KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

GAMBLERS' RISK OF RUIN AND OPTIMAL BETS ON FOOTBALL SCORES

BY

ANKOMAH, ROBERT KAY

BA. (Economics & Statistics)

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FUFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF M.PHIL ACTUARIAL SCIENCE

OCTOBER 2015

DECLARATION

I hereby declare that this thesis is my own work towards the award of the M.Phil degree (Actuarial Science) and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

ANKOMAH, Robert Kay

(20331563) ....

Student Signature Date

Certified by:

Dr. A.Y Omari-Sasu ....

Supervisor Signature Date

Certified by:

Prof. S. K AMPONSAH ....

Head of Department Signature Date

DEDICATION

I dedicate this work to my mother Mrs. Susana Baah Acquah.

ABSTRACT

Football scores which serves as the basis for wagering vis-a-vis the betting centres of *mybet.com sporting center, Supabet, Alfabet, Premier bet, Safari bet, Euro bet* in Ghana, is influenced by varied variables of both quantitative and qualitative making it uncertain, as the possibility of ruin stare gamblers. Model based on Poisson regression is adopted through bivariate poisson distribution to estimate the scoring rates of teams and probabilities of wins for both home and away using sports data of EPL 2013/2014 season and validated with 2014/2015 season. MLEs approach is adopted for parameter estimations. Red cards, Yellow cards, corner profile, shots on targets of teams as key scores determinants are analysed for purpose of determining the significant elements, while assessing the risk exposure of the gambler through gambler's ruin theory. The result indicated that, with a higher away scoring rates of teams, there remain higher probability of winning away than a home win, nonetheless making the point that, teams on average perform well playing at home. Corners profile and shots on target was found to significantly influence both home and away play of teams. The study concludes that, the optimally betting strategy for the gambler is to bet bold in a sub-fair situation, while taking a pulse caution in a super-fair situation relative to the expected fortune to reach from an initial bankroll.

ACKNOWLEDGEMENTS

I am most grateful to the Almighty God for the continuous guidance and protection over my colleagues and me through the course of this study.

I would also like to thank my parents, Mr. and Mrs. Edward Acquah and my siblings for all the encouragements, prayers and immeasurably supports they gave me throughout this study.

Further, I express much thanks to all teachers of Kyekyewere Salvation Army Primary School, Mampong district, especially the headmistress, Grace Serwaa for their support and commitment to this work. To Miss Mary Owusu, thanks for the special role you played in putting this work together.

Special thanks goes to my Supervisor Dr. A.Y Omari-Sasu of the Department of Statistics and Actuarial, KNUST for the special interest and great commitment He showed towards this work, and further thanks him for the suggestions and criticisms which have contributed to making this work a success.

Finally, enormous thanks go to all Lecturers at the Department of Mathematics, KNUST for the rich academic knowledge they gave me, and also to express gratitudes to all my year mates for the cordial relationships we shared in the entire period of this work.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS/ACRONYMS

SMTPs ..... Stationary Markov Transition Probabilities

MCMC ..... Markov Chain Monte Carlo

FLB ..... Favourite-Longshot Bias

MLEs ..... Maximum Likelihood Estimations

EPL ..... English Premier League

EFA ..... English Football Association

GLM ..... Generalized Linear Model

RSS ..... Residual Sum of Squares

PRM ..... Poisson Regression Model

PDF ..... Probability Density Function

PMF ..... Probability Mass Function

AIC ..... Akaike Information Criterion

PGR ..... Probability of Gambler's Ruin

RoR ..... Risk of Ruin

PoR ..... Probability of Ruin

CHAPTER 1

INTRODUCTION

This chapter is sets to introduce the theme of this research paper. Issues to be considered spans from Background of the Study, through Statement of the Problem, Research Objectives, Justification of the Research to Structure of the Research.

1.1 Background of the Study

Gambling and betting is an activity that has attracted mankind since time immemorial because of feature of uncertain outcome. And the phenomenon appears to be taking shape in Ghana steadily. The proliferation of betting companies with its outlets under the ownership, care and trademarks of *mybet.com, Supabet, premier betting, Safari bet* in the country is a testament of this assertion. The legal regime of the country seems to have been satisfied as per the tenets of the law vis-à-vis the open nature of its patronage. The economic implication of this recent phenomenon though not the focus of this research is enormous-offer of employment to the jobless, quick money for patrons, tax revenue for government and net resultant growth of the economy are the benefits of such enterprise. With the randomness or uncertainty of events coupled with mankind interest to knowing the outcome of an event in the next seconds, minutes, hours, days, weeks, months and years, one is often faced with binary choices of taking a decision or not for one reason or the other for fear of unpleasant circumstances. This leads us to make guesses or predictions which could be good or bad, right or wrong, correct or incorrect, favourable or unfavourable, beautiful or ugly. That is, inherent in these guesses and predictions in the form of decision making is the feature of gambling and are often predicated on

whatever information an individual has about the possibility, chances of occurrences and past happenings of similar events, if any. In probabilistic terms, the binary outcome of such a guess or prediction of a kind is indeterminate and could be 1 or 0 (1 for "positive response" and 0 for "negative response"). Such is the game of gambling and betting.

The activity of predicting sports results (any other events) and placing a wager on the outcome with the underlying intent of seeking an upsurge in income is described as sport betting and gambling. Sports result in this sense is *restricted* to association football scores. This outcome is normally not under the influence of the bookmaker nor the bettor. Unlike other sport bets, association football bets offer prospective bettors a number of opportunities for placing profitable bets for positive expected returns of their stakes.

Ranging from Fixed, Spread, Parimutuel, System, Sequential to Person-to-Person are repertoire of bets for the consideration of bettors in Ghana's gambling market. While sequential bet grants the bettor the opportunity to place single bet one at a time with the bookmaker, system bet requires the former to place a bet on joint occurrence of events regarding teams of interest to warrant proper win. Partial win of placed bets do not entitle a bettor to any odds win. Also, betting on any aspect of an event (without necessarily being the final scores) exist under spread bets. That is, first corner, goal and team to be carded, first-half score and other interest of pursuits exist for gamblers. For a pool of bets on a specific outcome, there arises the need to share the accrued income equally to only winners, and, differently in proportion to the stakes of the bettors after a fraction of bookmakers' risk-bearing cost has been deducted. This situation gives rise to the well known parimutuel betting. More familiarly, this type of betting is the Tote. Technically also, this betting allows a bettor to compete with other bettors under the supervision of a sanctioning bookmaker. Informal form of bets between persons exists, but mostly without the blessing of legally mandated institution to sanction such activity. Clearly, any of these bets provide a centre piece for the gambler at gaining positive expected returns.

Increasingly, betting and gambling or choosing to enter into a situation with an outcome that is not known by individuals on the light side is to enjoy excitement and pleasure.

