



UNIVERSITY OF
LEICESTER

UNIVERSITY OF LEICESTER

DOCTORAL THESIS

**Betting Markets:
Defining odds restrictions, exploring
market inefficiencies and measuring
bookmaker solvency**

Author:

Dominic Cortis

Supervisor:

Professor Jeremy Levesley

*A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy*

in the

Actuarial Science Research Group
Department of Mathematics

May 2016

Declaration of Authorship

I, Dominic CORTIS, declare that this thesis titled, ‘Betting Markets: Defining odds restrictions, exploring market inefficiencies and measuring bookmaker solvency’ and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

No time will be better

Family, exams, work, friendships, romance

Fiver on league win for Leicester city

Bond, Share, Derivative, Swap, annuity

for term or life, in arrears or advance

Poker, life, lottery, betting, finance

Certain, Likely, impossible, unlikely

Outcome and result unknown precisely

We all get a turn in the games of chance

Decisions to take: fast? for a lifetime?

Action forever we cannot adjourn

But most surely it cannot be a crime

To stop, ask, calculate, think, trust, test, doubt

Therefore we bite the time to take our turn

Yet seldom we win a chance to turn time.

UNIVERSITY OF LEICESTER

Abstract

College of Science and Engineering
Department of Mathematics

Doctor of Philosophy

**Betting Markets:
Defining odds restrictions, exploring market inefficiencies and measuring
bookmaker solvency**

by Dominic CORTIS

Betting markets have been of great interest to researchers as they represent a simpler set-up of financial markets. With an estimated Gross Gambling Revenue of 45bn yearly on betting on outcomes alone (excluding other gambling markets such as Casino, Poker and Lottery), these markets deserve attention on their own merit.

This thesis provides simple mathematical derivation of a number of key statements in setting odds. It estimates the expected bookmaker profit as a function of wagers placed and bookmaker margin. Moreover it shows that odds set by bookmakers should have implied probabilities that add up to at least one. Bookmakers do not require the exact probability of an outcome to have positive expected profits. They can increase profitability by having more accurate odds and offering more multiples/accumulators. Bookmakers can lower variation in payouts by maintaining the ratio of wagers and implied probability per outcome.

While bookmakers face significant regulatory pressures as well as increased taxes and levies, there is no standard industry model that can be applied to evaluate the minimum reserves for a bookmaker. Hence a bookmaker may be under/over-reserving funds required to conduct business. A solvency regime for bookmakers is presented in this work.

Furthermore a model is proposed to forecast soccer results – this focuses on goal differences in contrast to traditional models that predict outcome or goals scored per team.

Significant investigations are made on the inefficiencies of betting markets. The likelihood of Brazil reaching different stages of the 2014 World Cup, as perceived by odds, is compared to events on and outside the pitch to imply bias. An analysis of over 136,000 odds for European soccer matches shows evidence of the longshot bias. Finally this work investigates how it is possible to profit from market inefficiencies on betting exchanges during short tournaments by using a Monte Carlo simulation method as a quasi-arbitrage model.

Acknowledgements

I thank all those who helped me at the start, during or at the end of this journey by being behind me to push me, in front of me to show me the way, below me to hold me when I fell, above me to keep an eye out or beside me to encourage me.

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	v
Contents	vi
List of Figures	ix
List of Tables	x
1 Introduction	1
2 Expected values and variances in bookmaker payouts: A Theoretical approach towards setting limits on odds	5
2.1 Evaluating Odds	5
2.2 Bookmaker Margin	7
2.3 Outcomes on a Single Event	8
2.3.1 Restrictions on Setting Odds	8
2.3.2 Bookmaker Profit and Deviation	9
2.4 Multiples	12
2.5 Conclusion	15
3 Developing a solvency framework for bookmakers	16
3.1 Bookmaker's Portfolio	17
3.2 Bundling Technique	18
3.2.1 Variance per Bundle	18
3.2.2 Portfolio Variance	19
3.3 Determining Solvency	21
3.4 Challenges and Extensions	22
3.5 Conclusion	23
4 A public (mis)interpretation of Brazil's world cup performance	24
4.1 Introduction	24

4.2	Context: The World Cup Finals and Betting Markets	25
4.2.1	World Cup Finals	25
4.2.2	World Cup Betting Markets	26
4.3	Data and Trends	27
4.3.1	Data	27
4.3.2	Perception versus Performance	27
4.3.2.1	Group Stage – Croatia [3-1], Mexico [0-0], Cameroon [4-1]	27
4.3.2.2	Knock-out Stage – Chile [1-1], Colombia [2-1], Germany [1-7]	30
4.4	Discussion and Conclusion	30
5	Predicting goal difference	33
5.1	Introduction	33
5.2	Literature Review	34
5.3	Data and methodology	36
5.4	A four-factor linear model	38
5.5	Draws estimated using $m=0.25$	40
5.6	Conclusion and Discussion	42
6	Betting market efficiency in European soccer	44
6.1	Introduction	44
6.2	Literature Review	45
6.2.1	The Favourite-Longshot Bias	45
6.2.2	Exploring Other Biases	48
6.2.3	Testing for Insider Knowledge	49
6.2.4	Arbitrage Opportunities	50
6.3	Methodology	51
6.3.1	Data	51
6.3.2	Determining Probabilities and Other Metrics	52
6.3.3	Subdividing the ranges	52
6.3.4	Correlations	53
6.4	Analysis and Results	54
6.4.1	General	54
6.4.2	Conflicting results in Overs Market	55
6.5	Conclusion	57
7	Profiting on inefficiencies in betting derivative markets: The case of UEFA EURO 2012	61
7.1	Introduction	61
7.2	Betting Markets	63
7.3	Method	64
7.3.1	The Algorithm	64
7.3.2	Comparing Output with Market	66
7.4	Results	68
7.5	Conclusion	68
8	Conclusion	70
8.1	The Business of Bookmaking	70

8.2	Predicting Soccer outcomes	72
8.3	(In)efficient markets	72
8.4	Other Work	73
8.5	Implications on the Financial Industry	73
A	Payout Distribution for Constant Bookmaker Margin	75
B	Pricing risk through simulation: Revisiting Tilley bundling and least square Monte Carlo methods	76
B.1	Motivation	76
B.2	Introduction	76
B.2.1	Use of Monte Carlo methods in pricing derivatives	77
B.2.2	Risk profiles	78
B.2.3	European-style Option in the GBM one-factor model	78
B.3	The Two Monte Carlo methods	80
B.3.1	Steps Involved	80
B.3.2	A Numerical Example of Pricing: An American Call Option	81
B.3.2.1	LSM	82
B.3.2.2	Tilley Bundles	84
B.3.2.3	Comparison of the two systems	85
B.3.3	A Numerical Example of Risk Pricing: An American Put Option .	87
B.3.4	Comparing Results	89
B.4	Future Work	89
C	Data Adjustments	90
D	Comparison of Weibull Parameters with past Euro and African Cup of Nations Games	92
D.1	Introduction	92
D.2	Data	92
D.3	Data Treatment	93
D.4	Results	94
E	Sample Output for Group D	95
	References	97

List of Figures

1.1	Estimated Yearly Gross Gambling Revenues (H2 Gambling Capital, 2014)	3
2.1	European odds for two matches as shown on Betsonn.com as at 16:00 on 25th August 2014.	7
2.2	Bestseller Option on BWin as at 15:30 on 17th November 2014.	14
3.1	Odds available on Pinnacle as at 16:51 on 4th September 2014.	17
4.1	Market arranged by outcome per team.	26
4.2	Perceived Probabilities during match vs. Croatia	28
4.3	Perceived Probabilities between first and second match.	28
4.4	Perceived Probabilities during nine minutes of draw vs. Cameroon.	29
5.1	Frequency of model predictions compared to actual goal differences.	41
6.1	Relationship between p_s and p_o	54
8.1	Contingent bets for the Manchester Utd - Liverpool FC Europa League match shown on SkyBet as at 13:00 on 17th March 2016 (match was on same day). For example the bet on Anthony Martial (Manchester United player) and over 2.5 goals in a match at European odds of 4 is a bet contingent on both events which are related.	71
B.1	A sample generation of 20,000 stock price paths for a year with an initial price of 100, volatility of 20% and $r = 0.06$	79
B.2	Undiscounted EPE and PFE for a one year Euro Call Option	80
B.3	A 3D-pdf of a stock price generation for five years with an initial price of 100, volatility of 40% and $r = 0.06$	88
B.4	EPE and PFE for an American Call Option simulated using Tilley Bundling and LSM.	89

List of Tables

2.1	Multiple Wagers on Manchester City vs. Liverpool and Real Madrid vs. Cordoba	13
3.1	Wagers on a Portfolio of Bets with One Bookmaker	20
4.1	Probabilities at different time points	28
5.1	The average of the (%) correlations over the fifteen seasons	37
5.2	Regression Coefficients	38
5.3	Exact Prediction of Goal Difference	39
5.4	Results when using $m = 0.5$	39
5.5	Results when using $m = 0.25$	40
5.6	Cohen's κ	41
6.1	Number of Observed Odds	54
6.2	Bettor Losses for Unit Bets per Odds and Brier Scores	56
6.3	Correlation between Profits and Observed Proportions (p_o)	57
6.4	Correlation between Brier Scores and Profits	58
6.5	Correlation between Brier Scores and Observed Proportions (p_o)	59
6.6	Goals Scored per match	59
7.1	Number of Outcomes for Each Market	64
7.2	Used Funds and Profits Made (in thousands)	68
B.1	Generation of sixteen stock price paths at the end of each year with an initial price of 100, volatility of 30% and $r = 0.02$	82
B.2	LSM-method steps 2 to 4 at $t = 3$	83
B.3	LSM-method steps 2 to 4 at $t = 2$	83
B.4	LSM-method steps 2 to 4 at $t = 1$	84
B.5	Tilley Bundling method at $t = 3$	85
B.6	Tilley Bundling method at $t = 2$	86
B.7	Tilley Bundling method at $t = 1$	86
B.8	The value of an American Put stock option. The values in brackets show standard deviation.	87
D.1	Chi-Square Values for the EURO ¹	94
D.2	Chi-Square Values for the African Cup of Nations ²	94
D.3	Chi-Square Values for the World Cup ³	94
E.1	Calculation of for Group D using $\alpha = 0$	96

To my true supporter, Bigu

Chapter 1

Introduction

The concept of probability has been essential to academic achievement and industrial progress over the past four hundred years. Prior to its formalisation, outcomes were considered as an act of omnipotent being – a toss of a die is a five because god wills it. Probability and statistical techniques make modern science possible; both in measuring significance through hypothesis testing and in the belief that truth can be measured and explained. While statistics and probability are the forebearers of modern science, gambling was their muse.

Historically, probability and gambling have been closely linked ([Bernstein, 1996](#); [David, 1962](#)). Gerolamo Cardano (1501 - 1576) was an astrologer, physician, mathematician and gambler who, in an attempt to understand games of chance, first coined the classical definition of probability: given a total number n of equally possible outcomes of which m result in the occurrence of a given event, then the probability of that event occurring is $\frac{m}{n}$. However many recognise Chevalier de Méré's (1607 - 1684) questions on gambling games, specifically his continuous losses and the *problem of points* to Blaise Pascal (1623 - 1662) as the initial foundations to probability theory. Pascal famously discussed the *problem of points* in his correspondence with Pierre de Fermat and through their discussions these two mathematicians developed the initial stages for probability. Pascal was also instrumental in encouraging Christiann Huygens (1629 -1695) to write the first book on probability theory, *De Ratiociniis in Ludo Aleae* (On Reasoning in Games of Chance), in 1657.

Pierre de Laplace (1749 - 1827) introduced the application of probability theory to scientific and practical problems in the early nineteenth century. This spurred the developments of the theory of errors, Bayes' theory, statistics and actuarial mathematics during the same century. Subsequently Markov chains, which are extensively used in stochastic processes, were introduced in 1906 ([Markov, 1906, 1971](#)). Modern theory of probability

is based on Andrey Kolmogorov's (1903 - 1987) work on measure theory. Other significant theories during the last century include *Markowitz' theory of portfolio selection* ([Markowitz, 1952](#)), Monte-Carlo simulation, game theory and chaos theory.

During the past decades, research in probability theory has been mainly applied in financial and economic domains. Notwithstanding "games of chance" initiated this path and it may well be that gambling provides a setting for further advancements in probability theory.

Gambling can refer to a range of cases, ranging from lottery to poker. The term betting tends to refer to a subsection of gambling focused on betting on outcomes of events such as the results of sports or elections. Betting markets have been of great interest to economists ([Peel, 2008](#)) and other researchers mainly because they can be interpreted to be a simple alternative to financial markets ([Lessmann et al., 2009](#); [Sauer, 1998](#); [Shin, 1993](#); [Tirole, 1982](#)).

A crucial challenge to any reader in betting related research is the lack of consistency in the notation and the formalisation of the key concepts. Furthermore any derivations of expected profit tend to focus on discrete best-estimate values rather than providing a range of results. My first chapter attempts to address this by firstly providing an overview of how different types of odds are determined by their respective probabilities. This acts as an introduction to readers who are new to the area and leads to the derivations of the restrictions on, and profitability of, a bookmaker when setting odds including the occurrence of arbitrage. These two sections can be considered as an extension to [Levitt's \(2004\)](#) work on bookmakers' profitability when pricing events with binary outcomes to events with more than two. The same chapter extends these results to multiples, also referred to as accumulators, another type of bet which has become significantly popular. [Zafiris \(2014\)](#) gave special attention to this type of bet by showing the net expected profit from a bettor's perspective. This chapter evaluates the payout from a bookmaker's standpoint and extends it to a more general form. The findings presented here are mainly theoretical and include mathematical proofs but they are also backed up by simple examples throughout.

Another motivation for betting research is the actual size of the industry itself ([Levitt, 2004](#); [Peel, 2008](#)). In particular the gambling industry is a key contributor to the gross domestic product (GDP) of some jurisdictions. For instance it contributes to over 10% of the Maltese GDP ([Agius, 2014](#)). On a smaller relative scale, betting on outcomes, also referred to as sports betting, is a sub-category of gambling that contributed £2.3 billion (circa 0.16%) to the United Kingdom GDP in 2011 ([Deloitte, 2013](#)).

During 2014, the sports betting industry is estimated to have made gross winnings (GGR), evaluated as wagers less winning pay-outs, of around 45 billion euro ([H2 Gambling Capital, 2014](#)). Interactive betting, which includes internet and mobile betting, has been a key driver to growth together with deregulation to the increase in sports betting ([European Gaming and Betting Association, 2014](#); [Fawcett et al., 2014](#); [Griffiths, 2004](#)). There has been an increase of over 200% in GGR of interactive sports betting over the past ten years as shown in Figure 1.1. There are no signs of a change in the trend of a growing sports betting market as mobile and social gaming become more popular ([Torres & Goggin, 2014](#)) and governments are viewing the regulation of the industry as possible economic and taxation gain, rather than a vice to control ([Atkinson et al., 2000](#)).

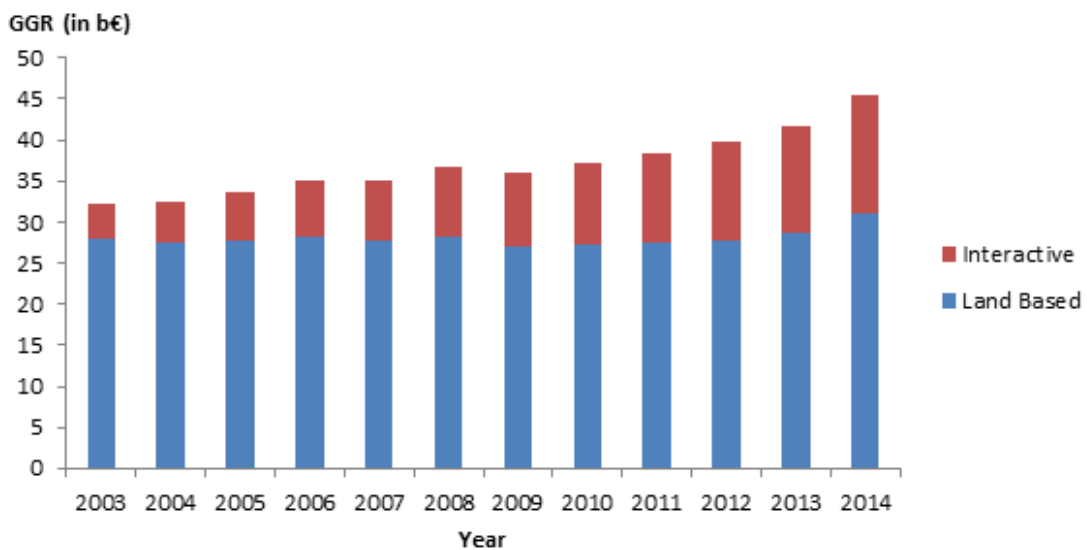


FIGURE 1.1: Estimated Yearly Gross Gambling Revenues ([H2 Gambling Capital, 2014](#))

Due to the size of the market, the solvency of bookmakers is an issue to be addressed with a similar rigour to that of financial institutions. Bookmakers are not immune to making losses (e.g. [iGaming Business, 2013, 2014](#)). In this respect, Chapter 3 introduces a method to evaluate risk metrics that can be extended to limit, grant or insure a bookmaker. This chapter also acts as a natural area where my actuarial training is applied directly.

Betting markets are usually used as a cross check of a predictive model. In essence, a predictive model is validated if it provides outcomes such that betting on them results in profitability. Hence a dissertation focused on betting markets needs to have a section focused on different types of models. Chapter 5 summarises the major developments in modelling soccer results and an alternative approach is proposed.

The speculative nature of betting markets has led to significant research such as their relationship with external forces ([Braun & Kvasnicka, 2008](#)), insider trading ([Schnytzer &](#)

Shilony, 1995; Shin, 1993; Vaughan Williams & Paton, 1997), herding behaviour (Law & Peel, 2002) and a comparison of group to individual decision making (Adams & Ferreira, 2010).

Specifically a great deal of research has focused on the market efficiency of betting markets (K. H. Brown & Abraham, 2002; Deschamps & Gergaud, 2012; Easton & Uylangco, n.d.; Gandar et al., 2004; Gil & Levitt, 2012; Hvattum, 2013; Paul & Weinbach, 2002; Terrell & Farmer, 1996; Woodland & Woodland, 2001). While techniques used within these market efficiency inquiries differ greatly, there is one key common question that is being asked: “Do odds on different outcomes represent the actual probabilities correctly?”.

Chapters 4, 6 and 7 tackle this question from different perspectives. Chapter 4 is a critical exploration on why the odds on Brazil’s prospects during the World Cup may have implied some form of irrationality. Chapter 4 is descriptive in nature and hence Chapter 6 elaborates on further by using a larger sample of bets to test inefficiencies in the European soccer betting market. Finally, in Chapter 7 a model is proposed to take advantage of the inconsistencies in betting derivatives market. This would be described as quasi-arbitrage between markets by Paton & Vaughan Williams (2005).

Chapter 2

Expected values and variances in bookmaker payouts: A Theoretical approach towards setting limits on odds

The rationale of this chapter is to fill a gap that formalises basic betting concepts, such as arbitrage, in an academic setting. While doing other research focused on market efficiency, I did not find an academic text that sets out the key underpinnings in setting odds. Furthermore most discussions focused on expectations rather than deviations. This chapter, most of which has been published ([Cortis, 2015a](#)), starts by defining different methods of evaluating odds and then proves six propositions related to the limitations on bookmakers when setting odds. It provides a prelude to a more concentrated effort at dealing with overall bets placed with a bookmaker as discussed in Chapter 3 and a basis towards the betting market research in the later chapters.

2.1 Evaluating Odds

The odds on an event represent the probability of different outcomes occurring and are used instead percentages as they display the return on a wager. There are three main styles of displaying odds: American, English or European odds and three other styles that are mainly popular in Asia: Hong Kong, Indonesian and Malaysian odds. The majority of online betting sites offer the facility for clients to display odds according to their preferred format between the three main styles. Throughout this dissertation, any reference to odds will be to European odds unless otherwise stated.

American odds, also known as Money Line odds, can be shown as positive or negative figures. A positive figure shows how much would be won on a wager of 100 units while negative odds show how much needs to be wagered in order to win 100 units. Positive figures are generally used for outcomes that are less than 50% likely to occur while negative figures are generally used for outcomes that are more than 50% likely to occur. English or Fractional odds show the amount that would be won excluding the wager as a ratio to the wager with both values being integers. They can be calculated as the ratio of the probability of an outcome not occurring to the probability of the outcome occurring. European odds are also called decimal odds and represent the odds of an outcome occurring as the inverse of its probability. Hong Kong odds show the winnings, net of the wager, made on a wager of one unit. Therefore it is always one less than the European odds. Indonesian odds are similar to American odds but relate to the amount won or to be wagered to win one unit. Finally Malaysian odds show the winnings, net of the wager, made on a wager of one in the case of likely outcomes and the amount to be wagered in order to win one unit, excluding own wager, in case of unlikely outcomes. In essence Hong Kong odds are equal to Indonesian and Malaysian odds for unlikely and likely outcomes respectively.

Mathematically the odds on outcome i with probability p_i to occur would be shown as below:

$$\text{American / Money Line Odds: } \begin{cases} +100 \frac{1-\pi_i}{\pi_i} : \pi_i \leq 0.5 \\ -100 \frac{\pi_i}{1-\pi_i} : \pi_i > 0.5 \end{cases} \quad (2.1a)$$

$$\text{English / Fractional Odds: } 1 - \pi_i / \pi_i \quad (2.1b)$$

$$\text{European / Decimal Odds: } \frac{1}{\pi_i} \quad (2.1c)$$

$$\text{Hong Kong Odds: } \frac{1 - \pi_i}{\pi_i} = \frac{1}{\pi_i} - 1 \quad (2.1d)$$

$$\text{Indonesian Odds: } \begin{cases} \frac{1-\pi_i}{\pi_i} : \pi_i \leq 0.5 \\ -\frac{\pi_i}{1-\pi_i} : \pi_i > 0.5 \end{cases} \quad (2.1e)$$

$$\text{Malaysian Odds: } \begin{cases} -\frac{\pi_i}{1-\pi_i} : \pi_i \leq 0.5 \\ \frac{1-\pi_i}{\pi_i} = \frac{1}{\pi_i} - 1 : \pi_i > 0.5 \end{cases} \quad (2.1f)$$

For example consider an outcome that has a 20% chance of occurring. The American, English, European, Hong Kong, Indonesian and Malaysian odds can be displayed as $100 \frac{1-0.2}{0.2} = +400$, $(1 - 0.2)/0.2 = 4/1$, $\frac{1}{0.2} = 5$, $\frac{0.8}{0.2} = 4$, $+4$ and $\frac{-0.2}{0.8} = -0.25$. This can be interpreted as a successful one unit wager on this outcome resulting in receiving five units back (European/Decimal), consisting of the own stake and a winning of four units

(English/Fractional, Hong Kong, Indonesian) resulting in a net stake of 25% of winnings (Malaysian) or a 400% profit on the stake (American/Money Line).

A negative money line such as -400 can be interpreted as needing to stake 400 units in order to win one hundred units. Similarly this would be -4 in Indonesian odds. The equivalent Fractional, Decimal and Hong Kong odds would be 1/4, 1.2, 0.2 respectively. This would represent an outcome that is 80% likely to occur.

2.2 Bookmaker Margin

The probabilities of all possible mutually exclusive outcomes within a probability space of an event add up to 100%. However bookmakers do not operate in a 100% scenario (Archontakis & Osborne, 2007; Cortis et al., 2013; Štrumbelj, 2014).

For example consider a probability space for an event with two equal outcomes. Under a fair market, the (European) odds would be equal to 2 for each outcome. However a bookmaker would price the odds a bit lower, say 1.9. This means that having received two wagers of one unit on each outcome, the bookmaker would pay out 1.9 units and make a profit of 0.1 units.

19:00	Real Madrid - Córdoba	1.05	14.50	30.00	live
20:00	Manchester City - Liverpool	1.86	3.75	4.05	live

FIGURE 2.1: European odds for two matches as shown on Betsonn.com as at 16:00 on 25th August 2014.

The inverse of European odds, being the implied probabilities, adds up to more than 100%, and the difference to 100% is generally described as the bookmaker margin (Cain et al., 2003; Peel & Thomas, 1992; Štrumbelj, 2014). Other similar terms include the over-round (Zafiris, 2014) and the vig (Peel & Thomas, 1992). A real-life example can be seen in Figure 2.1. Using the example of Manchester City – Liverpool, the state space consists of a home team (Manchester City) win, draw or away team win. The sum of the inverse of probabilities are equal to $1.86^{-1} + 3.75^{-1} + 4.05^{-1} = 1.05121\%$. The bookmaker margin in this case is 5.121%.

The fact that implied probabilities from a bookmaker do not necessary add up to one leads to two key concerns: is it necessary for implied probabilities to add up to more than 100%? Moreover how does this difference between real and implied probabilities affect bookmaker profitability?

2.3 Outcomes on a Single Event

As an initial examination, this section deals with setting odds for one event. Prior to examining the bookmaker's profit, the restrictions on setting odds are proved.

Assumption 2.1. All odds are provided by one bookmaker.

Assumption 2.2. All probabilities, implied or actual, relate to one event with a probability space of n mutually exclusive outcomes.

2.3.1 Restrictions on Setting Odds

Definition 2.1. Let π_i represent the implied probability, calculated as the inverse of the European Odds, of outcome i occurring where $i \in \{1, \dots, n\}$.

Definition 2.2. Let p_i represent the actual probability of outcome i occurring where $i \in \{1, \dots, n\}$.

Definition 2.3. Let w_i be the total value in units of wagers placed on outcome i .

Assumption 2.3. The bookmaker is aware of the real value of p_i and can control π_i but not $w_i \forall i \in \{1, \dots, n\}$.

Definition 2.4. Let P be the random variable representing the profit made by the bookmaker for bets made on one event.

$$\therefore E(P) = E(\text{Wagers}) - E(\text{Payouts}) = \sum_{i=1}^n w_i - \sum_{i=1}^n \frac{w_i p_i}{\pi_i} \quad (2.2)$$

Proposition 2.1. If the implied odds of an event add up to less than one, then arbitrage exists.

Proof. (By Contradiction) We need to find a strategy for a bettor that would result in a certain loss for the company [$P < 0$]. We know that $\sum_{i=1}^n \pi_i < 1$. Assume a strategy such that $w_i = \pi_i$. Therefore for any outcome the payout by the company to the bettor is $w_i \pi_i^{-1} = 1$. $\therefore P = \text{Wagers} - \text{Payouts} = \sum_{i=1}^n w_i - 1 = \sum_{i=1}^n \pi_i - 1 < 0$. $\square$

Proposition 2.2. A bookmaker would risk a negative expected profit if the implied probability of any outcome is lower than the actual probability.

Proof. (By Contradiction) Let $p_q > \pi_q$ for some $q \in \{1, \dots, n\}$. Assume a single wager of 1 on outcome q . $\therefore E(P) = E(\text{Wagers}) - E(\text{Payouts}) = 1 - \frac{p_q}{\pi_q} < 0$. $\square$

Each of these propositions lead to two important restrictions for bookmakers when setting odds for each outcome of an event, these must add up to at least one and should not be smaller than the actual probability. Therefore the relationship shown in equation (2.3) can be set.

$$\pi_i = p_i(1 + k_i) \text{ where } k_i > 0; i \in \{1, \dots, n\} \quad (2.3)$$

2.3.2 Bookmaker Profit and Deviation

It is very common for k to be treated as a constant, where the same ratio of implied to actual probability is maintained for all outcomes (e.g. Archontakis & Osborne, 2007; Cortis et al., 2013; Goddard & Asimakopoulos, 2004; Zafiris, 2014). Contrastingly Shin (1993) demonstrates how bookmakers can evaluate odds in order to minimize the effects of insider training and Štrumbelj (2014) further proves that using Shin's technique on implied odds resulted in a better forecasting accuracy of real probabilities of fixed odds over a series of 48,126 matches in five sports.

Koch & Shing (2008) explain that bookmakers are limited to a finite list of odds, especially if they operate in an English/Fractional odds environment since odds are either rounded up or possibly limited by regulation. They report finding 88 odds levels in UK horse racing betting markets between 2001 and 2006. A similar limitation, albeit to a lower extent, can be found for European/Decimal odds. For example, it will be impossible for an implied probability to be 75% since this would result in recurring decimal odds (1.3̄). An examination of 9,120 European odds provided by eight bookmakers¹ for the starting odds of the 380 matches² of all 2013/14 English Premiership results in only 588 unique odds. This may mean that bookmakers, given a particular bookmaker margin, are likely to round actual probabilities to implied probabilities to the nearest value in a grid or use particular set doubles and triples of odds, rather than adjusting solely for insider trading. For example, consider an event with two mutually-exclusive outcomes and a bookmaker margin of around 5%. Some likely doubles, representing the European/Decimal odds for the two outcomes, could be 1.8 – 2, 1.6 – 2.25 and 1.55 – 2.5.

The bookmaker's profit is dependent on the amount and spread of wagers, which are dependent on consumer preference, and the bookmaker margin, which is set internally. Although bookmakers face challenges in the deviation of possible profits, this has been

¹Bet365, BWin, Interwetten, Ladbrokes, Pinnacle Sports, Stan James, VC Bet, and William Hill.

²Each match has three possible outcomes: home team win, away team win or draw. Odds ranged from 1.1 to 29.

generally disregarded in academic work. This sub-section initially evaluates the distribution of profit in the case of a bookmaker margin that is variable for different outcomes and then simplifies it to a constant bookmaker margin applied to all outcomes.

Proposition 2.3. Equation (2.3) $\implies$

$$P \sim \mathcal{N}\left(\sum_{i=1}^n \frac{k_i w_i}{1+k_i}, \sum_{i=1}^n \frac{(k_i-1)w_i}{1+k_i} + \sum_{i=1}^n \frac{w_i^2}{\pi_i(1+k_i)} - \left(\sum_{i=1}^n \frac{k_i w_i}{1+k_i}\right)^2\right)$$

Proof.

$$\begin{aligned} E(P) &= E(Wagers) - E(Payouts) \\ &= \sum_{i=1}^n w_i - \sum_{i=1}^n \frac{w_i p_i}{\pi_i} \\ &= \sum_{i=1}^n w_i - \sum_{i=1}^n \frac{w_i p_i}{p_i(1+k_i)} = \sum_{i=1}^n \frac{k_i w_i}{1+k_i} \end{aligned}$$

$$\begin{aligned} E(P^2) &= E(Wagers^2) - 2E(Wagers)E(Payouts) - E(Payouts^2) \\ &= \left(\sum_{i=1}^n w_i\right)^2 - 2\sum_{i=1}^n w_i \sum_{i=1}^n \frac{w_i p_i}{\pi_i} + \sum_{i=1}^n \frac{w_i^2 p_i}{\pi_i^2} \\ &= \left(\sum_{i=1}^n w_i\right)^2 - 2\sum_{i=1}^n w_i \sum_{i=1}^n \frac{w_i}{1+k_i} + \sum_{i=1}^n \frac{w_i^2 p_i}{p_i^2(1+k_i)^2} \\ &= \left(\sum_{i=1}^n w_i\right)^2 - 2\sum_{i=1}^n w_i \sum_{i=1}^n \frac{w_i}{1+k_i} + \sum_{i=1}^n \frac{w_i^2}{\pi_i(1+k_i)} \\ &= \sum_{i=1}^n w_i \sum_{i=1}^n \frac{w_i(k_i-1)}{1+k_i} + \sum_{i=1}^n \frac{w_i^2}{\pi_i(1+k_i)} \end{aligned}$$

$$\begin{aligned} Var(P) &= E(P^2) - [E(P)]^2 \\ &= \sum_{i=1}^n w_i \sum_{i=1}^n \frac{(k_i-1)w_i}{1+k_i} + \sum_{i=1}^n \frac{w_i^2}{\pi_i(1+k_i)} - \left(\sum_{i=1}^n \frac{k_i w_i}{1+k_i}\right)^2 \end{aligned}$$

□

Proposition 2.4. If the implied probability of any outcome is lower or equal to the actual probability, then the bookmaker is expected to have a profit.

