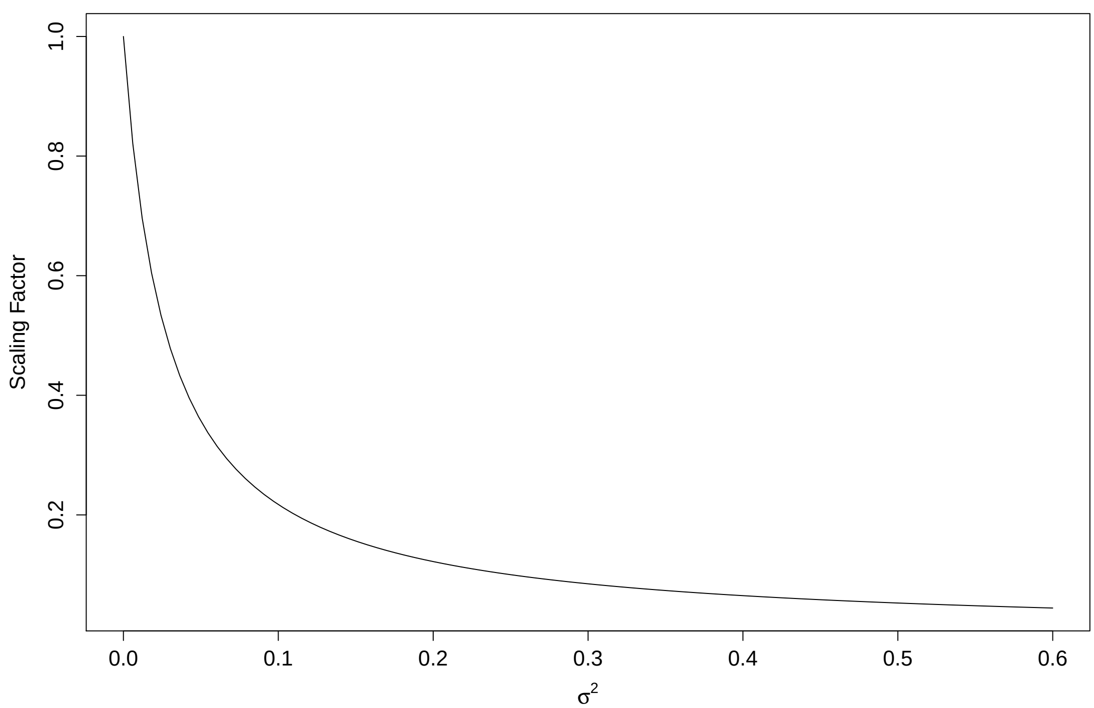


Figure 6.3.1: The value of λ^* from equation (6.3.5), given $p = 0.5$, $o_1 = 2$, $o_2 = 5$, over a range of variances, calculated via the Beta Method.

These findings show that betting under uncertainty, then, even a small amount of doubt regarding the accuracy of the probability estimate will result in a significant downscaling of the fraction wagered.

6.3.1 String Bets under Uncertainty

Imagine an investor has some probability-generating statistical model, which, as before, outputs the probability of some gambling event with a measure of uncertainty. Given some offered odds, they might bet some amount of their capital and again, as before, the odds might change between the time of the bet being placed and the beginning of the event. The problem is now confounded twice; what is the optimal bet size, given that one or many bets have already been placed on the event, and given some (potentially updating) information regarding the probability estimate?

This question will be investigated for the three main string-bet evaluation methods investigated in Section 6.4.

First, let $s^*(p)$ be the optimal bet size of a certain wagering strategy, but without taking uncertainty into account (in these cases, the $s^*(p)$ are linear in p, but this would not be true in general). In this way, $s^*_{\mathrm{GRM}}(p)$ signifies the function describing the optimal bet size, given that the investor is maximising the growth-rate at each stage (see Table 6.2.1).

6.3.2 Growth-Maximisation

When approaching betting via maximisation of some log-growth function, the optimal bet size given that the probability is known as p is

$$s^*_{GRM}(p) = \frac{po_2(1-f_1) - (1-p)(1+f_1o_1)}{o_2}$$

From equation (6.3.4), as $s^*_{\text{GRM}}(p)$ is linear in p, the optimal scaling can be found easily, as

$$\lambda^*_{GRM} = \frac{s^*_{GRM|f_1}(p)^2}{s^*_{GRM|f_1}(p)^2 + \sigma^2\left(\frac{o_2(1-f_1)+1+f_1o_1}{o_2}\right)^2}. \tag{6.3.6}$$

For small values of f_1 and some fixed variance, the relationship between the initial wager and the scaling factor is concave and decreasing. This gives further reason for the wager size to be scaled downwards, as this suggests that even a small amount of uncertainty regarding the event probability causes the optimal wager size to be shrunk significantly. For the classic Proebsting example, placing a variance on the initial estimate of the probability to be 0.05, gives the initial wager to be only 8.9% of the investor's wealth, in comparison to 25% from before. Further, the optimal size of the second wager is 20.7%, resulting in a total wager size of 29.6% of the initial wealth, in comparison to 47.5% before. Alternatively, for n previous wagers, this optimal wager size under uncertainty easily scales to

$$\lambda^*_{\text{GRM}} = \frac{s^*_{\text{GRM}}(p)^2}{s^*_{\text{GRM}}(p)^2 + \sigma^2\left(\frac{o_{n+1}(1-F_n)+1+F_nO_n}{o_{n+1}}\right)^2}. \tag{6.3.7}$$

6.3.3 Consolidated Betting

Within the framework of consolidated betting, the optimal wager given a previous Kelly-optimal bet will be calculated. The assumption regarding the previous wager is necessary, as allowing the first bet to be some general value gives a relationship (shown in equation (6.2.13)) which becomes intractable when assessing its properties

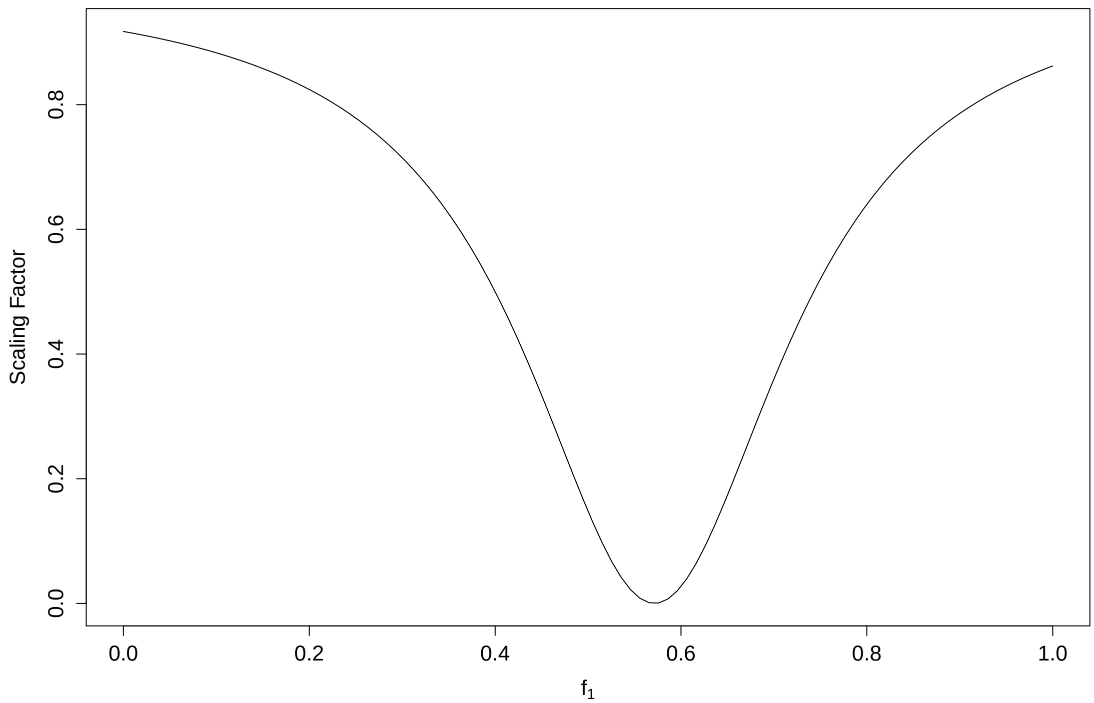


Figure 6.3.2: The value of $\lambda^*_{GRM|f_1}$, given $p = 0.5$, $o_1 = 2$, $o_2 = 5$, $\sigma^2 = 0.01$, over a range of sizes of the initial wager.