But for an individual with a non-concave utility function, with marginal utility increasing over some range of incomes; when the expected value of losses exceed that of gains, the benefits from the gains may exceeds the damage through the losses and that may be a cause for wants to gamble. For most gamblers however, taking the risk of gambling is a proximate avenue of resurrecting an insolvent financial position to raise enough funds to pay creditors and have surplus to meet other budget constraints. Truth be told, the individual may possibly loose the little money to be brought to total bankruptcy. This is because, such enterprise exposes the bettor to risks of substantial losses and rewards simultaneously. This makes gambling and betting a sweet-bitter game.

(Issues of believes that is skewed to superstition in winning a placed-bet based on colour of attire, soothsayers and others of similar nature which may be a motivation to place a bet are beyond the scope of this paper).

The outcome of gambling and betting is thus probable in nature and inherently very risky with money involved alongside accompanied instantaneous loss or gain which the gambler is prone to. In respect of this associated randomness, critical issues ought to be considered for the optimality of the available funds of the bettor, whiles minimizing the associated ruin exposure. This is necessary in any case to ensure the bettor's wealth is maximized potentially in the face of uncertain circumstances. And probability and statistics (the actuary's tools) provide elegant mathematical framework for evaluating such uncertainty and randomness. This is because such game as betting and gambling thus involves probabilistic decision-making strategies. And probability and statistics are content areas in mathematics and stochastic modelling that is well suited for explorations with and through the real-world context of games and events with uncertain outcome.

Beyond the need for quality probabilistic decision-making for the optimality of the bettor's fund and ruin possibility in the face of the associated volatility relative to betting and gambling; subtle though, but, important questions arise for address and discussions: At what time of revenue gain/loss level should the gambler reach to warrant a stop in betting? How much of the gambler's wealth is to be devoted to the gambling or betting? Whiles

the latter in many instances easily determined prior to betting and gambling so as to reduce the related uncertainty of the sport event and also as a protection against non-deterministic errors; the former sets in the course of the betting and gambling process as means of avoiding gradual dissipation and subsequent depletion of resources and/or as protection of short fall in odds accruing from the betting activities.

To address the latter, an appealing, compelling and well elaborated proposal by Kelly (1956) and further studied by others like Breiman (1961), Finklestein and Whitley (1981), Feller (1968), Ferguson (1965); is for the bettor to bet a fixed proportion of his/her bankroll on sequence of bets. The reason for it as described by Breiman (1961) is grounded on the fact that; Kelly criteria maximizes the exponential rate of growth and provides the minimal expected time to reach a pre-assigned balance (Ethier, 2004), while the others approached it in similar fashion but in slightly worded form. This is necessary at least from the layman's perspective to ensure the bettor has something to fall on in the event of loss of initial invested funds or bad fortune. Also, Sudderth(1971) opined that, the log—return of the fixed proportion of the bettor's wealth is to be maximized with positive expectation; expressed mathematically as $E(\log(1+\beta x))$ where β is the fixed proportion to bet and maximizes the expectation of the log-return, x is the bettor's wealth; with the constant of proportionality been equal to 1. This proposal as put forward by these eminent scholars is the widely publicized Kelly criterion betting strategy which results in maximum expected geometric rate of bankroll growth, but entails wild swings, which are not for the faint heart. The Kelly criterion is formally considered later in this paper.

The former, which is Optimal stopping time (theory) for the gambler to avoid ruin is an interesting area of stochastic process application to betting and gambling that has received considerable attention at seeking to provide and address the concern of the gambler relative to when to stop. Prominent of these researches is the authoritative works of Dubins and Savage (1965) and Sudderth (1971). Optimal stopping time problems concern the effect of a gambler's fortune over systems of events for deciding whether or not to stop playing a sequence of games to leap over ruin. The usual stopping problem

is made to a situation where the player is allowed to stop at certain times along a given path. This is made possible by filtration of information up to and including time t. It therefore suffices to reason that, the *optimality* or *sub-optimality* of a gambler's fortune is largely pivoted on knowing when to opt out on the balance of revenue loss and gains. This is typical of a situation where the gambler losses more than the amount he invests-*sub-optimality*. The reverse could be considered optimality all else remaining the same-*(ceteris paribus)*. The former is thus addressed through this optimization strategy which is well grounded in mathematical framework. The idea and focus of the stopping time to a very large extent is to ensure the gambler leap over ruin.

In the address of these subtle though, but, very important questions regarding the fraction of bet and the optimal stopping point of the gambler to avoid ruin, two strategies are thus critical to the game and should prompts the gambler—maximizing the expected utility of player's stake and minimizing the probability of ruin.

The maximization of the expected utility of a gambler relative to the available wealth, as well as the minimization of eventual probability of ruin (PoR) as indicated above are largely dependent on the prediction accuracies of bettors relative to the strengths and weaknesses of the teams' scoring prowess on the basis of which bets are placed. The prophetic potency and nature of these predictions against which bets may be placed are purely done on the back of available present information regarding the strength or lethargic nature of teams' departments (from defence to attack), and other external factors that are outside the pitch of play.

However, bettors have often relied on teams history of matches played as functional characters of conceding rate, scoring rates, foul rates, and number of corners earned or conceded, shot on targets, foul rates, cards received as well as final position on the league table for prediction. Much as previous match played may serve as basis of good bet for a bettor, that may not be a better baseline for the gambler without ascertainment of which variables most influence scores for teams. This stems from the fact that, other extraneous factors like weather altitude of a particular match, transfer or new player

and manager, pitch suitability, home and away play influence, decision and indecisions of referees and many others directly and indirectly goes to determine whether a particular game for which bet has been placed will soar up the bankroll of the bettor. Clearly, the betting and gambling enterprise ebbs above just having the desire to win on a team and money to wager on teams, but requires deep and critical observation and caution cognizance of key scores variables' influence on teams as enumerated above.

In the stance of this complexity regarding the factors, whether qualitative or quantitative and the direct bearing of influence on games, which partially informs bettors to place bets against bookmakers, the cerebral question is; what sort of help exist for bettors to consider?.

In the address of this question by previous researchers, simple to complicated models have accordingly been built to exploring the dynamics of the football game for the consideration and analysis of bettors to place profitable bets. This is necessary for gamblers of football scores considering the associated estimates of probabilities and odds by bookmakers, so as not to be exploited by them.

The natural questions to ask further are: How does the gambler seek to optimally place bets in the face of the variables of influence on a game? How does the gambler place bets on teams of interest relative to the variables discussed, and what betting strategy should the bettor adopt to avoid being brought to bankruptcy?

Whitrow (2007) considered two optimization algorithms for many bets using log-utility function. Whithrow (2007) thus considered a stochastic gradient-based approach as a generalization of Kelly staking to the case of multiple bets. In respect of predictions of football games, Moroney (1956), Reep et al (1971), Dixon and Coles (1997) accordingly developed models in that regard but without some indistinctness.

Cognizance of the existing works, this paper adopts a prediction model through Poisson Regression approach to seek to estimate the home and away strengths of teams from some key variables of football influence and further test their significance of some scores determinants. That is, prediction model is formulated to explore the possibilities of

chance for proper bets to be placed by the gambler through Poisson regression model. And assessing the risk ruin exposure of the gambler.