Proof. $p_q \leq \pi_q \implies \pi_i = p_i(1+k_i) \forall k_i > 0. \therefore E(P) = \sum_{i=1}^n \frac{k_i w_i}{1+k_i} \geq 0$

□

Proposition 2.4 leads to an extension over Proposition 2.2 since it shows that expected bookmaker profits are positive if implied probabilities are larger than actual probabilities. A special case, discussed earlier, is when k is constant. In this case Proposition 2.3 can

easily be extended as shown in Equation (2.4) (a proof from first principles is given in Appendix A).

$$P \sim \mathcal{N} \left(\frac{k}{1+k} \sum_{i=1}^n w_i, \sum_{i=1}^n \frac{w_i^2}{\pi_i (1+k)} - \left(\frac{\sum_{i=1}^n w_i}{1+k} \right)^2 \right) \quad (2.4)$$

The evaluation of the distribution for a constant bookmaker margin leads to three observations. Firstly Proposition 2.4 shows a relaxation to Assumption 2.3 in that the bookmaker does not need to know the exact value of the actual probability for each outcome but only make sure that the implied probability is higher than the actual probability.

Another important outcome is that the percentage expected profit for a betting company is $\frac{k}{1+k}$. For example, considering again an event that can result in one of two equally likely outcomes but is sold at odds of 1.8, the bookmaker margin $k = 1.8^{-1} + 1.8^{-1} - 1 = \frac{10}{9} - 1 = \frac{1}{9}$. We would expect equal amounts to be wagered on each outcome since they are equally likely. Assuming a wager of one unit on each outcome, the wagers add up to two while the payout is 1.8. This is a profit of 0.2 from two units which is 10%. This is equivalent to $\frac{k}{1+k} = \frac{1/9}{10/9} = \frac{1}{10}$.

In the English Premiership match example shown in Figure 2.1, the bookmaker margin was 5.121% and therefore the expected profit is 4.872%. This simple calculation is important to point out as at times the bookmaker margin may be confused with expected profits.

Proposition 2.5. Certain profit is made if wagers on each outcome are distributed in proportion to the probabilities of each outcome.

Proof. Assume $k > 0$, $\sum_{i=1}^n w_i > 0$, $w_i = b_i \sum_{i=1}^n w_i$ and $\sum_{i=1}^n b_i = 1$.

The risk is minimised if there are no fluctuations, therefore there is no variance.

$$\begin{aligned}
 Var(P) &= 0 \\
 \frac{1}{(1+k)^2} \left(\sum_{i=1}^n \frac{w_i^2}{p_i} - \left(\sum_{i=1}^n w_i \right)^2 \right) &= 0 \\
 \sum_{i=1}^n \frac{(b_i \sum_{i=1}^n w_i)^2}{p_i} - \left(\sum_{i=1}^n w_i \right)^2 &= 0 \\
 \left(\sum_{i=1}^n w_i \right)^2 \left(\sum_{i=1}^n \frac{b_i^2}{p_i} - 1 \right) &= 0 \\
 \sum_{i=1}^n \frac{b_i^2}{p_i} - \sum_{i=1}^n p_i &= 0 \\
 \sum_{i=1}^n \frac{b_i^2 - p_i^2}{p_i} &= 0
 \end{aligned}$$

One of the solutions for the above is $b_i = p_i$. □

Proposition 2.5 is equivalent to the first scenario described by Levitt (2004) in which bookmakers sustain profits by maintaining the proportion of money bet on each side of an equally likely binary outcome. Consider that in the Manchester City – Liverpool example, if a total of 51, 25 and 23 units are wagered on the home team winning, a draw or the away team winning respectively; the company would have received 99 units in total wagers but would pay a maximum of 94.86 units, resulting in certain profit.

Bookmakers may be tempted to adjust odds according to the volume of wagers made on different outcomes (Spoorendonk & Kristiansen, 2015). One can point out that there are two rationales for this. The first is that the general public may have better information and are therefore better at estimating actual probabilities. Secondly the bookmaker may wish to diminish the likelihood of creating arbitrage when its odds are compared to others. However this proposition provides a third important justification for bookmakers to adjust odds according to the volume of wagers made; it guarantees profit.

2.4 Multiples

A Multiple is a bet with a pay-out contingent on two or more independent outcomes. For example, as shown in Figure 2.1, a bettor can place a multiple wager on Real Madrid and Manchester City winning their respective matches for odds of $1.05 \times 1.86 = 1.953$.

This bet is paid off only if both Real Madrid and Manchester City win their respective matches.

Definition 2.5. Let $P^{(m)}$ be the random variable representing the profit made by the bookmaker on bets over a multiple with m events.

In essence a multiple of m events with n_m outcomes per event, can be represented as one outcome with $\prod_{j=1}^m n_m$ outcomes and a margin of $(1 + k_m)^m - 1$ where k_m is the geometric mean of the bookmaker margin for each event. This means that it is easy to prove that

$$P^{(m)} \sim \mathcal{N}\left(\frac{(1 + k_m)^m - 1}{(1 + k_m)^m} \sum W, \frac{1}{(1 + k_m)^m} \sum \frac{W^2}{\pi^{(m)}} - \left(\frac{\sum W}{(1 + k_m)^m}\right)^2\right) \quad (2.5)$$

$$\text{where } \sum W = \sum_{i_1=1}^{n_1} \left(\sum_{i_2=1}^{n_2} \left(\dots \left(\sum_{i_m=1}^{n_m} w_{1_{i_1} 2_{i_2} \dots m_{i_m}} \right) \right) \right)$$

$$\sum \frac{W^2}{\pi^{(m)}} = \sum_{i_1=1}^{n_1} \left(\sum_{i_2=1}^{n_2} \left(\dots \left(\sum_{i_m=1}^{n_m} \frac{w_{1_{i_1} 2_{i_2} \dots m_{i_m}}^2}{\pi_{1_{i_1}} \pi_{2_{i_2}} \dots \pi_{m_{i_m}}} \right) \right) \right)$$

and $w_{1_{i_1} 2_{i_2} \dots m_{i_m}}$ represents the total amount wagered on a multiple bet that pays out if all outcomes i_1, i_2 to i_m within independent events $1, 2, 3, \dots, m$ respectively occur.

Consider multiple wagers made on the two matches shown in Figure 2.1. Each match has three outcomes, and therefore in total there are nine outcomes to the combination of the two matches. As an example consider a total of 100 units wagered over these nine outcomes as shown Table 2.1. The distribution of company profits can be calculated to follow $P^{(2)} \sim \mathcal{N}(9.8039, 20449.56)$.

TABLE 2.1: Multiple Wagers on Manchester City vs. Liverpool and Real Madrid vs. Cordoba

$w_{1_{i_1} 2_{i_2}}$	M. City Win	Draw	Liverpool
Real M. win	20	10	20
Draw	10	5	10
Cordoba win	10	10	5

Proposition 2.6. The profit on multiples is greater than that on singles and a bookmaker is likely to make more profit if more wagers on multiples are made.

Proof. (By Induction) We need to prove that $E(P^{(x)}) > E(P^{(y)}) \forall x > y \in \mathbb{N}^+$.

Let $\sum W$ show the total amount of wagers made.

Let $m = 1$:

$$E(P^{(2)}) = \frac{(1+k)^2-1}{(1+k)^2} \sum W = \frac{k(2+k)}{1+k} \sum W > \frac{k}{1+k} \sum W = E(P^{(1)})$$

Let $m \geq 2$:

$$E(P^{(m+1)} - P^{(m)}) = \frac{(1+k)^{m+1}-1}{(1+k)^{m+1}} \sum W - \frac{(1+k)^m-1}{(1+k)^m} \sum W = \frac{k(2+k)}{1+k} \sum W > 0$$

□

Multiple bets are highly promoted by bookmakers. A case in point is BWin. This bookmaker promotes ‘Best Seller’, being a bet consisting the five most popular multiples, within their main online sports betting page (Figure 2.2). The allure of multiples is for both the supply and demand end of the market. Bettors are more easily enticed with higher odds (Cain et al., 2003; Direr, 2013) while Proposition 2.6 substantiates Zafiris’ (2014) suggestion that multiples result in higher expected bookmaker profits per bet placed.

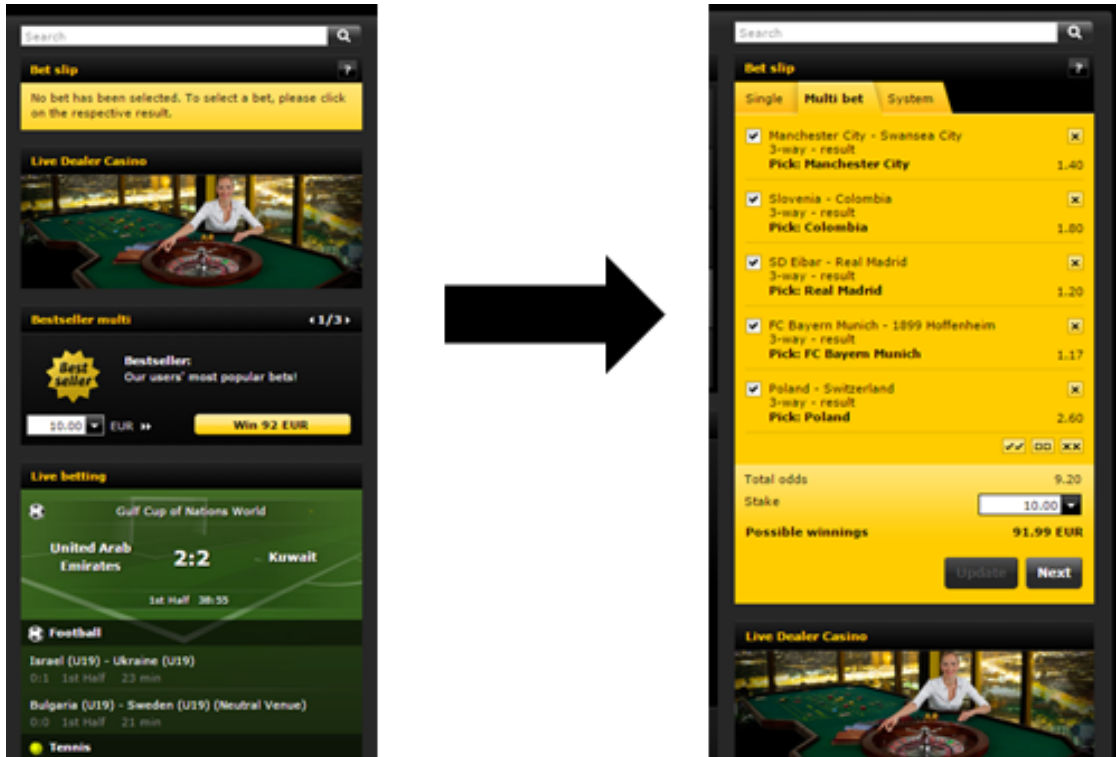


FIGURE 2.2: Bestseller Option on BWin as at 15:30 on 17th November 2014.

An illustration of the higher profits is a multiple bet on the outcomes of two coins (A & B). Consider a bookmaker offering odds of 1.8 on each outcome for each coin. Then the multiple odds would be $1.8^2 = 3.24$ on each of the four multiple outcomes³. If a total of four units are wagered on the single event of coin A (that is two units on each outcome), the expected pay-out is $1.8 \times 2 = 3.6$ units. The expected pay-out is lower at 3.24 units if four units are wagered over the multiple with two events and four multiple outcomes (that is one unit on each multiple outcome).

³Both Heads, Both Tails, Coin A Head and Coin B Tails, Coin A Tail and Coin B Heads.

2.5 Conclusion

Most investigations on odds offered in betting markets have concentrated on the efficiency displayed by these odds, usually from the perspective of the bettor. This chapter has provided some simple, but essential, proofs relating to the distribution of expected profit from a series of bets placed with a bookmaker.

Results point out two clear restrictions on the implied probabilities of market odds offered by bookmakers: the implied probabilities should be greater than the actual probabilities and should add up to at least 100%. The commercial implications for a bookmaker are also apparent as there are larger margins of profit on multiples and profit is guaranteed if wagers are kept in ratio with implied probabilities.

Chapter 3

Developing a solvency framework for bookmakers

When individuals place their savings in a bank or purchase an insurance product, they are safe in the knowledge that they are protected through a regulatory framework. Essentially, banks and insurance companies face significant regulation, especially with respect to their required reserves. However banks reserve a proportion of their depositor's funds as they assume that not all depositors will withdraw their investments at the same instant, and insurances do not reserve the sum insured of all policies since not all extreme adverse events are likely to occur at the same time.

Similarly an individual placing a bet should be safe in the knowledge that the bookmaker can pay its dues without requiring the bookmaker to maintain reserves necessary to pay out all bets. Regulation is disparate and, unlike financial services, the possibility of a bookmaker located in one country offering odds in another is restricted. While bookmakers also face significant regulatory pressures as well as increased taxes and levies, there is no standard industry model that can be applied to evaluate the minimum reserves for a bookmaker. This results in either the bookmaker not having significant funds to cover for adverse bets or the bookmaker purchasing additional protection for contingency purposes, the latter leading to inefficient use of capital.

This chapter presents an alternative to previous betting related research by focusing on the bookmaker solvency rather than the betting market itself or the clients. Variations of this chapter were presented at The 2015 MathSport Conference ([Cortis, 2015d](#)) and the 50th Actuarial Research Conference ([Cortis, 2015c](#)) hosted by the University of Toronto

where it [achieved an honorable mention](#)¹. It starts off by extending the work from Chapter 2 from mutually exclusive outcomes to a variety of mutually exclusive, independent and conditional outcomes. This is then followed by the technique used, showing an example and a discussion on how this technique can be used to measure solvency. Finally extensions, challenges and overall conclusions are provided.

3.1 Bookmaker's Portfolio

Bookmakers set odds on outcomes of different events on which customers can place wagers on. Typically these are set in different markets since there are various outcomes from one event. For example the 1X2 market in a football match (the event) might result into three outcomes (home team win, draw, away team win). The same event could be used to determine the half time 1X2 market. Furthermore odds can be displayed in a variety of manners as discussed in Chapter 2.

Rotation Number:

Lines effective:

9/4/2014 08:49:01 AM

Betting Menu

FIBA WORLD CUP - MEN BASKETBALL - THU 9/4

GAME		HANDICAP	MONEY LINE	TOTAL POINTS
Thu 9/4	2213 Iran	+12 1.990	6.880	Over 143.5 1.952
09:00 AM	2214 France	-12 1.917	1.137	Under 143.5 1.952
Thu 9/4	2215 Croatia	-7 1.943	1.309	Over 159.5 1.952
09:00 AM	2216 Puerto Rico	+7 1.961	3.830	Under 159.5 1.952
Thu 9/4	2217 Lithuania	+1 1.980	2.030	Over 153 1.934
12:30 PM	2218 Slovenia	-1 1.925	1.877	Under 153 1.970
Thu 9/4	2219 Turkey	-4 1.980	1.595	Over 139 1.961
12:30 PM	2220 Dominican Republic	+4 1.925	2.520	Under 139 1.943
Thu 9/4	2221 Spain	-16.5 1.917		Over 152 1.943
01:00 PM	2222 Serbia	+16.5 1.990		Under 152 1.961
Thu 9/4	2223 Argentina	+5.5 1.970	3.160	Over 154.5 1.961
01:00 PM	2224 Greece	-5.5 1.934	1.413	Under 154.5 1.943

FIGURE 3.1: Odds available on Pinnacle as at 16:51 on 4th September 2014.

In calculating the variance of company profits for multiples placed over the same set of m events, multiple outcomes are mutually exclusive and complete the probability space of all outcomes. However a company typically offers a significantly large number of events to place wagers on, such that little to no multiples refer to the same events. Consider the multiple/accumulator bets made on the International Basketball Federation (FIBA) World Cup as shown in Figure 3.1.

Bet α : Iran to win @ 6.880 & Croatia to win @ 1.309 for total odds of 9.006.

¹As a previous prize winner for the best presentation in 2013 for the work described in Appendix B, I was not able to compete for best presentation.

Bet β : Puerto Rico (3.830) & Argentina (3.160) to win for total odds 12.1028.

Bet γ : Argentina (3.160) & Turkey (1.595) to win for total odds 5.0402.

Bet δ : Lithuania to win @2.03.

In this case bets α and β are mutually exclusive since Croatia are playing Puerto Rico while bets β and γ have a positive correlation as they both depend on Argentina winning. Bet δ is independent of other bets. Assuming that these bets were made in the given order, bet β reduces the total variance in company profits since it acts as a hedge to bet α . On the other hand bet γ increases risk by a larger factor than bet δ .

The portfolio of bets placed with a bookmaker is constantly changing, especially with the popularity of live in-play betting (Hogg, 2013; Keogh & Rose, 2013) which is the possibility of placing wagers of outcomes while a match is being played (A. Brown, 2012; Croxson & James Reade, 2014). Therefore a simple estimate is required.

3.2 Bundling Technique

The proposition given here is to subdivide the portfolio of bets wagered with a bookmaker in bundles according to odds, evaluate the distribution of each bundle separately and then calculate the distribution of the portfolio. These bundles can be set according to the odds grids set up by the bookmaker but the odds offered to customers will increase significantly if multiples are used. Therefore a range of odds within each bundle would be preferable.

3.2.1 Variance per Bundle

For each bundle q , evaluate the distribution of wagers $W \sim \mathcal{N}(\overline{W}_q, \sigma_{w_q}^2)$; the number of different wagers², denoted n_q ; the geometric mean of the market spread k_q (weighted by n_q) over these bets; and the arithmetic mean of implied probabilities π_q . By simplifying Equation (2.4), the average variance per bet, denoted $\overline{Var}_q$, can be therefore estimated as shown in Equation (3.1) leading to a total bundle variance Var_q as per Equation (3.2)

$$\overline{Var}_q = \overline{W}_q^2 \left(\frac{1}{(1 + k_q) \pi_q} - \frac{1}{(1 + k_q)^2} \right) \quad (3.1)$$

²Bets on the same set of outcomes are to be counted as one wager. For example if two people bet different amounts on the same outcome, this is to be treated as one wager.

$$Var_q = n\overline{Var}_q + 2 \sum \left(r\sqrt{\overline{Var}_q}\sqrt{\overline{Var}_q} \right) = n\overline{Var}_q + 2 \binom{n}{2} r\overline{Var}_q \quad (3.2)$$

where r is the average correlation for all pairs of bets within this bundle.

The variance of a bundle can fluctuate to anywhere between zero if all bets are perfectly negatively correlated to each other to just under $n^2\overline{Var}_q$ if bets are positively correlated to each other. In an applicative scenario, it would be recommended to back-test the correlation between all bets made within a particular bundle. The value of $\sigma_{w_1}^2$ furthermore helps in controlling the dispersion in the amount of odds as a possible consideration for bundling could be the size of the wager. Assuming an infinite number of possibilities and a finite number of bets, an example of no correlation is used. In this case $Var_q = n\overline{Var}_q$, leading to the profitability of a bundle q , denoted B_q following the distribution:

$$B_q \sim \mathcal{N} \left(\frac{k_q}{1+k_q} n_q \overline{W}_q, \overline{W}_q^2 \left(\frac{1}{(1+k_q)\pi_q} - \frac{1}{(1+k_q)^2} \right) \right) \quad (3.3)$$

For instance, consider that a company has 300 bets of odds averaging at odds of 1.07, wagers at an average of 10 units and a bookmaker margin of 5%. The profits in this bundle follow the distribution $B_q \sim \mathcal{N}(142.86, 3360.55)$ and have a probability of a loss of less than 0.7%³. This result is highly sensitive to the correlation coefficient. If this was considered to be 0.5, then the variance of this bundle would be 502,401.36 and the probability of loss as high as 42%⁴.

3.2.2 Portfolio Variance

Having calculated the parameters for each bundle, one can finally calculate the joint distribution for the company's whole portfolio of bets. Assuming that b bundles are created, the total portfolio variance is $\sum_s Var_s + 2 \sum_{s>t} r_{s,t} \sqrt{Var_s Var_t}$ for $s, t \in \{1, 2, \dots, b\}$ where $r_{s,t}$ is the correlation coefficient between bundles s and t .

As an example, consider calculating the distribution of bookmaker profits for a portfolio of one hundred thousand bets uniformly subdivided over 20 bundles as shown in Table 3.1. It is further assumed that the bookmaker margin per bundle increases for low-likelihood outlooks in line with the longshot bias, which states that the gap between real and implied probabilities is higher for less likely outcomes (Vaughan Williams & Paton, 1997; Štrumbelj, 2014). This assumption is also consistent with the expectancies of a higher proportion of multiple/accumulator bets at higher odds.

³ $1 - \Phi(2.47)$.

⁴ $1 - \Phi(0.20)$.

TABLE 3.1: Wagers on a Portfolio of Bets with One Bookmaker

Bundle (q)	n_q	$\bar{W}_q$	k_q	π_q	$\mathbf{E}(B_q)$	$\sqrt{\text{Var}_q}$
1	5000	35	5.00%	97.50%	8333	654
2	5000	35	5.50%	92.50%	9123	879
3	5000	20	6.00%	87.50%	5660	613
4	5000	15	6.50%	82.50%	4577	537
5	5000	10	7.00%	77.50%	3271	408
6	5000	10	7.50%	72.50%	3488	457
7	5000	10	8.00%	67.50%	3704	507
8	5000	10	8.50%	62.50%	3917	559
9	5000	10	9.00%	57.50%	4128	614
10	5000	10	9.50%	52.50%	4338	673
11	5000	5	10.00%	47.50%	2273	369
12	5000	5	10.50%	42.50%	2376	405
13	5000	5	11.00%	37.50%	2477	446
14	5000	5	11.50%	32.50%	2578	494
15	5000	5	12.00%	27.50%	2679	553
16	5000	5	12.50%	22.50%	2778	629
17	5000	5	13.00%	17.50%	2876	731
18	5000	2	13.50%	12.50%	1189	354
19	5000	2	14.00%	7.50%	1228	467
20	5000	2	14.50%	2.50%	1266	827

A presumption that one can make is that closer bundles are more likely to be related than distant ones. The reasoning behind this is that low odds imply high likelihood while high odds imply low likelihood. Hence these are more likely to belong to a set of mutually exclusive outcomes. This might not be always true. For example a market with many possible outcomes is likely to produce many high mutually exclusive odds. Notwithstanding, the correlation between different bundles can again range between -1 and 1 but as a further specimen example it can be set as shown in Equation (3.4).

$$r_{s,t} = \begin{cases} 0.75 : |s - t| = 1, 2 \\ 0.50 : |s - t| = 3, 4 \\ 0.25 : |s - t| = 5, 6 \\ 0.00 : 6 < |s - t| < 14 \\ -0.25 : |s - t| = 14, 15 \\ -0.50 : |s - t| = 16, 17 \\ -0.75 : |s - t| = 18, 19 \end{cases} \quad (3.4)$$

The total variance is calculated as $Vector_{\sqrt{\text{Var}_q}} \times \text{Correlation Matrix} \times Vector_{\sqrt{\text{Var}_q}}^T$ where the correlation matrix is a 20×20 matrix showing the correlations between different

bundles and $Vector^T_{\sqrt{Var_q}}$ shows the standard deviation per bundle as calculated by Equation (3.3). Equation (3.5) shows these figures for the case examined in Table 3.1.

$$Vector^T_{\sqrt{Var_q}} = \begin{pmatrix} 653.72 \\ 879.43 \\ 613.47 \\ 537.16 \\ 407.72 \\ 457.03 \\ 507.15 \\ 559.11 \\ 613.94 \\ 672.87 \\ 368.69 \\ 404.72 \\ 445.92 \\ 494.37 \\ 553.35 \\ 628.54 \\ 730.90 \\ 354.18 \\ 467.47 \\ 826.70 \end{pmatrix} \quad (3.5)$$

The distribution of the portfolio of bets placed with the bookmaker, denoted $Port$, can therefore be evaluated as following the distribution:

$$Port \sim \mathcal{N}(72261, 30026410) \quad (3.6)$$

3.3 Determining Solvency

The estimation of the expected profit and variance of a portfolio of bets for a bookmaker can lead to further tools for regulators and bookmakers to manage their own risk, here described as the deviation from the expected. One application could be an extension of the six sigma method (Kwak & Anbari, 2006; Schroeder et al., 2008) whereby the expected profit of the portfolio needs to be at least four and a half times the standard deviation for the bookmaker to be able to operate.

Many financial regulatory regimes require a minimum of capital funds to be held by the financial operator to operate. Such regimes tend to determine this amount as a percentile measure, typically the Value-at-Risk (VaR) measure at some level of confidence. A clear example is the Solvency II insurance regime that is being introduced for the European insurance industry (Doff, 2008). The VaR evaluates the amount of capital required for the financial institution to be able to withstand extreme events. In layman terms, the VaR answers the question: What amount of capital is required for the company not to go bankrupt by $x\%$ probability?

This measure has been widely criticised from a theoretical perspective, as it seems to suggest that a small level $[(1 - x)\%]$ of failures are tolerated (Doff, 2008) and it does not have adequate mathematical properties such as sub-additivity (Acerbi & Tasche, 2002; Artzner et al., 1999; Dhaene et al., 2006).

Furthermore VaR does not consider the magnitude of the losses beyond its value. In this respect many recommend the use of the Expected Shortfall (ES), also referred to as the Tail-Value-of-Risk (Acerbi & Tasche, 2002; Yamai & Yoshida, 2005). ES can be described as the expected loss in the extreme percentile beyond the VaR.

3.4 Challenges and Extensions

One key challenge in applying this method relates to which timeframe of bets to use. Should a bookmaker apply risk measurement for all bets placed within a particular time-frame, open at a particular moment or closing at a particular time-frame? The per-bundle parameters for each bookmaker can be calculated at different times and used to project future capital requirements. Furthermore the use of different correlation parameters may result in evaluating best-estimate, pessimistic and optimistic measures which are then utilised to rate the riskiness of each bookmaker.

On the plus side, the method presented here is modular and can be easily enriched. For example, the portfolio of bets is subdivided in bundles by implied probability. The bundling technique may provide more accurate best-estimates if subdivided in more dimensions such as the bookmaker margin, size of wagers, type of event, type of market and timeframe.

The application of any model to measure risk is bound to be imperfect (Greenspan, 2008) and therefore should not be an end-all. Following a number of model failures, many point out that risk models act more as a false sense of security rather than a proven system for insolvency prevention (Colander et al., 2009; Danielsson, 2002). On a similar tone, many have investigated inadequacies in the models used to measure risk

in the financial sector (e.g. [Berkowitz & O'Brien, 2002](#)) or discussed methods on how to assess these (e.g. [Kupiec, 1995](#)). Fundamentally financial institutions operate in a market in which their actions affect the value of market securities ([Colander et al., 2009](#)) and financial models depend on the solvency of counter-parties since they operate in a highly leveraged environment with significant trade between the institutions.

Contrastingly to financial markets, the outcome of an event is not and should not be affected by a bookmaker since trade between bookmakers is currently minimal and bookmakers are not highly leveraged. The failures in application of mathematical developments in finance, such as risk models and derivatives, can also be blamed on misuse rather than simply design limitations ([Bezzina & Grima, 2012](#); [Salmon, 2012](#)). Moreover mathematical models add insight ([Geoffrion, 1976](#)) and provide a further understanding of the sensitivities of output to changes ([Thiele et al., 2014](#)). A model cannot be assumed to describe perfectly the behaviour of the underlying system but it gives us a chance to see how to react in case of extremities.

3.5 Conclusion

The setting of reserves of a portfolio of bets wagered with a bookmaker is therefore a lacuna that needs to be addressed. As the gambling sector continues to grow, a consolidated regulatory and internal approach to risk would promote active management of own risks, increase customer protection, enhance investors' ability to analyse bookmakers and reduce regulatory arbitrage.

This chapter introduces a method that measures the betting risk of a bookmaker by subdividing bets in different bundles. The model presented here is based on a number of assumptions such as distributions being mainly normal distributed, bundle sizes being equal and internal correlation for each bundle being zero. This leads to significant scope for further research which would focus on ideal bundling regime, actual experienced per-bundle correlations and challenges in setting up a risk-management framework that adapts this risk measurement technique.

Chapter 4

A public (mis)interpretation of Brazil's world cup performance

4.1 Introduction

The Brazil National Football team is arguably the World's most successful soccer¹ national team, being the only team to have played in all final World Cup tournaments since 1930, and having garnered five world cup wins. It is therefore not surprising that Brazil was touted to win the 2014 World Cup by several sport commentators and analysts (e.g. [Goldman Sachs, 2014](#); [McNulty, 2014](#)). Adding to high expectations for the Brazil National Team in the 2014 World Cup, was the fact that Brazil was hosting it. Six of the nineteen of previous world cup winners were also the World Cup hosts (Uruguay 1930, Italy 1934, England 1966, Germany 1974, Argentina 1978 and France 1998).

This chapter is an elaboration of a [presentation](#) at a conference ([Cortis, 2015b](#)) where I obtained a graduate scholar award. Following feedback during the conference, I have collaborated with an economist with expertise in behavioural economics in order to improve the paper which is now under review at the journal of the conference organiser ([Cortis & Briguglio, under review](#)). Throughout this chapter the perceived chances of Brazil winning the 2014 World Cup, as revealed in betting market data, are described. The next section describes the format of the World Cup Finals and how betting exchanges are set-up. Details of how odds actually fluctuated are then reviewed, in tandem with Brazil's actual performance.

¹The terms soccer, association football and football are used interchangeably throughout this thesis.

4.2 Context: The World Cup Finals and Betting Markets

4.2.1 World Cup Finals

World Cup Finals are held every four years during June and July and are typically hosted by one nation based in, or outside, Europe on an alternating basis. All nation members of FIFA participate in qualification rounds to determine the 32 teams that qualify for the final tournament². The method and number of teams that qualify varies within each of the six regions (Africa, Asia, Oceania, Europe, North and Central America and Caribbean, and South America). The host nation qualifies automatically.

The 32 teams are first subdivided into eight groups of four teams each, based on a semi-stratified random process and subject to the host nation being in group A. The process involves teams seeded into four pots, based on a FIFA/Coca-Cola World Ranking that measures the strength of a team based on recent results (FIFA, n.d.), prior to a random selection. This process, intended to avoid early elimination of the best teams, is still capable of generating a so-called “group of death”, that is a group with many potential World Cup winners (McHale & Davies, 2007). A case in point during the 2014 World Cup is Group B since it included Spain and the Netherlands, the two finalists of the previous competition.

Regulations of progress during the final tournament are well defined by the organising committee FIFA (FIFA, 2011). A league, held at group stage, identifies the two highest ranked teams, based on points, who then qualify for the knock-out stage. A number of additional hierarchical set of criteria are applied in the case of equal points to determine ranking. The first criterion is goal difference (which is the difference between goals scored and goals conceded), followed by the number of goals scored. If two or more teams are still equally ranked, then the same order of criteria are applied but based on direct encounters between the teams concerned. A ballot might also be the deciding factor, though this scenario has never occurred to date (Monkovic, 2014).

A total of 16 teams then participate in the knock-out, during which teams are eliminated after losing a match. In case of a draw after ninety minutes, an additional thirty-minute period is played, followed by penalties if result is still a draw³. The matches would have been determined well ahead of the groups being picked and are set such that top ranked teams play second ranked teams in the first round of the knock-out (also known as the Round of 16). The second round, or quarter-final (where eight teams play in four

²The total number of teams that qualify for the final tournament has been on a steady increase from thirteen teams in 1930 to thirty two teams in 2014. There have also been recent discussions of increasing this to 40 teams in future World Cups (Snowball, 2015).

³Unlike in other sports, draws are very likely in soccer (Van Calster et al., 2008).

separate matches) is followed by the semi-final (four teams) and the final (two teams), the winner of which is the World Cup Winner (FIFA, 2011).

4.2.2 World Cup Betting Markets

In this context, odds can be offered for each team to be eliminated at any stage, and a market existing for each team. For example Figure 4.1 shows a market for the final outcome for Brazil: Elimination at group stage, Elimination at Round of 16, . . . , Losing Finalist, Winner⁴. While several permutations are offered, the more popular bets are placed on a team reaching particular stages, each of which is conditional on previous performance⁵. An additional outcome, Group Stage Winner, is usually added to these. These markets, defined as derivative markets in Chapter 7, are either clustered by group (Group Winner, Round of 16) or with all teams listed (Round of 16, Quarter Final, Semi-Final, Finalist, Winner). Here, a market exists for reaching each elimination stage rather than for every team⁶.