under uncertainty. Let

$$s^*_{CW}(p) = \frac{(1-p)o_2 - po_1^2}{o_1 o_2}$$

then using the same technique as in Section 6.3.2 gives:

$$\lambda^*_{CW} = \frac{s^*_{CW}(p)^2}{s^*_{CW}(p)^2 + \sigma^2 \left(\frac{o_1^2 + o_2}{o_1 o_2}\right)^2}. \tag{6.3.8}$$

This, then is the recommended scaling amount for string bets under uncertainty. Using this will rapidly stunt the size of wagers placed on subsequent offered odds, and thus by taking the change of odds, and the uncertainty around the probability estimate into account, the investor who acts in such a way becomes very risk-averse. When considering a number of previous wagers, the ideal betting amount, adapted from equation (6.3.8) becomes

$$s^*_{\mathrm{CW}}(p, \sigma^2) = \frac{s^*_{\mathrm{CW}}(p)^3}{s^*_{\mathrm{CW}}(p)^2 + \sigma^2 \left(\frac{O_n^2 + o_{n+1}}{O_n o_{n+1}}\right)}$$

where the notation used for consolidated wagering has been adopted.

6.4 String Bets in Exchanges

As introduced in Section 6.2, string bets in exchanges occur when an investor's ideal betting amount is constrained by some price limit. These price limits are an enforced maximum allowable stake at some offered odds, and are a result of the odds-setters being other investors, who set limits to their potential losses.

Note that although betting exchanges allow investors to both back and lay events, this will not be explored in this work. On a simplistic level, betting and laying events are intrinsically the same thing; backing some event is the same as laying all other events, and vice versa. The exact nature of the placement of bets when laying is considered is an area of work in itself, and has been recently explored in Noon et al. (2013).

6.4.1 Restricted Markets

Let the set of available odds offered on a particular event, known as *price options*, be ordered such that for any pair of price options, indexed by i and i' be such that $o_i > o_{i'}$ if and only if $i < i'$. For each price option o_i, let l_i be its associated *price limit*; the maximum allowed wager at this price. In addition, define $\tilde{l}_i = l_i/W$, which is the size of the i'th pot limit as a fraction of the investors initial wealth W. The investors wealth can then be normalised to $W = 1$ without loss of generality.

As the odds have been ordered from high to low, the investor observes the odds to be decreasing as they are considered sequentially. Table 6.2.1 can then be used to choose an appropriate investment strategy. Decreasing odds correspond to the '5% decrease' rows, which registers many of the potential techniques as having equal efficacy in maximising long-term growth. However, given that betting exchanges are constructed from the offered odds of a large number of other investors, the chances of being skimmed by nefarious means are very small. In addition, this work concentrates on the investor choosing between a set of odds (with their associated price limits) at a single point in time. Therefore, the classic setup of Proebsting (see Section 6.2.1) is not relevant to this case.

Given these reasons, as well as the lure of the ease of calculations, the pure growth-rate maximisation approach will be used to analyse the optimal allocation of wagers in exchange betting. Let f_i^* and $\bar{f}_i^*$ be the general notation relating to some unconstrained and constrained optimal wager, in relation to the i'th price option respectively.

Lemma 6.4.1. *If a market has n different price options, $o_1 > o_2 > \cdots > o_n$ with normalised price limits $\tilde{l}_1, \tilde{l}_2, \cdots, \tilde{l}_n$, then it is always preferable to invest $\tilde{l}_i$ in price option o_i before considering any other option $o_{i'}$ with $i' > i$, given some strictly increasing utility of wealth.*

Proof. Assume that the total wager size is $F_n = \sum_{i=1}^{n} f_i$, and is known. This wager's associated expected return, given some general strictly-increasing utility of wealth u is

$$\mathbb{E}\{u[f_1, \cdots, f_n]\} = pu\left[1 + \sum_{i=1}^{n} f_i o_i\right] + (1-p)u[1 - F_n]. \tag{6.4.1}$$

As F_n is fixed, the maximisation of equation (6.4.1) is equivalent to the maximisation of

$$\max_{f_1, \cdots, f_n} \sum_{i=1}^{n} f_i o_i.$$

Let k be some constant with the property that

$$\sum_{i=1}^{k-1} \tilde{l}_i < F_n < \sum_{i=1}^{k} \tilde{l}_i.$$

The optimal solution to equation (6.4.1), given k is then

$$f_i^* = \begin{cases} \tilde{l}_i & \text{if } 1 \leq i \leq k-1 \\ F - \sum_{i=1}^{k-1} \tilde{l}_i & \text{if } i = k \\ 0 & \text{otherwise.} \end{cases}$$

As equation (6.4.1) is strictly increasing in each of the o_i, any deviation from this strategy would result in an expected utility being lower than that prescribed above.

$\square$

Define *ID-multiplicative* functions, U to be the class of functions with the property that the inverse of their derivative; $V = (U')^{-1}$ is multiplicative, i.e. $V(ab) = V(a)V(b)$. It is of use to clarify which functions conform to this ID-multiplicative property. In particular, the class of isoelastic utility functions shall be considered (and are equivalent to a one-parameter Box-Cox transformation) and are defined as

$$u(c) = \begin{cases} \frac{c^{1-\eta}-1}{1-\eta}, & \eta > 0, \, \eta \neq 1 \\ \\ \ln(c), & \eta = 1. \end{cases} \tag{6.4.2}$$

Within this class of utility functions, the parameter η represents the risk aversion. A choice of a larger η results in a more risk-averse strategy. The only utility functions which have a constant relative risk aversion are in this class (Arrow, 1971), meaning that decision making is not affected by scale, leading to the familiar fractional betting strategies of Kelly (see Section 3.2.2 for further details).

Theorem 6.4.2. *A twice-differentiable function is ID-multiplicative if and only if it belongs to the isoelastic utility family, or is constant.*

Proof. First note that

$$u'(c) = c^{-\eta}, \quad u''(c) = -\eta c^{-\eta-1}.$$

Given that $\eta > 0$, the second derivative is negative over its whole domain, and $u(c)$

is therefore concave on its whole domain. In addition, $(u')^{-1}(c) = c^{-\frac{1}{\eta}}$. It is clear to show that this inverse function is multiplicative. If the function is a constant, the inverse of its derivative is 0, which is trivially multiplicative.

On the other hand, a function which is ID-multiplicative has the feature that $u'^{-1}(ab) = u'^{-1}(a)u'^{-1}(b)$. This takes the form of the fourth of Cauchy's functional equations. It was shown, and reproduced in (Aczél, 1969), that for some function f,

$$f(ab) = f(a)f(b) \Rightarrow f(a) = a^k$$

for some constant k, or $f(a) \equiv 0$. If $u'^{-1}(a) = a^k \Rightarrow u'(a) = a^{k'}$ for $\tilde{k} = k^{-1}$. Then $u(a) = \frac{1}{\tilde{k}+1}a^{\tilde{k}+1} + \lambda$, where λ is another constant. Let $\tilde{k}+1 = 1-\eta$ and $\lambda = -(1-\eta)^{-1}$, then $u(a)$ conforms to the definition of the isoelastic family in equation (6.4.2). □

It shall now be shown that reasonably simple inference can be made regarding the optimal wager allocation upon the n betting options, as long as the utility of wealth is ID-multiplicative, or equivalently, as long as the utility of wealth belongs to the isoelastic family. Note that the isoelastic family of utilities are strictly increasing (and thus constant functions shall be ignored).

Lemma 6.4.3. *Consider an unrestricted market with underlying probability of p and with some price option o_i, then the optimal fractional wager sizes f_i^*, given some ID-multiplicative utility of wealth U are*

$$f_i^* = \frac{V(1-p) - V(po_i)}{V(1-p) + o_iV(po_i)} \tag{6.4.3}$$

where $V = (U')^{-1}$.