1.2 Betting and Gambling Terminologies

To better understand gambling and betting as in the foregoing research paper, one needs to apprise himself/herself with the needed lexicons to make meaning of terms herein referred to. The following are key terminologies worth knowing them relative to the underlying subject under discussion:

- **Bettor**: Often referred to as the gambler or player of the game of betting and gambling; The bettor is the individual who makes the single decision to wager portions of his bankroll or available wealth for additional revenue in the form of bets to meet budget constraints and sometimes for pleasure and excitement.

- **Pick**: This refers to the selection among all the possible outcomes on which the bettor places a bet. It is fundamentally akin to what the proposition(s) of the gambler is as against the various possibilities the game offers or the bookmaker gives for selection. Example is predicting a Chelsea lead over Man United before half-time with whatever scores or wagering on a first corner for Chelsea instead of Man United in the case of spread betting. And picking a win for Chelsea over Man United for single bet.

- **Event**: This describes the specific sporting event(s); and generates the subset of possible outcomes for selection. It is the on-going or yet to-be started football match that presents the opportunity to the gambler for picks to be made. Example of event is a Chelsea vs. Leicester of match day two in the 2014/2015 Barclays premiership season.

- **Result**: The eventual outcome of the event. When the pick of the gambler coincides with the event(s); the bettor is said to have won the bet(s) in which case he/she

is paid an amount proportional to his stake times the odds such pick comes with. On the contrary, when the picks of the gambler is at variance with eventual result (s), the bettor loses his stake and odds that could have been received if positive or favourable outcome was the case.

- **Stake**: This is the amount of money being wagered in a single bet, the amount to a large extent represents a bettor's level of probability or level of certainty cognizance of the event's climate and subsequent picks. That is, a higher probability of win often determines a higher amount in monetary value or any underlying asset of interest for wagering and ultimately determines the amount of odds to receive. The reverse is true.

- **Odds**: This represents the payout or payoff to be received by the bettor if pick and result coincides or prediction turns out to be correct. This is slightly different from result. It is the occurrence of result either positive or negative that determines the odds to be received or none. In other words, there must be an occurrence of results for the odds buttons to be pressed, if any. The odds are often expressed in decimal points where the amount staked is multiplied by it to determine the actual payout bookmakers pay to gamblers. Odds are set by bookmakers but differently from each other. Odds offered by bookmakers for an event outcome are implicitly estimates of the game probabilities regarding a selected pick. And the possibility of inaccurate estimates either deliberately or as an oblivion is highly expected which this paper explore later under issues in wagering markets.

- **Bookmaker**: : Often described as a bookie, the bookmaker is the business entity that provides the sport betting and gambling services. Mybet.com sports betting, Supa bets, Premier betting, Eurobets are examples of bookmakers in Ghana. Bookmakers crate the platform for prospective gamblers to seek to trade a "given stake" for bigger odds if events work favourably in the bettor's interest. Records of all betting and gambling transactions are done for events of such kinds and at times

placed online for public analysis. These entities are mostly legally recognized and licensed to sanction betting and gambling activities.

- **Profit**: This is the excess of odds over stakes in a sequence of picks in an event or better still collection of win amount over stakes in betting. That is, it is the additional amount the gambler receives over and above the stake when bet is won and is over.

(Moya, 2012)

1.3 Statement of the Problem

Gambling and betting issues have excited interesting and compelling prepositions among Actuaries, Economists, Mathematicians and Statisticians respectively because of the inherent risk of its nature, economic gain and/or drain on households scare resources and accompanied utility maximization, its complexity and mathematical abstractness, and, its dynamics and volatile nature. Much attention however has been devoted to particularly addressing inherent risk and volatile nature of it and modelling of scenarios and strategies for optimality of funds adoption in the process of gambling and betting. Economics literatures also abound on testing inefficiencies in wagering markets but in lesser proportion to assessing the inherent risks.

The modelling of football prediction models for placing bets over the period has served as a guide for sports prediction results but without little indistinctness. This emanates mainly from the randomness, poor and partial mathematical representation of certain variable of interest and influence. This is further compounded by the fact that, the bettor just like the bookmaker does not influence the outcome of football scores.

More importantly, variables that ranges from mindset of players, nature of pitch, referee' decisions and/or indecisions, weather, team selection by a coach to momentarily loss of concentration of players covertly goes to determine the outcome (i.e. win, draw or loss)

against the picks of a gambler. These are exogenous to the gambler control and with some being qualitative in nature for proper evaluation and formalization. For those that are quantitative and formalized, how significant are they in influencing scores? Consequently, an individual gambler's estimate of a probability of win or loss amidst these difficulties cannot be a good representative of the bettor's trust or measure to commit a given stake to wager.

In essence, the edge, desire and decision to bet a Swansea win over Man United is an expression of higher probability of win in return of higher odds cannot be independent to the happenings on the pitch or team selection by a coach. Rosett (1965) thus makes the point lucidly that 'if risk-neutral bettors have unbiased expectations of win probabilities, then the proportion of money bet on a horse (football) will equal the probability', which is unfortunately not correctly determined because of non-quantitative variables that influences the outcome of football events. In essence, Rosett (1965) asserted that, the amount to place at a bet beyond the pulse rate or blood pressure of a bettor depends on the certainty or otherwise of positive expectation which could deviate from normality for reason stated above, and is expressed by Hausch and Ziemba (1995) who stated that, 'Betting to win usually leads to losses except in extreme favourites'. That is, statistical models that give precise probabilities of win or loss must seek to introduce more variables while seeking to assess the level of influence of these variables on scores for critical predictions. To address this lapse, careful consideration should be taken at arriving at a true probability of win when a gambler decides to wager a stake on a bet. At best, the implied probabilities in offered odds should be well calculated prior to placed bets. In summary, model of good prediction for teams' win or lose is worth considering.

In wagering markets, the concern of good wealth allocation for placing bets remains a daunting task. Plausible strategies like equal allocation of stakes in the case of (multiple bets) on each pick, considering favourable bets to stake on either *fixed, system, parimutuel, spread bets* adoption are considered. This seeming difficulty stems from the fact that, the number of picks to make at each point (in the case of system or sequential bets)

largely hinge on characteristics or intricacies of the games in question and the associated multiplicative payout to be received in the parlays. The associated return and riskiness of each pick is thus a vital consideration in this regard vis-à-vis the dynamics of football games and the revolving variables of influence on the outcome. This amount allocation issue and its importance stems from the eminent ruin scare the gambler is prone to.