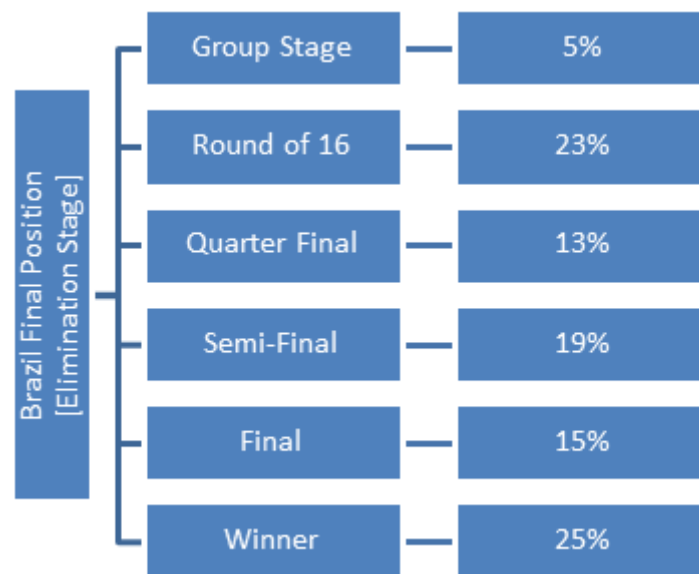


FIGURE 4.1: Market arranged by outcome per team.

As discussed in Chapter 2, odds can be displayed in a variety of formats, with the three most popular being European/Decimal, English/Fractional and American/Money-Line. They can also be offered to customers in a variety of settings with the main three being traditional, pari-mutuel, and betting exchanges. In a traditional setting, a bookmaker sets odds and takes risks based on customer bets, while in pari-mutuel winnings are shared according in ratio to the amount of money wagered on each outcome. For example

⁴Note that these add up to one.

⁵For example in order for a team to win the World Cup, it needs to reach the final and all previous stages.

⁶As seen in Table 4.1.

if a total of \$1,000 were wagered on Brazil winning the World Cup from a total of \$10,000 wagered on the World Cup winner (and \$9,000 wagered on other teams), then anyone who wagered on Brazil would win \$10 for every \$1 wagered. This would be represented as odds of 10 if no commission is taken by the organiser.

The key difference in a betting exchange is that individuals can either accept particular odds or propose other ones. Consider someone who wants to bet on Brazil to win a match (backing) but get twice her/his money back. This constitutes odds of 3 and s/he can place this request on the exchange. Someone who fancies Brazil not winning the match (laying) so says that s/he is willing to sell odds at evens (2). If there are only these two bettors on the market, then there is no agreement. However a third individual may wish to wager a bet at evens on Brazil winning the World Cup and may accept the former offer by backing it. At this stage, the bet was matched at odds of 2. In every case the amount being wagered is also considered. So if the layer considered a \$500 bet while the backer a \$200 bet, \$300 would still be available to be matched at these odds.

The betting exchange charges commission on any profits made but, unlike a bookmaker, it does not set odds or take risks of having too many bets on one particular outcome (Koning & van Velzen, 2009). Rather, it acts like a stock market exchange offering binary options - a type of derivative. The markets are offered, and hence fluctuate, continuously. The market is sufficiently large to also involve professional traders as well algorithmic trades. At each point a bet is matched, there is an agreement at least by two parties, of the perceived probability of an outcome.

4.3 Data and Trends

4.3.1 Data

The data used in Table 4.1 below are drawn from Betfair, the world's largest betting exchange. They consist of all matched odds related to Brazil on their success in reaching each stage and winning their group. The data available were for the first and last time that odds were matched, resulting in over 970 data points that describe the price movements. As particular odds could have been traded more than two times (beyond the first and last), it is not possible to attribute a precise time stamp, though the trends are nonetheless evident.

Figures 4.2, 4.3 and 4.4 show the movements of implied probabilities at different stages of the group stage. Table 4.1 shows the implied probabilities before and after each Brazil match. At the start of the tournament Brazil were considered 25% likely to win the

TABLE 4.1: Probabilities at different time points

Opposition			Round of 16	Group Winner	Quarter Final	Semi Final	Final	Winner
Group Stage	Croatia	<i>Before</i>	95%	82%	72%	59%	40%	25%
		<i>After</i>	99%	92%	72%	59%	40%	26%
	Mexico	<i>Before</i>	99%	78%	72%	59%	40%	26%
		<i>After</i>	99%	87%	65%	51%	40%	19%
	Cameroon	<i>Before</i>	99%	90%	65%	51%	40%	20%
		<i>After</i>			73%	55%	37%	18%
Round of 16	Chile	<i>Before</i>			83%	58%	42%	27%
		<i>During</i>			47%	55-40%	20%	19-15%
		<i>After</i>				74%	43%	22%
Quarter Final	Colombia	<i>Before</i>				85%	50%	31%
		<i>During</i>					57%	33%
		<i>After</i>					52%	31%

World Cup and 40% likely to reach the final. The team was perceived as having a 19% likelihood of losing at the semi-final stage of the competition.

4.3.2 Perception versus Performance

4.3.2.1 Group Stage – Croatia [3-1], Mexico [0-0], Cameroon [4-1]

During the first stage, Brazil was drawn to play against Croatia, Mexico and Cameroon. Brazil started off on the wrong foot, as they conceded an own goal at the start of the game. This dropped their chance of winning the group from 82% to 76% and to qualify from 95% to 93% as shown in Figure 4.2. By the twentieth minute they had obtained the draw and by the end of the match they were winning 3-1. Their chances of either winning the group or qualifying had shot up to 92% and 99% respectively. In fact, by the time Brazil's star player, Neymar, scored the first goal, the probability of winning the world cup was perceived as 1% higher. During this time the other markets were very illiquid, exhibiting little to no changes.

Were Mexico to draw Cameroon in the same group, the chances of Brazil being ranked first in the group would increase. On the contrary Mexico impressed and won, thereby depressing the chance of Brazil winning the group from 92% to 78%, even though the chances of Brazil passing the stage remained intact and the chances of reaching higher stages were only slightly affected.

Mexico impressed during their match versus the hosts and the result was a draw – leading the odds for Brazil winning the group back to 87%. At this stage Brazil held joint first

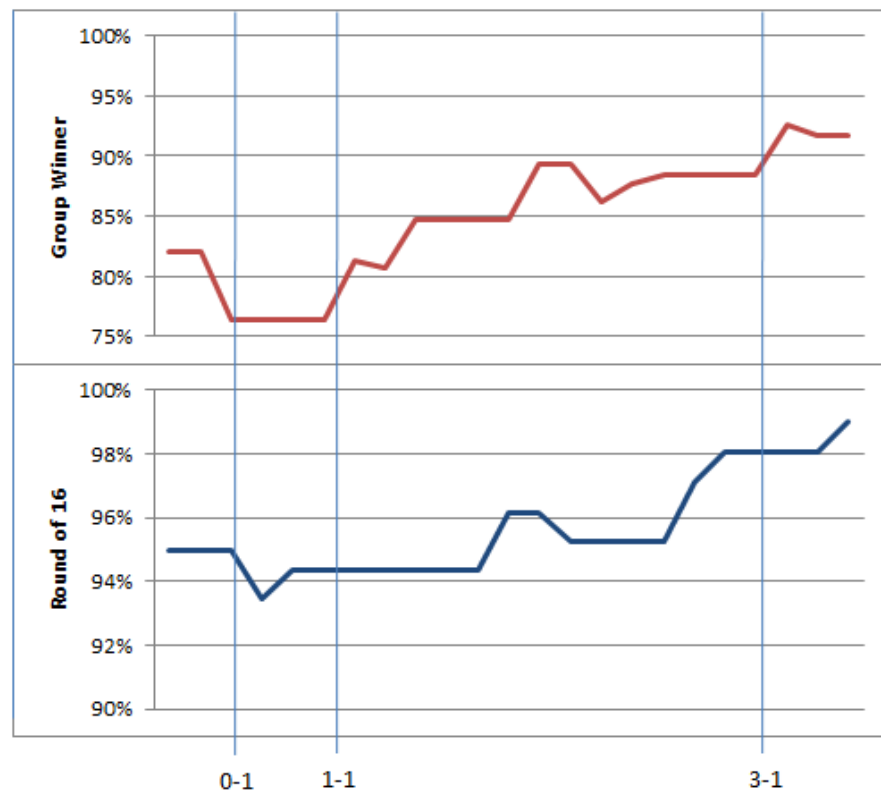


FIGURE 4.2: Perceived Probabilities during match vs. Croatia

position with Mexico (4 points each), while Croatia and Cameroon were still at zero points. By the time Brazil played the last match with Cameroon, Croatia had won against Cameroon. Brazil were therefore playing what possibly was the weakest team in the group – certainly one with no prospect of advancing further. A draw would have secured Brazil reaching the next stage and possibly even winning the group.

Brazil's chances of winning the group had been already rising from 87% to 90% prior to kick-off. They scored the first goal but then conceded a draw. Figure 4.4 highlights the nine minutes of pain in which Brazil were held to a draw. Yet, by the time they scored the third goal, markets were treating them as a definite group winners. In fact, Brazil went on to win the group and reached the next round. Although still favourites, their likelihood to win was far lower than at the start, and now perceived to be at 18%.

4.3.2.2 Knock-out Stage – Chile [1-1], Colombia [2-1], Germany [1-7]

At the start of the Round of 16 match, the pre-match odds were at their highest to date. Bettors seemed to be oblivious to the prior results which had clearly shown that Brazil were not overpowering the other teams. During the match itself, Brazil suffered significantly and only won on penalties. This diminished their perceived odds by about 40% while the game was progressing. Immediately post-match, odds did not show an improvement over pre-tournament expectations.

At the quarter final (Brazil versus Columbia), the same pattern emerged again. Pre-match perceptions showed improved expectations over Brazil's performance than the post-match odds of the previous match imply. Brazil were immediately winning versus Colombia, although not being significantly superior from a performance point of view. At one point during the match, the score was 2-0 and the probabilities of winning the cup shot up to 33%, the highest recorded value for Brazil during the entire tournament. Brazil did, in fact go on to win the quarter final, but their star player, Neymar, suffered an injury. Still, at the end of the quarter final match, Brazil remained favourites to win the semi-final against Germany.

By the time the semi-final match kicked off, the information of the previous games seemed to have seeped in, and that the odds started to be considered at evens. This was, actually, the first time that odds on Brazil reduced likelihood during non-playing time. In hindsight, these odds were still possibly optimistic, considering that Germany eventually annihilated Brazil - with a score of 7-1.

4.4 Discussion and Conclusion

Examining Brazil's odds during the World Cup reveals a trend: the post-match odds showed a persistently lower probability than the pre-tournament odds. Brazil consistently underperformed yet, at almost each stage, by the time it kicked off the next match, there seemed to be an expectation of the team picking up and transforming itself into the legendary team bettors seemed to expect it to be at the start. As the outcome finally revealed, that legendary team was not Brazil at all: it was Germany. It was only when the shocking development of Neymar's injury unfolded, that the dynamics of the market finally changed, lowering the perceived probability, and raising the odds.

In normal circumstances, the average direction of bettors' valuation should be zero. Randomness and deviations will occur, but the expectation is that a team should be both under-valued and over-valued (Malkiel, 2007). Instead, this data indicates that

bettors persistently over-valued Brazil. What could have caused this misinterpretation of Brazil's performance? Why did bettors seem to persistently lose sight of the true value and prospects of the home team?

One conjecture is that the notion of Brazil as a legendary team anchored bettors to their initial high expectations and continued to gravely influence the interpretation of events. Anchoring bias occurs when people put too much emphasis on a first piece of information or impression when making decisions (Tversky & Kahneman, 1974). Failure to shake the anchor, or to adjust, results in poor estimates of values and probabilities (Epley & Gilovich, 2001). While team records provide a reference point, that record, by itself, may be an inaccurate assessment of *current* team talent and performance. The fact that Brazil have such a strong positive directional deviation could suggest bettor overreliance on team record in their team valuation.

That Brazil were playing home may also have resulted in a process whereby bettors over-valued the probability of a win. Judgement of probabilities tends to be influenced by the ease with which such instances come to mind (availability bias). People tend to believe that events are more frequent or probable if examples of them are easier to remember. In turn, the main source of this error is salience and events that have been well-publicized or prominent tend to be estimated with exaggerated probabilities (Tversky & Kahneman, 1973, 1974). All sports games, moments and teams do not receive equal coverage in sports media. Brazil would anyway have been the subject of considerable media coverage, but even more so when they were playing home. Similarly, games and moments with the most potential for accompanying analysis and fan viewership (e.g. unexpected events) are generally overplayed and overemphasized, possibly influencing how bettors behave in subsequent weeks. This may also have been the reason why the injury finally nudged bets to be closer to realistic expectations.

More generally, sports bettors may overestimate their skill at understanding sports (overconfidence bias). In investor behaviour, experience and mistakes have been found to be linked through overconfidence (Barber & Odean, 2000) and men tend to be more confident than women (Barber & Odean, 2001). Constant exposure, discussion and coverage (rarely based statistical inference, analysis or averages, but rather on views of former players, stories, highlights, emotions, flukes) may lead (predominantly male) sports bettors to deem themselves above average predictors of games and appraisers of talent.

The discussion presented in this chapter is based on a simple eye test of data drawn from the largest betting exchange whose characteristics are close to a derivatives market exchange, and where odds change based on beliefs. While many of the favourites at the start of the tournament went on to win the group stage, a divergence between Brazil's world performance and the public's bets can be documented. Anchoring, availability

and overconfidence biases are well-researched behavioural heuristics, and the discussion above suggests that they may be present in this context. This presents an interesting agenda for further analysis, particularly given that sport betting exchanges replicate the environment of financial markets in a simpler environment which is one where the outcome is known sooner.

Chapter 5

Predicting goal difference

5.1 Introduction

The prediction of soccer matches has been the subject of many researchers (e.g. [Dixon & Coles, 1997](#); [Goddard & Asimakopulos, 2004](#); [Greenhough et al., 2002](#); [Hvattum & Arntzen, 2010](#); [Koning et al., 2003](#)) but in the majority of cases, these can be subdivided into two categories; those that try to model goals scored and those that try to predict the result directly ([Goddard, 2005](#))¹. Both methods have their relative advantages but in essence they are trying to balance out the magnitude of a win, mostly through predicting goals scored, and the final outcome. By forecasting outcomes directly, models treat a 1-0 and 4-0 as similar. On the other hand predicting goals directly results in the significant number of complications discussed in the literature review.

This chapter, co-authored with Haidar Haidar, proposes an approach attempting to predict the magnitude of a win by directly forecasting the goal difference in a match. In simple terms, by trying to find an equilibrium between deciding which is the best team (predicting outcome) and how good a team is (predicting goals scored), the focus presented here is on how much better/worse a team is than its opponent (goal difference). The English Premier League is used as an example and a good measure of success is noted when applying a four-factor linear model.

The next section is a review of different types of models used to predict soccer performance. This is followed by a description of our methodology, in which the data is described and a rationale for using a four-factor model is provided. Subsequently, the

¹[Goddard \(2005\)](#) found minor differences in a comparison of performance of four different models even if one that uses goal-based team performance variables to predict the outcome directly was reported as a slightly better model than the other three.

results of applying our model are investigated. Finally these findings are compared to others and future areas of work are proposed.

5.2 Literature Review

Albeit some novel qualitative approaches have been used to analyse soccer performance; such as Gréhaigne et al. (2001) and Sarmiento et al. (2013); the majority of proposed techniques tend to be of a quantitative, rather than qualitative, nature (Sarmiento et al., 2014). For example Hughes & Bartlett (2002) recommend key performance indicators (KPIs) that are classified into four domains: match classification, including score and shots on/off target; biomechanical, which in soccer have mainly focused on kicking and passing (e.g. Lees & Nolan, 1998); tactical, such as possession (e.g. P. D. Jones et al., 2004); and technical such as tackles won and successful dribbles (e.g. Luhtanen et al., 2001). Specifically, when faced with many indicators, principal components analysis has been recommended in sports performance analysis to reduce these to a smaller set of independent KPIs (Choi et al., 2008; Federolf et al., 2012; O'Donoghue, 2008, 2013).

On a similar pattern, many models have been developed in order to predict the winner of a soccer match. Although their aim is to predict outcome rather than measure performance, it can be easily argued that any input parameters used in such models are good KPIs of a team's future performance. However some of these models require multidimensional team specific data in order to run (Demir et al., 2012) and therefore this has led to many models using easily accessible data, such as goals scored, to predict match outcomes.

In predicting the number of goals scored, a significant number of papers point out that this is a skewed distribution (Bittner et al., 2007, 2009; Greenhough et al., 2002; Heuer et al., 2010). Many tend to simulate the goals scored per team to follow a Poisson distribution (Dixon & Coles, 1997; Koning et al., 2003). However there have been numerous discussions that this may not be the case for a number of reasons including that a team has an increased chance of scoring if it has already scored (Bittner et al., 2007, 2009; Heuer et al., 2010), the number of goals scored is positively correlated with the number of spectators (Van Calster et al., 2008) and red cards increasing the scoring intensity by 88% for the winning team and 5% for the losing one (Ridder et al., 1994). There is also an additional intricacy on whether to treat the goals scored by each team as independent.

Prediction in soccer has the increased complexity that the outcome of a draw is not only possible but also much more common than other sports (Archontakis & Osborne, 2007). In fact around 9% of all games end up in goalless draws (Ridder et al., 1994).

The number of goals scored per game has dropped significantly since the 1960s due to game advancements (Gaviria, 2000) and less expected goals promotes the chances of more draws. Furthermore soccer is considered more competitive than other sports, including American sports (Ben-Naim et al., 2007), hence resulting in more balanced games. Particularly Ridder et al. (1994) state that teams of equal strength have a 25% chance of ending in a draw. Furthermore Archontakis & Osborne (2007) report that 29.76% of World Cup matches end in a draw, ranging from 28.75% during first round matches to 32.29% in second round matches as better teams of similar strength remain and there is increased risk aversion to taking risks at the knock-out stage. This risk-aversion leads to average and average-to-bad teams to be more likely to end up in goalless matches (Ridder et al., 1994).

Out of the three possible outcomes of a soccer match (home team win, away team win and draw), the draw is reported to be the hardest to predict (Pope & Peel, 1989; Kuk, 1995). Draws also result in divergent results when evaluating betting markets. For example, draw odds displayed a negative longshot-bias² in the English soccer and backing draws provided lower losses than backing home or away wins³ (Snowberg & Wolfers, 2010). Similar results were reported by Dixon & Coles (1997). Pope & Peel (1989) note that the draw is the outcome, out of the three possible results of a match, that experiences the least movement across all bookmakers. The standard deviation of the home and away wins differs by around 4% from the standard deviation of the draw outcome. In fact Fibonacci techniques in which wagers are made on draws (Archontakis & Osborne, 2007; Demir et al., 2012) have been examined to be relatively successful as bookmakers face difficulties in pricing draws efficiently.

Although the number of models that predict the outcome of a match directly are numerous (Kuypers, 2000; Goddard & Asimakopoulous, 2004; Goddard, 2005; Hvattum & Arntzen, 2010)- there are three main developments that have focused on estimating draws. Glenn & David (1960) proposed a model that applies the probability of a win to home or away team, and predicts a draw if the probability is not within a certain threshold. An improvement by Henery (1981) involved changing the single draw parameter to two, that is one for each team. Finally Kuk (1995) extended this further by considering separate parameters for each home and away team while still using four parameters: one for home team games playing home (not the same if the same team is playing away), away team games playing away, home team draw and away team draw.

²A positive longshot-bias means that outcomes that are perceived as highly likely by betting companies occur more often than the markets imply while outcome that are perceived less likely to win occur less often than expected (Cain et al., 2000). This occurs because typically bettors are risk-seekers who prefer to wager on unlikely outcomes that would provide significant wins.

³This will also be reported in Chapter 6.

A third method of approximating results is put forward in Chapter 7. Although that chapter focuses on profiting of market inefficiency of betting derivatives during Euro 2012, a Monte-Carlo simulation is used to predict firstly match outcome, then goal difference and subsequently goal scored. The argument presented is that these three have the given pecking order when ranking teams.

O'Donoghue (2013) provides an example of how regression can be used to predict goal difference for soccer national teams based on their international ranking. In this research, this method is extended by predicting directly goal differences, that is the winning margin of the winning team, rather than the traditional models that focus either on goals scored or match outcome.

5.3 Data and methodology

The top English soccer league, known as the premier league, is the subject of the study used in this chapter. The study covers fifteen consecutive seasons from 2000/2001 to 2014/15. In each season, each team plays 38 matches, with twenty teams playing in the league, totaling 380 matches in the season. Each team plays as many home games as away games. Based on the sample of the 5,700 games played over the fifteen seasons under investigation, the relative frequency of not scoring in home games is 23.4%, while it is 34.6% in away games. It is also noted that the total number of goals scored by home teams exceeds the total number of goals scored by away teams by 154.2 goals each season (average of 0.4058 goals / match).

The aim of this chapter is to find indicators that explain goal differences, which represent the number of goals scored by the away team subtracted from the number of goals scored by the home team⁴. The goal difference is our dependent variable and is denoted by Y throughout. Another reason for studying the goal difference is that it can be used to study the predictability of a draw in a match that is when the goal difference is zero. The goal difference is expected to be explained by two sets of performance indicators, short-term indicators and long-term indicators. The pre-match points per game of the playing teams are used as long-term indicators that explain some of the variation of our dependent variable, Y . The pre-match point of the home team and away team are denoted by P_h and P_a respectively (or alternatively P_I where $I \in \{Home, Away\}$). These are calculated as

$$P_I = \frac{3 * (\text{Number of wins to date in season}) + \text{Number of draws to date in season}}{\text{Number of games played to date}} \quad (5.1)$$

⁴For example a score of 3-1 is a goal difference of 2.

The short-term indicators are calculated over the previous n matches and it only takes into account the information captured in the previous n matches. The previous n matches that are used in the calculations for the home team are based on the previous home matches only, while the previous n matches that are used in the calculations for the away team are based on the previous away matches of the away team. The chosen short-term indicators in this study are to consider the effect of playing home or away onto the final score, and the short-term indicators selected are listed below:

$X_1^{[n]}$ The number of goals scored for the home team in the previous n games played home.

$X_2^{[n]}$ The number of goals scored against the home team in the previous n games played home.

$X_3^{[n]}$ The number of goals scored for the away team in the previous n games played away.

$X_4^{[n]}$ The number of goals scored against the away team in the previous n games played away.

[Dambrauskaite et al. \(2012\)](#) defined PCA as a statistical variable reduction technique that linearly transforms a set of variables into another set of new uncorrelated variables (called components) that preserve as much variation as the original variables presented. The transformation allocates as much variation as possible in the first component, followed by the second component, and leaving the last components with as small variation as possible. We perform PCA on our six variables (P_h , P_a , $X_1^{[n]}$, $X_2^{[n]}$, $X_3^{[n]}$ and $X_4^{[n]}$) for a better understanding of the variation and collinearity in our data. We, however, do not use the principle components in our regression model, but rather reduce the set of original variables in order to ease understanding. At least four components are needed to explain more than 86% of the variation of the original six variables in all the fifteen seasons under consideration. Hence a regression model can be used that excludes two variables (we do not use $X_1^{[n]}$ and $X_3^{[n]}$). The below table represents the pairwise average of the fourteen correlations between each pair of variables.

TABLE 5.1: The average of the (%) correlations over the fifteen seasons

Variable	$X_1^{[n]}$	$X_2^{[n]}$	$X_3^{[n]}$	$X_4^{[n]}$	P_h	P_a
$X_1^{[n]}$	100	-19.7	1.6	-2.0	58.4	0.6
$X_2^{[n]}$	-19.7	100	-2.3	2.8	-48.9	-1.1
$X_3^{[n]}$	1.6	-2.3	100	-19.5	1.9	55.5
$X_4^{[n]}$	-2.0	2.8	-19.5	100	-1.7	-51.4
P_h	58.4	-48.9	1.9	-1.7	100	-0.4
P_a	0.6	-1.1	55.5	-51.4	-0.4	100

5.4 A four-factor linear model

A four-factor linear model for goal differences is constructed by regressing the variable Y against P_h , P_a , $X_2^{[n]}$ and $X_4^{[n]}$. The model is written as

$$Y_j = \alpha_0 + \alpha_1 X_2^{[n]} + \alpha_2 X_4^{[n]} + \alpha_3 P_{h_j} + \alpha_4 P_{a_j} + \epsilon_j, \quad \forall j \in [1, 2, \dots, N^{[n]}], \quad (5.2)$$

where ϵ_j represents the noise and $N^{[n]} = 380 - 20n$ is the number of observations that are used in the model. The number of observations could be less than $380 - 20n$, when a match is postponed until a later time of the season. The coefficients of the regression applied to the fifteen seasons are given in Table 5.2.

TABLE 5.2: Regression Coefficients

	α_0	α_1	α_2	α_3	α_4
2000/01	-0.9603	+0.0003	+0.0628**	1.2777*	-0.4904
2001/02	-0.0854	+0.0815***	+0.0031	1.0593*	-1.0962*
2002/03	+0.5784	-0.0039	-0.0704	1.4258*	-1.1349*
2003/04	+1.0406	-0.0165	-0.0119	0.6829*	-1.0369*
2004/05	+0.2906	-0.0110	+0.0551	0.7424*	-0.9070*
2005/06	+1.1428	-0.0467	-0.0652	1.0258*	-0.9215*
2006/07	+0.6013	+0.0168	-0.0271	1.0814*	-1.0865*
2007/08	+0.8506	-0.0084	-0.0089	1.0102*	-1.2601*
2008/09	+1.7344	-0.0693***	+0.0165	0.5683*	-1.3951*
2009/10	+0.1001	+0.0243	+0.013	1.778*	-1.5156*
2010/11	+0.8952	-0.0339	-0.0203	0.8764*	-0.9393*
2011/12	+1.5176	-0.0595	-0.0712	1.0607*	-1.1938*
2012/13	+0.9688	-0.0007	-0.0769***	1.2014*	-1.227*
2013/14	+0.9743	-0.0624	-0.0063	1.139*	-1.2409*
2014/15	-0.0976	+0.0094	-0.0051	1.1493*	-0.7781*

* significant at $p \leq 0.01$, ** significant at $p \leq 0.05$, *** significant at $p \leq 0.1$
All values show evidence of significance using a two-tailed test.

The error, E_j , is then defined as the difference between the actual goal difference, Y_j , and the predicted value, $E[Y_j]$, which is $\hat{Y}_j$ rounded to the nearest integer when $m = 0.5$ as shown in Equation 5.3. Table 5.3 shows the percentage of games for which the model predicted the exact goal difference in each of the seasons. The model gives the exact prediction of the goal difference of over 20% of the matches, varying from 20% to 31.07% between seasons.

$$E[Y_j]^{(m)} = \begin{cases} 0 & \text{if } \text{abs}(\hat{Y}_j) < m \\ 1 & \text{if } m \leq \text{abs}(\hat{Y}_j) < 0.5 \\ \text{round}(\hat{Y}_j) & \text{if } \text{abs}(\hat{Y}_j) \geq 0.5 \end{cases}, \quad (5.3)$$

TABLE 5.3: Exact Prediction of Goal Difference

Season	00-01	01-02	02-03	03-04	04-05	05-06	06-07	
$\%(E_j = 0)$	28.32%	29.09%	25.55%	31.07%	24.73%	23.10%	29.75%	
Season	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15
$\%(E_j = 0)$	22.26%	32.26%	26.09%	28.78%	24.73%	28.52%	20.00%	25.71%

For $o \in \{\text{Home, Draw, Away}\}$, the variables used to capture the performance of the model are defined as

n_o showing the actual number of matches ending in a particular outcome. For example n_h counts the number of games in which the home team won (i.e. $Y_j > 0$)

$n_o^{(m)}$ showing the number of matches predicted to end in a particular outcome. For example $n_a^{(0.5)}$ counts the number of games predicted to be an away win when using the nearest integer predicting value (i.e. $E[Y_j]^{(m)} < 0$)

D_o showing the number of correct predictions. For example D_d shows the percentage of correct draw predictions.

TABLE 5.4: Results when using $m = 0.5$.

	Variable/Season												
Season	N	n _h	n _d	n _a	n _h ^(0.5)	n _d ^(0.5)	n _a ^(0.5)	$\frac{n_h^{(0.5)}}{n_h}$	$\frac{n_d^{(0.5)}}{n_d}$	$\frac{n_a^{(0.5)}}{n_a}$	D _h %	D _d %	D _a %
00-01	279	141	68	70	146	126	7	1.04	1.85	0.10	63.8	30.2	7.1
01-02	275	128	70	80	115	133	27	0.90	1.90	0.34	56.0	27.8	18.7
02-03	274	136	67	71	126	122	26	0.93	1.82	0.37	55.9	23.8	19.7
03-04	280	126	79	75	122	133	25	0.97	1.68	0.33	53.2	34.6	20.0
04-05	279	126	78	75	130	126	23	1.03	1.62	0.31	58.7	26.2	20.0
05-06	277	149	52	76	141	113	23	0.95	2.17	0.30	65.1	16.8	19.7
06-07	279	131	71	77	134	122	23	1.02	1.72	0.30	61.1	31.2	20.8
07-08	274	125	74	75	144	85	45	1.15	1.15	0.60	71.2	23.5	30.7
08-09	279	126	77	76	148	91	40	1.17	1.18	0.53	70.6	37.4	38.2
09-10	276	136	79	61	146	90	40	1.07	1.14	0.66	70.6	38.9	31.2
10-11	278	134	77	67	114	159	5	0.85	2.06	0.07	52.2	30.8	3.0
11-12	279	127	68	84	120	124	35	0.94	1.82	0.42	55.9	25.8	26.2
12-13	270	121	70	79	100	136	34	0.83	1.94	0.43	52.9	28.7	20.3
13-14	280	133	56	91	120	111	49	0.90	1.98	0.54	58.7	19.8	33.0
14-15	280	132	63	85	121	142	17	0.92	2.25	0.20	56.1	25.3	12.9

N shows the number of matches predicted.

5.5 Draws estimated using $m=0.25$

Table 5.4 shows that the model is underestimating the number of away wins and over estimating the number of draws. Take for example the most recent season where 85 away wins were experienced but the model only predicts 17 away wins (20% of this value). In the same season 142 draws are predicted but only 63 were experienced (225%) of this value.

Hence it is envisaged that the model would need a shortening of the interval, for which a draw is predicted, by changing the value of m in the definition of $E[Y_j]^{[m]}$. The value of m is therefore reduced from 0.5 to 0.25, and the prediction performance of the model is shown in Table 5.5. In such a case if a match is predicted to have a goal difference of 0.15, it is set as a draw. However a prediction of 0.3 would be considered as a +1 home team win. The probability of guessing the outcome of a match (i.e Win, Loss and Draw) has increased in all fifteen seasons that are subject to the test.

TABLE 5.5: Results when using $m = 0.25$.

Season	Variable/Season						$D_h\%$	$D_d\%$	$D_a\%$
	$n_h^{(0.5)}$	$n_d^{(0.5)}$	$n_a^{(0.5)}$	$\frac{n_h^{(0.5)}}{n_h}$	$\frac{n_d^{(0.5)}}{n_d}$	$\frac{n_a^{(0.5)}}{n_a}$			
00-01	195	56	28	1.38	0.82	0.40	80.1	30.4	22.9
01-02	165	63	47	1.29	0.90	0.59	75.2	31.8	33.8
02-03	160	70	44	1.18	1.04	0.62	69.9	27.1	32.3
03-04	177	66	37	1.40	0.84	0.49	76.2	39.4	28
04-05	172	62	47	1.37	0.79	0.63	73	26.7	36
05-06	177	65	35	1.19	1.25	0.46	74.5	16.9	29
06-07	167	70	42	1.27	0.99	0.55	69.5	35.7	31.2
07-08	164	42	68	1.31	0.57	0.91	78.4	19.1	46.7
08-09	184	41	54	1.46	0.53	0.71	80.1	43.4	46.1
09-10	164	54	58	1.21	0.68	0.95	75.7	44.4	42.6
10-11	183	73	22	1.37	0.95	0.33	74.6	32.9	11.9
11-12	157	62	60	1.24	0.91	0.71	69.3	35.5	41.7
12-13	143	71	56	1.18	1.01	0.71	65.3	19.7	31.7
13-14	165	51	64	1.24	0.91	0.70	77.4	17.7	38.5
14-15	161	82	37	1.22	1.30	0.44	66.7	23.2	21.2

The model gives the correct outcome of 52% of all games in our sample when compared to 41.7% when using $m = 0.5$. The correct predictions increased for all home and away wins as well as most matches ending in a draw. Figure 5.1 shows that the number of correct predictions increase significantly. The majority of games end up in a draw or a one goal difference hence showing a peak around the values of 0, 1 and -1.