Proof. The expected utility of a wager of size f_i can be written as

$$\mathbb{E}(U[f_i]) = pU[1 + f_i o_i] + (1 - p)U[1 - f_i]$$

with its derivative being

$$\begin{aligned}\frac{\partial \mathbb{E}(U[f_i])}{\partial f_i} &= po_i U'[1 + f_i o_i] - (1 - p)U'[1 - f_i] \qquad (= 0 \text{ at } f_i = f_i^*)\\ &\Rightarrow po_i U'[1 + f_i^* o_i] = (1 - p)U'[1 - f_i^*].\end{aligned}$$

Applying the multiplicative function $V = (U')^{-1}$ to both sides results in

$$[1 + f_i o_i]V(po_i) = [1 - f_i]V(1 - p) \tag{6.4.4}$$

with the rearrangement of equation (6.4.4) giving the solution shown in equation (6.4.3). □

Lemma 6.4.4. *Given a price option* o_i *with associated normalised price limit* $\tilde{l}_i$ *and known probability* p*, then the wager size required to maximise some ID-multiplicative and concave utility function* U*, is*

$$\bar{f}_i^* = \min\left[\tilde{l}_i, \frac{V(1 - p) - V(po_i)}{V(1 - p) + o_i V(po_i)}\right]. \tag{6.4.5}$$

Proof. It is known that the unconstrained maximum of $\mathbb{E}\{U[f_i]\} = pU[1 + o_i f_i] +$

$(1-p)U[1-f_i]$ gives the f_i^* from from Lemma 6.4.3. Given that in addition, these isoelastic utilities are concave, the second derivative is negative at this point and the function $\mathbb{E}\{U[f_i]\}$ is increasing up to f_i^* , implying that if $l_i < f_i^*$, then betting as much as possible maximises the log utility at every step, in turn implying the general result. $\square$

Given some ordered list of price options, $o_1, \cdots, o_n$ with their associated normalised price limits $\tilde{l}_1, \cdots, \tilde{l}_n$, the optimal strategy for a general betting scenario shall be derived. This strategy will be dependent upon the underlying probability of the event occurring, p being known, as well as some measure of the investor's attitude towards risk aversion, parameterised by the risk aversion measure η from the isoelastic utility family shown in equation (6.4.2).

Firstly, let

$$\alpha_i(\eta) := \left(\frac{1-p}{po_i}\right)^{\frac{1}{\eta}},$$

which is some measure of the profitability of a wager, given the measure of risk-aversion found in the isoelastic family of utility functions, in equation (6.4.2).

Lemma 6.4.1 then simplifies the problem of allocating fractional wagers to a general market to a large degree. Given that the investor is betting into the i'th price option, it guarantees that the previous $i-1$ price options must already be fully utilised. The optimal size of wager in the i'th, and current price option can be found by optimising $\mathbb{E}[u(f_i)]$ over f_i where:

$$\mathbb{E}(u[f_i]) = \frac{1}{1-\eta}\left[p\left(1+\sum_{j=1}^{i-1}\tilde{l}_j o_j + f_i o_i\right)^{1-\eta} + (1-p)\left(1-\sum_{j=1}^{i-1}\tilde{l}_j - f_i\right)^{1-\eta} - 2\right].$$

Thanks to Lemma 6.4.4, this achieves its constrained maximum at

$$\bar{f}_i^* = \min\left\{\tilde{l}_i, \max\left[0, \frac{1-\sum_{j=1}^{i-1}\tilde{l}_j - \alpha_i(\eta)\left(1+\sum_{j=1}^{i-1}\tilde{l}_j o_j\right)}{1+o_i\alpha_i(\eta)}\right]\right\}. \tag{6.4.6}$$

Given η, define $\omega_i^0(\eta)$ and $\omega_i^1(\eta)$ as the minimum wealth required bet into the i'th price option to be profitable, and the minimum wealth required to invest the entire price limit in the i'th option, respectively. Equation (6.4.6) can be rearranged to find these values, by setting $\tilde{f}_j^*$ equal to 0, then $\tilde{l}_i$, and removing the normalisation condition $W = 1$:

$$\omega_i^0(\eta) = \frac{\sum_{j=1}^{i-1}\tilde{l}_j + \alpha_i(\eta)\sum_{j=1}^{i-1}\tilde{l}_j o_j}{1-\alpha_i(\eta)} \tag{6.4.7}$$

$$\omega_i^1(\eta) = \frac{\sum_{j=1}^{i-1}\tilde{l}_j + \alpha_i(\eta)\sum_{j=1}^{i-1}\tilde{l}_j o_j + \tilde{l}_i}{1-\alpha_i(\eta)} \tag{6.4.8}$$

Given these results, the only additional information required to find the optimal wagers for each of the n price options is the number of options which shall be fully invested into. This can be found, and utilised by the following algorithm:

Algorithm 6 Method to calculate optimal stake in general number of betting options and price limits

Input: Set of available odds $o_1, \cdots, o_n$. Set of odds' normalised price limits $\tilde{l}_1, \cdots, \tilde{l}_n$.
Probability of event p. Measure of risk-aversion η.
Set $i = 1$, $j = \omega_1^1(\eta)$ (from equation (6.4.8)). Create empty vector $f[]$.
while $j < W$ **do**
  $f[i] = \tilde{l}_i$.
  $i = i + 1$.
  $j = \omega_i^1(\eta)$.
end while
$f[i] = \bar{f}_i^*$ (from equation (6.4.6)).

6.4.2 The Two-Option Example

The solution offered in Algorithm 6 is demonstrated in a simple example. Imagine that there are only two price options, 1.25 and 1.1, with both pot limits equal to £75 and an estimated probability of success of 0.5. Clearly $\tilde{l}_i$, $i = 1, 2$ depends on the wealth of the investor, W. Choose $\eta = 1$, such that the investor's risk aversion is the same as one who bets as Kelly recommends, maximising the logarithm of wealth over the wagers. Given this choice of risk aversion, equations (6.4.7) and (6.4.8) simplify to:

$$\omega_2^0(1) = \frac{\tilde{l}_1(p + (1-p)o_1 o_2^{-1})}{K(p, o_2)}, \qquad \omega_2^1(1) = \frac{\tilde{l}_1(p + (1-p)o_1 o_2^{-1}) + \tilde{l}_2}{K(p, o_2)}.$$

In addition, under this simplification, the unrestricted optimal wager reduces to

$$f_2^* = K(p, o_2) - \tilde{l}_i \left(\frac{p o_2 + (1-p) o_1}{o_2} \right) = K(p, o_2)[1 - \omega_2^0(1)]. \tag{6.4.9}$$

Equation (6.4.9) demonstrates the nature of the optimal betting strategy, which is closely tied to the form of $\omega_i^0(\eta)$. By following Algorithm 6 there is a relationship between an investor's total wealth available, and their total wager size (as this is used to calculate $\tilde{l}_1, \cdots, \tilde{l}_n$ which makes the importance of this term clear. This is demonstrated in Figure 6.4.1.