More to the point, in a single bet in the form and feature of sequential betting, the much publicized Kelly criterion provides an optimal resource apportionment for consideration and subsequent adoption. The criterion asserts that, the individual bets a fixed fraction of his/her bankroll at a time when faced with sequential bets. Clearly, this is very much documented in the annals and journals of risk and investment, finance and portfolio markets. More difficulties however arise in the case of system and/or multiples bets where more than an event are to be gambled on; and the optimal strategies to adopt to increase return on the bankroll of the gambler is not easily determined. This is because; such a scenario is not elegantly been addressed as there are lingering issue regarding bankroll allocations on various picks which beg for answers. In fact, Whitrow (2007) adopted gradient-based and algorithms approaches at seeking to estimate the optimality of such a situation, by analysing the optimality of simultaneous games. In respect of these, the questions will be: how does a gamble seek to redistribute wealth from a non-yielding event to another when confronted with similar events? Clearly, such situation is practically difficulty for evaluation.

Akin to the above issues lie the possibility of the gambler getting ruined or rich as stated already. Ruin occurs when the gambler has nothing to wager after a number of losses for events of a kind. The risk of ruin exposure of the gambler for number of events brings cold pulse for gamblers as he/she in many respects seek to optimize the initial bankroll. In respect of this lie the optimal betting strategy for adoption to (at worst) strike balance from bets without necessarily being brought to bankruptcy.

1.4 Objectives of the Study

1.4.1 General Objective

The objective of the research is to analyse the ruin possibilities of the gambler for optimal betting and gambling on football scores.

1.4.2 Specific Objectives

In respect of this objective, specifically, this research seeks to:

- Formulate prediction model on football scores to estimate probabilities of home win, draw and away win.
- Determine the influence of some variables as significantly influencing both home and away scores, if any.
- Examine the Kelly criterion from bookmaker's offered odds
- Determine the risk of ruin (RoR) of the gambler, and the optimal betting strategy for adoption.

1.5 Justification of the Study

Classical study of betting and gambling largely argues that, the individual is able to make more revenue from a fraction of income staked in wagering*ceteris paribus*; with utility of pleasure for participation being minimal. This explains why unlike insurance where people resort to reduce risk, gambling is sought to bear risk.

Heuristically, the gain in revenue is a proximate cause for a person to place a bet to buy and bear a risk of such fluctuated nature. It suffices therefore to come to the understanding of proper and plausible strategies that yields more revenue from small fraction of money to be adopted amidst the dynamics of football events.

The steady progression of betting and gambling outlets and its subsidiaries of such kinds under trademarks alluded to in paragraph one, with its attendant economic implications for the country is huge, and accordingly requires the need for assessing the optimality of a gambler's fraction of wealth wagered on a particular event. This is because, the continuous survival or insolvency of these betting centres and its subsidiaries is largely pivoted on good fund management couple with the patronage of services by patrons. Spree of huge income loses without (somewhat) percentage gain in bankroll that often results from poor prediction of matches vis-à-vis improper reading and/or appreciation of variables of influence in games is dire for the game of betting and gambling in Ghana where more people remain in the penury bracket. This is against the backdrop that, bookmakers make their income (sales) from bets by bettors who fail to accurately predict events to attract a penalty of a sort and commission often described in the gambling and betting circles as vigorish (vig) for short.

Previous researchers like Kuypers (2000) and Levitt(2004) have alluded to the reason of different pricing of similar events by different bookmakers. Conjecturing regarding the source of this variation of odds pricing do not leave out bookmakers *inadequate knowledge* of the dynamics involve in the game of football. That is, bookmakers do not equally have the same appreciation of the dynamics an event presents, left alone to know the impact it would have on the house financial situation. Minimum and maximum threshold bets amount are thus sets by different houses to indirectly deal with this lapse. A modelling of football scores will therefore serve as an important basis for apt reading of the emerging wagering markets in Ghana for appropriate odds to be assigned to events.

Furthermore, knowing when and how to go about to place bets for increase in revenue is a critical aspect of the game of betting and gambling. This is because, beyond the expressed desire to seek pleasure (good laughs), upsurge in income and meet budget constraints out of which a bettor may necessarily wants to gamble, how to go about it, i.e. knowing correctly how to allocate the amount to bet to avoid ruin, good interpretation of probabilities coupled with adoption of better strategies are very key ingredients worthy of

consideration. The complexity in this regard especially in system bets is very uncommon except for the work of Whitrow (2007), that sought to analysed simultaneously placed bets on events. That is, having a positive expected returns for a placed bet among other things, is for the gambler to be sufficiently accurate regarding the probability of win and the associated odds from the bookmaker in order to overcome the latter's edge (Dixon and Coles,1997).

Betting heavily in favourable situations increases the expected winnings in favourable games, but it also increases the fluctuations of the resources (Ferguson, 2000). This is synonymous to saying that volatile nature of gambling should prompt the individual to adopt the best of strategies even in the face of obvious probability of win. This is because; the much publicized Kelly criterion beyond the determination of the fraction to bet at each stage of betting leaves large degree of freedom regarding the strategy for winning to the gambler. Differently stated, the Kelly criterion does not teaches the gambler the strategy to be adopted except the amount to bet at each trial to avoid ruin. In spite of this lapse of Kelly criterion, good strategies could be adopted for optimization, though not risk-arbitrage strategies. Evidently, this illustrates the need for the gambler to spread the scare available resources (wealth) to maximize the expected utility. This is necessary to minimize the risk exposure the gambler from the initial bankroll and the expected fortune to reach.

The actuary's preoccupation has always been to use mathematics and statistics to assess the risk inherent in any enterprise, and to provide the needed risk caution accordingly. And this paper is not a departure from such task either, as it seeks to come out with a prediction model for bettors to consider to appropriately placing optimal bets, and also assesses the ruin possibilities of the gambler. The motivation of the prediction model for football scores is in two fold. To aid bettors develop an increased mindset of the dynamics of the game of football, and also to provide a guidance of probabilities estimates of unknown future outcomes for 'vulnerable' bettors as against the bookmakers edge in this respect. And further through the idea of gambler's ruin theory evaluate the actuarial risks of the

gambler getting ruined or rich over time of the staked amount for the consideration of bettors.

1.6 Structure of the Research

This research paper is structured into 5 chapters together with appendices and reference sections. Chapter 1 of this research gives introductory information of the study with clear statement of the problem, research objectives and justification for the study. Chapter 2 reviews the various perspectives shared on the subject in the foregoing paper (theoretical and empirical reviews) with emphases on scores modelling, Optimality of bets, Gambler's ruin and key issues worthy of attention in betting and gambling. Chapter 3 considers the methodology for the study regarding the methods or procedures, model, theory or distributions needed to establish proper understanding of issues as well as the stated objectives in the research paper. Chapter 4 looks at the Estimation, Analysis and Discussions, while Chapter 5 looks at Summary and Conclusions of Results and Recommendations.

CHAPTER 2

LITERATURE REVIEW

This chapter considers relevant literatures on the wagering markets optimality strategy of placing a bet and seeks to further review works on prediction models and application of Gambler's ruin theory and Kelly criterion. Specifically, the chapter surveys works of other researchers regarding the focus, coverage, methodology and other key issues worthy of discussion in the betting and gambling markets for football as relevant this paper. In this regard, the review is considered from two fronts- theoretical and empirical reviews.