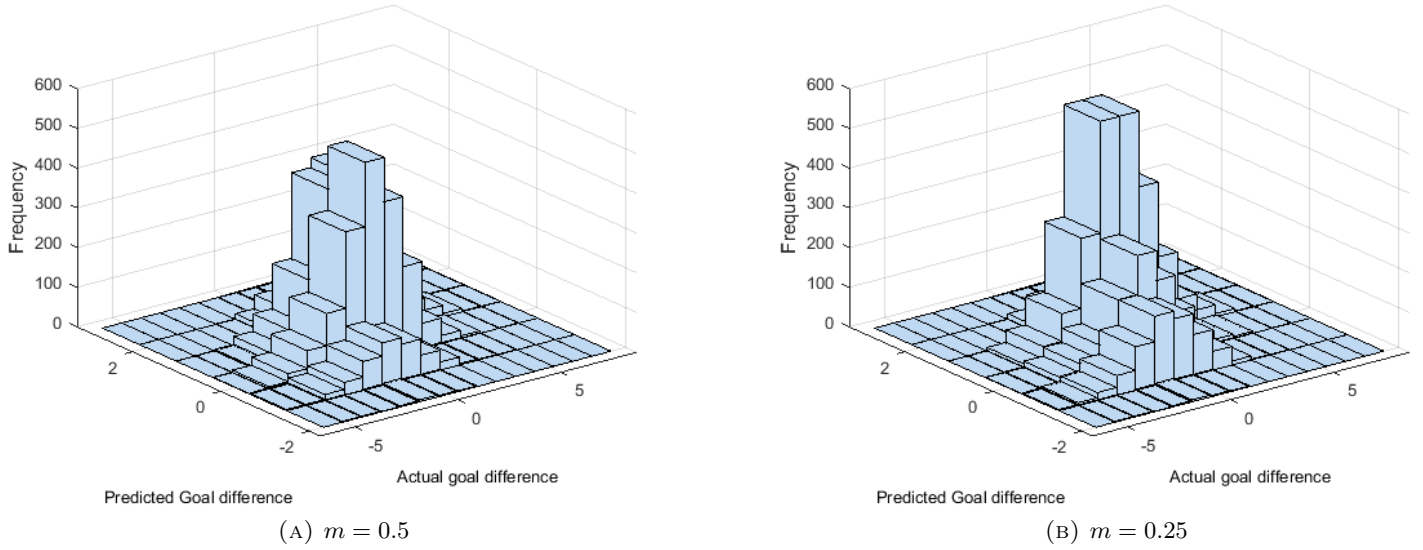


FIGURE 5.1: Frequency of model predictions compared to actual goal differences.

TABLE 5.6: Cohen's κ

Season	00-01	01-02	02-03	03-04	04-05
m = 0.5	0.0506	0.0599	0.0394	0.0546	0.0497
m = 0.25	0.1926	0.2344	0.1690	0.2317	0.1888
Season	05-06	06-07	07-08	08-09	09-10
m = 0.5	0.1015	0.0990	0.1620	0.2506	0.2328
m = 0.25	0.1562	0.1880	0.2495	0.3525	0.3290
Season	10-11	11-12	12-13	13-14	14-15
m = 0.5	-0.0236	0.0850	0.0554	0.1291	0.0368
m = 0.25	0.1173	0.2429	0.1117	0.2264	0.0884

Table 5.6 shows Cohen's Kappa (Cohen, 1960) evaluated for every season using both approaches. This value is used to measure the level of improvement over random guesses⁵. In essence a positive value shows that the model is an improvement over random guesses. Both models provide an improvement over random guesses except for the 2010-11 season where using $m = 0.5$ the model guessed 96 matches correctly while random guesses from the 114 home wins, 159 draws and 5 away wins would have been 100.19 correct guesses.

Using Altman's interpretation (Altman, 1990), values of κ between 0.2 and 0.4 can be treated as a fair level of improvement. We note that only two seasons show a fair level of improvement using $m = 0.5$ but almost half (seven from fifteen) when using $m = 0.25$.

⁵For example in 2010-11, the number of random correct home win guesses when using $m = 0.5$ would be $\frac{141}{279} \times 146 = 73.78$ while the number of total random guesses for win, draws and losses are 106.25.

5.6 Conclusion and Discussion

These results show that a simple model can be applied to determine goal difference. Our aim has been to promote prediction of goal difference rather than match outcomes or exact scores. This intermediary forecast leads to match outcomes by applying a range m such that if the goal difference is lower in magnitude than this value, the game is predicted as a draw. Otherwise positive values are considered as a home team win and negative values an away team win. By varying m , the model prediction improved significantly. However the prediction rate of draws is generally low. This supports the previous studies that the draw follows a random walk and cannot be predicted (Pope & Peel, 1989; Kuk, 1995; Štrumbelj, 2014).

The prediction of goals scored could be extended in a similar fashion to Cortis et al. (2013)⁶ who used a conditional simulation of goals scored depending on the goal difference.

It was also determined that a model that includes four components, being average points per match for each team, as well as the goals conceded in the last five matches at home by the home team and away by the away team are predictors of a match outcome. The model was also tested using goals conceded in the last six, seven and eight matches with similar results.

This research is restricted by being solely applied to one league and by the simplicity of the model. However the aim is to promote an alternative form of prediction rather than model precision. Future endeavours may focus on different leagues, comparisons with other models and possibly betting markets.

The model can be subject to many improvements and tests. The long-term indicator $P = P_h - P_a$, representing the difference in average points per match, could replace the two indicators P_a and P_h , without affecting the predictability of the model.

Post-ante data is used to test the model. A season finishes before regression coefficients are evaluated to calculate the accuracy of the matches in that particular season. It would be interesting to evaluate how the model performs ex-ante.

This research is restricted by being solely applied to one league and by the simplicity of the model. However our aim is to promote a new aim of prediction rather than model precision. Future endeavors may focus on different leagues, comparisons with other models and possibly betting markets.

⁶i.e. Chapter 7.

Chapter 6

Betting market efficiency in European soccer

6.1 Introduction

In Chapter 2, the relationship between the true probability (p_i) and the implied probability (π_i) of an outcome was defined by Equation (2.3). Furthermore Chapter 4 showed that odds, although showing the perceived probability of an outcome, may still not be correct. The betting market is sufficiently liquid such that bookmakers cannot offer odds that create arbitrage when considering all odds offered on the market as this would lead to a significant element of anti-selection. For example even if one bookmaker is certain that the probability of a two-outcome event is equally likely, it may not offer odds at even if the rest of the market is pricing it at 3:1 probability ratio for outcomes A and B. This is because a clever bettor would place a wager on Outcome A (at odds of 4) at a generic bookmaker and a wager on Outcome B (at odds of 2) with the bookmaker in question in a similar fashion to the proof in Proposition 2.1. This would result in the particular bookmaker having all wagers on Outcome B. Although this may be profitable in the long run for the bookmaker, it is more likely to result in default in the short run.

Hence odds may not be realistic, even after adjusting for the bookmaker margin. The value of the prospective prize and the wager result in a subjective probability, explained as the likelihood of this outcome as perceived by betting markets. The aim of this chapter¹ is to compare the subjective probability to real outcomes in major soccer markets, thus examining whether betting markets are correctly pricing events. Our literature review points out four main possible determinants for unfair odds: the favourite-longshot

¹This is joint work with Ranier Buhagiar.

bias, other biases, insider information and arbitrage opportunities². This is followed by our methodology and results which lead to our conclusion of evidence of the favourite-longshot bias. In order to examine this, three hypotheses are defined.

Hypothesis 1. The profits made by bettors are positively correlated with the probability of an outcome.

Hypothesis 2. The profits made by bettors are positively correlated with the accuracy of the prediction.

Hypothesis 3. The accuracy of predictions are positively correlated with the probability of an outcome.

6.2 Literature Review

6.2.1 The Favourite-Longshot Bias

The favourite-longshot bias has been in the eye of numerous authors who test the market efficiency in betting markets. It was first discovered by psychologist Griffith (1949), who concluded that the subjective probabilities do, on average, reflect the objective probabilities, yet a systematic undervaluation of odds for high-probability outcomes, and consequently an overvaluation of odds for low-probability outcomes, does exist. It is the most familiar bias in the betting industry, especially in the horse race betting (Thaler & Ziemba, 1988). Horse race betting is normally held in a parimutuel setting which differs from traditional bookmaking as it involves the pooling of funds, where the winners of the event have a share in the pool and the amount won depends on the total amount wagered on the horse or greyhound, after deducting any taxes and organiser fees. On the other hand, successful bets placed with bookmakers pay out at the odds offered when the stake was wagered and accepted. Nevertheless, in both forms of betting, low probability, high odds outcomes tend to be preferred by bettors, causing this high demand to decrease odds and consequently causing favoured outcomes to be less profitable than others. This means that there is a tendency for gamblers to over-estimate longshots and under-estimate favourites (Ali, 1977).

In theory, only negative returns can be expected from all odds categories, implying the grouping of odds representing specific ranges of probabilities (Chapter 2). Nonetheless losses can be minimised by focussing on high probability outcomes or other near-favourites (Smith, 2010). Gandar et al. (2001) explored that odds of the top four

²Vaughan Williams (1999) provides a very detailed overview of information efficiency in betting markets. This chapter focuses on the main themes related to our hypothesis tests.

favourites at the opening of on-track horse betting subjective probabilities are much lower than actual winning probabilities. Furthermore, horses between the sixth favourite and the 17th ranked, except for the horse ranked 16th, experience subjective probabilities that are higher than the objective probabilities. The bias diminishes at closing of bets, since odds at this period provide a more accurate and true measure of the horses' winning probabilities (Crafts, 1985). It is thus anticipated that a movement in odds will cater for a more informed market environment (Bird & McCrae, 1987).

Ali (1977) compared subjective and objective probabilities of 20,247 horse races and found that except for the most favourite horse in every race, the subjective probability surpasses the matching objective probability. Moreover, this difference for the 4th ranked horse and descending, were proved statistically significant. On the other hand, an underestimation of the high winning probability horse was established as the subjective probabilities were lower than the actual probability. Similar findings were discovered by Asch et al. (1982) with the first favourite having an objective probability of 0.361, exceeding the corresponding subjective probability of 0.325. Conflicting probabilities were found for the least favourite horses with the 9th ranked horse facing an implied probability of a mere 0.006 against the real probability of 0.018 of winning.

In an attempt to analyse reasons behind this favourite-longshot bias, Ali (1977) developed a wealth utility function for Mr B, representing all bettors that establish final odds of bookmakers. The convex function implied a risk loving relationship where, as wealth decreases, bettors tend to undertake higher risks and bet on longshots, albeit having minimal chances of winning. Similar conclusions were found by Hausch et al. (1981). The trend to under-bet favourites and over-bet low probability outcomes is augmented in the last few races of the day. On average, bettors end up losing up to 20% of their initial capital by the end of the betting period (Asch et al., 1982). Struggling to break even by the closing of the day may be the logic behind this risk seeking attitude of the betting public.

In contrast to the Expected Utility Theory, the favourite-longshot bias illustrates that bettors prefer to place money on bets with low expected returns and numerous researchers have attempted to explore possible solutions as to why this is the case (Suhonen & Linden, 2010). Hausch et al. (1981) provided a second explanation where bettors simply prefer low probability horses for elevated satisfaction if the horse ends up winning. Sauer (1998) further evaluated that the favourite-longshot bias may be due to bettors overestimating probabilities of longshots, whilst Thaler & Ziemba (1988) state that horses are chosen for reasons other than those relating to their winning probabilities, such as the horse's name. It is argued that the bookmaking market model in the betting industry stands between a pure monopoly and the Bertrand competition (Hurley & McDonough, 2013).

No better preferences develop in the latter model, but this is not the case in reality, since any imperfections performed by bookmakers will lead to a bias in the market.

An extensive literature on the longshot bias is available, with the majority of these works exploring a positive relationship where, as odds increase, the number of relative bets wagered rises, and vice-versa (Griffith, 1949; Ali, 1977; Hausch et al., 1981; Bird & McCrae, 1987; Gandar et al., 2001)³. Nevertheless, some exceptions have been found. Certain studies show that this bias is reversed in particular areas. Examining data from Happy Valley and Shatin racetracks in Hong Kong, Busche & Hall (1988) managed to find a betting population who over-bet on favourites and under-bet on longshots. Analogous results were obtained by Walls & Busche (2003) testing Japanese and Hong Kong racecourses. They observed that only the 4th ranked horse is significantly over-bet, while other horses show no signs of being significantly under-bet. Contrasting biases may be located within the same market. The English soccer betting market showed a positive-longshot bias in the home and away wins, whilst the draw odds display a negative bias (Deschamps & Gergaud, 2012). Furthermore, backing draws provided a much higher average return of -7% compared to home win and away win probabilities, yielding -11% and -14% respectively.

Evidence shows that the favourite-longshot bias may be associated with geographical regions (Kukuk et al., 2008). The bias has been found in western countries such as in the US (Ali, 1977), New Zealand (Gandar et al., 2001), Australia (Bird & McCrae, 1987), and Germany (Winter & Kukuk, 2006). A reversed bias appeared in Asian markets, yet no reasonably justified explanation as to why this differs has been pinned down (Coleman, 2004). These were apparent in Hong Kong (Busche & Hall, 1988; Walls & Busche, 2003) and in Japan (Walls & Busche, 2003). Attempting to analyse these dissimilarities in the favourite-longshot bias, Koch & Shing (2008) place responsibility on variations in odds grids adopted by bookmakers (as described in Section 2.3.2). Applying different odds levels in the same betting market may automatically give rise to a favourite-longshot bias⁴.

Moreover, modern bettors consider betting to be an act of consumption rather than investment (Paul & Weinbach, 2012). The betting volume in the 2008 NFL season seems to depend on quality of opponents, time of play, television coverage, uncertainty of outcome and other factors attracting bettors who wager for non-monetary reasons

³Vaughan Williams et al. (forthcoming) provide an interesting investigation on the favourite-longshot bias in online poker markets.

⁴For example consider two outcomes, one with probability of 86.7% and the other with probability 0.62% . Assuming no profit margins, the odds would be 1.1534 and 161.29 . A betting company tends to have odds grids such that odds offered are nicer figures and in this case it is more likely to offer the odds at 1.15 and 150 . While there is less than 1% discrepancy in the first odds, there is more than 6% in the second. Given that the grid levels are wider for low probability events with higher odds, there is more likelihood of a gap between subjective and objective probabilities for these cases.

rather than the traditional economic theory of expected returns (Scott & Gulley, 1995). This may lead to an increase in bets backing longshots for the sole reason of bettors having the ‘best story’ to tell (Thaler & Ziemba, 1988).

6.2.2 Exploring Other Biases

The betting industry is characterised by different market makers, mostly betting exchanges and traditional bookmakers (Franck et al., 2010). Whilst the former solely provides a trading platform and earns a small commission on each transaction, bookmaking companies undertake high risks as they take opposing sides for every bet placed (Bernile & Lyandres, 2011; Cortis et al., 2013). Bookmakers operate on a high cost basis yet they are still able to compete with bet exchanges having lower costs and better prices. This is made possible through the exploitation of bettors’ sentimental preferences and price elasticity of their demand (Franck et al., 2011). Basically, with price insensitive bettors, bookies are able to lower odds (increase prices) with little or no effect on sales (bets being played). Price sensitive bettors, on the contrary, must be defined between elastic and inelastic. Elastic price sensitive gamblers are faced with higher odds whilst the inelastic betting public encounter lower odds. Kuypers (2000) explains how bookmakers are capable in abusing bettors’ behaviour and their loyalty bias to manipulate odds and earn greater profits without affecting the punter’s decision to bet.

Financial markets have been recognised to show a degree of preference towards domestic investments and financial instruments located close to home (Coval & Moskowitz, 1999). Likewise in gambling markets, bettors are accustomed to exhibit a prejudice towards teams or participants from their same country, *ceteris paribus* (Chincarini et al., 2010). Local bookmakers overestimate winning capabilities of domestic competitors, whilst foreign bookmaking companies provide better prices for the same domestic participant. Correspondingly, Braun & Kvasnicka (2008) established two diverse national preferences, the perception bias and the loyalty bias. When domestic bettors have a perception bias towards home countries, bookmakers find it profitable to decrease odds for a home country win. A bettor’s overestimation of their national team’s ability will perceive odds to be inaccurate and cheaply priced and local punters will still wager on their national team at lower odds, providing a greater profit opportunity for bookmaking companies. On the contrary, a loyalty bias will render the betting population to either wager on their national team, or on nothing at all. Past studies found conflicting results, ranging from more favourable odds on the more popular Spanish La Liga teams (Forrest & Simmons, 2008) to finding abnormally high losses on the more popular NFL teams (Avery & Chevalier, 1999).

The over/under betting market for soccer matches is a feature of the gambling industry relating to the number of goals scored within the match. Two goals or less define a match to finish under 2.5 goals, whilst three goals or more certifies a game ending with over 2.5 goals. Studies have been published documenting the relationship between time and goals scored, where later stages of matches tend to provide more goals than during any other time bracket primarily due to fatigue, lack of tactical discipline, and lapses in concentration (Armatas et al., 2007). Throughout time there has been a shift in the number of goals scored per match at which point modern games offer a greater spectacle. Statistical data has shown that the change in regulation to award 3 points instead of 2 for wins has encouraged away teams to be more aggressive leading to a substantial increase in away goals (Dobson et al., 2001). Kain & Logan (2014) explored that the over/under market is more cumbersome for bettors to predict than on betting lines used as forecasts of sporting event outcomes, mainly due to the lack of information available regarding the sum of scores. Moreover, considering the NFL totals market, skewness was identified in the forecast errors proving inefficiency in the market (Paul & Weinbach, 2002). Paul & Weinbach (2009) revealed a bias towards high scoring events in soccer, a relatively low scoring sport. Notwithstanding this, it remains one of the most popular sport, and hence they pointed out that its attractiveness must be due to reasons other than the level of scoring.

6.2.3 Testing for Insider Knowledge

Insider knowledge is information relevant to the event available only to a limited number of the betting population (Malkiel & Fama, 1970). The general public and bookmaking companies do not have access to such material and accordingly accurate odds are problematic to achieve given the limited knowledge available to betting companies. Opportunities develop and monopolistic bettors accompanied by rare information are able to use it to their advantage and earn higher returns from odds they believe to be inaccurate.

Malkiel & Fama (1970) raised awareness of insider information when proposing to test market efficiency and whether prices *fully reflect* available information. Strong form tests analyse the ability of insider traders to achieve a positive expected return on their investments. Betting markets are one of the few financial markets where insider trading can be quantified since all information is fully understood and can be effortlessly accessible. These characteristics have enabled Coleman (2013) to explore this market and confirm that insider trading does not only exist, but is also on the rise.

It is observed that money wagered at the last few minutes of the betting period are more accurate bets than at any earlier time (Sauer, 1998). Asch et al. (1982) shed light on the

possibility that late bettors may be insider traders who prefer to bet at last instances so that others have minimal time to recognise and reproduce their selections and hence earn a larger share of the winning pool.

Crafts (1985) investigated movements of odds in racetrack betting for a four month period. Odds are recorded at their initial forecast price and at their starting price once they are returned back to the track. A profitable opportunity for people with inside information was explored since odds that shorten throughout the period experience positive expected return at an average of £0.64 for every £1 wagered. Analogous results were obtained by Bird & McCrae (1987), suggesting that prior knowledge of price movements may be adopted as a betting strategy producing positive returns. Betting companies were observed to be unable to incorporate private information in odds formation at the beginning of betting periods, which however improves at starting prices.

The risk of buyers having private information is constantly apparent to bookmakers and the mispricing of games can prove hazardous to companies. Theoretical models have been developed to delve into the optimal price that should be charged by market makers (Schnytzer & Shilony, 1995). The degree of insider trading in the 1998 Australian horserace betting season was analysed through the model and it was discovered to be between 20% and 33% for each weighting category assigned to market bets. Coleman (2013) investigated the same Australian racetrack betting market thoroughly in search of insider traders. He concluded that it is estimated that 2% of investments in the betting markets are undertaken by bettors with private information. Empirical measures have proved that insider trading has been amplified over the past decade. It is considered to be a significant concern in betting markets as these monopolies will remain present to enjoy substantial profits.

6.2.4 Arbitrage Opportunities

Arbitrage in finance relates the advantageous opportunity of trading the same security that is priced differently in separate markets (Sharpe & Alexander, 1995). The riskless prospect of buying cheap and selling dear is a fundamental feature in investigating financial markets since it propels bettors to breakdown the prices to their valuable components thus keeping the market efficient (Shleifer & Vishny, 1997). Betting markets apply a relatively low bid-offer price and arbitrage opportunities may likewise be available (Chincarini et al., 2010).

The betting market comprises of several bookmaking companies offering odds for the same set of sporting events. A market characterised with complete efficiency would provide similar odds for identical events (Pope & Peel, 1989). Yet, it seems that all

bookmakers' odds vary slightly providing punters with the opportunity to reduce one's exposure to risk and minimise the expected losses. [Chincarini et al. \(2010\)](#) take advantage of the home bias found in the market to search for possible arbitrage opportunities using two different methods. Nevertheless, neither applying the optimal weightings to a '1X2' bet nor wagering for a team and against it, if possible, resulted in any arbitrage profits for the European Cup of 2008.

Arbitrage could feature in in-play betting when odds change after new information, possibly goals or having a player sent off, is available. [Gil & Levitt \(2012\)](#) explored that within 10 to 15 minutes after the changes in price, odds do not accurately reflect the probabilities. Notwithstanding this inefficiency, arbitrage opportunities are quite rare. Statistical models dependent on timing and price volatility have been developed to facilitate arbitrage ([A. Brown, 2012](#)). In spite of these measures, the betting market often renders itself efficient and allows minimal contingencies for arbitrage. In Chapter 7, evidence of quasi-arbitrage situations are discovered during short tournaments in betting derivative markets.

6.3 Methodology

6.3.1 Data

The data available for this research was extracted from www.football-data.co.uk for some of the most popular European leagues between the 2005/06 and 2014/15 seasons. Specifically the Belgian Pro League, Dutch Eredivisie, English Premiership, English Championship, French Ligue 1, German Bundesliga, Italian Serie A, Portuguese Primeira Liga, Scottish Premiership and Spanish La Liga are considered. All these form part of the top ten European leagues as measured by the Union of European Football Associations' (UEFA) Ranking for club competitions as per first January 2016. Comparable odds data for the Russian Soccer Premier League and the Ukrainian Premier League was not available even if these two form part of the top ten leagues. Instead the English Championship (second tier) and the Scottish Premiership were included as the UK is considered the largest betting market in the continent ([Hudson, 2014](#)).

The data at hand consisted of the full-time results, the average '1X2' odds (where '1' signifies a home team win, 'X' a draw and '2' implies an away team win) and average 'Over and Under 2.5 Goals' odds (where a match is considered to end as an under if 2 goals or less are scored and over if 3 goals or more are scored). As the over and under odds are mutually exclusive, the focus is based solely on the over markets since the results

on the unders markets will be their inverse. A list of excluded matches are included in Appendix C. The data consisted of over 136,000 odds observations.

6.3.2 Determining Probabilities and Other Metrics

Like any other business, bookmakers aim to make profits. In a traditional scenario of a bookmaker setting odds, the profit is incorporated within the odds (Kuyppers, 2000) by offering odds that are lower than those implied by the actual probability (Pope & Peel, 1989).

In this respect, there are three possible interpretations of probabilities for each outcome: the probability implied by betting odds (π) which is the inverse of the European Odds, the subjective probability of an outcome (p_s) which is the implied probability but adjusted to any bookmaker margin (k) as per Equation (2.3), and the actual objective probability of an outcome (p_o). For any range of bets, the objective probability can be estimated to be the number of actual occurrences as a proportion of the number of observations made.

The expected profit of a bookmaker is defined as per Proposition 2.3⁵. The measure of profitability will be also measured by counting the amount of percentage profit if a unit bet is placed on each observed odds. This can be defined as $\frac{\sum(\pi^{-1}-1)}{n}$ where I is a Boolean variable indicating whether the item occurred and n is the number of observations made within the range.

A simplistic interpretation of the Brier Score as a measure of accuracy of a particular prediction. The Brier Score is the mean square difference between the subjective probability and actual outcome. In our scenario for any range of subjective probabilities, the Brier Score is evaluated as $\frac{\sum(p_s-I)^2}{n}$. This metric can range from 0 in case of an exact prediction to 2. For example if three matches were predicted to be a home win with subjective probabilities 0.22, 0.23 and 0.24 while only the last one ended with the home team winning, the Brier Score for this range of bets is 0.6789.

6.3.3 Subdividing the ranges

Given the significant number of data points available, data points needed to be grouped by different ranges of subjective probability. Twenty ranges were considered as ideal since a smaller number of ranges would result into small sample sizes while larger ranges may not show any trends. If the ranges were all equal, that is subjective probabilities

⁵Also described by (Newall, 2015).

of ranges of 5%, some ranges would have significantly high sample sizes (for example there are 31,045 odds in the ranges of 25% to 30%) as well as significantly small values (no subjective probabilities above 95%). At a total of 136,321 odds observed, the ideal would be to have 6,817 odds per range. We use an iterative procedure such that each range will be between half and twice this value, that way no range will have more than four times the sample size of any other range.

The iterative sequence used is as follows:

1. Subdivide into 20 equal ranges of subjective probabilities.
2. Find the range with the largest number of odds observations.
 - If this is more than 13,633, subdivide the largest range in two by taking the mid-point of the subjective probability range and merge the smallest range with the nearest smallest range⁶.
 - If the largest range has 13,633 observations or less, find the smallest range. If this has less than 3,408 observations then merge this range to the smallest adjacent range and subdivide the largest range in two by taking the mid-point of the subjective probability range.
3. The previous step is repeated until all odds ranges have between 3,408 and 13,633 total odds observations.

The final ranges of subjective probabilities produced together with the number of observations are shown in Table 6.1. This process does have its limitations in that not all outcomes have equal dispersion. For example the subjective probabilities for draws ($\mu_{ps} = 0.265, \sigma_{ps} = 0.040$) and overs ($\mu_{ps} = 0.497, \sigma_{ps} = 0.069$) are less dispersed than home ($\mu_{ps} = 0.447, \sigma_{ps} = 0.157$) and away ($\mu_{ps} = 0.288, \sigma_{ps} = 0.139$). Nonetheless using the same odds ranges would make comparisons much easier.

6.3.4 Correlations

The hypotheses being tested are measured by three metrics. The accuracy of a prediction is measured by the Brier Score, the probability of an outcome by the observed probability while the profitability by the percentage profits. A higher Brier Score represents a lower amount of accuracy and hence a positive correlation in results indicates a negative correlation between bettors profits and accuracy (or alternatively a positive correlation

⁶For example, in the first iteration the largest range was the 25% to 30% range with 31,045 observations. This was split into a 25% – 27.5% and 27.5% – 30% range while the smallest range of 95% – 100% was added to the adjacent range 90% – 95%

TABLE 6.1: Number of Observed Odds

p_s	Total	Home	Draw	Away	Over 2.5
$0\% \leq p_s < 15\%$	7,374	1,137	816	5,421	
$15\% \leq p_s < 20\%$	7,179	1,131	2,151	3,897	
$20\% \leq p_s < 25\%$	10,732	1,323	4,489	4,920	
$25\% \leq p_s < 26\frac{1}{4}\%$	4,354	381	2,522	1,451	
$26\frac{1}{4}\% \leq p_s < 27\frac{1}{2}\%$	6,721	427	4,746	1,548	
$27\frac{1}{2}\% \leq p_s < 28\frac{1}{8}\%$	5,433	215	4,496	722	
$28\frac{1}{8}\% \leq p_s < 28\frac{3}{4}\%$	7,497	248	6,553	696	
$28\frac{3}{4}\% \leq p_s < 30\%$	7,040	544	5,135	1,361	
$30\% \leq p_s < 32\frac{1}{2}\%$	7,147	1,333	3,040	2,774	
$32\frac{1}{2}\% \leq p_s < 35\%$	4,049	1,558	60	2,431	
$35\% \leq p_s < 40\%$	9,320	4,826	37	2,927	1,530
$40\% \leq p_s < 42\frac{1}{2}\%$	6,597	2,570	9	1,044	2,974
$42\frac{1}{2}\% \leq p_s < 45\%$	8,291	2,638	10	817	4,826
$45\% \leq p_s < 47\frac{1}{2}\%$	8,615	2,350	4	611	5,650
$47\frac{1}{2}\% \leq p_s < 50\%$	7,304	1,995	4	538	4,767
$50\% \leq p_s < 52\frac{1}{2}\%$	5,868	1,808	3	482	3,575
$52\frac{1}{2}\% \leq p_s < 55\%$	5,384	1,636	3	444	3,301
$55\% \leq p_s < 60\%$	7,863	2,419	4	774	4,666
$60\% \leq p_s < 65\%$	4,098	1,675	3	549	1,871
$65\% \leq p_s < 100\%$	5,455	3,872	1	679	903
Total	136,321	34,086	34,086	34,086	34,063

between bookmaker profits and accuracy). We use two-tailed hypothesis tests for each relationship and also consider each league separately. In order to limit the effects of ranges with a small number of observations, a two-tailed hypothesis test on the correlation was applied twice: on all ranges and on ranges of at least 50 observations.

6.4 Analysis and Results

6.4.1 General

Figure 6.1 shows the subjective and objective probabilities for all odds, and the four outcomes (Home Team Win, Draw, Away Team Win, Over 2.5 goals). In an efficient market, subjective and objective probabilities would be equal or not having any particular discrepancies and hence would fit around a 45 degree line passing through the origin. The graph demonstrates a trend of high odds with low probabilities are characterised by subjective probabilities that is slightly higher than objectives probabilities. On the contrary, low odds signifying higher probabilities feature implied probabilities that are lower than objective probabilities. This means that bettors tend to overestimate bets with low probability and high returns [Gandar et al. \(2001\)](#). Conversely, punters underestimate favourites and wager less money on high probability events.