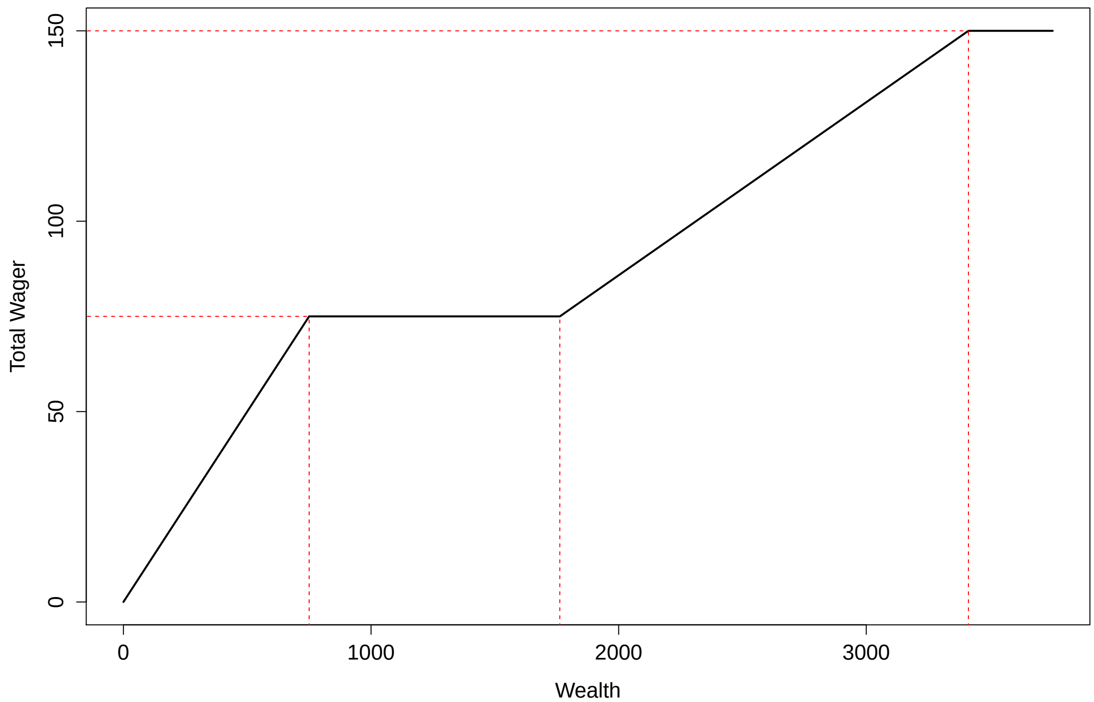


Figure 6.4.1: The wealth required to bet into two options, with odds of 1.25 and 1.1 respectively, bet limits on both options being £75, and an event probability of 0.5.

The first dotted vertical line can be calculated as the first price limit divided by $K(p, o_1)$. Between £0 and this wealth level the strategy is simple; bet the Kelly fraction multiplied by the current wealth. After this point, there is a period where no more capital is invested. This is due to this added investment being regarded as 'over-betting'.

The range of wealth over which the total wager is unchanged is determined by the difference between the size of the two price options, o_1 and o_2. As o_2 becomes closer to o_1, the term $p + (1-p)o_1o_2^{-1}$ becomes closer to 1, and f_2^* from equation (6.4.9) tends towards $K(p, o_2) - \tilde{l}_1$.

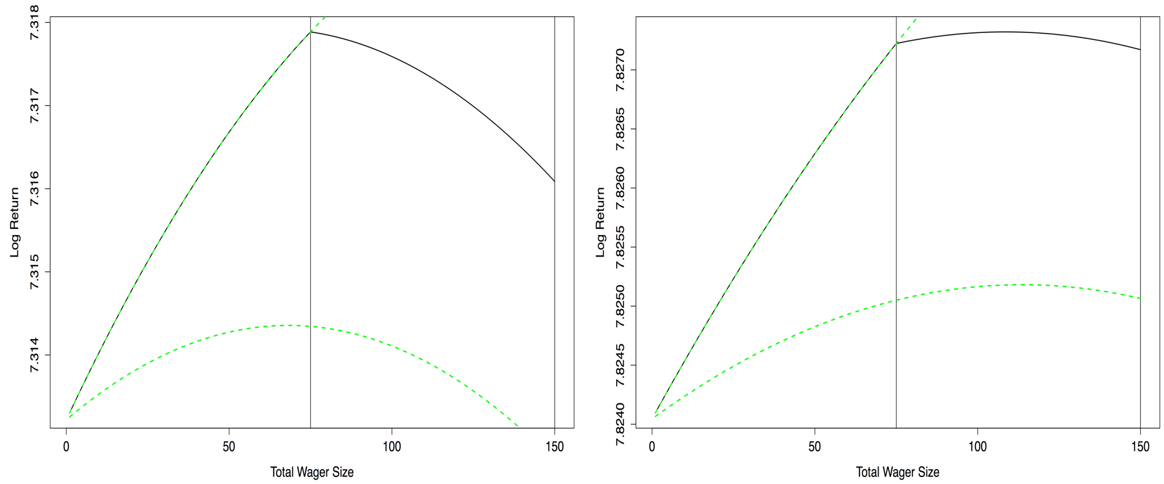


Figure 6.4.2: The log-return of wagers of various sizes, from a wealth of £1500 on the left and £2500 on the right. The green dashed lines represent the log-return of unconstrained wagers, given the two price options $o_1 = 1.25$ & $o_2 = 1.1$, with $p = 0.5$.

Figure 6.4.2 shows the log-return for an increasing investment size, given the same odds, probability and price limits as that seen in Figure 6.4.1. The graph on the left shows the situation where the first price option is fully utilised, and the second price option is left unused. What is important is the gradient of the unconstrained log-return for the second price option, at the price limit of the first option. On the left, where the second price option is not utilised the gradient is negative, whilst on the right, where the second price option is utilised, the gradient is positive. This indicates that there is additional growth rate available to the investor by betting into the second price option.

By considering the gradient of the lower wealth curve, and comparing it to $\omega_2^0(1)$, the wealth difference between them varies by a factor of $d = p + (1-p)o_1o_2 - 1$. An alternative way, then, to find the wealth level needed to invest into the second pot would be to calculate the Kelly wager of the second price option, given some wealth, and multiply is by this factor d.

6.5 Conclusions

This chapter has considered the problem of betting on multiple odds, given a single event. It has explored the reasons why this problem may be important to investors, and has considered the contributions of the literature to this point.

Given the recent observation of the exposure to skimming, a new betting approach, dubbed 'consolidated wagering' has been introduced, which has the property of avoiding the possibility of skimming, whilst also achieving more favourable growth rates than other candidate methods, often with a smaller wealth commitment.

The most prevalent occasion where string bets are encountered, that of limited sized markets in betting exchanges was tackled. A general solution was given to calculate which, if any, of the inferior markets should be utilised, and how much should be invested into each of the price options.

Finally, the implication of uncertainty regarding knowledge of the event probability was investigated, building on the work made on single wagers. This gives very risk-averse betting strategies, which nethertheless captures elements of the decision-making which are lost without this analysis.

Chapter 7

Conclusion

The work contained in this thesis pertains to three quite separate goals. Firstly, it is suggested that predictive models in sports, and elsewhere, can exhibit some form of systematic bias. The aim of this work was to describe how this bias manifests itself in a model's predictions, and more importantly, how to reverse engineer this bias' effect on the predictions to create methods to correct future predictions such that they are unbiased.

This work was derived under the assumption that the model's outputs were independent predictions of Bernoulli events, and successfully derived estimates of the nature of the underlying model bias, both when the bias was static and time-varying. This method was used to investigate the bias shown by bookmakers' odds, and also the change in bias for a simple predictive model for the outcome of football matches.

The recommended areas of exploration for further work are as follows:

- The model inference was only carried out whilst considering Bernoulli outcomes. It is envisaged that other types of outcomes will provide the bias-estimation methods with data that provides more information, and thus should result in more accurate estimates of the bias's form.

- The method's chosen time-varying structure for the model bias is shown to be inefficient for describing some real-life situations. It would therefore be of use to investigate other options for the form of the bias' time-varying structure, and the impact of these different structures of the success of inference.

The second section of work considered the problem of predicting the movement of certain pre-match betting market features as they evolve through time, which occurs in response to the release of information to the betting public's attention. The increase in the amount of money invested into markets is highlighted as an important variate for the prediction of the other market features (such as the betting odds), and thus the prediction of the rate of money entering into markets is considered.

The approach created in this work utilises two separate approaches; one which considered the problem from a top-down perspective, and matched the currently observed observations to a historic collection of similar markets, and uses the similarities to extrapolate the market size forwards, based on how similar markets have evolved historically. The complementary approach considers the problem from the ground up, and breaks the market movements into: those caused by information entering the market, which are detected using changepoint analysis; and other movements which represent a general increase in interest in the betting market, which causes the market size to increase linearly on the log scale.

It is shown that by combining both of these approaches, an accurate estimate for how the size of a market will increase over time is formed, which performs robustly, no matter how long before the event's commencement a prediction is formed. Given that the model created in this work was the first attempt in literature of its type, future work would be to attempt to improve the performance of the estimates, and to investigate the ability of this, and future methods to predict different forms of betting markets.

Finally, string bets were considered, in which an investor is forced to bet multiple times at different odds on the same event. The first scenario where string bets are encountered is where the prediction of some pre-match market is poor, and thus an investor may bet prematurely, at sub-optimal odds. This situation was shown to depart from the classic betting systems recommended by Kelly, and new systems of betting were created, which were shown to perform better than all other methods proposed in the literature.