2.1 Theoretical Literature

2.1.1 Understanding Optimality of a Bet

In one key respect, betting and gambling markets provide a natural environment for testing the optimality of a strategy relative to the amount risked in an event. This is because; the prospect of realizing an increase in revenue base when result coincides with selection is enough motivation for a bettor to adopt the best of strategies to optimize a given bankroll. That notwithstanding, the 'optimality strategy of a bet' is a composite phrase that has not been easily explained, as there remains no simple and easy way(s) of determining the optimality of an amount.

In the stance of this complexity regarding the determination of an optimality of a bet relative to an amount risked, many researchers have through dissimilar approaches adopted some criteria for determining the optimality strategy of risking a fraction in a wager. Loosely speaking, optimal strategy of a bet refers to the adoption of good line of action to achieving the most out of a penny that is wagered in a bet so as to avoid

ruin and maximize utility of wealth, while maintaining the level of accrued income, if any. Optimality of system bets therefore seeks to answer the question: how does a gambler concurrently seek to maximize returns on events for which bets are placed?

Relative to the existence of dissimilar approaches at optimizing a bet, Haigh (2000) through his work affirms this preposition as he reviews the work of Kelly, and, interpreted it in the context of spread betting under several optimal criteria. Gottlieb (1985) for example, approached an optimality criterion from the minimization of expected exit time to wealth from an interval, as he finds a strategy for an infinite sequence of wagers and the needed optimality criterion to adopt. Kelly (1956) and Whitrow (2007), each also adopted different strategies at seeking to optimize the respective bankroll of bettors. More to the point, Browne and Whitt (1996) under several criteria proposed how Kelly risking of proportional amount of a bettor's wealth is found to be optimal.

Truelove (1970) writes that, a betting strategy for a gambler is a set b of functions $\{b_j\}$, where at each game j, b_j associates each bet with partial history, that is, $b_j(x_0, p_1, x_1, ..., x_{j-1}, p_j)$ where $0 < b_j < x_{j-1}$. Hartvigsen (2009) points out that, a wagering strategy is a rule dictating how much of the gamblers bankroll should be risked for each bet to ensure optimality. He obtained, through the idea of nonlinear programming technique, an expression that characterises the strategy for placing equal sized wagers on number of bets. That is, to him the optimal betting strategy in the case of number of bets is equal sized wager.

Optimality strategy of a bet whether (System, Sequential, Fixed or Pari-mutuel) bets thus seek to ensure that, the amount of stake wagered being small or big yields a return that puts the financial situation of the bettor in a good shape as a reward for taking and bearing risk in an uncertainty market. Thus, strategy of optimality in wagering market could be fixed, sequential, parimutuel or system bets. Whitrow (2007) asserted that, simultaneous games of such nature "enables and affords the gambler the opportunity to spread resources accordingly across events simultaneously with the intent of increasing expected returns while reducing the volatility of risk effects as is the ensuing issue of

some bettors against the much publicized Kelly criterion. This is particularly so, because, when faced with pool of concurrent events to place a bet, the criterion opines that, the bettor only places a bet at a time (sequentially) than seeking to optimize simultaneously. In the light of the above, it suffices to explain the optimality strategy of bets as an approach that ensures maximum expected returns of a fraction of amount wagered or spread concurrently on events of sports. And the adoption of such an approach to bring about maximized returns is very much varied in the wagering markets, as many theorems and criteria have long been proposed.

Clearly, optimality in the context of betting and gambling requires that, the best of utils is derived from a fraction wagered vis-à-vis the maximization of the log returns of the gambler's utility as promulgated by Kelly (1965). In fact, to Kelly, the optimal strategy for a gambler to adopt to increase returns and avoid eventual ruin is for the gambler to bet a fixed percentage of the bankroll at every stage of the way when faced with comparable favourable bets. MaClean et al. (1987) further state that, fractional Kelly strategies are effective in that betting a fixed fraction (for example a half or a quarter) of the Kelly proportion reduces returns monotonically, but curvilinearly reduces risk and therefore offers potentially interesting strategy trade off. However, using full Kelly proportions offers the highest return for a strategy and thus gives the maximum likelihood of rejecting efficiency. Ethier (2004) argues that, under fairly general conditions the Kelly criterion maximizes the median of terminal wealth.

Bremain (1961) in seeking to further study the seminal work of Kelly showed that, fractional Kelly for sequences of favourable bets one at a time, given the assigned odds will minimize the expected number of trials to reach a preassigned fortune, a suggested criterion of optimality for the gambler.

However, the possession of a strategy without a favourable game to place favourable bets render the bettor's plan of seeking an upsurge in revenue void. Favourable bets are bets in which there exists opportunities for optimal returns from a particular favourable game. Truelove (1970) asserts that, favourable game is one in which there exists betting strategy

for which the probability of eventual ruin is less than one for some portion of the bettor's bankroll. This stems from the fact that, betting heavily on favourable situations increases the expected winnings in favourable game, but also increases the volatility of the resources. This requires the gambler to rationally bet lightly when odds are not in favour of him or her, but, also bets heavily when odds ply off to his advantage (Ferguson, 1965).

Rotando and Thorp (1969) posit that, a favourable game to place a bet is a game in which there exist a strategy such that $P_r(\lim_{n\to\infty} S_n = +\infty) > 0$, where S_n is the wealth or capital of the gambler after n trials of bets. Similarly, but theoretically, Ferguson (1965)defines a favourable game to be as one in which there exists a betting system for which the probability of ruin is less than one for some value $x > 0$ of the initial wealth x_0.

Ferguson (1965) further advanced that, a betting system tells an individual bettor how much to bet in each game. He further asserted that, the amount b_j to bet on an event is dependent on other extraneous factors, other than just money. He introduced and established a Markov betting system, as being the situation when an amount b_j depends on the current resource envelope and accompanied present probability of win. Ferguson (1965) was quick to add that, in tandem with the stationary Markov transition probabilities (SMTP) no theorems along these lines are available.

Dubins and Savage (1965) under the assumption of countable additivity established a result that states the optimal strategy for a gambler confronted by two moves; either he/she gambles his wealth on a given game, or stays with the current wealth if the bookmaker in this instance allows it. This allows the gambler to determine the best point of bet to maximize the needed returns. They also analyse strategies for gamblers or investors to increase their chances of reaching a certain monetary and/or survival goals while facing a losing proposition. This he did by engaging in bet doubling so as to reach the probability of winning an amount.

Furthermore, Keller (1994) through the approach of dynamic programs simulation modelled a continuous optimal betting strategy for backgammon. Within a certain

confidence level of winning, He opines for the player to double up the amount risked in a wager for optimize funds.

Evidently, optimality is not attained through a single approach, however, with a single purpose of seeking to increase wealth over a number of bets if possible to remain solvent so as to meet budget constraints as the above survey reveals, all things being equal (*ceteris paribus*). The existence of a favourable game is very much amenable to optimize strategy for adoption to revenue stability.