TABLE 6.2: Bettor Losses for Unit Bets per Odds and Brier Scores

Subjective Probability	Total (incl Overs)			Home			Draw			Away		
	Pr/Lss	Exp P/L	Brier	Pr/Lss	Exp P/L	Brier	Pr/Lss	Exp P/L	Brier	Pr/Lss	Exp P/L	Brier
$0\% \leq p_s < 15\%$	-25.3%	-7.3%	0.082	-8%	-7.2%	0.103	-35.3%	-6.7%	0.084	-27.5%	-7.4%	0.078
$15\% \leq p_s < 20\%$	-13.1%	-7.5%	0.138	-10.8%	-7.5%	0.14	-21.2%	-7.3%	0.128	-9.3%	-7.6%	0.143
$20\% \leq p_s < 25\%$	-12.7%	-7.4%	0.168	-9.9%	-7.4%	0.171	-12.6%	-7.2%	0.169	-13.5%	-7.6%	0.167
$25\% \leq p_s < 26\frac{1}{4}\%$	-14.4%	-7.4%	0.181	-5.2%	-7.3%	0.194	-15.8%	-7.4%	0.18	-14.3%	-7.4%	0.181
$26\frac{1}{4}\% \leq p_s < 27\frac{1}{2}\%$	-11%	-7.5%	0.192	-6.3%	-7.3%	0.198	-12%	-7.6%	0.191	-9.5%	-7.5%	0.194
$27\frac{1}{2}\% \leq p_s < 28\frac{1}{4}\%$	-9.7%	-7.7%	0.198	-8.3%	-7.1%	0.199	-8.4%	-7.8%	0.2	-18.1%	-7.5%	0.187
$28\frac{1}{4}\% \leq p_s < 28\frac{3}{8}\%$	-6.5%	-7.6%	0.205	1.4%	-7.3%	0.215	-6.4%	-7.7%	0.205	-10.3%	-7.5%	0.2
$28\frac{3}{8}\% \leq p_s < 30\%$	-5.6%	-7.2%	0.209	-18.3%	-7.2%	0.193	-2.3%	-7.1%	0.213	-13.1%	-7.5%	0.2
$30\% \leq p_s < 32\frac{1}{2}\%$	-10.5%	-7.7%	0.211	-8.5%	-7.4%	0.214	-7.1%	-7.9%	0.214	-15.2%	-7.6%	0.205
$32\frac{1}{2}\% \leq p_s < 35\%$	-7.8%	-7.6%	0.223	-8.6%	-7.5%	0.222	-3.5%*	-8.6%	0.229	-7.3%	-7.7%	0.224
$35\% \leq p_s < 40\%$	-7.4%	-7.5%	0.235	-7.3%	-7.6%	0.235	-14.7%**	-8.4%	0.227	-6%	-7.5%	0.235
$40\% \leq p_s < 42\frac{1}{2}\%$	-8.8%	-7.3%	0.241	-5.5%	-7.6%	0.244	-27.2%***	-9.5%	0.228	-3.1%	-7.4%	0.245
$42\frac{1}{2}\% \leq p_s < 45\%$	-7.8%	-7.1%	0.246	-7.4%	-7.5%	0.246	4.7%**	-9.2%	0.254	-2.9%	-7.2%	0.248
$45\% \leq p_s < 47\frac{1}{2}\%$	-6.9%	-6.9%	0.248	-4.2%	-7.5%	0.25	-2.0%**	-8.7%	0.249	-6.6%	-7.3%	0.249
$47\frac{1}{2}\% \leq p_s < 50\%$	-5.2%	-6.8%	0.25	-8.1%	-7.6%	0.25	-6.5%***	-10%	0.249	-2.7%	-7.4%	0.25
$50\% \leq p_s < 52\frac{1}{2}\%$	-4.5%	-6.7%	0.25	-3.8%	-7.6%	0.249	-41.7%***	-9.4%	0.253	-4.2%	-7.6%	0.25
$52\frac{1}{2}\% \leq p_s < 55\%$	-6.6%	-6.9%	0.249	-3.7%	-7.6%	0.247	70.0%***	-9.3%	0.217	-1.8%	-7.5%	0.246
$55\% \leq p_s < 60\%$	-6.1%	-7.2%	0.244	-4.2%	-7.6%	0.242	-21.8%***	-9.9%	0.248	-5.7%	-7.5%	0.242
$60\% \leq p_s < 65\%$	-5.4%	-7.2%	0.231	-4.8%	-7.4%	0.229	44.0%***	-9.9%	0.14	-3.6%	-7.3%	0.229
$65\% \leq p_s < 100\%$	-2.5%	-7.1%	0.18	-1.4%	-7.3%	0.171	29.0%***	-14.4%	0.113	-4.3%	-6.9%	0.196
All Odds	-9.1%	-7.3%	0.208	-6.1%	-7.5%	0.219	-10%	-7.5%	0.19	-12.5%	-7.5%	0.18

* ≤ 100 Observations, ** ≤ 50 observations, *** ≤ 10 observations

All values for 'Over 2.5 Goals' odds were based on ranges with more than 50 observations as per Table 6.1. The loss for all odds was 7.7% while the expected profit 6.7% while the Brier score was 0.244. Values can be deduced from Tables 6.1 and 6.2.

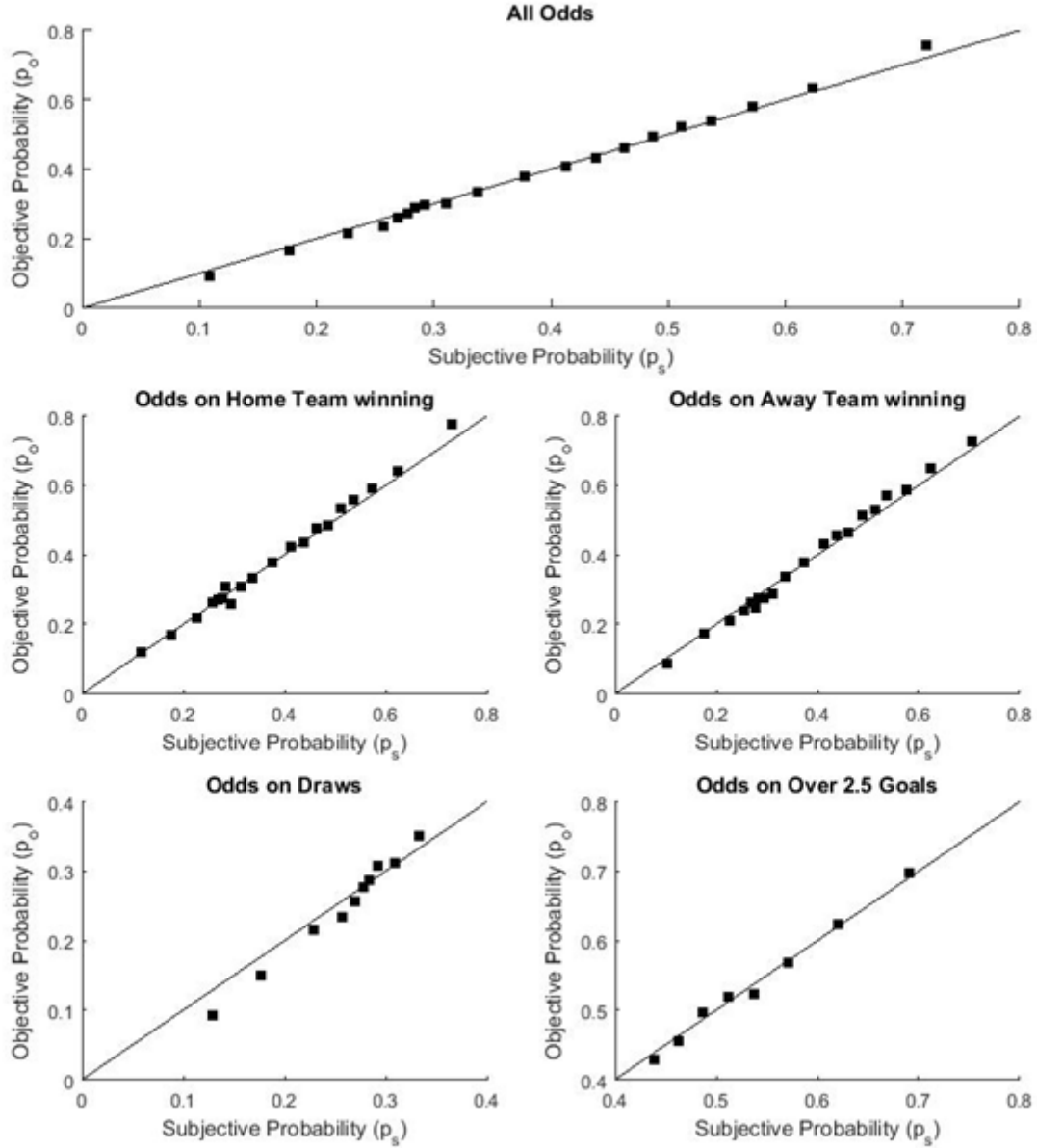
FIGURE 6.1: Relationship between p_s and p_o .

Table 6.2, showing the actual and expected losses as defined in Chapter 2 and the Brier Score for different ranges and different outcomes, implies that our Hypothesis (1) maybe true⁷. Although the expected loss, determined by the betting margin, is somewhat similar for all subjective probability ranges, the losses made in lower probability outcomes are higher than higher probability outcomes, implying a longshot bias. This is corroborated by the majority of significant correlations between profits and observed probabilities

⁷Note that results are similar to Deschamps & Gergaud (2012) as losses on draws are the lowest and a more optimistic value than the expected loss.

TABLE 6.3: Correlation between Profits and Observed Proportions (p_o)

		Overall	Home	Draw	Away	Over
All Leagues	All >50	0.778*	0.474**	0.529** 0.946*	0.773*	0.594*** 0.563***
Belgian Pro League	All >50	0.491**	0.143 0.630**	-0.371 0.618	0.383*** 0.219	-0.619*** -0.254
Dutch Eredivisie	All >50	0.447**	-0.181 -0.105	-0.131 0.476	0.522** 0.520**	-0.483 0.081
English Premiership	All >50	0.691*	-0.118 0.109	0.780** 0.935*		-0.248 0.215
English Championship	All >50	0.154	0.219 0.375	0.830* -0.272	0.524** 0.133	0.163 0.757**
French Ligue 1	All >50	0.579*	0.141 0.283	0.648** -0.025	0.509** 0.573**	0.607** 0.266
German Bundesliga	All >50	0.107	-0.039 -0.237	0.870* 0.848*	-0.15 -0.033	0.555*** 0.41
Italian Serie A	All >50	0.662*	0.738* 0.727*	0.454** 0.135	0.652* 0.615*	-0.147 -0.052
Portuguese Prim. Liga	All >50	0.731*	0.586* 0.814*	0.886*	0.583* 0.624**	0.429 0.323
Scottish Premier	All >50	0.430***	-0.101 0.456***	0.467 0.566	0.032 0.323	-0.22 0.149
Spanish La Liga	All >50	0.792*	0.061 0.206	0.502 0.908*	0.284 0.102	0.415 0.331

* significant at $p \leq 0.01$, ** significant at $p \leq 0.05$, *** significant at $p \leq 0.1$
All values proven significant using a two-tailed test.

being positive as shown in Table 6.3. This means that the higher probability events result in a higher profit (lower loss) - the definition of the longshot bias.

All significant correlations are between Brier Scores and Profits (Table 6.4) are positive. This implies that less accurate odds may lead to more profits (lower losses) by bettors. The results for our third hypothesis, shown in Table 6.5, also imply that the Brier Score is higher for higher probability events. In layman terms, the accuracy implied by betting markets of unlikely events is better than the accuracy of likely events. This does not prove our third hypothesis, indeed there is evidence of a contradictory hypothesis.

6.4.2 Conflicting results in Overs Market

The Overs market, showing the odds offered on over 2.5 goals in a match, is characterised by a short deviation in subjective probabilities. This has resulted in some conflicting results.

TABLE 6.4: Correlation between Brier Scores and Profits

		Overall	Home	Draw	Away	Over
All Leagues	All >50	0.808*	0.301	-0.095 0.974*	0.851*	0.821* 0.067
Belgian Pro League	All >50	0.520**	0.086 0.256	0.764** 0.865**	0.494** 0.494***	0.798* -0.138
Dutch Eredivisie	All >50	0.662*	0.429*** -0.058	0.829* 0.692***	0.760* 0.772*	0.508 -0.351
English Premiership	All >50	0.723*	0.233 0.008	0.899* 0.968*	0.641*	0.584*** 0.750**
English Championship	All >50	0.333	0.808* 0.673*	0.969* 0.479	-0.138 0.316	0.516 0.853**
French Ligue 1	All >50	0.690*	0.396*** 0.618*	0.908* 0.403	0.684* 0.724*	0.606** 0.792**
German Bundesliga	All >50	0.072	-0.019 -0.323	0.941* 0.933*	0.185 0.335	0.673** -0.554
Italian Serie A	All >50	0.785*	0.842* 0.839*	-0.172 0.618***	0.701* 0.752*	0.008 -0.111
Portuguese Prim. Liga	All >50	0.811*	0.757* 0.807*	0.940*	0.611* 0.740*	0.896* 0.051
Scottish Premiership	All >50	0.566*	0.544** 0.362	0.972* 0.839*	0.432*** 0.557**	-0.447 -0.17
Spanish La Liga	All >50	0.674*	0.259 0.182	0.915* 0.965*	0.406*** 0.211	-0.027 -0.136

* significant at $p \leq 0.01$, ** significant at $p \leq 0.05$, *** significant at $p \leq 0.1$
 All values proven significant using a two-tailed test.

The first conflicting result is the evidence of a negative correlation between profits and observed proportions in the Belgian league. On further investigation, this is characterised by small observations in the 35% to 42.5% range. Indeed there were only eight observations, with six of these matches ending up with a total of three or more goals. This would have actually ended up in a profit being made if wagering on over 2.5 goals. We envisage this might be a blimp due to the small number of observations made, especially since there is no similar trend in other leagues and no evidence of the Belgian Pro League having any different propensity of scoring than other leagues.

In comparing the accuracy of results measured as the Brier Score with the observed proportion, the Belgian, Dutch, English (Premiership), German and Spanish leagues showed evidence of a negative correlation. This indicates that bookmakers tend to be less accurate in predicting unlikely outcomes than likely ones. Some of the difficulty lies in the fact that these five leagues have the higher standard deviation of goals scored per match from the ten leagues sampled as shown in Table 6.6.

TABLE 6.5: Correlation between Brier Scores and Observed Proportions (p_o)

		Overall	Home	Draw	Away	Over
All Leagues	All >50	0.653*	0.590*	0.284 0.993*	0.716*	0.163 -0.522
Belgian Pro League	All >50	0.625*	0.554** 0.382	0.314 0.928*	0.725* 0.895*	-0.776* -0.38
Dutch Eredivisie	All >50	0.686*	0.392*** 0.544**	0.445 0.964*	0.764* 0.757*	-0.928* -0.903**
English Premiership	All >50	0.645*	0.457** 0.566**	0.975* 0.993*	0.678*	-0.676** -0.115
English Championship	All >50	0.795*	0.637* 0.638**	0.940* 0.711	0.554** 0.925*	0.735** 0.569
French Ligue 1	All >50	0.709*	0.685* 0.586**	0.897* 0.903*	0.816* 0.953*	0.096 0.717**
German Bundesliga	All >50	0.666*	0.632* 0.592**	0.984* 0.982*	0.634* 0.578**	0.025 -0.891*
Italian Serie A	All >50	0.671*	0.653* 0.639*	0.256 0.860*	0.691* 0.889*	-0.503 -0.229
Portuguese Prim. Liga	All >50	0.591*	0.565* 0.581**	0.990*	0.659* 0.619**	0.622** 0.374
Scottish Premiership	All >50	0.645*	0.390*** 0.611**	0.661*** 0.923*	0.560** 0.560**	0.764** 0.303
Spanish La Liga	All >50	0.599*	0.428*** 0.458***	0.790* 0.986*	0.707* 0.665*	-0.415 -0.700**

* significant at $p \leq 0.01$, ** significant at $p \leq 0.05$, *** significant at $p \leq 0.1$
All values proven significant using a two-tailed test.

TABLE 6.6: Goals Scored per match

League	Goals/match
Belgian Pro League	2.72(± 1.65)
Dutch Eredivisie	2.87(± 1.71)
English Prem	2.66(± 1.67)
English Cham	2.59(± 1.60)
French Ligue 1	2.37(± 1.54)
German Bundesliga	2.61(± 1.64)
Italian Serie A	3.08(± 1.77)
Portuguese Prim. Liga	2.43(± 1.58)
Scottish Prem	2.62(± 1.64)
Spanish La Liga	2.70(± 1.70)
All Leagues	2.65(± 1.66)

6.5 Conclusion

The main objective of this study was to investigate the European betting market in order to examine the longshot bias and how this is related to accuracy of the odds on offer. We found evidence to sustain all our first two hypotheses: bettors make higher losses on highly unlikely events and accurate odds.

Yet the results for the third hypothesis are most likely due to the limitations of the Brier Score itself as it tends to overcompensate for highly likely outcomes. For example if someone was predicted to have a 1% probability but occurs 2% of the time would result in an expected Brier Score of 0.0197. On the other hand an outcome with a subjective probability of 90% but an actual occurrence of 95%, would have an expected Brier Score of 0.5. The latter is a higher value even if the former represents a mis-estimation of 100%.

The implications on the rationale why bettors are biased can be explained by the risk seeking nature of a typical bettor. Indeed a typical bettor is characterised to consider more the amount to be won rather than the likelihood of winning. Bookmakers are very well aware of this and possibly explains the recent push to sell more accumulators, that is one bet placed on a series of outcomes that result in an extremely high payout that is very unlikely. We believe that further research examining the risk profile of bettors, possibly by examining their selections versus certainty equivalents, may give further insights on the existence of the longshot bias.

An implication on bookmakers is also derived from our second hypothesis. As Brier scores are correlated to bettors' profitability, this translates to bookmakers making higher profits when odds are more accurate. Chapter 2 states that the exact probability of an outcome may not be necessary known by a bookmaker to achieve expected positive returns. Here this is extended: although not strictly necessary to guarantee profit, our results indicate that greater accuracy implies higher bookmaking profits.

Chapter 7

Profiting on inefficiencies in betting derivative markets: The case of UEFA EURO 2012

7.1 Introduction

In Chapter 4, a descriptive evaluation of how odds on outcomes related to Brazil's performance changed during the World Cup and how these were not necessarily correct was provided. This was extended in Chapter 6 where an evaluation on over 136,000 odds showed that there is evidence of inefficiencies in the betting market. This chapter takes it step further by investigating an approach that capitalises on such inefficiencies. The procedure used here not only pinpoints the most likely odds that are incorrectly priced but also proposes an investment approach that modifies the Kelly criterion in order to determine the percentage of own funds to be invested. The majority of this chapter is adapted from a joint paper ([Cortis et al., 2013](#)) published in 2013.

Association football has a significant following throughout the world and is widely considered to be the most popular team sport in most countries. It is therefore not surprising that major association football tournaments are some of the most popular sports betting markets. This empirical study investigates whether it is possible to profit from market inefficiencies on betting exchanges during short tournaments. Specifically, this study will focus exclusively on the UEFA European national football team tournament held over 24 days in mid-2012 (Euro 2012).

Typically betting companies offer odds to clients payable on the occurrence of a predefined event. A betting exchange operates differently from a betting company since it does

not take on the risk and additionally, it does not price odds. As discussed in Chapter 4, betting exchanges function very much like financial markets as bettors can ‘back’ (buy / bet on) an outcome or ‘lay’ (sell / bet against) the particular outcome. Contrary to the odds offered by betting companies, the odds on betting exchanges directly reflect the betting demand and supply. Betting exchanges offer a significant advantage by generally providing higher odds (Griffits, 2005; P. Jones et al., 2004; Koning & van Velzen, 2009). One must note that betting exchanges need to be liquid to offer these advantages to a bettor over traditional bookmakers. Furthermore the charge on winnings made by betting exchanges may severely diminish any advantage over the odds offered by betting companies. However betting exchanges tend to be very liquid, particularly in popular markets.

Such popular markets¹ on betting exchanges, specifically Betfair, are characterized by a large trading volume and odds that are easily accessible by the public and can be ‘backed’ (bought) or ‘laid’ (sold) without having a major effect on the odds. Effectively, this makes such markets deep, liquid and transparent. Moreover, as Law & Peel (2002) put it, gambling markets have a substantial advantage over financial markets as they generate the final outcome and rate of return within a shorter time frame.

Sports betting markets have been mainly analyzed in two different ways. For instance, Lessmann et al. (2009) use the output from a model and compare it with market odds. On the other hand, other studies focused on analyzing the market odds directly. For example, Vlastakis et al. (2009) found cases of arbitrage in a sample of odds from five bookmakers while Egon et al. (2009) also found such cases when using a sample of odds from eight bookmakers and one betting exchange. Similarly in Chapter 6 we found evidence of the longshot bias by directly considering odds.

The model proposed in this paper is an innovative combination of the two approaches - it uses odds from particular betting markets as the sole input for the model, applies a noise parameter and creates odds for other related markets. In cases where the model output is lower than the market odds, there is scope to wager against the market. However the amount of the wager is a contentious issue, especially due to the dependent structure of the markets being analyzed. Therefore, a second important decision on how to determine the amount of wager is made.

The aim of this model is to firstly determine how the amount of noise, which can be interpreted as the market’s knowledge, affects the odds on other related markets and whether there is scope to take advantage of possible pricing errors. Furthermore, by

¹Popular markets relate mainly to popular football events such as the World Cup, Euro Cup, UEFA Champions League and the English Premier League.

applying a noise parameter of zero, the model seeks to investigate possible quasi-arbitrage opportunities during short tournaments.

The model is applied to the UEFA European national football team tournament held over 24 days in mid-2012 (Euro 2012). This tournament (European Cup) is held every four years and, since 1996, a total of 16 European national teams participate in the final tournament. This tournament has provided a number of surprises in past, most notably when Denmark and Greece won the tournament in 1992 and 2004 respectively. Due to the short-term nature of the tournament and the time lag between every game played by national teams during the two year pre-tournament qualifying period (less than two games per month), evaluating odds for such tournaments is considered harder than that for club tournaments.

This chapter is subdivided into five sections. After this introduction, Section 7.2 will describe briefly the sports betting markets being examined in this chapter. The model, which includes the algorithm creating odds and the wager allocation criteria, is described in Section 7.3. The final two sections present the results of the model and discuss the implications of these findings.

7.2 Betting Markets

The format of EURO 2012 described here is similar to the World Cup format described in Chapter 4. The first stage of the European Cup, referred to as the group stage, consists of four groups of four teams. During this stage, all teams within each group play each other once. The two highest ranked teams qualify for the second stage of the tournament.

The betting markets of interest for this paper during the group stage are '1X2', 'Straight Forecast', 'Dual Forecast', 'Group Winner', 'To Qualify' and 'Rock Bottom' markets. The '1X2' market refers to the outcome of each game; where 'X' refers to the likelihood of a draw while '1' and '2' refer to the likelihood of the first or second team winning respectively.

The 'Straight Forecast' market refers to all possible permutations of the first two teams in the group after all games have been played. This is very much related to the 'Dual Forecast' market which relates to all possible combinations of the first two teams at the end of the group stage. For example, a wager on England/France in the 'Straight Forecast' market would result in a win if England win the group while France are ranked second. However it would result in a win in the 'Dual Forecast' market if England and France are ranked top two irrespective of whether France or England top the group. As the name implies, the 'Group Winner', 'To Qualify' and 'Rock Bottom' markets represent

TABLE 7.1: Number of Outcomes for Each Market

Market	Rock Bottom	Straight Forecast	Dual Forecast	Group Winner	To Qualify
No. of Outcomes	4	12	6	4	4

the likelihoods of each team being ranked first, in the top two or last after all games have been played respectively. The number of outcomes for each market are shown in Table 7.1.

It is quite apparent that the likelihoods for the ‘Straight Forecast’, ‘Dual Forecast’, ‘Group Winner’, ‘To Qualify’ and ‘Rock Bottom’ markets are related to each other and are dependent on the ‘1X2’ markets. In this respect, these five markets will be referred to as derivative markets.

7.3 Method

The aim of this chapter is to assess the effect of small changes in the probabilities of the ‘1X2’ markets on the derivative markets as described in Section 7.2. This is assessed in a two distinct processes: an algorithm that generates a simulation of 100,000 paths for the probabilities of each derivative market; then a comparison of these results with the market odds and placing of bets according to specific criteria when there is a mismatch.

Each group within the tournament consists of four teams, with each team playing each other team once resulting in a total of six games played per group. Teams play on the same day at different times except for the last two games that are played at the same time to avoid match fixing. Mismatches between the algorithm’s outcomes and the derivative markets were examined at five time intervals for each group: prior to the group games start and after each of the first four matches. Any market odds referred to were collected from Betfair between games.

7.3.1 The Algorithm

The data required for the algorithm are the scores for any group matches already played and the ‘1X2’ market odds for all remaining matches. ‘1X2’ Market odds for outstanding games are transposed into probabilities and adjusted proportionally in order for these to add up to one. A noise parameter, representing the error in market odds, is applied to these probabilities in order to obtain adjusted probabilities. Finally the outcome of

unplayed matches is simulated from these adjusted probabilities. This process is repeated in order to obtain 100,000 paths and can be summarized in the steps shown below².

- 1) For each outcome probability, create a new adjusted outcome probability

$$p_{i,k}^{(2)} = \frac{p_{i,k}(1 + F_{i,k})}{\sum_{i=w,d,l} p_{i,k}(1 + F_{i,k})} \quad (7.1)$$

where $p_{W,k}$, $p_{D,k}$, $p_{L,k}$ represent the probability of the k^{th} match ending with a win, draw or loss for the first team respectively such that $p_{W,k} + p_{D,k} + p_{L,k} = 1$; $i \in (W, D, L)$ and $F_{i,k} \sim N(0, \alpha)$.

- 2) Use $p_{i,k}^{(2)}$ to simulate the outcome (W,D,L) for all remaining k matches.
- 3) Simulate the goal difference for all remaining matches. This is equal to zero in case of a draw but assume it follows a Weibull Distribution with parameters 1.30 and 1.47 if otherwise.
- 4) Simulate the number of goals scored.
 - (a) In case of a draw, use a Weibull Distribution with parameters 1.64 and 1.75.
 - (b) In case of a win by one goal, use a Weibull Distribution with parameters 1.42 and 1.76 to find the number of goals scored by the winning team.
 - (c) In case of a win by two or three goals, use a Weibull Distribution with parameters 0.86 and 1.60 or 0.83 and 1.60 respectively less one goal to find the number of goals scored by the losing team.
 - (d) In case of a win by more than three goals, assume that the loser did not score any goals.

For each simulated scenario, the predicted results are aggregated to the known results in order to derive the probabilities for the derivative markets. The final standing of a team is dependent on the number of points earned (three points for each win, one point for each draw, no points for each loss). However in case of a tie, the final standings are determined using the following criteria in the given order:

²All the parameter values have been extracted as the best fit from group matches of past UEFA European national football team tournaments since 1996. Euro 1996 was the first such European national football team tournament to award three points for each win at group stage. The order of the generation is to simulate the goal difference and then the number of goals scored as the former is given priority over the latter when deciding the final position of a team. Appendix D shows an audit trial comparing the fit of the Weibull Distribution to the more popular Poisson Distribution.

- a) higher number of points obtained in the matches among the teams in question;*
- b) superior goal difference in the matches among the teams in question (if more than two teams finish equal on points);*
- c) higher number of goals scored in the matches among the teams in question (if more than two teams finish equal on points);*
- d) superior goal difference in all the group matches;*
- e) higher number of goals scored in all the group matches;*
- f) position in the UEFA national team coefficient ranking system;*
- g) fair play conduct of the teams (final tournament);*
- h) drawing of lots.*

UEFA (2010)

The UEFA national team coefficient ranking system is mainly used for the allocation of teams in groups such that each group contains teams with varying levels of proficiency. The points earned by each team are a result of their performance in the preceding three major tournaments (European Cup and World Cup). However, the calculation of this coefficient is beyond the scope of this paper. It suffices to say that no two teams participating in this tournament had the same number of points under this system, resulting in no need to use fair play conduct or drawing of lots as a final means in deciding the final group position.

7.3.2 Comparing Output with Market

Once the probabilities for the derivative markets are evaluated, these can be compared to the market odds observed on the Betfair betting exchange. If the algorithm is predicting correct probabilities, then it would be profitable to place bets when the market odds are higher than predicted and lay bets if otherwise. This paper takes the former approach by applying an adjusted Kelly Strategy on such favourable bets. The Kelly criterion (Kelly, 1956) sets the percentage of own funds placed on favourable bets that would maximize the logarithm of own funds. This criterion has been used in numerous scenarios (Piotrowski & Schroeder, 2007) such as financial markets (Rotando & Thorp, 1992; Thorp, 1997), card games such as blackjack (Thorp, 1997), and betting on outcomes (Lessmann et al., 2009; Thorp, 1997).

The Kelly criterion is ideally utilized in a series of independent favourable bets. Yet, the derivative markets explained earlier in Section 7.2 are not independent. For example, if the market odds for the ‘Straight Forecast’ are providing favourable bets for Czech Republic winning Group A followed by Greece in second position, it is very likely the

‘Group Winner’ market is also providing favourable bets for Czech Republic winning the group. Furthermore, if a team’s market odds are the result of speculation or lower following, there would be favourable bets for a certain outcome consistently after each game. Therefore, it would be naïve to apply the Kelly criterion blindly. Instead the betting strategy will apply two filters; firstly depending on time of wager; secondly, subsidiary outcomes. A number of assumptions are made:

- an initial fund of \$100,000 for each group,
- no transaction charges, and
- the possibility of wagers to the nearest infinitesimally factor.

At each time interval, the Kelly criterion is evaluated for cases where the predicted odds are lower than the market odds. However the percentage of the fund wagered on this market is adjusted by deducting this percentage from the adjusted Kelly criterion that imply the same condition at the same or previous time, subject to a minimum adjusted Kelly criterion of zero.

For example, if prior to the start of the games ($t = 0$), the Kelly criterion is 5% for wagering on Czech Republic/Greece in the ‘Straight Forecast’ market and 8% for wagering on Czech Republic winning the group; the adjusted Kelly criterion for wagering on the Czech Republic winning the group is 3%, assuming no further possibilities for favourable bets. Furthermore, if at the second time interval Kelly criteria of 4% and 17% are evaluated for the two outcomes in the given order, the adjusted Kelly criterion is 0% for wagering on the ‘Straight Forecast’ result, since 5% have already been wagered on this outcome; and 9% ($17\% - 5\% - 3\%$) on the ‘Group Winner’ outcome assuming no further possibilities for favourable bets. Therefore at $t = 1$ an additional \$9,000 is placed on Czech Republic winning Group A.

In this respect, we can assume that the Kelly criterion for the ‘Straight Forecast’ market is given priority over the same result for the ‘Dual Forecast’ and ‘Group Winner’ markets, while these two are given priority over the ‘To Qualify’ market. In other words, the ‘To Qualify’ market (eg. Czech Republic to qualify) is considered as a subsidiary market to the ‘Group Winner’ market (eg. Czech Republic to win group) and the ‘Dual Forecast’ market (eg. Czech Republic and Greece to qualify) and all these three are considered subsidiary to the ‘Straight Forecast’ markets (eg. Czech Republic first and Greece second). This method applied to Group D is shown in Appendix E.

7.4 Results

Model odds have been created and compared to market odds as described in previous section for four values of α . This value determines the standard deviation of noise in creating the adjusted outcome probabilities. At $\alpha = 0$, the model is assuming no noise. In this scenario, one would expect that the odds in the derivative markets would closely match the model odds and hence no bets against the market to be made. However, the results shown in Table 7.2 suggest otherwise, implying the existence of arbitrage. Furthermore, there are no significant changes in profit levels made by applying different noise parameters.

TABLE 7.2: Used Funds and Profits Made (in thousands)

	k	0	4	8	20	Average
Group A	# of bets	21	20	21	20	20.50
	Used Funds	\$44.80	\$46.59	\$44.33	\$45.68	\$45.35
	Profit(Loss)	\$49.04	\$41.66	\$47.99	\$47.23	\$46.48
Group B	# of bets	24	25	26	25	25.00
	Used Funds	\$49.61	\$49.06	\$49.64	\$49.14	\$49.37
	Profit(Loss)	\$47.97	\$46.27	\$49.99	\$48.91	\$48.28
Group C	# of bets	29	28	28	30	28.75
	Used Funds	\$32.16	\$33.57	\$32.05	\$33.43	\$32.80
	Profit(Loss)	(\$16.69)	(\$16.22)	(\$15.65)	(\$16.55)	(\$16.28)
Group D	# of bets	20	22	22	23	21.75
	Used Funds	\$52.58	\$54.00	\$53.23	\$54.86	\$53.67
	Profit(Loss)	(\$27.59)	(\$28.47)	(\$28.36)	(\$27.62)	(\$28.01)
Total	Used Funds	\$179.16	\$183.22	\$179.26	\$183.11	\$181.19
	Profit(Loss)	\$52.72	\$43.25	\$53.97	\$51.96	\$50.48
Profit	as a % of allocation	13.18%	10.81%	13.49%	12.99%	12.62%
	as a % of usage	29.43%	23.60%	30.11%	28.38%	27.86%

The allocated funds for each group amounted to \$100,000. However the maximum used funds in a group amounted to \$54,865 and the total used funds is less than half of the allocated funds. The total profit made amounted to an average of 12.62% of total allocated funds and 27.86% of used funds.