The second scenario where string bets are considered is in betting exchanges, where bet sizes may be subject to some upper limit. In this case, the optimal betting strategy was derived, under Kelly's growth-rate maximisation approach. This strategy is found for a general set of utility functions, and is illustrated via a simple scenario where only two odds are considered.

Appendix A

Rival Bias-Estimation Techniques

Naïve Calculation

A simple approach simply attempts to reform equation (4.2.4), such that each element is directly observable. Given that $\hat{\boldsymbol{\theta}}$ is indirectly observed whenever new odds are seen, $\hat{\boldsymbol{\ell}}$ can be found via a simple transformation. Note that $\boldsymbol{\ell}$ is not observed, but can be estimated via observations of $\boldsymbol{y}$, under a number of assumptions, and the bias mean can be estimated crudely via:

$$\mu \approx \overline{\log\left(\frac{1-\hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}}\right)} - \log\left(\frac{1-\overline{\boldsymbol{y}}}{\overline{\boldsymbol{y}}}\right). \tag{A.0.1}$$

where $\overline{\boldsymbol{y}}$ signifies the arithmetic mean of the elements of $\boldsymbol{y}$, etc. This approximation will only produces unbiased estimates when the probabilities are distributed around 0.5; at this value $\log[(1-\theta_i)/\theta_i] \approx 0$ thus $\log[(1-\bar{\boldsymbol{y}})/\bar{\boldsymbol{y}}] \approx 0$, and

$$\begin{aligned}\mu &\approx \overline{\log\left(\frac{1-\hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}}\right)} - \overline{\log\left(\frac{1-\boldsymbol{\theta}}{\boldsymbol{\theta}}\right)} \\ &\overset{\bar{\boldsymbol{\theta}}\approx 0.5}{\approx} \overline{\log\left(\frac{1-\hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}}\right)}. \qquad \text{(A.0.2)}\end{aligned}$$

When the probabilities have a mean different to 0.5, then the second term from equation (A.0.1), $\log[(1-\bar{\boldsymbol{y}})/\bar{\boldsymbol{y}}]$ acts as a crude correction.

Stochastic Approximation

Stochastic Approximation is a method used to solve problems of the form $f(x) = \alpha$, where $f(x)$ is not directly observable, but is instead inferred via another function $g(x)$, where $\mathbb{E}g(x) = f(x)$. The method, introduced in Robbins and Monro (1951), uses an iterative process of the form $x_{n+1} - x_n = c_n[\alpha - g(x_n)]$, to approximate the solution in terms of x, given some feasible initial value x_0 and where c_n is some chosen discounting factor, with $c_n \to 0$ as $n \to \infty$. Under certain conditions related to the monotonicity and boundedness of f and g, as well as conditions on the speed of convergence of c_n to 0 this algorithm is guaranteed to converge to the true solution (given an asymptotically large amount of data). A common choice for the discounting factor is $c_n = c/n$, for some choice for $c > 0$.

This approach is highly sensitive to the choice of c and is thus tricky to implement successfully in practice. In order to mitigate this risk, Polyak and Juditsky (1992) altered the structure of the algorithm, such that

$$x_{n+1} - x_n = d_n[\alpha - g(x_n)] \tag{A.0.3}$$

as before (but with some new choice of discounting factor d_n). Inference is based not on x_n, but $\bar{x}_n$, the mean of the previous n estimated values. Here, d_n must conform a certain (but different) set of conditions; a common choice is $d_n = n^{-\beta}$ with $0 < \beta < 1$, with β being chosen based on previous experimentation.

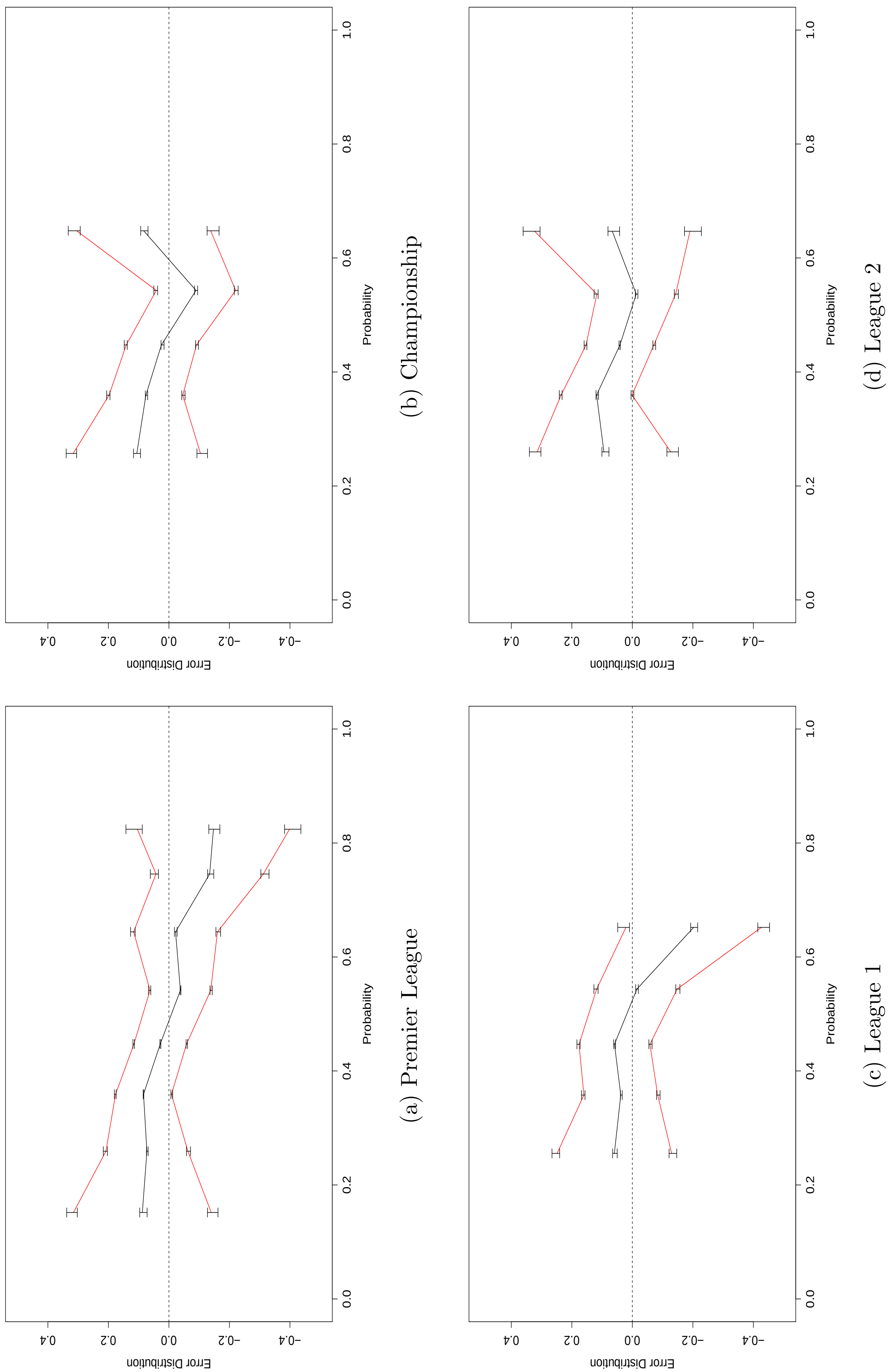
When this process is applied to the problem outlined in Section 4.2, α from equation (A.0.3) is y_t, the t'th observed Bernoulli outcome, and the iterative approach becomes:

$$\mu_t = \mu_{t-1} + t^{-\beta}\left[y_{t-1} - \frac{1}{1 + \exp(\ell_{t-1} - \mu_{t-1})}\right], \quad t \geq 2$$

and inference is taken from $\bar{\mu}_t = \frac{1}{T}\sum_{t=1}^{T} \mu_t$, as in Polyak and Juditsky (1992).