2.1.2 Models of Predictions for Football Scores

Gamblers place bets for or against a particular team's (win, draw or lose) on a game or event on the weakness and strengths of the teams, vis-à-vis history of previous good performances or otherwise as already stated. Accordingly, variables of new manager in charge and tactics, transfer or purchase of new players, away and home play influences, impact of influential player, weather among others are said to determine the likelihood of a gambler's bet against and/or for the win of a team. This makes sense, because, the selection or pick of an individual is overtly and covertly determined through the scoring potency of teams which are functional features of the above variables of interest in every game.

This preposition has led to statistical models from simple to complicated, that seek to accurately predict probabilities of match outcomes via statistical tools that analyses the strengths and weaknesses of teams to serve as guide for profitable bets to be placed. According to Dixon and Coles (1997), the goal of such models for prediction is also to outperform the predictions of bookmakers, who use them to set odds on the outcome of football matches.

In the light of seeking edge over bookmakers in the setting of odds, Dixon and Coles (1997) derived a method for estimating the probabilities of football results through simple bivariate Poisson distribution for the number of goals scored in a match by each opposing

teams. In this regard, Poisson regression model technique was employed cognizance of the seemingly difficulties posed by the dynamics of teams' performances. They further used their model to illustrate the strategy for placing a bet on all outcomes for which the ratio of the model probability to that of bookmakers' probabilities exceeds a specified level. For sufficiently high levels, the duo's model show that this strategy yields a positive expected return, even allowing for the in-built bias in the bookmakers' odds. A modified independent Poisson model was also obtained by Rue and Salvesen (2000), in which case separate attacks and defence strengths were taken care of. Also, a psychological factor was incorporated into the model to reflect the overall differences between opposing teams. The pioneer works of Moroney (1956), and Reep and Benjamin (1968) in these areas are very much documented. Poisson and negative binomial distributions were employed to model goals per match resulting from the strengths and weaknesses of teams. Reep and Benjamin (1968) analysed the possession and distribution of passes among players of a team. To them, these passes among players are time-dependent as it evolves from lost-of-ball to possession-of-ball from time to time, approximately making the 'pass-success' a negative binomial. The drawback however, in the duo's work stemmed from their treatment of goal shot (like penalty) as part of passing moves, and was subsequently used as a count of 0-pass moves. That notwithstanding, an exclusion of goal shot will thus improve the negative binomial fit of the model. Also, Moroney (1956) show that the number of goals scored by a team in a football game is better represented through the negative binomial approach, contrary to the Maher (1982) claim of Poisson distribution being a far better way of quantitatively accounting for the scores of teams in a football match. And through this approach recounted that, the relative strengths of teams whether playing at home or away are almost the same. He premised his argument further on the grounds that, the expected mean of Poisson is constant which gives an indication that the quality of the number of goals might be constant from game to game with such factors as the quality of opposing team's strengths, injuries in teams and others. In fact, unlike Maher (1982) the former strongly argues that, the goals scored in a football game are

better fitted through 'modified Poisson distribution' (Negative binomial) than that of Poisson distribution.

In a similar fashion to account for the strengths and weakness of teams of stochastic nature over time, Glickman and Stern (1998) modelled a state-space model on the premise that, the strengths and weakness of teams follow a first-order autoregressive (AR) process, while asserting the stochastic nature of events on the field. Markov Chain Monte Carlo (MCMC) approach was further used in arriving at the objective of obtaining a plausible inference concerning teams and accounted for parameters of influence on games. Specifically, Gibbs sampling method was adopted in this regard to fit and summarise the data used. Also, Maher (1982) by means of the maximum likelihood method estimated the offensive strength for scoring goals and defensive strength against conceding goals for teams using the idea of maximum likelihood for bivariate Poisson distribution.

On the assumption that teams' defence and attack strengths change randomly over time rather than being static, Crowder et al (2002) also estimated teams' attack and defence capabilities through some approximate computation (in matrix form) which results are favourably to the duo's work (Dixon and Coles (1997)). On refining the independent Poisson model of Dixon and Coles (1997), Crowder et al (2002), represented respectively, α_{it} and α_{jt} as the attack rate of teams i and j at time t, β_{it} and β_{jt} being their respective defence rates with an incorporated home advantage factor η to model goals scored by teams to aid bettors' to appropriately place a bet. (Crowder et al. 2002).

In essence, the gambler ought to consider among all others, the strength of not just one team but opposing team to place a bet for optimality of funds.

In contrast to considering the strengths and weakness of opposing teams to model goal prediction for the consideration of bettors' to place a bet, Knorr-Held (2000) modelled one-sided team strength by the use of team's win, lose or draw results. The attack and defence strengths and weakness of both teams, unlike other works, were inseparable in Knorr-Held (2000) work. Similar to the work of Knorr-Held (2000), Koning (2000) derived point estimates for the variance parameters. Respectively, the latter applied the idea of

maximum likelihood while the former used an extended Kalman filter together with an ad-hoc method for a variance parameter.

The impact of some variables (exogenous as they are) on match outcome has in recent times be a focus of research for proper evaluations, if any, to be made in that regard. For example, evaluating the effect of a red carded player on a team's win or lose; and others of such. Ridder et al (1994) indicate how a player sent off (either red carded or out of poor tactical decision by a coach) affects a team performance and the overall score outcome of the event. Together with the characteristics of teams strength via-a-vis home advantage, offensive and defensive strengths and their interaction by taking into account not only goals but also possession of the ball, Hirotsu and Wright (2003) derived a model to assess the rate of goals scored and conceded as being affected by these characteristics.

Moreover, the temperature differences for teams (both home and away) of the cities recognised as home (of any team) have much influences on football scores. McSharry (2007) in this respect, investigated the effect of altitude difference on teams' performance for a number of countries empirically. And subsequently concluded that, altitude has a significant impact on physiological performance as it favours high altitude teams mostly playing a game at both high and low altitudes through the estimates of probabilities of home team winning by altitude difference.

Call it most controversial, Hill (1973) makes the assertion that, " anyone who had ever watched a football match could easily make the observation to conclude that, the game was either all skill or all chance". He justified this position by calculating the correlation between analysts opinions and the final league tables, finding that even though chance was involved, these were also the result significant amount of talent affecting to the final outcome of the match.

In the light of these enumerated models of previous researchers on models predictions about scores for the consideration of bettors to place a bet, the nagging question cynics ask is; how accurate are these models? The inherent variability of models built around Poisson and other distributions couple with dynamics of the football game leaves match

to be desired, but also provides a much better potential basis for proper and optimal bets to be place.

This study builds on other works and incorporates other factors to assess their significances in prediction for bettors.

2.1.3 The Gambler's Ruin

Gambler Ruin theory was developed by Feller, W.(1968) based on the idea of probability theory and particularly on Markov property, where a gambler wins or loses money by chance specified in probabilistic terms. Since the pioneering work of Feller (1968), the mathematical theory of gambler's ruin has received attention from other researchers through the extension and application of the concept and idea.