7.5 Conclusion

In this chapter, the efficiency of different related markets and the effect of participants knowledge have been examined by using the odds of base markets as a prediction for derivative markets. The results obtained suggest the possibility of contradicting views.

In general, the mismatches between the market and the model odds were mostly in cases of extremely unlikely events at the earlier time points and extremely likely events at the last time point. The winner of the Euro 2012 was not a surprise as favourites Spain won, beating Italy 4-1 in the final. However, not all teams performed as expected. In Group A, the two teams that passed to the next round were treated as the least likely pair to make it to the next round while bettors had high hopes on The Netherlands but this team obtained the last position in Group B. On the other hand, there were no significant surprises in Groups C and D.

Traditional literature indicated that odds on unlikely events tend to be under-priced when compared to likely events (Cain et al., 2003; Snowberg & Wolfers, 2010). On the other hand, the results of this paper support the possibility of higher odds for unlikely events for derivative markets at the earlier stages of the tournament. Other than incorrect pricing, a possible explanation to this is that the market is expecting a dynamic rather than stochastic process. For example, the probability of Spain losing all three games is being treated as less than the multiplication of a loss in each of the single games, because in case of an initial loss, the market is expecting Spain to try harder in the second one.

The simple model described in this paper can be used to actively seek miss-pricing opportunities as well as the risk faced by a betting company on accepting certain bets. The adjusted Kelly strategy proposed is easy to implement, and can be applied for events and markets that are dependent. Further research is required on both of these, as the model may be adjusted to a dynamic model, such that the probabilities of game outcomes are altered after each game is simulated, and more advanced Kelly strategies, such as those proposed by Buchen & Grant (2012), are applied.

Finally, the proposed method can also be investigated in other sports. This extension is however dependent on having all game odds available, thus requiring that the tournament is played within a short time frame, and that all teams (or individuals) within each group play directly against each other. Tournaments that fit these criteria are numerous and include the FIFA World Cup, Rugby World Cup and Olympic Basketball.

Chapter 8

Conclusion

Research on betting markets will continuously increase due to the availability of data sets and the importance of the sector (Peel, 2008). This is accentuated by academic journals solely dedicated to related topics such as The Journal of Gambling Business and Economics, The Journal of Prediction Markets and the STN Journal of Sports Modelling and Trading.

8.1 The Business of Bookmaking

This thesis focused on trying to explain inconsistencies within the market. While there are excellent literature reviews on the efficiency of betting markets¹, a formal derivation of the key quantitative restrictions, as shown in Chapter 2, was missing from academic work. Here the returns on a series of wagers on an outcome are defined together with their deviation. This chapter includes proofs and discussions of the necessary conditions for a market to be operational.

- The sum of probabilities implied by odds, for all mutually exclusive outcomes of an event, should add up to more than one as otherwise there would be the existence of arbitrage.
- The probability implied by odds should be greater than the actual probability for each possible outcome in order for a bookmaker to have positive expected profits.
- The expected profitability for a bookmaker (or loss for a bettor) is $\frac{k}{1+k}\%$ of the total wagers places, where k is the bookmaker margin/over-round.

¹See Vaughan Williams (1999).

- Bookmaker profit is not dependent on knowing the exact probability of an outcome. This is in line with Kuypers (2000) who demonstrated that bookmakers' expected profits increase if they move from efficient to inefficient odds in order to capitalize on bettors' sentimental bias. However this is limited by the first two findings in this list and evidence of increased bookmaker profitability when setting more accurate odds (Chapter 6).
- Profitability on multiples/accumulators is higher than the profitability on single outcome bets.
- Profitability is guaranteed if the wagers on each outcome are placed in ratio to the implied probabilities.

Chapter 2 is complete in its current form but does warrant further research. Profitability is guaranteed if the variance of outcomes is zero, a special case being when wagers on different outcomes are placed in same proportion as their proportion of implied probabilities. However it would be ideal to prove that this is an 'if and only if' scenario.

Furthermore multiples/accumulators warrant more academic attention. While this work focused on a bet placed on a series of independent outcomes, contingent bets are currently offered as the example shown in Figure 8.1. It would also be interesting to investigate how to allocate profits (or losses) to the different outcomes – possibly by applying a Shapley, Marginal, Proportional or Percentile Method. For example, if a bet based on five outcomes is not paid out because of one outcome not occurring, should the profit be attributable to the non-occurring outcome.

If the necessary conditions for a bookmaker in setting odds on an event are known, the next step would be to extend them to bookmaker solvency. Having held discussions with industry experts, regulated entities in the betting industries do not have a standard model or regulatory set-up that define how much capital should be reserved. In fact, in most cases, they are asked to have a bank guarantee to pay out in case of liquidity issues. This adds credit-counter party risk and may result in not having enough cover (i.e. more risk) or having paid too much which leads to an inefficient use of capital. A novel approach to measure bookmaker solvency shown in Chapter 3 involves subdividing all bets received by bookmaker in odds ranges in order to calculate overall risk. This model is scalable but still theoretical. Hence it is subject to significant more research on how to best be adapted – by adjusting the number of bundles, using sport as a possible bundle, the width of each bundle as well as the assumptions used. The next step is to use a real odds, together with simulated bettor profiles to examine the use of the model as well as sensitivity test the different assumptions.

Player & Goal Specials 
Anthony Martial to score and Over 2.5 Goals in match 3/1
Anthony Martial to score, Both teams to score and Over 2.5 Goals in match 4/1
Anthony Martial to score, Man Utd to win and Over 2.5 Goals in match 6/1
Anthony Martial to score and Man Utd to win to nil 8/1
Anthony Martial to score, Man Utd to win and Both teams to score 8/1
Juan Mata to score and Over 2.5 Goals in match 9/2
Juan Mata to score, Both teams to score and Over 2.5 Goals in match 11/2
Juan Mata to score, Man Utd to win and Over 2.5 Goals in match 17/2
Juan Mata to score, Man Utd to win and Both teams to score 11/1
Juan Mata to score and Man Utd to win to nil 11/1
Memphis Depay to score and Over 2.5 Goals in match 4/1
Memphis Depay to score, Both teams to score and Over 2.5 Goals in match 9/2
Memphis Depay to score, Man Utd to win and Over 2.5 Goals in match 7/1
Memphis Depay to score and Man Utd to win to nil 9/1
Memphis Depay to score, Man Utd to win and Both teams to score 9/1

FIGURE 8.1: Contingent bets for the Manchester Utd - Liverpool FC Europa League match shown on SkyBet as at 13:00 on 17th March 2016 (match was on same day). For example the bet on Anthony Martial (Manchester United player) and over 2.5 goals in a match at European odds of 4 is a bet contingent on both events which are related.

8.2 Predicting Soccer outcomes

Most soccer prediction models focus on either determining the outcome or the number of goals scored. The model presented in Chapter 5 tries to balance these two by predicting the goal difference in a match with very strong results.

Draws are the most challenging out come to predict from the three outcomes of a soccer match² (Pope & Peel, 1989; Kuk, 1995). The number of expected draws during a season was very high when a draw was defined as any match predicted with a goal difference of less than 0.5 goals in magnitude. An improvement was noted when this was reduced

²The other two being a home team win or an away team win.

to 0.25. Further investigation would be an evaluation of the ideal cut-off point, which might not be symmetric around zero.

A limitation of the model is that it used statistics from each season to predict the results of that particular season. A natural extension is to use one season's parameters to estimate the following season's results as well as testing the model in leagues other than the English Premiership. Additionally, the model would also be ideally tested against other benchmark models and the betting market.

8.3 (In)efficient markets

Chapter 4 showed that odds on Brazil were fluctuating between matches based on Brazil's World Cup successes of past years even if the team was not performing brilliantly. In essence this chapter showed that betting on the favourite does not always work. Future work will investigate how the changes in Brazil's perceived outcomes were affected by the other teams (such as Germany) and a synopsis of similar scenarios.

The English Premiership this year has also shown significant signs of anchoring to past years. For example on 7th February 2016, Tottenham and Arsenal were on equal points placed second in the table. Notwithstanding a similar set of future fixtures, Arsenal were being touted as twice as likely to win the Premiership as Tottenham.

In no way was this intended to state that betting on unlikely events is preferable. In fact, Chapter 6 examined over 136,000 odds from Europe's popular leagues to show that expected losses made by bettors made on unlikely events are much higher. Future intended work is to introduce measures other than the Brier Score as a measure of accuracy, ideally ones which can also be applied in a continuous environment to test the accuracy of live odds³.

The model introduced in Chapter 7 showed that inconsistencies within the market can be taken advantage of. While typical models use external data, such as past scores, to predict an outcome and then bet on it, the model presented here used betting odds to examine contradictions between markets that lead to quasi-arbitrage. Future iterations of this model should consider the human bias as observed in Chapters 4 and 6.

³Odds offered while a match is in progress.

8.4 Other Work

This collection of work would be enriched by greater interdisciplinary collaboration. While evidence of bias in the market was found and also taken advantage of, there is little discussion on why this occurs. [Cain et al. \(2008\)](#) improve on the standard parametric explanation of Cumulative Prospect Theory to show that agents with differing assumptions can optimally bet over a range of both likely and unlikely events⁴. An interesting investigation would be to cross link this work with utility functions by hopefully adding more heterogeneous risk preferences as in [Ali \(1977\)](#) in an attempt to describe the whole demand and supply curve for odds.

This can be supplemented by research determining the traits of risk-taking behaviour of a bettor. The main approach would be to source bettors, either through a betting company or by creating a betting pool, and ask them to fill in a questionnaire about their traits. The answers can then be adapted to use The 3M Model ([Mowen, 2000](#)). The dependent variable, being risk, would be measured as the discrepancy between one's option and the betting markets' expected outcome.

8.5 Implications on the Financial Industry

In essence a bet is a binary option, with a single known pay out made if an outcome occurs. Moreover, betting and insurance share the same common theme with only a few differences. Firstly there must be insurable interest for one to open up an insurance policy while one must be independent of an outcome to bet on it. A contract that pays me £1,000 on Queen Elizabeth II's death is a bet not an insurance policy given that I have no insurable interest in the queen's survival. Secondly the value of a pay out on an insurance claim can fluctuate, especially in a general business scenario while that of a bet tends to be fixed. For example a car accident may be a few pounds in case of a scratch to a few hundred thousands if it results in deaths. There are alternative situations, such as the fluctuating payout on a spread-bet⁵ or the fixed amount paid on a life policy in case of death. Finally the timing of a bet is usually known as it is paid at the end of an event while the trigger for an insurance pay-out may fluctuate.

Bets are a simpler representation of financial contracts. If betting markets are known to be inefficient, then the implication is that financial markets follow suit.

⁴The standard parametric approach only explains the long shot bias.

⁵This acts just like a forward/future.

Appendix A

Payout Distribution for Constant Bookmaker Margin

Proposition A.1. If $\pi_i = p_i(1+k)$ for some constant k , then $P \sim N\left(\frac{k}{1+k} \sum_{i=1}^n w_i, \frac{1}{(1+k)^2} \left(\sum_{i=1}^n \frac{w_i^2}{p_i} - (\sum_{i=1}^n w_i)^2\right)\right)$

Proof. $E(P) = E(\text{Wagers less Payouts}) = \sum_{i=1}^n \left(p_i \sum_{i=1}^n w_i - w_i \times \frac{1}{\pi_i} \times p_i\right)$

$$E(P) = \sum_{i=1}^n w_i - \sum_{i=1}^n \frac{w_i}{1+k} = \frac{k}{1+k} \sum_{i=1}^n w_i$$

$$E(P^2) = E((\text{Wagers less Payouts})^2) = \sum_{i=1}^n \left(\left(\sum_{i=1}^n w_i - w_i \times \frac{1}{\pi_i} \right)^2 \times p_i \right)$$

$$E(P^2) = \sum_{i=1}^n \left[\left(\sum_{i=1}^n w_i \right)^2 p_i - \frac{2w_i}{1+k} \sum_{i=1}^n w_i + \frac{1}{p_i} \left(\frac{w_i}{1+k} \right)^2 \right]$$

$$E(P^2) = \left(1 - \frac{2}{1+k}\right) \left(\sum_{i=1}^n w_i\right)^2 + \frac{1}{(1+k)^2} \sum_{i=1}^n \frac{w_i^2}{p_i}$$

$$E(P^2) = \frac{k-1}{1+k} \left(\sum_{i=1}^n w_i\right)^2 + \frac{1}{(1+k)^2} \sum_{i=1}^n \frac{w_i^2}{p_i}$$

$$\text{Var}(P) = E(P^2) - [E(P)]^2 = \frac{k-1}{1+k} \left(\sum_{i=1}^n w_i\right)^2 + \frac{1}{(1+k)^2} \sum_{i=1}^n \frac{w_i^2}{p_i} - \frac{k^2}{(1+k)^2} \left(\sum_{i=1}^n w_i\right)^2$$

$$\text{Var}(P) = \frac{1}{(1+k)^2} \left(\sum_{i=1}^n \frac{w_i^2}{p_i} - (\sum_{i=1}^n w_i)^2 \right) \quad \square$$

Appendix B

Pricing risk through simulation: Revisiting Tilley bundling and least square Monte Carlo methods

B.1 Motivation

This appendix marks the start of my PhD research and was developed in conjunction with an investment bank and a fellow PhD student. I was awarded [best presentation](#) when presenting this work at the 2013 Actuarial Research Conference hosted by Fox Business School at Temple University in Philadelphia ([Branvall et al., 2014](#)). It helped form my later work related to betting derivatives in two respects: a deeper understanding of issues relating to risk (Chapter [2](#)) and learning on how to use Monte Carlo techniques (Chapter [7](#))

B.2 Introduction

This appendix is subdivided into five sections. Other than the motivation, I discuss the key developments of the use of Monte-Carlo techniques in pricing derivatives as well as define the key risk measures of interest giving the example for a European-style Option. Finally I present two popular Monte Carlo techniques and compare the results of pricing the best estimate value and the risk metrics of American-style Options.

B.2.1 Use of Monte Carlo methods in pricing derivatives

In the early 90s, Monte Carlo simulation was deemed not useful to price derivatives with early exercise features. Indeed [Hull \(1993\)](#) states that “*Monte Carlo simulation can only be used for European-Style Options*” in his second edition of the popular text book “Options, Futures and Other Derivatives”.

However that same year [Tilley \(1993\)](#), an actuary, presented a paper with a goal “*to dispel the prevailing belief that American style options cannot be valued efficiently in a simulation model*”. This was the first Monte Carlo method to price such a derivative but was quickly followed by many others. Monte Carlo methods vary in their approach but [Fu et al. \(2001\)](#) show that they can be summarised in three major categories:

- methods that mimic the backwards induction algorithm,
- methods that parametrise the early exercise curve, and
- methods that find the efficient upper and lower bounds.

A major development was made by [Longstaff & Schwartz \(2001\)](#) as their approach was simple and effective. Indeed [Hull \(2012\)](#) updated his wording by the eighth edition to state “*Monte Carlo simulation is well suited to valuing path-dependent option*”.

The technological progress, especially with the introduction of parallel computing, has enabled practitioners to obtain results from simulation techniques much faster. Moreover Monte Carlo techniques are now used extensively within the financial sector as they provide significant advantages over closed form solutions such as partial differential equations. They tend to be simpler to solve in higher dimensional problems and provide risk metrics, usually being percentile measures, required by financial regulators worldwide. For example, the new European insurance Solvency regime ‘Solvency II’ that has come in effect on 1st January 2016 sets the solvency capital requirement as a 99.5% Value-at-Risk (VaR) ([Doff, 2008](#)).

Within the banking sector, the convergence of financial institutions has resulted in risk not only when a financial security loses its value but also when it is deeply in the money as the counter-party may not be able to pay this. Indeed percentile measures other than the VaR have been developed by Basel II/III regulation as discussed in the next subsection.

B.2.2 Risk profiles

Risk can be defined as deviation from the expected and there are various measures of risk. A key measure for market risk is the Value-at-Risk (VaR) (Linsmeier & Pearson, 1996) while counterparty credit risk is usually measured as the Potential Future Exposure (PFE) or the Expected Positive Exposure (EPE). VaR and PFE are percentiles but VaR would measure the amount a holder would pay out if a derivative is out-of-the-money while PFE relates to the risk of a counterparty not paying their dues when a derivative is in-the-money.

EPE is defined as the average exposure over time, considering only positive exposures over time (Gregory, 2010). Therefore PFE and EPE are related with a gain (exposure) while VaR is associated with losses. Furthermore PFE and EPE tend to be normally associated with a point far in the future (Gregory, 2010).

B.2.3 European-style Option in the GBM one-factor model

Throughout this appendix, pricing and risk measures are evaluated for a stock price simulated over a period of time using the geometric Brownian motion (GBM) method which is explained in any standard financial engineering textbook, such as Glasserman (2003) and (Hull, 2012). As an example, consider a European call on an equity exercisable in a year's time. This equity follows the GBM one-factor model, has a strike price and initial price of 100, a constant volatility of 20% and a constant drift rate of 6%. In this scenario, no distinction is made between the calculations in the risk-neutral and the real measures. Furthermore no dividend payment is assumed.

The present value of this contract is 10.9851. The risk measures at the start of the contract, all subject to being a minimum of zero, for this European call stock option can be calculated in a closed form solution as shown below since we are aware that the equity is assumed to follow a log-normal distribution. One can note that the VaR for an option tends to be zero as it is the minimum possible value since the holder of a European call option has the right but not the obligation to purchase the equity for a price of 100 in a year's time. Thus if the equity price is 80, the holder of the option would not exercise it and hence result in a value of 0. The VaR of 32.6091 would apply for a long forward(/future) since this hedge would commit the buyer to purchase the equity at a price of 100.

$$\bullet \int_0^{e^{0.06} PFE_{0.99}} \frac{1}{x\sigma\sqrt{2\pi}} \exp \left[-\frac{\left(\log(x) - \log(100) - \left(\mu - \frac{\sigma^2}{2} \right) \right)^2}{2\sigma^2} \right] dx = 0.99 \Rightarrow PFE_{0.99} = 65.0075.$$

- $$\int_{e^{0.06}VaR_{0.01}}^{\infty} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{\left(\log(x)-\log(100)-\left(\mu-\frac{\sigma^2}{2}\right)\right)^2}{2\sigma^2}\right] dx = 0.99 \Rightarrow VaR_{0.01} = -32.6091 \Rightarrow VaR_{0.01} = 0.$$
- $$EPE = e^{-0.06} \left(\frac{\int_{100}^{\infty} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{\left(\log(x)-\log(100)-\left(\mu-\frac{\sigma^2}{2}\right)\right)^2}{2\sigma^2}\right] dx}{\int_{1000}^{\infty} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{\left(\log(x)-\log(100)-\left(\mu-\frac{\sigma^2}{2}\right)\right)^2}{2\sigma^2}\right] dx} - 100 \right) = 18.9229.$$

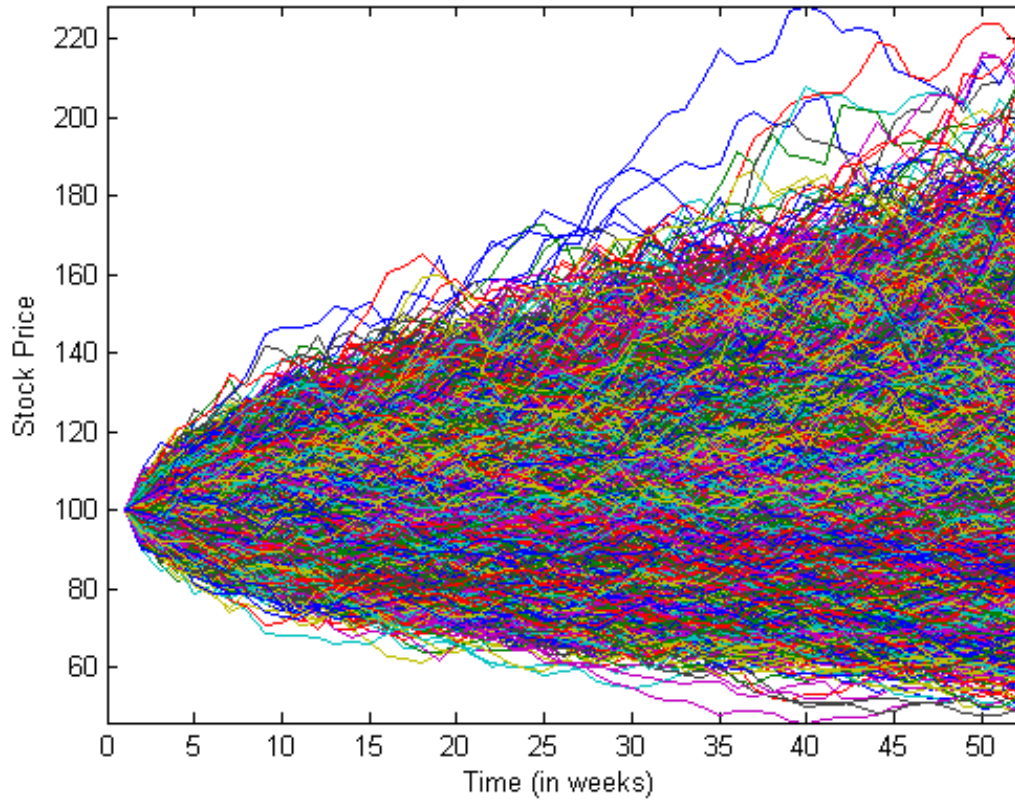


FIGURE B.1: A sample generation of 20,000 stock price paths for a year with an initial price of 100, volatility of 20% and $r = 0.06$.

The results can also be obtained easily by simulation. The steps involved are simulate a number of equity prices over a period of one year (as per shown in Figure B.1), deduct 100 from each simulated price, set all negative values as zero and discount to the present value (by multiplying by $e^{-0.06}$). Then the metrics can be calculated as follows:

- Present Value: Average of all values.
- PFE(0.99): Find the 99th percentile value when placed in ascending order.
- VaR(0.01): Find the 1st percentile value when placed in ascending order.
- EPE: Find the average of all non-negative values.

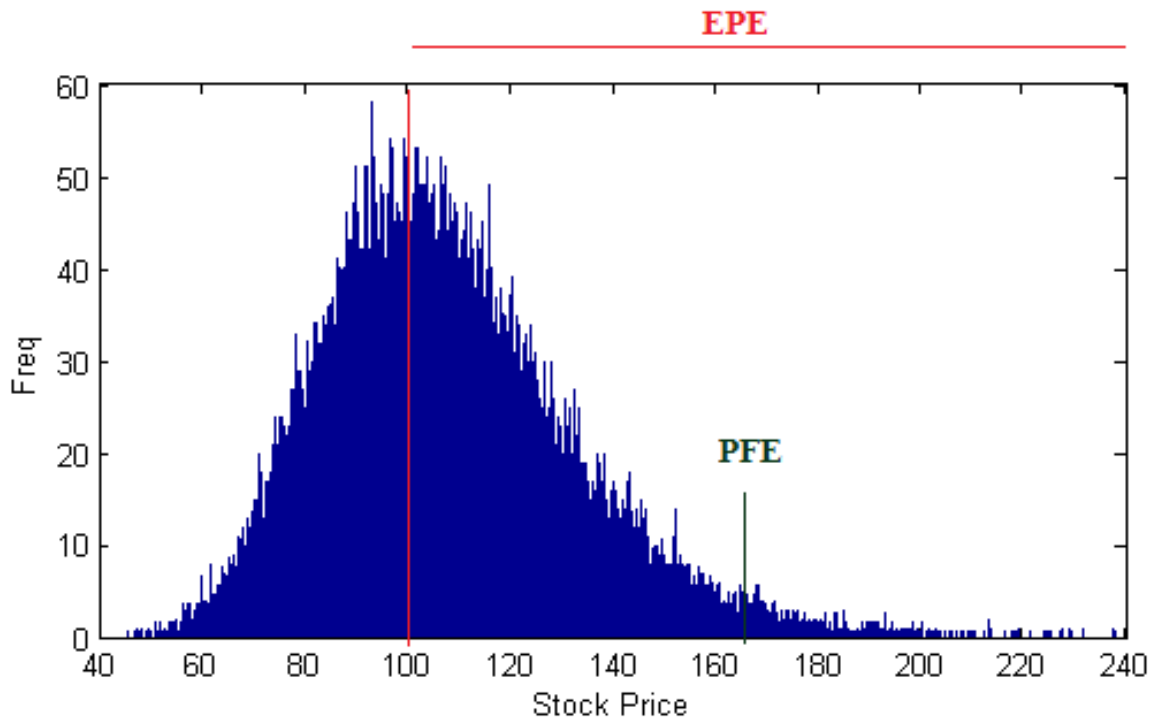


FIGURE B.2: Undiscounted EPE and PFE for a one year Euro Call Option

B.3 The Two Monte Carlo methods

B.3.1 Steps Involved

The two simulation approaches compared here are the Least-Squares Monte Carlo [LSM-Method] ([Longstaff & Schwartz, 2001](#)) and Tilley Bundling ([Tilley, 1993](#)). In both approaches, a sample path of stock prices are first produced using the GBM method. One must be aware that the holder of an American option may exercise the option if in-the-money at predefined times up to expiry of the contract. Therefore at any of these predefined times, a decision criteria is applied as to whether the derivative should be exercised or otherwise. The option is exercised if the expected future pay-off from continuation is lower than the current payoff at any time.

The LSM-method estimates the expected payoff from continuation by a regression formula applied on all in-the-money stock price paths. The steps for this method applied on an American stock option, assuming K times during which this option can be exercised, are summarised below.

Step 1: Determine the value of the future pay-off at time $t = K$ for every stock price path. This would be equal to the pay-off for a European-style option.

Step 2: Determine which stock price paths are in-the-money at the previous time step (t_m). Denote the value of the pay-offs at this timestep as X and the discounted future pay-off values for these paths determined in Step 1 as Y .

Step 3: Produce a regression function in which the independent variables are X and the dependent variables are Y . Use this regression function to estimate the expected discounted future pay-off Y' for each value of X .

Step 4: If $X > Y'$, the option will be exercised at time t_m . In this case the future pay-off value for these price paths is set as X , otherwise Y is maintained.

Step 5: Repeat steps 2 to 4 until $t = 1$.

Step 6: Determine the value of the American option as the average final discounted future pay-off.

The Tilley Bundling method sets the decision criteria by dividing the paths into a number of bundles. Assuming that $N = A \times B$ paths are simulated for K time-steps. The valuation method for an American call stock option can be summarised in the steps given below.

Step 1: Determine the value of the future pay-off at time $t = K$ for every stock price path. This would be equal to the pay-off for a European-style option.

Step 2: Re-order the stock price paths by stock price at the previous timestep (t_m) in ascending order for a call option.

Step 3: Divide the N paths into A bundles of B paths each in the order given in step 3.

Step 4: Set the holding value of each path as the discounted mean of all future pay-off values within each bundle.

Step 5: For each path, calculate whether the present pay-off value is higher than the holding value. Set the Boolean variable X as one if this is the case or zero if otherwise. This will produce a string of ones and zeros.

Step 6: Determine one boundary that decides whether to execute the option at time t_m by finding the first string of ones whose length is bigger than each subsequent string of zeros.

Step 7: All current pay-offs which are higher than the boundary evaluated in step 7 would be exercised at time t_m . Record whether each path would be exercised at this time. Update the future pay-off as the current pay-off for these cases and as the holding value for all other cases.

Step 8: Repeat steps 2 to 7 until $t = 1$.

Step 9: Determine the value of the American Option as the mean discounted pay-off of the earlier exercise time (evaluated in step 8) for each path.

B.3.2 A Numerical Example of Pricing: An American Call Option

As a further explanation, the two methods described above will be applied on an American call option exercisable at the end of each year for the next four years for a strike price of 100. Sixteen paths are generated using the GBM method are shown in table B.1. At the end of the fourth year, eight paths are in-the-money and their (future) pay-offs are shown in the table. This is the first step for both methods.

TABLE B.1: Generation of sixteen stock price paths at the end of each year with an initial price of 100, volatility of 30% and $r = 0.02$

Path	t=0	t=1	t=2	t=3	t=4	Step 1
A	100.00	99.53	110.58	157.73	298.48	198.48
B	100.00	44.01	36.96	25.81	32.72	0.00
C	100.00	88.83	103.66	135.62	194.44	94.44
D	100.00	48.16	59.81	63.23	44.42	0.00
E	100.00	127.21	195.09	283.07	248.94	148.94
F	100.00	75.36	72.48	31.29	29.36	0.00
G	100.00	105.00	39.74	38.18	51.53	0.00
H	100.00	85.66	65.80	43.27	40.53	0.00
I	100.00	104.47	97.05	59.55	8.79	0.00
J	100.00	159.99	192.00	257.60	198.98	98.98
K	100.00	115.17	100.54	73.57	44.61	0.00
L	100.00	155.84	124.39	122.25	163.50	63.50
M	100.00	76.79	38.89	22.90	26.58	0.00
N	100.00	128.64	138.71	223.08	245.77	145.77
O	100.00	99.00	164.48	177.53	122.58	22.58
P	100.00	118.34	150.51	208.06	229.56	129.56

B.3.2.1 LSM

At the second step, one needs to examine which paths are in-the-money at the previous time step ($t = 3$). In this scenario these are the same paths as those in-the-money at time $t = 4$ (A, C, E, J, K, L, N, O and P). The pay-off at this time step and the discounted pay-off for each of these paths are shown in table B.2.

The LSM-method applies a regression function to evaluate the estimated discounted future value. In this case, the basis functions to be used are the pay-off, its square and a constant. The least-squares best fit for this case is $Y' = -0.00121X^2 + 0.52209X +$

75.26894. At this stage step 3 can be finalised by finding the the estimated discounted future pay-off. These are compared to the current pay-offs as shown (step 4). In this example, at $t = 3$ only paths E, J and N are exercised since their current pay-off (183.07, 157.60, 123.08) is higher than their estimated discounted future pay-off (146.02, 97.04, 142.91). The future pay-off is set as the current pay-off for these three paths.

TABLE B.2: LSM-method steps 2 to 4 at $t = 3$.

Path	Step 2		Step 3	Step 4	
	X	Y	Y'	Exercised?	Fut Pay-off
A	57.73	194.59	101.36	No	194.59
B	-	-	-	-	-
C	35.62	92.59	92.33	No	92.59
D	-	-	-	-	-
E	183.07	146.02	130.14	Yes	183.07
F-I	-	-	-	-	-
J	157.60	97.04	127.38	Yes	157.60
K	-	-	-	-	-
L	22.25	62.26	86.28	No	62.26
M	-	-	-	-	-
N	123.08	142.91	121.13	Yes	123.08
O	77.53	22.14	108.45	No	22.14
P	108.06	127.02	117.50	No	127.02

TABLE B.3: LSM-method steps 2 to 4 at $t = 2$.

Path	Step 2		Step 3	Step 4	
	X	Y	Y'	Exercised?	Fut Pay-off
A	10.58	190.77	83.63	No	190.77
B	-	-	-	-	-
C	3.66	90.77	84.45	No	90.77
D	-	-	-	-	-
E	95.09	179.48	154.79	No	179.48
F-I	-	-	-	-	-
J	92.00	154.51	149.56	No	154.51
K	0.54	-	85.15	No	-
L	24.39	61.04	85.00	No	61.04
M	-	-	-	-	-
N	38.71	120.66	90.65	No	120.66
O	64.48	21.70	111.69	No	21.70
P	50.51	124.53	98.55	No	124.53

Given the updated future pay-offs, the in-the-money paths at $t = 2$ can be investigated (step 2) and a quadratic regression function applied $Y' = -0.01051X^2 + 0.26806X + 85.28985$ (step 3). In this scenario, no path would be exercised at this time as one can see in table B.3.

TABLE B.4: LSM-method steps 2 to 4 at $t = 1$.