Appendix B

Comparing Modelling Bias



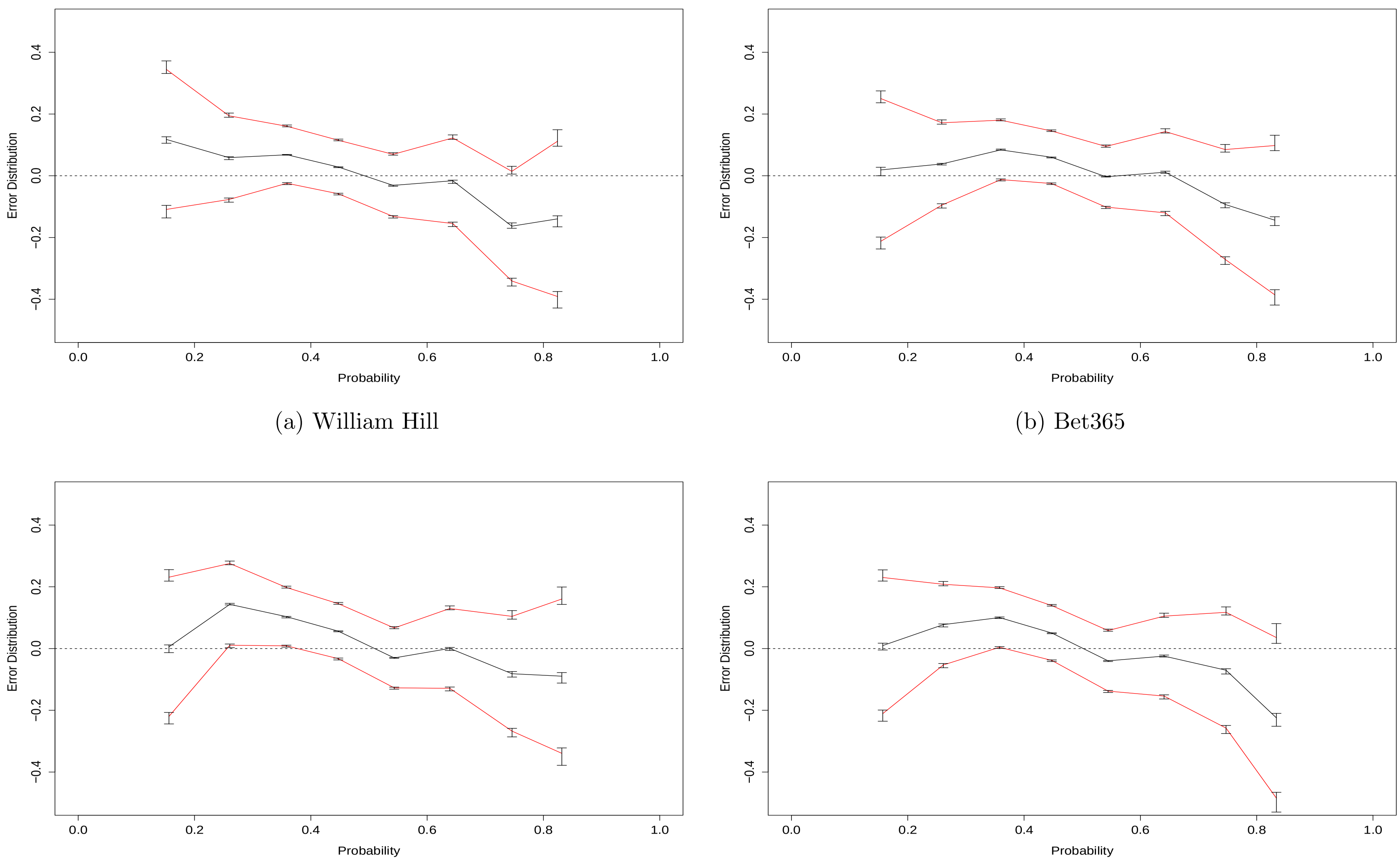
(a) Premier League

(b) Championship

(c) League 1

(d) League 2

Figure B.0.1: The modelling error in bookmaker's odds of home wins, broken down into subsets, dependent upon the estimated probability for each event, and compared against the leagues from which the odds are derived.



(a) William Hill

(b) Bet365

(c) Average Bookmaker's Odds

(d) Maximum Bookmaker's Odds

Figure B.0.2: The modelling error in bookmaker's odds of home wins, broken down into subsets, dependent upon the estimated probability for each event, and compared against different bookmakers, along with the average odds offered and maximum odds amongst all bookmakers.

Appendix C

Changepoint Detection

Changepoints can be defined as being "*points within a data set where the statistical properties change*" (Killick et al., 2012). More precisely, if there are m changepoints within a particular time series, with positions $\tau_{1:m} = (\tau_1, \cdots, \tau_m)$, then the series of observations $y_{1:n} = (y_1, \cdots, y_n)$ are segmented into $m+1$ parts where the statistical properties of the data within segments are unchanging. By convention, the set of changepoints are flanked at the start and end of the time series, by others, i.e. $\tau_0 = 0$ and $\tau_{m+1} = n$. The i'th such interval will contain the data $y_{(\tau_{i-1}+1):\tau_i}$.

A commonly used structure for forming objective functions, used for identifying the optimal set of changepoints is

$$\sum_{i=1}^{m+1} [\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i})] + \beta f(m). \tag{C.0.1}$$

The choice of cost function $\mathcal{C}(.)$, and penalty β to guard against the overfitting of changepoint locations is an important area of changepoint literature. This choice will

Algorithm 7 Pruned Exact Linear Time (PELT)

Input: Set of data of the form $\boldsymbol{y}_{1:n} = (y_1, \cdots, y_n)$.
A measure of fit $\mathcal{C}(.)$ dependent on the data.
A penalty β which does not depend on the number or location of the changepoints.
A constant K that satisfies: $\mathcal{C}(\boldsymbol{y}_{(t+1):s}) + \mathcal{C}(\boldsymbol{y}_{(s+1):T}) + K < \mathcal{C}(\boldsymbol{y}_{(t+1):T})$.

1: Let $n =$ the length of the data and set $F(0) = -\beta$, $cp(0) = 0$, $R_1 = \{0\}$.
2: **for** $\tau^* = 1, \cdots, n$ **do**
3: Calculate $F(\tau^*) = \min_{\tau \in \mathbb{R}_{\tau^*}} [F(\tau) + \mathcal{C}(\boldsymbol{y}_{(\tau+1):\tau^*}) + \beta]$;
Let $\tau' = \arg\min_{\tau \in \mathbb{R}_{\tau^*}} [F(\tau) + \mathcal{C}(\boldsymbol{y}_{(\tau+1):\tau^*}) + \beta]$;
Set $cp(\tau^*) = (cp(\tau'), \tau')$;
Set $R_{\tau^*+1} = \{\tau \in \{R_{\tau^*} \bigcup \{\tau^*\}\} : F(\tau^*) + \mathcal{C}(\boldsymbol{y}_{(\tau+1):\tau^*}) + K \leq F(\tau^*)\}$.
4: **end for**

Output: The changepoints recorded in $cp(n)$.

not be dwelt upon here, but the most common choice for cost function is twice the negative log likelihood (see Horváth (1993)), whilst the most common choice for the penalty is simply $\beta f(m) = \beta m$ (see Haynes et al. (2014) for a thorough treatment).

There are many techniques in the literature which find the optimal changepoint locations for a choice of cost function and penalty, such as *Segment Neighbourhood* (see Auger and Lawrence (1989)) and *Optimal Partitioning* (see Jackson et al. (2005)), however such methods are relatively slow, being at best $\mathcal{O}(n \log n)$. In comparison, *Pruned Exact Linear Time*, or 'PELT' (Killick et al., 2012), is shown to take only $\mathcal{O}(n)$ via the utilisation of pruning step before the minimisation of the objective is approached.

The process of PELT is shown in Algorithm 7. Here, the notation $F(\tau)$ simply represents the minimised form of equation (C.0.1) for the subset of the data $y_{1:\tau}$. In addition $cp(.)$ is a vector of detected changepoint locations.

The speed improvement to the changepoint detection process stems, in part to the fourth input requirement from Algorithm 7. The condition asserts that, given a subset of the data, $y_{(t+1):T}$, the overall cost decreases with the introduction of a changepoint somewhere within the sequence (which is shown to be true for almost all cost functions in Killick et al. (2012)). Given this being true, then if the condition

$$F(t) + \mathcal{C}(y_{(t+1):s}) + K \geq F(s)$$

holds, t can never be the optimal last changepoint before the end time T, effectively pruning the search-space for changepoints which may improve the objective function.