Wilcox (1971), used the gambler's ruin to develop a framework that predicts the risk of ruin (ROR). His model assumed that the firm's financial state could be defined as its adjusted cash position at any time. The gambler's ruin idea thus assesses the bankruptcy of entities based on the inflows and outflows of liquid resources. Wilcox viewed the firm as a gambler who begins the game with an amount of money equal to its net assets. The firm is then assumed to win an incremental amount of assets with probability p or losses it with probability $1 - p = q$ and bankruptcy occurs when the firm's asset falls to 0. The dynamics of a firm's net assets can thus be described by a stochastic process estimated on the time series of net asset, Wilcox (1971) pointed out.

Also, Coad et al (2012), modelled the bankruptcy of firms by theorising the gambler's ruin framework through arguing that, a firm's performance is best modelled as a random walk process, and further argued that survival is non random and depends primarily on the stock of accumulated resources. In analysing the growth of firms, Coad and his colleagues averred that, the growth of firms is best captured by the gambler's ruin model and postulated that critical firm size might be analogous to how many chips the gambler has when they start, and hence how long they stay at the table and the likelihood of

reaching an outcome that is positive. The gambler ruin theory is in this instance used to assess the survival of firms relative to how long they survive in a competitive environment given the firm's reserves.

According to Harick et al (1999), gambler's ruin—which in their words is a 'classical solution to random walk' can be used to model population sizing to predict the quality of genetic algorithms based on the increased population as a point in time.

Mohan (1995), on the assumption of constant probability of winning any game, with each game being independence of the other adopted the gambler ruin theory to derive a working formula and calculated the correlation between the results of two games. By conditioning probabilities of the number of wins or lose on the next game, Mohan (1995), obtained a transition matrix between the win or lose of placed bets.

Clearly, the application of gambler ruin as postulated by Feller, and used by Wilcox to pioneer the bankruptcy prediction of firms has helped. Regrettably though, its applications to the wagering market has been very minimal except for theoretical derivation and documentation.

2.1.4 The Kelly Criterion

Originally designed for the transmission of information, the Kelly criterion, a much publicized criterion for placing bets is said to have been an avenue for widening the frontier of bets in the wagering markets, and very much so in portfolio market analysis for optimised funds. The inherent feature of the criterion as many have described it leans credence to its popularity and acceptance in other markets like the financial and portfolio markets. And the criterion has further inspired many scholars of write-ups and researches in academia regarding the amount of money to invest on a game or portfolio. The problem of exactly how to apportion wealth between various random variables is the focal problem of portfolio selection; and so it is correct to suppose that, the results of optimal allocation of wealth are of considerable interests of all economists and financial

analysts, (Finkelstein and Whitley, 1981) argues.

The criterion argues that, faced with a series of favourable games to place bet, the gambler should wager a fixed percentage of the bankroll at a time to maximize returns and avoid eventual ruin. This makes much sense because, committing an entire bankroll into a game for which the outcome is improbable like bets is very unsafe. Also, putting all bankroll on each event is too risky because each bet looms the danger to lose everything and as the number n of bets goes to ∞ a gambler could eventually be brought to bankruptcy.

The fixed percentage of the bankroll to place bet on each trial according to Kelly (1956) is to ensure the log geometric growth-rate of the bankroll. Also, Kelly showed that in order to achieve maximum growth of wealth, at every bet a gambler should maximize the expected value of the logarithm of his capital, because it is the logarithm of his capital which is additive in repeated bets and to which the law of large numbers applies. In this regard, many money management systems which maximizes the expected value of the capital are said to employ Kelly criterion.

Kelly (1965) further stated that, for a super-fair situation, where the probability of winning a certain event exceeds the probability of losing the event, i.e. $p > q$, the optimal wager (the fraction to bet at each time is $\alpha^* = p - q$), but $\alpha^* = 0$ when $p < q$ in which case the sub-fair situation is established.

Bremains (1961) studied this proposal and developed theoretical underpinnings for the validity of the Kelly system, by generalising the criterion to number of distinct distributions with specified probabilities of win and loss. In furthering the tentacle of Kelly's work; Thorp (1969) applied the work on other favourable games including casino blackjack, barracat, wheel of fortunes and portfolio markets; clearly stating the robustness and adaptability of the criterion. Fundamentally, the Kelly criterion according to Thorp (1969) is sought after as a protection against adverse fluctuations that achieves positive eventual ruin.

Piotriowski and Schroeder (2007) posit and affirm that, the criterion maximises the expectation value of logarithmic wealth for bookmakers' bets as originally postulated

by Kelly, and further reveal that, the criterion gives an advantage over different class of strategies. Also, under asymptotic sense, Ethier (1998 and 2004) argues that, the criterion maximises the median of the proportional bettor's fortune. This position and observation is also shared by Maslov and Zhang (1998) but on the controversial grounds by the later that, the mean and the median of a sum of independent and identically distributed (iid) random variables are equal.

Bellman and Kalaba (1957), considered the role of dynamic programming in statistical communication theory and generalized and further extended the Kelly's result. The first to introduce the Kelly criteria in an economic context, Latane (1959), showed that investors should maximize the geometric mean of their portfolio. Also, Browne and Whitt (1996), considered the Bayesian version of gambling and investment problems, where the underlying stochastic process has parameter values that are unobserved random variables, and derived a generalization of the Kelly criterion.

In a fascinating twist to academic brilliance, Centikaya and Parlar (1997), provided a critique of the simple logarithm assumption for the utility of terminal wealth and solved the problem with a more general utility function. The duo showed that in general case the optimal policy is not Myopic.

Clearly, the Kelly criterion has been widely used in gambling and investment circles and excited more write ups. Confronted with a "favourable game" to place bet by the gambler, the criterion aver that, with a known probability of win $0 < p < 1$ with odds $\theta > 1$, the optimal percentage of the bankroll to bet on games as postulated by Kelly is given as:

$$f^* = \frac{p\theta - q}{\theta} \tag{2.1}$$

provided $p > \frac{1}{\theta}$ and $\theta \neq 1$

Mathematical derivation of this fraction is deferred to section 3.5.1

2.1.5 Key Issues in Wagering Market

Some occurrences and existence of frictions within the wagering markets between bookmakers and bettors on one hand, as well as bettor's orientation regarding the level of expectation of wins and losses on another hand, remains a lingering issue to be address, and has accordingly attracted the attention and interest of other researchers. While some of these win expectations are delusional, others remain very achievable and are all very much dependent on wagering market dynamics. Of particular interests are the utility function and efficiency and inefficiency situations, and these have largely been discussed.