Path	Step 2		Step 3	Step 4	
	X	Y	Y'	Exercised?	Fut Pay-off
A	-	-	-	-	187.03
B	-	-	-	-	-
C	-	-	-	-	88.99
D	-	-	-	-	-
E	27.21	175.96	126.65	No	175.96
F	-	-	-	-	-
G	5.00	-	(4.72)	Yes	5.00
H	-	-	-	-	-
I	4.47	-	(9.29)	Yes	4.47
J	59.99	151.48	104.21	No	151.48
K	15.17	-	70.11	No	-
L	55.84	59.84	121.31	No	59.84
M	-	-	-	-	-
N	28.64	118.30	131.05	No	118.30
O	-	-	-	-	21.28
P	18.34	122.09	88.35	No	122.09

Finally, the same steps are re-traced for $t = 1$. The regression function applied at this time is $Y = -0.12002X^2 + 9.78184X - 50.6562$. At this time, only paths G and I would be exercised as these two have a negative expected future pay-off (even if in reality the American stock option cannot have a negative pay-off). The future pay-offs for these two cases are updated as the pay-off at time $t = 1$.

The value of the American option, priced using the LSM-method, is the discounted average of all the final future pay-offs $\left(\frac{187.03+0+88.99+\dots+122.09}{16 \times 1.02} \right)$. This is equal to 57.26.

B.3.2.2 Tilley Bundles

The same set of paths will be used to price an American call option using the Tilley bundling technique with four bundles having four paths each.

At time $t = 3$, the paths are first set in ascending order of price with path M having the lowest price and E having the highest price (Step 2). These paths can then be subdivided into four bundles (First Bundle: paths M, B, F and G; Second Bundle: paths H, I, D and K; Third Bundle: paths L, C, A and O; Fourth Bundle: P, N, J, E). The bundles are subdivided using a dotted line in table B.5.

The holding value of each bundle is the mean discounted future pay-off of all four paths within a bundle. The first two bundles have a holding value of zero while the

third and fourth bundles have holding values of $\frac{63.50+94.44+198.48+22.58}{1.02 \times 4} = 92.89$ and $\frac{129.56+145.77+98.98+148.94}{1.02 \times 4} = 128.25$ respectively (step 4).

The present pay-offs, that is the pay-off of the path if exercised at $t = 3$, is only higher than the holding values for paths J and E (step 5). Therefore, the boundary condition for step 6 is that all stock prices that have a value of at least 257.60 at time $t = 3$ are exercised. The future pay-off is adjusted as the pay-off at $t = 3$ for paths J and E and the holding values for all other paths (step 7).

The same steps are applied at $t = 2$. It can be noticed in table B.6 that only path K has a present pay-off value higher than the holding value (step 5). However there is no string of ones that exceeds the length of each subsequent string of zeros and therefore no paths would be exercised at this time (step 6). On a similar note no paths have a present pay-off value higher than the holding value and therefore no path is exercised at $t = 1$.

In conclusion, paths E and J have an earliest exercise point at $t = 3$ while all other paths are exercised at expiry if in-the-money. The value of this American call stock option is the mean discounted pay-off is

$$\frac{1}{16} \left[\frac{(283.07 - 100) + (257.60 - 100)}{1.02^3} + \frac{198.48 + 94.44 + 63.50 + 145.77 + 22.58 + 129.56}{1.02^4} \right] = 57.85 \quad (\text{B.1})$$

TABLE B.5: Tilley Bundling method at $t = 3$.

Path	Step 2	Step 4		Step 5		Step 7	
	Stock Price	Disc Fut	Pay-off	Holding Value	Pres Pay-off	X	Fut Pay-off
M	22.90		0.00	0.00	0.00	0	0
B	25.81		0.00	0.00	0.00	0	0
F	31.29		0.00	0.00	0.00	0	0
G	38.18		0.00	0.00	0.00	0	0
H	43.27		0.00	0.00	0.00	0	0
I	59.55		0.00	0.00	0.00	0	0
D	63.23		0.00	0.00	0.00	0	0
K	73.57		0.00	0.00	0.00	0	0
L	122.25		63.50	92.89	22.25	0	92.89
C	135.62		94.44	92.89	35.62	0	92.89
A	157.73		198.48	92.89	57.73	0	92.89
O	177.53		22.58	92.89	77.53	0	92.89
P	208.06		129.56	128.25	108.06	0	128.25
N	223.08		145.77	128.25	123.08	0	128.25
J	257.60		98.98	128.25	157.60	1	157.60
E	283.07		148.94	128.25	183.07	1	183.07

TABLE B.6: Tilley Bundling method at $t = 2$.

Path	Step 2	Step 4		Step 5		Step 7	
	Stock Price	Disc Fut Pay-off	Holding Value	Pres Pay-off	X	Fut Pay-off	
B	36.96	0.00	0.00	0.00	0	0	
M	38.89	0.00	0.00	0.00	0	0	
G	39.74	0.00	0.00	0.00	0	0	
D	59.81	0.00	0.00	0.00	0	0	
H	65.80	0.00	0.00	0.00	0	0	
F	72.48	0.00	0.00	0.00	0	0	
I	97.05	0.00	0.00	0.00	0	0	
K	100.54	0.00	0.00	0.54	1	0	
C	103.66	92.89	99.74	3.66	0	99.74	
A	110.58	92.89	99.74	10.58	0	99.74	
L	124.39	92.89	99.74	24.39	0	99.74	
N	138.71	128.25	99.74	38.71	0	99.74	
P	150.51	128.25	137.70	50.51	0	137.70	
O	164.48	92.89	137.70	64.48	0	137.70	
J	192.00	157.60	137.70	92.00	0	137.70	
E	195.09	183.07	137.70	95.09	0	137.70	

TABLE B.7: Tilley Bundling method at $t = 1$.

Path	Step 2	Step 4		Step 5		Step 7	
	Stock Price	Disc Fut Pay-off	Holding Value	Pres Pay-off	X	Fut Pay-off	
B	44.01	0.00	0.00	0.00	0	0	
D	48.16	0.00	0.00	0.00	0	0	
F	75.36	0.00	0.00	0.00	0	0	
M	76.79	0.00	0.00	0.00	0	0	
H	85.66	0.00	81.02	0.00	0	81.02	
C	88.83	97.78	81.02	0.00	0	81.02	
O	99.00	135.00	81.02	0.00	0	81.02	
A	99.53	97.78	81.02	0.00	0	81.02	
I	104.47	0.00	33.09	4.47	0	33.09	
G	105.00	0.00	33.09	5.00	0	33.09	
K	115.17	0.00	33.09	15.17	0	33.09	
P	118.34	135.00	33.09	18.34	0	33.09	
E	127.21	135.00	114.11	27.21	0	114.11	
N	128.64	97.78	114.11	28.64	0	114.11	
L	155.84	97.78	114.11	55.84	0	114.11	
J	159.99	135.00	114.11	59.99	0	114.11	

B.3.2.3 Comparison of the two systems

The price of an American put stock option at a strike price of 40 exercisable 50 times a year was evaluated through these two Monte-Carlo methods as shown in table B.8. This result is based on 100,000 paths (half of which are antithetic to reduce variance) based on a yearly interest rate of 6%, a number of initial stock prices ranging from 36 to 44, a volatility measure of 20% or 40%, and a one or two year time period.

TABLE B.8: The value of an American Put stock option. The values in brackets show standard deviation.

S	σ	T	FD	Tilley Bundling		LSM-method	
				I-S	O-S	I-S	O-S
36	0.2	1	4.487	4.479 (0.009)	4.478 (0.009)	4.461 (0.009)	4.472 (0.009)
36	0.2	2	4.848	4.822 (0.011)	4.832 (0.011)	4.829 (0.011)	4.832 (0.011)
36	0.4	1	7.109	7.097 (0.019)	7.107 (0.019)	7.063 (0.019)	7.033 (0.019)
36	0.4	2	8.509	8.493 (0.023)	8.507 (0.023)	8.474 (0.023)	8.489 (0.023)
38	0.2	1	3.257	3.248 (0.009)	3.250 (0.009)	3.231 (0.009)	3.233 (0.009)
38	0.2	2	3.751	3.748 (0.011)	3.736 (0.011)	3.744 (0.011)	3.742 (0.011)
38	0.4	1	6.154	6.145 (0.019)	6.145 (0.019)	6.105 (0.019)	6.108 (0.019)
38	0.4	2	7.667	7.672 (0.022)	7.653 (0.022)	7.646 (0.022)	7.655 (0.022)
40	0.2	1	2.320	2.309 (0.009)	2.310 (0.009)	2.300 (0.009)	2.289 (0.009)
40	0.2	2	2.890	2.885 (0.011)	2.889 (0.011)	2.879 (0.010)	2.880 (0.010)
40	0.4	1	5.318	5.306 (0.018)	5.310 (0.018)	5.254 (0.018)	5.283 (0.018)
40	0.4	2	6.913	6.906 (0.022)	6.915 (0.022)	6.896 (0.022)	6.866 (0.022)
42	0.2	1	1.621	1.617 (0.008)	1.607 (0.008)	1.601 (0.008)	1.600 (0.008)
42	0.2	2	2.217	2.210 (0.010)	2.199 (0.010)	2.196 (0.010)	2.199 (0.010)
42	0.4	1	4.588	4.566 (0.017)	4.570 (0.017)	4.528 (0.017)	4.554 (0.017)
42	0.4	2	6.236	6.233 (0.021)	6.233 (0.021)	6.215 (0.021)	6.217 (0.021)
44	0.2	1	1.113	1.104 (0.007)	1.102 (0.007)	1.096 (0.006)	1.096 (0.006)
44	0.2	2	1.693	1.670 (0.009)	1.687 (0.009)	1.679 (0.009)	1.675 (0.009)
44	0.4	1	3.952	3.944 (0.017)	3.949 (0.017)	3.898 (0.016)	3.895 (0.016)
44	0.4	2	5.627	5.633 (0.021)	5.655 (0.021)	5.621 (0.021)	5.617 (0.021)

The Tilley bundling technique used consisted of 250 bundles of 400 paths each as this value maintains a low ratio of bundles to paths as recommended in the original paper (Tilley, 1993). A cubic regression function was used for the LSM-method.

The price was evaluated also on an out of sample (O-S) generation for both methods. The value priced under O-S is a random generation of an additional 100,000 paths (half of which antithetic) but using the regression function as evaluated for the in sample (I-S) set under a LSM-method (step 3) or using the same boundary condition as evaluated for the I-S generation under the Tilley Bundling method (step 6).

Both methods compare favourably with the price set by the finite difference method (Hull, 2012) with 40,000 time steps per year over 1,000 steps for the stock price. The

Tilley Bundling method is slightly superior to the LSM-method however the latter can be improved by applying more basis functions to the regression formula.

B.3.3 A Numerical Example of Risk Pricing: An American Put Option

The evaluation of risk measures for an American option directly from an equation is a trickier process than for a European style option due to the early exercise features of the former. One can envisage that in this circumstance, the probability distribution function of the stock price is dependent on time and stock price (Hull, 2012) as the example shown in Figure B.3.

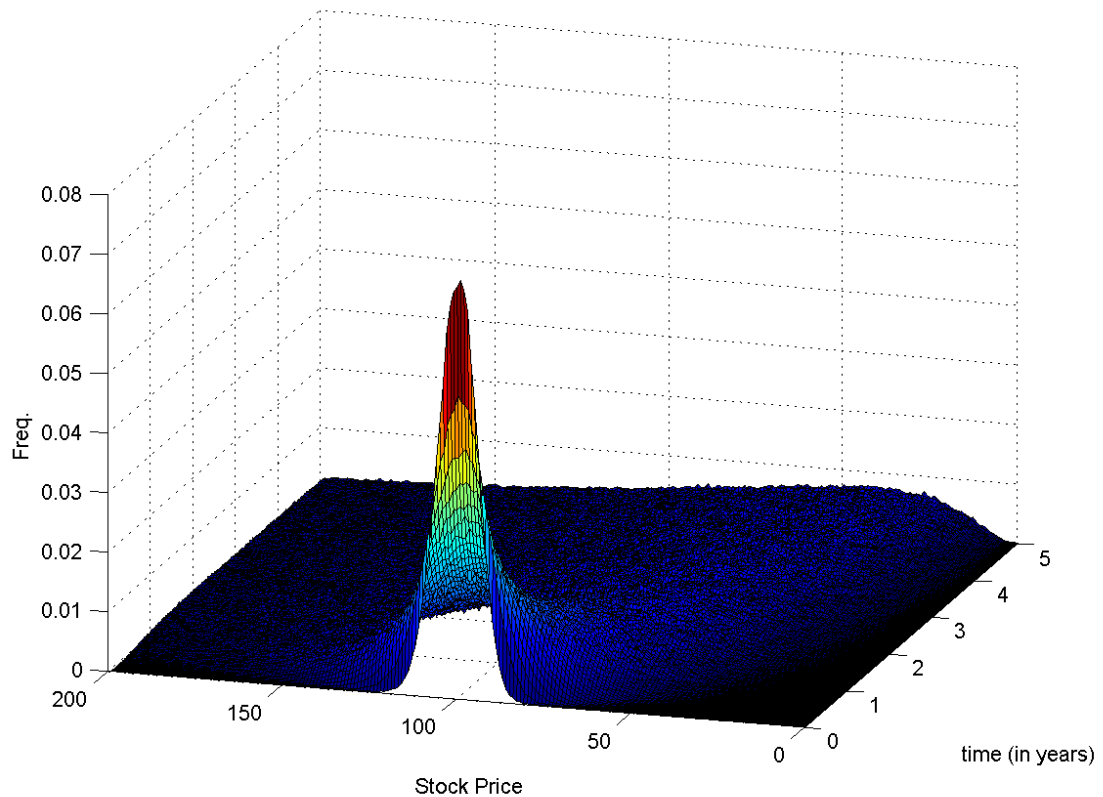


FIGURE B.3: A 3D-pdf of a stock price generation for five years with an initial price of 100, volatility of 40% and $r = 0.06$.

In this appendix, I have evaluated the PFE(99%) and EPE values of a one-year American put option with a strike price that is equal to the initial price of 100, a drift of 6% and volatility of 40% with 50 exercise points per year. The VaR for this contract is simply zero for all cases. Throughout this exercise, the risk measures are shown at a present value at the start of the contract.

The values of EPE and PFE are not easily obtainable to be able to make comparisons. Hence one possible approach is to define a possible minimum and maximum value in order to be able to make comparisons. The minimum values could be set as the metrics for a European Option ending at the same time. An American Option has early exercise features and hence should carry more credit-counter party risk than a European style option.

Maximum possible risk metrics can also be evaluated by taking into consideration the log-normal distribution at each of the 50 exercise points. The PFE can therefore be found as that value that lies in the extreme 1% in-the-money value of the option over all remaining exercise points by solving Equation B.2. Similarly a possible maximum approximation for the EPE is finding the average expected positive exposure of all time points (Equation B.3). One can consider this as if dealing with a series of European Options each exercised at a different time point.

$$\sum_{t=k}^{50} [F_2(x, t, 100 - PFE_{0.99} e^{0.0012t})] = 0.01(51 - k) \quad (B.2)$$

$$EPE_{k^{th} \text{ time}} = \frac{\sum_{t=k}^{50} \left[e^{-0.0012t} F_2(x, t, 100) \left(100 - \frac{F_3(x, t, 100)}{F_2(x, t, 100)} \right) \right]}{(51 - k) \sum_{t=k}^{50} F_2(x, t, 100)} \quad (B.3)$$

where

$$F_2(x, t, a) = \int_0^a \left[\left((x\sigma)^{-1} \sqrt{\frac{25}{t\pi}} \right) \exp \left[-\frac{\left(\log(x) - \log(S_0) - \left(\mu - \frac{\sigma^2}{2} \right) \frac{t}{50} \right)^2}{\sigma^2 \frac{t}{25}} \right] \right] dx$$

$$F_3(x, t, a) = \int_0^a \left[\left(\sigma^{-1} \sqrt{\frac{25}{t\pi}} \right) \exp \left[-\frac{\left(\log(x) - \log(S_0) - \left(\mu - \frac{\sigma^2}{2} \right) \frac{t}{50} \right)^2}{\sigma^2 \frac{t}{25}} \right] \right] dx$$

B.3.4 Comparing Results

The EPE and PFE evaluated by the two Monte Carlo methods are more straightforward. These are evaluated as the mean of all positive values and the ninety ninth percentile of the expected future pay offs calculated at steps 4 (LSM-method) and 7 (Tilley Bundling). One hundred thousand simulations are used to compare our results as shown in Figure B.4.

In every simulated scenario, a number of observations can be made. Firstly all metrics are larger than the European style equivalent which is a trivial result. However any estimate is closer to the European style option when nearing the end of the option since the American style option would have a lower number of exercise points. LSM produces less stable results since this is dependent on the regression formula produced.

The maximum PFE is significantly larger than the options since in cases of an option being in-the-money by a large present pay-off value, it is very likely that this is exercised early and therefore reducing the likelihoods of later high in-the-money pay-off values.

Secondly the maximum EPE is lower for a similar reason. At the start of the contract, a contract with a low value would not be exercised as it would have a higher expected future value than the current value. In other words, the early exercise feature diminishes the likelihoods of either very extreme or very low pay-off values and thus decreases maximum PFEs and EPEs when compared to case where the option is exercised at a random time step as the direct method indicates.

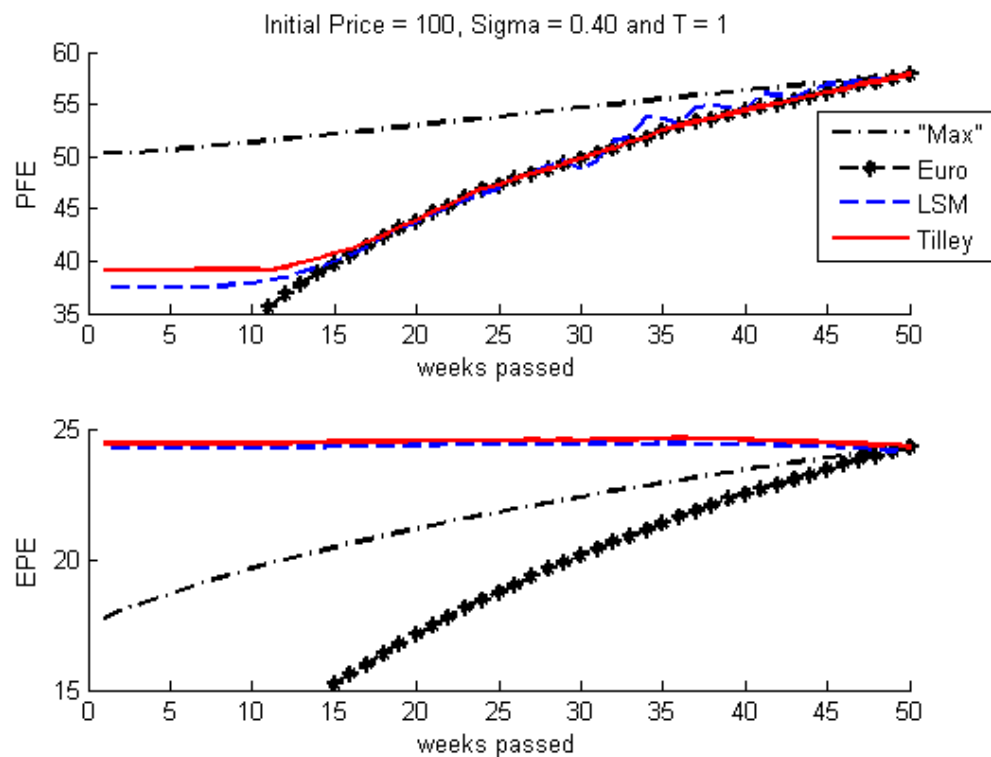


FIGURE B.4: EPE and PFE for an American Call Option simulated using Tilley Bundling and LSM.

B.4 Future Work

The two methods described here have been in extensive use for twenty years but most published applications focus on pricing the value of a derivative rather than the attached risk. In future I hope to investigate more hybrid methods such as applying piecewise regression (as in essence Tilley Bundling is a piecewise regression with the regression function being a constant), logistic regression applied to both methods, and multi-dimensional sorting under a Tilley bundling scheme (such as KNN) for more exotic derivatives.

Appendix C

Data Adjustments

Matches for which odds were not available:

- Leiria vs. Nacional (Portuguese Prem. Liga) on 12th May 2012. The reason for this omission is that Leiria was facing financial difficulties following the squad not being paid for several months.
- All 2009/10 Mouscron (Belgian Pro League team) matches as the team was declared bankrupt.
- All 2009/10 Belgian Pro League playoffs. Playoffs were only used for this season.

Matches for which match odds were not available:

- Messina vs. Lazio (Italian Serie A) match on 18 January 2006. The 1X2 odds for this match were excluded as average odds implied the existence arbitrage.

Matches for which overs/unders odds were not available.

- Cagliari vs Fiorentina (Italian Serie A) match on 22nd March 2006.
- Porto vs. Leiria (Portuguese Prem. Liga) match on 25th August 2006.*
- Sporting Lisbon vs. Boavista (Portuguese Prem. Liga) on 26th August 2006.*
- Beira Mar vs. Aves (Portuguese Prem. Liga) on 26th August 2006.
- Burnley vs. Stoke (English Championship) on 23rd January 2007.
- Rangers vs. Falkirk (Scottish Premiership) on 17th January 2009.*

- Aberdeen vs. Celtic (Scottish Premiership) on 18th January 2009.*
- Kilmarnock vs. Falkirk (Scottish Premiership) on 9th May 2009.*
- St Mirren vs Motherwell (Scottish Premiership) on 9th May 2009.*
- Inverness vs. Hamilton (Scottish Premiership) on 10th May 2009.
- Lazio vs. Reggina (Italian Seria A) on 20th May 2009.*
- Deportivo La Coruna vs. Valencia (Spanish La Liga) on 20th December 2009.
- Benfica vs. Leiria (Portuguese Prem. Liga) on 5th May 2012.
- Real Madrid vs. Villareal (Spanish La Liga) on 8th February 2014.

Matches for which match odds were not available but approximated using Bet365.com odds

- All matches marked *
- Biera Mar vs. Aves Pacos (Portuguese Prem. Liga) on 26th August 2006.
- Sporting Braga vs. Ferreira (Portuguese Prem. Liga) on 26th August 2006.

Appendix D

Comparison of Weibull Parameters with past Euro and African Cup of Nations Games

D.1 Introduction

The purpose of this appendix is to compare Poisson and Weibull parameters in estimating goal scored and goal differences for the EURO 2012, African Cup of Nations 2013 and 2014 World Cup soccer tournaments. Euro games are played every four years while African Cup of Nations games have been played on even years. From 2013, the latter will be played every odd year in order not to clash with the World Cup.

The inclusion of the African Cup of Nations is justified as the model was also intended to be applied to this tournament. However the liquidity in the market was too low to find any reasonable odds to make comparisons with. Moreover not all the odds for all the matches were offered at the outset of the tournament.

D.2 Data

EURO All group matches played during the 1996, 2000, 2004 and 2008 tournaments (96 matches).

African Cup of Nations Group matches, in which four teams played, from 1996 to 2012 (204 matches).

World Cup All group matches played during the 1998, 2002, 2006 and 2010 tournament (192 matches)

The rationale for using matches since 1996 is that these were the first tournaments for which a win contributed three points. Prior to 1996, a win during these short tournaments contributed two points. Both tournaments consist of four groups with four teams within each group. As all teams play each other during the group stages, there are six matches per group.

Only data for three groups was collected for The African Cup of Nations in 1996 and 2010 as Nigeria and Togo had withdrawn from each tournament respectively leaving a group with three teams. Having three teams may have changed the dynamics of the group.

D.3 Data Treatment

The Weibull distribution is a continuous fit to the goal difference and number of goals scored, which are discrete in nature. Therefore data was slightly altered such as to estimate the goals scored by the losing team plus 0.5. In cases of a win, the best fit for goal differences is estimated on the goal differences less 0.5.

The Poisson distribution assumes the probability of zero as likely. The average parameter of the number of goals scored by the losing team, given a particular score difference, was estimated directly. However, in cases of a win, the goal difference average was evaluated as number of goals less one in order to obtain the likelihoods of obtaining a zero. Therefore if four matches did not end in a draw, and the goal difference was +1, +1, +2, +1 then $\lambda = 0.25$.

After having evaluated the parameters for the two distributions, the two models are back tested by comparing the expected number of matches with the given goals scored or goal difference with the actual values using Chi-Square Testing.

There are two issues related to these tests:

- The last category was set as the highest number of goals scored or goal difference. For example, the goal difference for matches that did not end in a draw was between 1 and 5 for the European Cup (with no matches with a goal difference of four). In this case, the expected matches ending with +1, +2, +3, +4 and ≥ 5 were evaluated.
- This leads to expected values being lower than 5, which is a restriction of Chi Square.

TABLE D.1: Chi-Square Values for the EURO¹

	Poisson	Weibull	df	# matches
Goal Diff/not a draw	7.38	7.45	4	74
No of Goals scored by better team if:				
draw	2.20	1.99	3	22
one goal diff	1.41	0.10	3	41
two goal diff	0.56	1.75	2	19
three goal diff	0.22	0.84	1	13

TABLE D.2: Chi-Square Values for the African Cup of Nations²

	Poisson	Weibull	df	# matches
Goal Diff/not a draw	2.07	2.35	4	151
No of Goals scored by better team if:				
draw	0.60	1.71	4	53
one goal diff	4.70	1.36	4	77
two goal diff	0.39	0.79	3	47
three goal diff	0.56	1.75	2	19

TABLE D.3: Chi-Square Values for the World Cup³

	Poisson	Weibull	df	# matches
Goal Diff/not a draw	6.38	3.84	4	137
No of Goals scored by better team if:				
draw	4.65	1.02	4	55
one goal diff	1.77	0.75	3	70
two goal diff	0.96	0.10	3	41
three goal diff	0.94	1.78	2	14

D.4 Results

In all cases, there were no significant differences between expected and observed values. The Poisson model is a slightly better fit for eight from fifteen cases as shown by a lower Chi-Square Value in Tables D.1, D.2 and D.3. However there are many cases for which the Weibull is a far superior fit to the data, such as the predicting the number of goals scored given that goal difference is 1, justifying the use of a Weibull Distribution.

¹One game ended 5-0.

²Seven matches ended with a goal difference of 4 and 1 game with a goal difference of 5. In half of these matches, the losing team scored one goal. No goal was scored by the losing team in the other four matches.

³Twelve matches ended with a goal difference greater than 4 goals [six 4-0, two 5-0, 6-1, 6-0, 7-0, 8-0]

Appendix E

Sample Output for Group D

Table [E.1](#) shows the calculation of profit for Group D using $\alpha = 0$. The teams in this group were England (EN), France (FR), Sweden (SW) and Ukraine (UK). At each time point (TP), the simulated odds from the model (Sim Odds) are tabulated if higher than the market odds (Mrkt Odds). The Kelly criterion is calculated using the standard method in the column KC. The stake, as a percentage of 100,000 units, on the outcome (RB - ‘Rock Bottom’, SF - ‘Straight Forecast’, DF - ‘Dual Forecast’, GW - ‘Group Winner’, TQ - ‘To Qualify’) is determined by the adjusted Kelly Criterion (FK). Outcomes that occurred are given in italics.

TABLE E.1: Calculation of for Group D using $\alpha = 0$

TP	M	O	Sim Odds	Mrkt Odds	KC	Prev All	FK	Stake (in \$00s)
0	RB	EN	5.57	5.8	0.9%		0.9%	\$8.5
0	RB	FR	6.82	7.8	2.1%		2.1%	\$21.2
0	SF	EN/SW	12.92	14	0.6%		0.6%	\$6.4
0	SF	EN/UK	11.76	13	0.9%		0.9%	\$8.8
0	SF	SW/EN	17.85	21	0.9%		0.9%	\$8.8
0	SF	SW/FR	15.54	16.5	0.4%		0.4%	\$4.0
0	SF	SW/UK	27.47	38	1.0%		1.0%	\$10.4
0	SF	UK/EN	16.55	18	0.5%		0.5%	\$5.2
0	SF	UK/FR	13.82	16	1.0%		1.0%	\$10.5
0	SF	UK/SW	26.12	34	0.9%		0.9%	\$9.1
0	DF	EN/SW	7.50	8.2	1.3%	1.5%	0.0%	\$0.0
0	DF	EN/UK	6.87	7.6	1.6%	1.4%	0.2%	\$2.0
0	DF	FR/SW	6.08	6.2	0.4%	0.4%	0.0%	\$0.0
0	DF	FR/UK	5.64	5.9	1.0%	1.0%	0.0%	\$0.0
0	DF	SW/UK	13.39	17.5	1.9%	2.0%	0.0%	\$0.0
0	GW	EN	3.30	3.35	0.6%	1.5%	0.0%	\$0.0
0	GW	SW	6.38	6.8	1.1%	2.3%	0.0%	\$0.0
0	TQ	SW	2.68	2.78	2.0%	3.9%	0.0%	\$0.0
0	TQ	UK	2.51	2.76	5.5%	4.6%	0.9%	\$9.3
1	GW	SW	5.41	5.7	1.1%	2.3%	0.0%	\$0.0
1	GW	UK	5.15	5.4	1.1%	2.5%	0.0%	\$0.0
1	TQ	SW	2.61	2.74	2.9%	3.9%	0.0%	\$0.0
1	TQ	UK	2.52	2.66	3.2%	5.5%	0.0%	\$0.0
2	RB	FR	6.68	8	2.8%	2.1%	0.7%	\$7.1
2	SF	UK/EN	7.21	7.6	0.8%	0.5%	0.3%	\$3.0
2	SF	UK/SW	11.64	14.5	1.8%	0.9%	0.9%	\$9.1
2	DF	SW/UK	10.33	11	0.6%	2.9%	0.0%	\$0.0
2	GW	UK	2.73	3.2	7.9%	3.7%	4.2%	\$42.0
2	TQ	SW	6.12	6.5	1.1%	4.8%	0.0%	\$0.0
2	TQ	UK	1.58	1.74	14.0%	10.9%	3.0%	\$30.4
3	RB	SW	1.44	1.48	5.5%	0.0%	5.5%	\$54.8
3	RB	UK	11.36	13.5	1.5%	0.0%	1.5%	\$15.0
3	SF	FR/UK	3.52	6.4	15.2%	0.0%	15.2%	\$151.9
3	TQ	FR	1.09	1.13	25.1%	16.6%	8.5%	\$84.7
4	SF	EN/FR	2.95	3.1	2.4%	0.0%	2.4%	\$23.6
4	GW	EN	2.95	3.05	1.6%	3.9%	0.0%	\$0.0
4	GW	UK	8.99	9.2	0.3%	7.9%	0.0%	\$0.0
4	GW	UK	9.2	0.2%	5.9%	0.0%	0.0%	\$0.0

References

- Acerbi, C., & Tasche, D. (2002). Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes*, 31(2), 379–388. Retrieved from <http://dx.doi.org/10.1111/1468-0300.00091> doi: 10.1111/1468-0300.00091
- Adams, R., & Ferreira, D. (2010). Moderation in groups: Evidence from betting on ice break-ups in alaska. *The Review of Economic Studies*, 77(3), 882–913.
- Agius, M. (2014, 11 6). *Remote gaming is second biggest contributor to gdp*. Retrieved 2014-11-11, from http://www.maltatoday.com.mt/news/national/45886/remote_gaming_is_second_biggest_contributor_to_gdp#.VHivkfmsUm4
- Ali, M. M. (1977). Probability and utility estimates for racetrack bettors. *The Journal of Political Economy*, 803–815.
- Altman, D. G. (1990). *Practical statistics for medical research*. CRC press.
- Archontakis, F., & Osborne, E. (2007). Playing it safe? A Fibonacci strategy for soccer betting. *Journal of Sports Economics*, 8(3), 295. Retrieved from <http://jse.sagepub.com/cgi/content/abstract/8/3/295>
- Armatas, V., Yiannakos, A., & Sileloglou, P. (2007). Relationship between time and goal scoring in soccer games: Analysis of three world cups. *International Journal of Performance Analysis in Sport*, 7(2), 48–58.
- Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228. Retrieved from <http://dx.doi.org/10.1111/1467-9965.00068> doi: 10.1111/1467-9965.00068
- Asch, P., Malkiel, B. G., & Quandt, R. E. (1982). Racetrack betting and informed behavior. *Journal of Financial Economics*, 10(2), 187–194.
- Atkinson, G., Nichols, M., & Oleson, T. (2000). The menace of competition and gambling deregulation. *Journal of economic issues*, 34(3), 621–634.
- Avery, C., & Chevalier, J. (1999). Identifying investor sentiment from price paths: The case of football betting (digest summary). *Journal of Business*, 72(4), 493–520.