Bibliography

An examination of the empirical derivatives of the favourite-longshot bias in racetrack betting.

J Aczél. *On Applications and Theory of Functional Equations.* Elemente Mathematik vom Höheren Standpunkt aus. Birkhäuser, Basel, 1969.

PH Algoet and T Cover. Asymptotic Optimality and Asymptotic Equipartition Properties of Log-Optimum Investment. *The Annals of Probability*, pages 876–898, 1988.

MM Ali. Probability and utility estimates for racetrack bettors. *The Journal of Political Economy*, 85(4):803, 1977.

KJ Arrow. The Theory of Risk Aversion. *Essays in the Theory of Risk-Bearing*, pages 90–120, 1971.

P Asch, BG Malkiel, and RE Quandt. Market efficiency in racetrack betting. *Journal of Business*, 57(2):165, 1984.

E Asinof. *Eight men out: the Black Sox and the 1919 World Series.* Macmillan, 2011.

IE Auger and CE Lawrence. Algorithms for the Optimal Identification of Segment Neighborhoods. *Bulletin of Mathematical Biology*, 51(1):39–54, 1989.

RD Baker and IG McHale. Optimal Betting under Parameter Uncertainty: Improving the Kelly Criterion. *Decision Analysis*, 10(3):189–199, 2013.

V Barnett and S Hilditch. The Effect of an Artificial Pitch Surface on Home Team Performance in Football (Soccer). *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 39–50, 1993.

R Bell and TM Cover. Competitive optimality of logarithmic investment. *Mathematics of Operations Research*, 5(2):161–166, 1980.

R Bell and TM Cover. Game-Theoretic Optimal Portfolios. *Management Science*, 34 (6):724–733, 1988.

R Bellman and R. Kalaba. On Adaptive Control Processes. *Automatic Control, IRE Transactions on*, 4(2):1–9, Nov 1959. ISSN 0096-199X. doi: 10.1109/TAC.1959.1104847.

S Benninga. Financial modeling. *MIT Press Books*, 1, 2008.

D Bernoulli. Exposition of a New Theory on the Measurement of Risk. *Econometrica: Journal of the Econometric Society*, pages 23–36, 1954.

N Bernoulli. Correspondence of Nicolas Bernoulli concerning the St Petersburg game. *Petersburg Game. Letter to Pierre Raymond de Montmart*, 1713.

RH Boyko, AR Boyko, and MG Boyko. Referee Bias Contributes to Home Advantage in English Premiership Football. *Journal of Sports Sciences*, 25(11):1185–1194, 2007.

M Boyle. *The Art of Bookmaking*. High Stakes, 2006.

L Breiman. Investment Policies for Expanding Businesses Optimal in a Long-Run Sense. *Naval Research Logistics Quarterly*, 7(4):647–651, 1960.

L Breiman et al. Optimal Gambling Systems for Favorable Games. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961.

S Browne and W Whitt. Portfolio Choice and the Bayesian Kelly Criterion. *Advances in Applied Probability*, pages 1145–1176, 1996.

M Cain, D Law, and D Peel. The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy*, 47(1):25–36, 2000.

M Cain, D Law, and D Peel. The Favourite-Longshot Bias, Bookmaker Margins and Insider Trading in a Variety of Betting Markets. *Bulletin of Economic Research*, 55(3):263–273, 2003.

G Cardano. Liber de Ludo Aleae, 1565/1663. *Published in*, 1663.

MC Chalabi. UK's Gambling Habits: What's Really happening?,

2014. URL http://www.theguardian.com/news/datablog/2014/jan/08/uks-gambling-habits-whats-really-happening.

SJ Chapman. The Kelly Criterion for Spread Bets. *IMA Journal of Applied Mathematics*, 2006.

M Dixon and M Robinson. A Birth Process Model for Association Football Matches. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):523–538, 1998.

MJ Dixon and SG Coles. Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2):265–280, 1997.

D Edelman. On the Financial Value of Information. *Annals of Operations Research*, 100(1-4):123–132, 2000.

EF Fama. Efficient capital markets: A review of theory and empirical work*. *The journal of Finance*, 25(2):383–417, 1970.

TS Ferguson. Betting Systems which Minimize the Probability of Ruin. *Journal of the Society for Industrial & Applied Mathematics*, 13(3):795–818, 1965.

M Finkelstein and R Whitley. Optimal Strategies for Repeated Games. *Advances in Applied Probability*, pages 415–428, 1981.

AD Fitt, CJ Howls, and M Kabelka. Valuation of Soccer Spread Bets. *Journal of the Operational Research Society*, 57(8):975–985, 2006.

D Forrest and I McHale. Anyone for tennis (betting)? *The European Journal of Finance*, 13(8):751–768, 2007.

E Franck, E Verbeek, and S Nüesch. Prediction Accuracy of Different Market Structures Bookmakers versus a Betting Exchange. *International Journal of Forecasting*, 26(3):448 – 459, 2010. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2010.01.004. URL `http://www.sciencedirect.com/science/article/pii/S0169207010000105`. Sports Forecasting.

G Friesen and PA Weller. Quantifying Cognitive Biases in Analyst Earnings Forecasts. *Journal of Financial Markets*, 9(4):333–365, 2006.

JM Gandar, RA Zuber, and RP Lamb. The Home Field Advantage Revisited: A Search for the Bias in Other Sports Betting Markets. *Journal of Economics and Business*, 53(4):439–453, 2001.

WR Gilks and GO Roberts. Strategies for Improving MCMC. In *Markov Chain Monte Carlo in Practice*, pages 89–114. Springer, 1996.

J Goddard and I Asimakopoulos. Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23(1):51–66, 2004.

X Golay et al. A New Correlation-Based Fuzzy Logic Clustering Algorithm for FMRI. *Magnetic Resonance in Medicine*, 40(2):249–260, 1998.

A Grant and PW Buchen. A Comparison of Simultaneous Kelly Betting Strategies. *Journal of Gambling Business and Economics*, 6(2):1–28, 2012.

A Grant, D Johnstone, and OK Kwon. Optimal Betting Strategies for Simultaneous Games. *Decision Analysis*, 5(1):10–18, 2008.

J Haigh. The Kelly Criterion and Bet Comparisons in Spread Betting. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(4):531–539, 2000.

NH Hakansson. Optimal Investment and Consumption Strategies under Risk for a Class of Utility Functions. *Econometrica: Journal of the Econometric Society*, pages 587–607, 1970.

DB Hausch, WT Ziemba, and M Rubinstein. Efficiency of the market for racetrack betting. *Management science*, 27(12):1435–1452, 1981.

K Haynes, IA Eckley, and P Fearnhead. Efficient Penalty Search for Multiple Changepoint Problems. *arXiv preprint arXiv:1412.3617*, 2014.

E Höög. Modelling Prices of In-Play Football Betting. 2014. Student Paper.

L Horváth. The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations. *The Annals of Statistics*, pages 671–680, 1993.

B Jackson et al. An Algorithm for Optimal Partitioning of Data on an Interval. *Signal Processing Letters, IEEE*, 12(2):105–108, 2005.

DA Jackson. Index Betting on Sports. *The Statistician*, pages 309–315, 1994.

GJ Jiang, I Lo, and A Verdelhan. Information Shocks, Liquidity Shocks, Jumps, and Price Discovery: Evidence from the US Treasury Market. *Journal of Financial and Quantitative Analysis*, 46(02):527–551, 2011.

M Jun, R Knutti, and DW Nychka. Spatial Analysis to Quantify Numerical Model Bias and Dependence: How many Climate Models are there? *Journal of the American Statistical Association*, 103(483):934–947, 2008.

JL Kelly. A New Interpretation of Information Rate. *Information Theory, IRE Transactions on*, 2(3):185–189, 1956.

R Killick, P Fearnhead, and IA Eckley. Optimal Detection of Changepoints with a Linear Computational Cost. *Journal of the American Statistical Association*, 107 (500):1590–1598, 2012.

R Killick, IA Eckley, K Haynes, and P Fearnhead. Package changepoint. *R package version 0.4.-2011.-http://cran. rproject. org/web/packages/changepoint/index. html*, 2014.

T Kuypers. Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics*, 32(11):1353–1363, 2000.

HA Latane. Criteria for Choice among Risky Ventures. *The Journal of Political Economy*, pages 144–155, 1959.

PM Lee. *Bayesian Statistics: An Introduction*. John Wiley & Sons, 2012.

Y Lin, JP Bagrow, and D Lazer. More Voices Than Ever? Quantifying Media Bias in Networks. *ICWSM*, 1(arXiv: 1111.1227):1, 2011.

LC MacLean, WT Ziemba, and G Blazenko. Growth versus Security in Dynamic Investment Analysis. *Management Science*, 38(11):1562–1585, 1992.

LC MacLean, R Sanegre, Y Zhao, and WT Ziemba. Capital Growth with Security. *Journal of Economic Dynamics and Control*, 28(5):937–954, 2004.

LC MacLean, Y Zhao, and WT Ziemba. Optimal Capital Growth with Convex Loss Penalties. Technical report, Working Paper, Dalhousie University, 2009.

LC MacLean, EO Thorp, Y Zhao, and WT Ziemba. Medium Term Simulations of the Full Kelly and Fractional Kelly Investment Strategies. *The Kelly Capital Growth Investment Criterion: Theory and Practice*, 2010a.

LC Maclean, EO Thorp, and WT Ziemba. Long-Term Capital Growth: The Good and Bad Properties of the Kelly and Fractional Kelly Capital Growth Criteria. *Quantitative Finance*, 10(7):681–687, 2010b.

LC MacLean, EO Thorp, and WT Ziemba. *The Kelly Capital Growth Investment Criterion: Theory and Practice*, volume 3. world scientific, 2011.

MJ Maher. Modelling Association Football Scores. *Statistica Neerlandica*, 36(3): 109–118, 1982.

H Markowitz. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, 1952.

HM Markowitz. *Portfolio Selection: Efficient Diversification of Investments*, volume 16. Yale University Press, 1968.

DG Mayer and DG Butler. Statistical Validation. *Ecological Modelling*, 68(1):21–32, 1993.

M Medo, YM Pismak, and YC Zhang. Diversification and Limited Information in the Kelly Game. *Physica A: Statistical Mechanics and its Applications*, 387(24): 6151–6158, 2008.

CF Menezes and DL Hanson. On the Theory of Risk Aversion. *International Economic Review*, pages 481–487, 1970.

MJ Moroney. Facts from Figures, 1956.

M Müller. Dynamic Time Warping. In *Information Retrieval for Music and Motion*, pages 69–84. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-74047-6. doi: 10.1007/978-3-540-74048-3_4. URL `http://dx.doi.org/10.1007/978-3-540-74048-3_4`.

E Noon. Extending Kelly Staking Strategies to Peer-to-Peer Betting Exchanges. 2014.

E Noon, WJ Knottenbelt, and D Kuhn. Kelly's Fractional Staking Updated for Betting Exchanges. *IMA Journal of Management Mathematics*, 24(3):283–299, 2013.

N O'Connor. A short history of the betting exchange industry, 2015. URL `http://www.bettingmarket.com/refraichir010388.htm`.

R Pollard. Home Advantage in Soccer: A Retrospective Analysis. *Journal of Sports Sciences*, 4(3):237–248, 1986.

R Pollard. Home Advantage in Football: A Current Review of an Unsolved Puzzle. *The Open Sports Sciences Journal*, 1(1):12–14, 2008.

B Polyak and A Juditsky. Acceleration of Stochastic Approximation by Averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855, 1992.

W Poundstone. *Fortune's Formula: The untold story of the scientific betting system that beat the casinos and wall street*. Macmillan, 2010.

JW Pratt. Risk Aversion in the Small and in the Large. *Econometrica: Journal of the Econometric Society*, pages 122–136, 1964.

Cheeky Punter. What is the Maximum Possible Bet Amount?, 2014. URL `http://www.cheekypunter.com/faq/what-is-the-maximum-possible-bet-amount`.

C Reep and B Benjamin. Skill and Chance in Association Football. *Journal of the Royal Statistical Society. Series A (General)*, pages 581–585, 1968.

G Ridder, JS Cramer, and P Hopstaken. Down to Ten: Estimating the Effect of a Red Card in Soccer. *Journal of the American Statistical Association*, 89(427): 1124–1127, 1994.

H Robbins and S Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, pages 400–407, 1951.

PA Samuelson. St. Petersburg Paradoxes: Defanged, Dissected, and Historically Described. *Journal of Economic Literature*, pages 24–55, 1977.

PA Samuelson. Why we Should not Make Mean Log of Wealth Big though Years to Act are Long. *Journal of Banking & Finance*, 3(4):305–307, 1979.

Paul A Samuelson. Lifetime Portfolio Selection by Dynamic Stochastic Programming. *The Review of Economics and Statistics*, pages 239–246, 1969.

Paul A Samuelson and Robert C Merton. Generalized Mean-Variance Tradeoffs for Best Perturbation Corrections to Approximate Portfolio Decision. *The Journal of Finance*, 29(1):27–40, 1974.

DG Schwartz. The History of Gambling: Roll the Bones, 2006.

HS Shin. Optimal Betting Odds against Insider Traders. *The Economic Journal*, pages 1179–1185, 1991.

E Sinclair. Confidence Intervals for the Kelly Criterion. *Available at SSRN 2457368*, 2014.

MA Smith, D Paton, and LV Williams. Market Efficiency in Person-to-Person Betting. *Economica*, 73(292):673–689, 2006.

WW Snyder. Horse racing: Testing the efficient markets model. *The Journal of finance*, 33(4):1109–1118, 1978.

D Spiegelhalter and YL Ng. One Match to Go! *Significance*, 6(4):151–153, 2009.

RT Stefani. Observed Betting Tendencies and Suggested Betting Strategies for European Football Pools. *The Statistician*, pages 319–329, 1983.

EO Thorp. *Beat the Dealer: A Winning Strategy for the Game of Twenty-One*. Random House LLC, 1966.

EO Thorp. Optimal Gambling Systems for Favorable Games. *Revue de l'Institut International de Statistique*, pages 273–293, 1969.

EO Thorp. The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market. *Finding the Edge: Mathematical Analysis of Casino Games*, 1(6), 1998.

EO Thorp. Understanding the Kelly Criterion. *The Kelly Capital Growth Investment Criterion: Theory and Practice, World Scientific Press, Singapore*, 2010.

A Titman, D Costain, G Ridall, and K Gregory. Joint Modelling of Goals and Bookings in Association Football matches. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, pages 1–30, 2015.

T Tjur. A Warning Concerning Random Effects and Random Coefficients in Logistic Regression Models for Binary Data. 2003.

P Tsirimpas. Specification and Performance Optimisation of Real-time Trading Strategies for Betting Exchange Platforms. 2015.

L von Bortkiewicz. *Das Gesetz der kleinen Zahlen*. BG Teubner, 1898.

V Vovk and F Zhdanov. Prediction with Expert Advice for the Brier Game. *The Journal of Machine Learning Research*, 10:2445–2471, 2009.

H Wardle. *British gambling prevalence survey 2007*. The Stationery Office, 2007.

H Wardle. *British gambling prevalence survey 2010*. The Stationery Office, 2010.

H Wardle and A Moody. Gambling Industry Statistics 2009–2014, 2014. URL `http://gamblingcommission.org.uk`.

C Whitrow. Algorithms for Optimal Allocation of Bets on Many Simultaneous Events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(5):607–623, 2007.

LV Williams. Information Efficiency in Betting Markets: a Survey. *Bulletin of Economic Research*, 51(1), 1999.

LV Williams and D Paton. Why is There a Favourite-Longshot Bias in British Racetrack Betting Markets? *The Economic Journal*, pages 150–158, 1997.

LM Woodland and BM Woodland. Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market. *The Journal of Finance*, 49(1):269–279, 1994.

A Wykes. *The Complete Illustrated Guide to Gambling*. Doubleday Books, 1964.

E Zambrano. Subtle Price Discrimination and Surplus Extraction under Uncertainty. *Journal of Mathematical Economics*, 52:153–161, 2014.

WT Ziemba and DB Hausch. The Dr. Z betting system in England. *Efficiency of Racetrack Betting Markets*, 2008.