- Barber, B., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, 55(2), 773-806. doi: 10.1111/0022-1082.00226
- Barber, B., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1), 261-292. Retrieved from <http://EconPapers.repec.org/RePEc:oup:qjecon:v:116:y:2001:i:1:p:261-292>.
- Ben-Naim, E., Vazquez, F., & Redner, S. (2007). What is the most competitive sport? *J. Korean Phys. Soc.*, 50(1), 124-126. Retrieved from <http://dx.doi.org/10.1080/135048500351672> doi: 10.1080/135048500351672
- Berkowitz, J., & O'Brien, J. (2002). How accurate are value-at-risk models at commercial banks? *The Journal of Finance*, 57(3), 1093-1111. Retrieved from <http://dx.doi.org/10.1111/1540-6261.00455> doi: 10.1111/1540-6261.00455
- Bernile, G., & Lyandres, E. (2011). Understanding investor sentiment: The case of soccer. *Financial Management*, 40(2), 357-380.
- Bernstein, P. L. (1996). *Against the gods: The remarkable story of risk*. Wiley New York.
- Bezzina, F. H., & Grima, S. (2012). Exploring factors affecting the proper use of derivatives. *Managerial Finance*, 38(4), 414-435. Retrieved from <http://dx.doi.org/10.1108/03074351211207554> doi: 10.1108/03074351211207554
- Bird, R., & McCrae, M. (1987). Tests of the efficiency of racetrack betting using bookmaker odds. *Management Science*, 33(12), 1552-1562.
- Bittner, E., Nußbaumer, A., Janke, W., & Weigel, M. (2009). Football fever: goal distributions and non-gaussian statistics. *The European Physical Journal B - Condensed Matter and Complex Systems*, 67(3), 459-471. Retrieved from <http://dx.doi.org/10.1140/epjb/e2008-00396-1>
- Bittner, E., Nussbaumer, A., Janke, W., & Weigel, M. (2007). Self-affirmation model for football goal distributions. *Europhys. Lett.*, 78, 58002.
- Branvall, R., Cortis, D., & Li, J. (2014). *Pricing risk through simulation: Revisiting tilley bundling and least square monte carlo methods* (Vol. 4) (No. 1). Philadelphia, USA. Retrieved from <https://prezi.com/0vuwyg6dshcg/arc-169/>
- Braun, S., & Kvasnicka, M. (2008). Against all odds?-national sentiment and wagering on european football. *National Sentiment and Wagering On European Football (March 2008)*. *Ruhr Economic Paper*(42).

- Brown, A. (2012). Evidence of in-play insider trading on a uk betting exchange. *Applied Economics*, 44(9), 1169-1175. Retrieved from <http://dx.doi.org/10.1080/00036846.2010.537644> doi: 10.1080/00036846.2010.537644
- Brown, K. H., & Abraham, F. J. (2002). Testing market efficiency in the major league baseball over-under betting market. *Journal of Sports Economics*, 3(4), 311-319.
- Buchen, P. W., & Grant, A. (2012). A comparison of simultaneous kelly betting strategies. *The Journal of Gambling Business and Economics*, 6(2), 1-28. Retrieved from <http://www.ingentaconnect.com/content/ubpl/jgbe/2012/00000006/00000002/art00001>
- Busche, K., & Hall, C. D. (1988). An exception to the risk preference anomaly. *Journal of business*, 337-346.
- Cain, M., Law, D., & Peel, D. (2000). The favourite-longshot bias and market efficiency in uk football betting. *Scottish Journal of Political Economy*, 47(1), 25-36. Retrieved from <http://dx.doi.org/10.1111/1467-9485.00151> doi: 10.1111/1467-9485.00151
- Cain, M., Law, D., & Peel, D. (2003). The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research*, 55(3), 263-273. Retrieved from <http://EconPapers.repec.org/RePEc:bla:buecrs:v:55:y:2003:i:3:p:263-273>
- Cain, M., Law, D., & Peel, D. A. (2008). Bounded cumulative prospect theory: some implications for gambling outcomes. *Applied Economics*, 40(1), 5-15.
- Chincarini, L., Contreras, C., et al. (2010). Home bias and international betting markets: Can institutional constraints and behavioral biases lead to arbitrage profits? *Journal of Gambling Business and Economics*, 4(3), 20.
- Choi, M., O'Donoghue, P. G., & Hughes, M. (2008). The identification of an optimal set of performance indicators for real-time analysis using principle components analysis. In A. Hokelmann & M. Brummond (Eds.), *Performance analysis of sport viii* (p. 295-301). Magdeburg: Universität Magdeburg, Department of Sport Science.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. Retrieved from <http://epm.sagepub.com/content/20/1/37.short> doi: 10.1177/001316446002000104
- Colander, D., Goldberg, M., Haas, A., Juselius, K., Kirman, A., Lux, T., & Sloth, B. (2009). The financial crisis and the systemic failure of the economics profession. *Critical Review*, 21(2-3), 249-267. Retrieved from <http://dx.doi.org/10.1080/08913810902934109> doi: 10.1080/08913810902934109

- Coleman, L. (2004). New light on the longshot bias. *Applied Economics*, 36(4), 315–326.
- Coleman, L. (2013). Just how serious is insider trading? an evaluation using thoroughbred wagering markets. *The Journal of Gambling Business and Economics*, 1(1), 31–55.
- Cortis, D. (2015a). Expected values and variances in bookmaker payouts: A theoretical approach towards setting limits on odds. *Journal of Prediction Markets*, 9(1), 1–14.
- Cortis, D. (2015b). A public interpretation of brazil’s world cup performance. In *Sixth international conference on sport and society*.
- Cortis, D. (2015c, August). Reserving for the dark side. In S. Broverman, A. Badescu, A. Courtemanche, S. Jaimungal, S. Lin, & V. Zhang (Eds.), *50th actuarial research conference*.
- Cortis, D. (2015d). A safe bet - or is it? In *Mathsport international 2015 conference*.
- Cortis, D., & Briguglio, M. L. (under review). A public (mis)interpretation of brazil’s world cup performance. *The Journal of Sports and Society*.
- Cortis, D., Hales, S., & Bezzina, F. (2013). Profiting on inefficiencies in betting derivative markets: The case of UEFA EURO 2012. *Journal of Gambling Business and Economics*, 7(1), 39–51. Retrieved from <http://EconPapers.repec.org/RePEc:buc:jgbeco:v:7:y:2013:i:1:p:39-51>
- Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6), 2045–2073.
- Crafts, N. F. (1985). Some evidence of insider knowledge in horse race betting in britain. *Economica*, 52(207), 295–304.
- Croxson, K., & James Reade, J. (2014). Information and efficiency: Goal arrival in soccer betting. *The Economic Journal*, 124(575), 62–91. Retrieved from <http://dx.doi.org/10.1111/ecoj.12033> doi: 10.1111/ecoj.12033
- Dambrauskaite, J., Haidar, H., Minsky, B., & Tang, Q. (2012). Quantitative insight into management of funds of hedge funds and consequences on fund alpha performance. *Reconsidering Funds of Hedge Funds: The Financial Crisis and Best Practices in UCITS, Tail Risk, Performance, and Due Diligence*, 331.
- Danielsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking and Finance*, 26(7), 1273–1296. Retrieved from <http://EconPapers.repec.org/RePEc:eee:jbfina:v:26:y:2002:i:7:p:1273-1296>

- David, F. N. (1962). Games, gods and gambling: The origins and history of probability and statistical ideas from the earliest times to the newtonian era.
- Deloitte. (2013). *The full picture (second edition): Measuring the economic contribution of the british betting industry*. The Association of British Bookmakers.
- Demir, E., Danis, H., & Rigoni, U. (2012). Is the soccer betting market efficient? a cross-country investigation using the fibonacci strategy. *Journal of Gambling Business and Economics*, 6(2), 29-49. Retrieved from <http://www.ubplj.org/index.php/jgbe/article/view/580>
- Deschamps, B., & Gergaud, O. (2012). Efficiency in betting markets: evidence from english football. *The Journal of Prediction Markets*, 1(1), 61–73.
- Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., & Vyncke, D. (2006). Risk measures and comonotonicity: A review. *Stochastic Models*, 22(4), 573-606. Retrieved from <http://dx.doi.org/10.1080/15326340600878016> doi: 10.1080/15326340600878016
- Direr, A. (2013). Are betting markets efficient? evidence from european football championships. *Applied Economics*, 45(3), 343-356. Retrieved from <http://dx.doi.org/10.1080/00036846.2011.602010> doi: 10.1080/00036846.2011.602010
- Dixon, M. J., & Coles, S. G. (1997). 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. Retrieved from <http://dx.doi.org/10.1111/1467-9876.00065> doi: 10.1111/1467-9876.00065
- Dobson, S., Goddard, J. A., & Dobson, S. (2001). *The economics of football*. Cambridge University Press Cambridge.
- Doff, R. (2008). A critical analysis of the solvency ii proposals. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 33(2), 193-206. Retrieved from <http://EconPapers.repec.org/RePEc:pal:gpprii:v:33:y:2008:i:2:p:193-206>
- Easton, S. A., & Uylangco, K. (n.d.). An examination of in-play sports betting using one-day cricket matches. *Journal of Prediction Markets*, 1, 93–109.
- Egon, F., Verbeek, E., & Nüesch, S. (2009). Inter-market arbitrage in sports betting. *NCER Working Paper Series: Sportometrics Program*. Retrieved from <http://www.ncer.edu.au/papers/documents/WPNo48.pdf>
- Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, 12(5), 391–396.

- European Gaming and Betting Association. (2014). *Market reality factsheet*. European Gaming and Betting Association.
- Fawcett, M., et al. (2014). Are the odds in your favour?: A film-production approach to exploring sports-betting advertising in the classroom.
- Federolf, P., Reid, R., Gilgien, M., Haugen, P., & Smith, G. (2012). The application of principal component analysis to quantify technique in sports. *Scand J Med Sci Sports*. Retrieved from dx.doi.org/10.1111/j.1600-0838.2012.01455.x
- FIFA. (n.d.). *Men's ranking procedure*. Retrieved 2016-01-15, from <http://www.fifa.com/fifa-world-ranking/procedure/men.html>
- FIFA. (2011). *Regulations: 2014 FIFA world cup* (Tech. Rep.). Retrieved from http://www.fifa.com/mm/document/tournament/competition/01/47/38/17/regulationsfwcbrazil2014_update_e_neutral.pdf
- Forrest, D., & Simmons, R. (2008). Sentiment in the betting market on spanish football. *Applied Economics*, 40(1), 119–126.
- Franck, E., Verbeek, E., & Nüesch, S. (2010). Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting*, 26(3), 448–459.
- Franck, E., Verbeek, E., & Nüesch, S. (2011). Sentimental preferences and the organizational regime of betting markets. *Southern Economic Journal*, 78(2), 502–518.
- Fu, M. C., Laprise, S. B., Madan, D. B., Su, Y., & Wu, R. (2001). Pricing american options: A comparison of monte carlo simulation approaches. *Journal of Computational Finance*, 4(3), 39–88.
- Gandar, J. M., Zuber, R. A., & Johnson, R. S. (2001). Searching for the favourite-longshot bias down under: an examination of the new zealand pari-mutuel betting market. *Applied Economics*, 33(13), 1621–1629.
- Gandar, J. M., Zuber, R. A., & Johnson, R. S. (2004). A reexamination of the efficiency of the betting market on national hockey league games. *Journal of Sports Economics*, 5(2), 152–168.
- Gaviria, A. (2000). Is soccer dying? a time series approach. *Applied Economics Letters*, 7(4), 275–278. Retrieved from <http://dx.doi.org/10.1080/135048500351672> doi: 10.1080/135048500351672
- Geoffrion, A. M. (1976). The purpose of mathematical programming is insight, not numbers. *Interfaces*, 7(1), 81–92. Retrieved from <http://dx.doi.org/10.1287/inte.7.1.81> doi: 10.1287/inte.7.1.81

- Gil, R. G. R., & Levitt, S. D. (2012). Testing the efficiency of markets in the 2002 world cup. *The Journal of Prediction Markets*, 1(3), 255–270.
- Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering (Stochastic Modelling and Applied Probability) (v. 53)* (1st ed.). Springer. Hardcover. Retrieved from <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387004513>
- Glenn, W. A., & David, H. A. (1960). Ties in paired comparison experiments using a modified thurstone-mosteller model. *Biometrics*, 16, 86-109.
- Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21(2), 331 - 340. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0169207004000676> doi: <http://dx.doi.org/10.1016/j.ijforecast.2004.08.002>
- Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23(1), 51–66. Retrieved from <http://dx.doi.org/10.1002/for.877> doi: 10.1002/for.877
- Goldman Sachs. (2014, May 30). *The world cup and economics 2014: A statistical model of the 2014 world cup*. (Tech. Rep.). Goldman Sachs.
- Greenhough, J., Birch, P., Chapman, S., & Rowlands, G. (2002). Football goal distributions and extremal statistics. *Physica A: Statistical Mechanics and its Applications*, 316(1–4), 615 - 624. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0378437102010300> doi: [http://dx.doi.org/10.1016/S0378-4371\(02\)01030-0](http://dx.doi.org/10.1016/S0378-4371(02)01030-0)
- Greenspan, A. (2008, 3 16). *We will never have a perfect model of risk*. Retrieved 2015-07-10, from <http://www.ft.com/intl/cms/s/0/edbdbc6f-f360-11dc-b6bc-0000779fd2ac.html#axzz3fTYm96IN>
- Gregory, J. (2010). *Counterparty credit risk: The new challenge for global financial markets*. John Wiley & Sons. Retrieved from http://books.google.com.mt/books?id=WZ_vbGGx1z4C
- Gréhaigne, J., Mahut, B., & Fernandez, A. (2001). Qualitative observation tools to analyse soccer. *International Journal of Performance Analysis in Sport*, 1(1), 52-61. Retrieved from <http://www.ingentaconnect.com/content/uwic/ujpa/2001/00000001/00000001/art00004>
- Griffith, R. M. (1949). Odds adjustments by american horse-race bettors. *The American Journal of Psychology*, 62(2), 290–294.

- Griffiths, M. (2004). Betting your life on it. *BMJ*, 329(7474), 1055–1056.
- Griffiths, M. (2005). Online betting exchanges: A brief overview. *Youth Gambling International Newsletter*, 238–327.
- H2 Gambling Capital. (2014, 11). *Global gambling data and intelligence subscription service, global summary dataset*.
- Hausch, D. B., Ziemba, W. T., & Rubinstein, M. (1981). Efficiency of the market for racetrack betting. *Management science*, 27(12), 1435–1452.
- Henery, R. J. (1981). Permutation Probabilities as Models for Horse Races. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(1), 86–91. Retrieved from <http://dx.doi.org/10.2307/2985154> doi: 10.2307/2985154
- Heuer, A., Müller, C., & Rubner, O. (2010, February 01). Soccer: Is scoring goals a predictable Poissonian process? *EPL (Europhysics Letters)*, 89(3), 38007+. Retrieved from <http://dx.doi.org/10.1209/0295-5075/89/38007> doi: 10.1209/0295-5075/89/38007
- Hogg, R. (2013). Egr report: Sports betting. In R. Roocroft (Ed.), (p. 30). London: Pageant Gaming Media.
- Hudson, J. (2014, 1 12). *Uk sports betting market continues to grow driven by gen y and mobile betting apps*. Retrieved 2016-03-03, from <http://www.favourit.com/en/content/uk-sports-betting-stats-mobile>
- Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*, 20(10), 739–754. Retrieved from <http://dx.doi.org/10.1080/026404102320675602> (PMID: 12363292) doi: 10.1080/026404102320675602
- Hull, J. C. (1993). *Options, futures and other derivatives (2nd edition)*. Prentice Hall.
- Hull, J. C. (2012). *Options, futures and other derivatives (8th edition)*. Prentice Hall.
- Hurley, W., & McDonough, L. (2013). Imperfect market-maker competition, heterogeneous expectations, and the favourite-longshot bias in wagering markets. *The Journal of Gambling Business and Economics*, 1(1), 3–12.
- Hvattum, L. M. (2013). Analyzing information efficiency in the betting market for association football league winners. *The Journal of Prediction Markets*, 7(2), 55–70.
- Hvattum, L. M., & Arntzen, H. (2010). Using {ELO} ratings for match result prediction in association football. *International Journal of Forecasting*, 26(3),

- 460 - 470. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0169207009001708> doi: <http://dx.doi.org/10.1016/j.ijforecast.2009.10.002>
- iGaming Business. (2013, 9 18). *Sky bet suffers“disastrous” champions league opening night*. Retrieved from <http://www.igamingbusiness.com/news/sky-bet-suffers-disastrous-champions-league-opening-night>
- iGaming Business. (2014, 1 15). *Epl weekend results cost uk bookmakers ‘up to £30m’*. Retrieved from <http://www.igamingbusiness.com/news/epl-weekend-results-cost-uk-bookmakers-30m>
- Jones, P., Hillier, D., Turner, D., & Comfort, D. (2004). Betting on the exchanges: Changing customer relationships in the sports betting market in the uk. *Management Research News*, 27(1/2), 95-103.
- Jones, P. D., James, N., & Mellalieu, S. D. (2004). Possession as a performance indicator in soccer. *International Journal of Performance Analysis in Sport*, 4(1), 98-102. Retrieved from <http://www.ingentaconnect.com/content/uwic/ujpa/2004/00000004/00000001/art00010>
- Kain, K. J., & Logan, T. D. (2014). Are sports betting markets prediction markets? evidence from a new test. *Journal of Sports Economics*, 15(1), 45-63.
- Kelly, J. L. J. (1956). A new interpretation of information rate. *Bell System Technical Journal*, 35, 917-926.
- Keogh, F., & Rose, G. (2013). *Football betting - the global gambling industry worth billions*. Retrieved 2010-11-18, from <http://www.bbc.co.uk/sport/0/football/24354124>
- Koch, A., & Shing, H.-F. (2008). Bookmaker and pari-mutuel betting: Is a (reverse) favourite-longshot bias built-in? *Journal of Prediction Markets*, 2(2), 29-50.
- Koning, R. H., Koolhaas, M., Renes, G., & Ridder, G. (2003, July). A simulation model for football championships. *European Journal of Operational Research, Elsevier*, 148(2), 268-276. Retrieved from <http://ideas.repec.org/a/eee/ejores/v148y2003i2p268-276.html>
- Koning, R. H., & van Velzen, B. (2009). Betting exchanges: The future of sports betting? *International Journal of Sport Finance*, 4(1), 42-62. Retrieved from <http://EconPapers.repec.org/RePEc:jsf:intjsf:v:4:y:2009:i:1:p:42-62>
- Kuk, A. Y. C. (1995). Modelling Paired Comparison Data with Large Numbers of Draws and Large Variability of Draw Percentages Among Players. *Journal of the*

- Royal Statistical Society. *Series D (The Statistician)*, 44(4). Retrieved from <http://dx.doi.org/10.2307/2348900> doi: 10.2307/2348900
- Kukuk, M., Winter, S., et al. (2008). An alternative explanation of the favorite-longshot bias. *The Journal of Gambling Business and Economics*, 2(2), 79–96.
- Kupiec, P. (1995). *Techniques for verifying the accuracy of risk measurement models* (Finance and Economics Discussion Series No. 95-24). Board of Governors of the Federal Reserve System (U.S.). Retrieved from <http://EconPapers.repec.org/RePEc:fip:fedgfe:95-24>
- Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics*, 32(11), 1353–1363.
- Kwak, Y. H., & Anbari, F. T. (2006). Benefits, obstacles, and future of six sigma approach. *Technovation*, 26(5–6), 708 – 715. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0166497204001828> doi: <http://dx.doi.org/10.1016/j.technovation.2004.10.003>
- Law, D., & Peel, D. A. (2002). Insider trading, herding behaviour and market plungers in the british horse-race betting market. *Economica*, 69(274), 327–38. Retrieved from <http://EconPapers.repec.org/RePEc:bla:econom:v:69:y:2002:i:274:p:327-38>
- Lees, A., & Nolan, L. (1998). The biomechanics of soccer: A review. *Journal of Sports Sciences*, 16(3), 211–234. Retrieved from <http://dx.doi.org/10.1080/026404198366740> (PMID: 9596356) doi: 10.1080/026404198366740
- Lessmann, S., Sung, M., & Johnson, J. E. (2009). Identifying winners of competitive events: A svm-based classification model for horserace prediction. *European Journal of Operational Research*, 196(2), 569–577. Retrieved from <http://eprints.soton.ac.uk/51338/>
- Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets?*. *The Economic Journal*, 114(495), 223–246. Retrieved from <http://dx.doi.org/10.1111/j.1468-0297.2004.00207.x> doi: 10.1111/j.1468-0297.2004.00207.x
- Linsmeier, T. J., & Pearson, N. D. (1996, September). *Risk measurement: An introduction to value at risk* (Finance No. 9609004). EconWPA. Retrieved from <http://ideas.repec.org/p/wpa/wuwpfi/9609004.html>
- Longstaff, F. A., & Schwartz, E. S. (2001). Valuing american options by simulation: a simple least-squares approach. *Review of Financial studies*, 14(1), 113–147.

- Luhtanen, P., Belinskij, A., Häyrynen, M., & Väättinen, T. (2001). A comparative tournament analysis between the euro 1996 and 2000 in soccer. *International Journal of Performance Analysis in Sport*, 1(1), 74-82. Retrieved from <http://www.ingentaconnect.com/content/uwic/ujpa/2001/00000001/00000001/art00006>
- Malkiel, B. G. (2007). *A random walk down wall street: The time-tested strategy for successful investing*.
- Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Markov, A. (1906). Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fiziko-matematicheskogo obshchestva pri Kazanskom universitete*, 15(135-156), 18.
- Markov, A. (1971). Extension of the limit theorems of probability theory to a sum of variables connected in a chain.
- Markowitz, H. (1952). Portfolio selection. *The Journal of finance*, 7(1), 77-91.
- McHale, I., & Davies, S. (2007). Statistical thinking in sports. In J. Albert & R. H. Koningt (Eds.), (p. 77-90). Florida: CRC Press.
- McNulty, P. (2014, 6 27). *World cup 2014: Brazil favourites but netherlands the ones to watch*. Retrieved 2016-01-10, from <http://www.bbc.co.uk/sport/0/football/28067841>
- Monkovic, T. (2014, 6). *Drawing lots at world cup? there must be a better way*. Retrieved 2016-01-03, from http://www.nytimes.com/2014/06/24/upshot/drawing-lots-at-world-cup-there-must-be-a-better-way.html?_r=0
- Mowen, J. C. (2000). *The 3m model of motivation and personality: Theory and empirical applications to consumer behavior*. Springer Science & Business Media.
- Newall, P. W. (2015). How bookies make your money. *Judgment and Decision Making*, 10(3), 225.
- O'Donoghue, P. (2008). Principal components analysis in the selection of key performance indicators in sport. *International Journal of Performance Analysis in Sport*, 8(3), 145-155. Retrieved from "<http://www.ingentaconnect.com/content/uwic/ujpa/2008/00000008/00000003/art00015>"
- O'Donoghue, P. (2013). *Statistics for sport and exercise studies: An introduction*. Taylor & Francis. Retrieved from <http://books.google.co.uk/books?id=s-1-y9oVLA0C>

- Paton, D., & Vaughan Williams, L. (2005). Forecasting outcomes in spread betting markets: can bettors use 'quarbs' to beat the book? *Journal of Forecasting*, 24(2), 139–154.
- Paul, R. J., & Weinbach, A. P. (2002). Market efficiency and a profitable betting rule evidence from totals on professional football. *Journal of Sports Economics*, 3(3), 256–263.
- Paul, R. J., & Weinbach, A. P. (2009). Are behavioral biases consistent across the atlantic?: the over/under market for european soccer. *Journal of Gambling Business & Economics*, 3(2).
- Paul, R. J., & Weinbach, A. P. (2012). Wagering preferences of nfl bettors: Determinants of betting volume. *Journal of Prediction Markets*, 6(1).
- Peel, D. (2008). Introduction: economics of betting markets. *Applied Economics*, 40(1), 1–3.
- Peel, D., & Thomas, D. (1992). The demand for football: Some evidence on outcome uncertainty. *Empirical Economics*, 17(2), 323–331. Retrieved from <http://dx.doi.org/10.1007/BF01206291> doi: 10.1007/BF01206291
- Piotrowski, E. W., & Schroeder, M. (2007). Kelly criterion revisited: optimal bets. *The European Physical Journal B - Condensed Matter and Complex Systems*, 57, 201–203. Retrieved from <http://dx.doi.org/10.1140/epjb/e2007-00126-3> (10.1140/epjb/e2007-00126-3)
- Pope, P. F., & Peel, D. A. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 323–341.
- Ridder, G., Cramer, J. S., & Hopstaken, P. (1994). Down to ten: Estimating the effect of a red card in soccer. *Journal of the American Statistical Association*, 89(427), 1124–1127. Retrieved from <http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1994.10476850> doi: 10.1080/01621459.1994.10476850
- Rotando, L. M., & Thorp, E. O. (1992). The Kelly Criterion and the Stock Market. *The American Mathematical Monthly*, 99(10), 922–931. Retrieved from <http://www.jstor.org/stable/2324484>
- Salmon, F. (2012). The formula that killed wall street. *Significance*, 9(1), 16–20.
- Sarmiento, H., Anguera, M. T., Pereira, A., Marques, A., Campaniço, J., & Leitão, J. (2014). Patterns of play in the counterattack of elite football teams - a mixed

- method approach. *International Journal of Performance Analysis in Sport*, 14(2), 411-427. Retrieved from <http://www.ingentaconnect.com/content/uwic/ujpa/2014/00000014/00000002/art00007>
- Sarmiento, H., Pereira, A., Matos, N., Campaniço, J., Anguera, M. T., & Leitão, J. (2013). English premier league, spanish la liga and italy's serie a; what's different? *International Journal of Performance Analysis in Sport*, 13(3), 773-789. Retrieved from <http://www.ingentaconnect.com/content/uwic/ujpa/2013/00000013/00000003/art00018>
- Sauer, R. D. (1998). The economics of wagering markets. *Journal of economic Literature*, 36(4), 2021-2064.
- Schnytzer, A., & Shilony, Y. (1995). Inside information in a betting market. *The Economic Journal*, 963-971.
- Schroeder, R. G., Linderman, K., Liedtke, C., & Choo, A. S. (2008). Six sigma: Definition and underlying theory. *Journal of Operations Management*, 26(4), 536 - 554. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0272696307000897> (Special Issue: Research in Supply Chain Quality) doi: <http://dx.doi.org/10.1016/j.jom.2007.06.007>
- Scott, F. A., & Gulley, O. D. (1995). Testing for efficiency in lotto markets. *Economic Inquiry*, 33(2), 175-188.
- Sharpe, W., & Alexander, G. (1995). *Investments*. Prentice Hall.
- Shin, H. S. (1993). Measuring the incidence of insider trading in a market for state-contingent claims. *Economic Journal*, 103(420), 1141-53. Retrieved from <http://EconPapers.repec.org/RePEc:ecj:econjl:v:103:y:1993:i:420:p:1141-53>
- Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1), 35-55.
- Smith, M. A. (2010). Exploring variations in odds bias across betting media. *Journal of Gambling Business & Economics*, 4(1).
- Snowball, B. (2015, 12 03). *Fifa consider expanding world cup to 40 teams*. Retrieved 2016-03-03, from http://www.eurosport.co.uk/football/world-cup/2014/fifa-could-expand-world-cup-to-40-teams_sto5013176/story.shtml
- Snowberg, E., & Wolfers, J. (2010, April). *Explaining the favorite-longshot bias: Is it risk-love or misperceptions?* (Working Paper No. 15923). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w15923>

- Spoorendonk, S., & Kristiansen, S. (2015, July). Optimise market margins using a conditional value-at-risk model in association football. In *Mathsport international 2015 conference*.
- Suhonen, N., & Linden, M. (2010). Unfair games, subjective probabilities, and favourite-longshot bias in finnish horse racing. *Journal of Gambling Business & Economics*, 4(2).
- Terrell, D., & Farmer, A. (1996). Optimal betting and efficiency in parimutuel betting markets with information costs. *The Economic Journal*, 846–868.
- Thaler, R. H., & Ziemba, W. T. (1988). Parimutuel betting markets: Racetracks and lotteries. *Journal of Economics Perspective*, 2, 161–174.
- Thiele, J. C., Kurth, W., & Grimm, V. (2014). Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using netlogo and 'r'. *Journal of Artificial Societies and Social Simulation*, 17(3). Retrieved from <http://EconPapers.repec.org/RePEc:jas:jasssj:2013-150-2>
- Thorp, E. O. (1997). The kelly criterion in blackjack, sports betting, and the stock market. In *The 10th international conference on gambling and risk taking*.
- Tilley, J. A. (1993). Valuing american options in a path simulation model. *Transactions of the Society of Actuaries*, 45(83), 104.
- Tirole, J. (1982). On the possibility of speculation under rational expectations. *Econometrica: Journal of the Econometric Society*, 1163–1181.
- Torres, C. A., & Goggin, G. (2014). Mobile social gambling: Poker's next frontier. *Mobile Media & Communication*, 2(1), 94–109.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207 - 232. Retrieved from <http://www.sciencedirect.com/science/article/pii/0010028573900339> doi: [http://dx.doi.org/10.1016/0010-0285\(73\)90033-9](http://dx.doi.org/10.1016/0010-0285(73)90033-9)
- Tversky, A., & Kahneman, D. (1974, September 27). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. Retrieved from <http://dx.doi.org/10.1126/science.185.4157.1124> doi: 10.1126/science.185.4157.1124
- UEFA. (2010). *Regulations of the UEFA european football championship 2010-12* (Tech. Rep.). Retrieved from http://www.uefa.com/multimediafiles/download/competitions/euro/91/87/57/918757_download.pdf

- Van Calster, B., Tim, S., & Van Huffel, S. (2008). The curse of scoreless draws in soccer: The relationship with a team's offensive, defensive, and overall performance. *Journal of Quantitative Analysis in Sports*, 4(1), 1-24. Retrieved from <http://EconPapers.repec.org/RePEc:bpj:jqsprt:v:4:y:2008:i:1:n:4>
- Vaughan Williams, L. (1999). Information efficiency in betting markets: A survey. *Bulletin of Economic Research*, 51(1), 1-39.
- Vaughan Williams, L., & Paton, D. (1997). Why Is There a Favourite-Longshot Bias in British Racetrack Betting Markets? *Economic Journal*, 107(440), 150-58. Retrieved from <http://ideas.repec.org/a/ecj/econjl/v107y1997i440p150-58.html>
- Vaughan Williams, L., Sung, M., Fraser-Mackenzie, P., Peirson, J., & Johnson, J. (forthcoming). Towards an understanding of the origins of the favourite-longshot bias: Evidence from online poker markets, a real-money natural laboratory. *Economica*.
- Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the european football betting market? evidence from arbitrage and trading strategies. *Journal of Forecasting*, 28(5), 426-444. Retrieved from <http://ideas.repec.org/a/jof/jforec/v28y2009i5p426-444.html>
- Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4), 934 - 943. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0169207014000533> doi: <http://dx.doi.org/10.1016/j.ijforecast.2014.02.008>
- Walls, D. W., & Busche, K. (2003). Broken odds and the favourite-longshot bias in parimutuel betting: a direct test. *Applied Economics Letters*, 10(5), 311-314.
- Winter, S., & Kukuk, M. (2006). Risk love and the favorite-longshot bias: Evidence from german harness horse racing. *Available at SSRN 940368*.
- Woodland, L. M., & Woodland, B. M. (2001). Market efficiency and profitable wagering in the national hockey league: Can bettors score on longshots? *Southern Economic Journal*, 983-995.
- Yamai, Y., & Yoshida, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking and Finance*, 29(4), 997 - 1015. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0378426604001499> (Risk Measurement) doi: <http://dx.doi.org/10.1016/j.jbankfin.2004.08.010>
- Zafiridis, N. (2014). When is a multiple bet better than a single? *Journal of Gambling Business and Economics*, 8(2), 1-15. Retrieved from <http://www.ubplj.org/index.php/jgbe/article/view/803>