- **Logarithmic Utility Function**

  The questions respectively, of how happy and sad a bettor will be if wealth was to increased exponentially than expected and/or reduced abnormally than projected after a sequence of placed bets, are better represented for good analysis through utility functions. The classical Kelly's work in this regard as he sought to maximize the log-returns of the utility of a gambler is a testament of this assertion. Also, under expected utility, a special preference for risk less outcomes is defined as risk aversion and modelled through concave utility (Diecidue et. al 2004).

  The preference of which utility functions better represents the interest of a gambler to avoid ruin has however excited interesting suggestions and research. While some researchers adopt for general utility functions for reason of easy adaptability to change, others like Bremain (1961) roots for logarithmic utility functions. Bremain (1961) thus argues that, in the framework and elegant environment of a mathematical criterion, log-utility give rise to a good strategy that increases expected returns. Whitrow (2007) mildly do not fancy logarithmic-utility function because of its vulnerability to high volatility of returns to which most investors are averse, but was quick to state that, log-utility functions are optimal for growth and capable of avoiding ruin.

Conlisk (1993) appended a utility function of gambling to the preference function of an otherwise standard theory which serves as an expected utility theory for a risk-averse individual. And further questioned the work of Friedman and Savage (1948)—simultaneous purchase of insurance and gambling within the framework of expected utility theory.

- **Betting and Gambling Markets Efficiency and Inefficiency**

  Good strategy somewhat yields the optimality of gambler's wealth cognizance of the fraction wagered in a bet. However, the inefficient and efficient nature of the betting markets, being system, sequential, parimutuel or fixed in nature has been a major focal point of discussions for many researchers because of the potential variations such existence brings to gamblers. These dynamics in betting and gambling markets have at times led to dashed expectations of bettors, and in some instances brought about accompanied but unintended consequences for bookmakers also. This originates potentially from the variation of considered probability of win by bettor and the corresponding probability as determined and estimated by the bookmakers' with an associated odd. That is, when the market odds compared to the correct odds are at variance, then the chances of winning are said not to be so good for the gambler.

  The argument for this reservation of the betting markets read as this: 'unlike most casino based games like roulette, black jack, slot machine and others where the house (bookmakers) has 'somewhat absolute' statistical edge over patrons; the bets on football scores excites more than it meets the eyes" This is because, bettors can gain substantially over bookmakers and (vice-versa) on the strengths of the formers' ability to identify odds of event that do not fully reflect true odds relative to the offered odds by the latter.

  Angie and Filippos (2011) point out in their research that, market efficiency exists when the quoted odds by a bookmaker reflects all publicly available information

related to the assessment of the match outcome probabilities. On the contrary, when there is information deficit or inconsistencies relative to the quoted odds by a bookmaker and true match outcome probabilities; the market is said to be inefficient. To them, bias which is a feature of inefficiency is a test of what a price — setting behaviour of the bookmaker is, an expressed view of bettors' about the bookmaker. Kuypers (2000) and Levitt (2004) in this regard aptly describe bookmakers as being more skilled at predicting the outcome of matches than bettors, that is to say, bookmakers optimally set inefficient odds in order to exploits biases of bettors (cited in Angie and Filippos, 2011).

Conventionally, depending on the likely outcome of events bookmakers may set odds to reflect the impulsive behaviour of bettors, since most gamblers bet impulsively based on an assigned odds to an events. This behaviour of bettors is often predicated on their biased perception of win probabilities. And information asymmetries that exist in wagering markets are, in one key respect responsible for this inefficiency existence and explain the working dynamics of the football betting markets compared to other betting markets.

Sinkey and Logan (2012) in their paper sought to investigate how behavioural strategies have an implication on wagering markets efficiency. To them, betting houses potentially make the market inefficient when some strategies are priced out at the expense of leaving profitable strategies to the detriment of gamblers. Statistically and economically, such action tends to affect the average earnings of bettors, if so, and exposes the bettor to easily being brought to eventual ruin. By this, "inefficiency is a measurable and predictable mispricing of events that leads to exploitations of bettors" *(sportsInsights.com)*.

Cain et al (2000) in their study identified two sources of betting and gambling inefficiencies which result in bias. They made the observation that, unlike horse-racing where odds are continuously varied with betting time, the workings of the

football betting markets is quite different where posted odds remain 'fixed' at least for the start of a game and sometimes till finality is brought to the game or event in question. They further makes the point that, akin to insider trading in some betting markets, bettors may use privileged information about a game to alter changes to the detriment of bookmakers but to their advantages; it would therefore not be out of place for bookmakers to insulate themselves from this possibility as would be the case in the event of insider trading by quoting odds that differ from the outcome probabilities, the trio observed. Secondly, they pointed out that, the odds of a bookmaker is largely dependent on the team winning or losing, nonetheless teams of wins possibility dispositions for win or loss fortunes are offered somewhat same odds against a particular score relative to the differential strengths and weakness they may both have vis-à-vis the various departments of the game.

Levitt (2004), in a similar fashion showed the manipulations bettors are subjected to by bookmakers, thus reinforcing the point earlier alluded to as the skilled way of bookmakers at calculating odds for consideration of bettors to place a bet. Evidently, it is not out of place to generally state that, the preoccupation of the bookmakers is not to predict without error the outcome of a given contest or events. Heuristically, this crates biases in the betting market and thus making it inefficient.

In one main regard, and in much documented literatures regarding market inefficiency in wagering market is the favourite—long shot bias (Ottaviani and Norman, 2007).

Favourite—longshot bias (FLB), can be describe as the observed tendency for the expected return to bets placed at lower odds to exceed that to bets placed at higher odds. That is favourite win more than the subjective market probabilities with longshot being less, and has been explained in terms of risk-affine gamblers who prefer a lower probabilities higher return in the face of obvious correct odds of favourable result. Levitt (2004) find a mild systematic bias in wagering market in

the form and nature of FLB, and further opined that, this bias is driven mainly by underbetting of scores results and overbettting of longshot draw outcomes.

Superficially, this widely known bias in betting and gambling market is a complete departure from the market efficiency hypothesis. According to Ottaviani and Norman (2007), the bias arises when the expected returns on long shot bets thoroughly tend to be lower than on favourite bets. They further asserted that, in simultaneous betting, this bias arises if last minute bet is placed by bettors who are privileged to have private information without necessarily knowing the distribution of other gamblers.

Cain et al (2000) however avow that, FLB is traditionally found in horse-racing but has gradually snowballed to become a feature of the football betting and gambling game. And appears when a gambler bets on result being home win, away win or draw and sometimes placing a bet on precise score of margin. The resultant effect of this known bias causes betting on favourites to yield significantly higher returns than betting on longshots.

In contrast to the above enumerated inefficiencies in the betting and gambling markets which are well documented in the body of literatures, equally much literature abounds to indicate that, efficiency exists in betting markets. In essence, there is a mixed view of statisticians and economists regarding the smooth operations of the wagering markets. Thaler and Ziemba (1988) posit that, wagering markets are best cut out for the exploration of market efficiency than other portfolio markets like stocks and assets. A priori, the wagering markets offer quick repeated feedback thus providing an environment of thrives for most gamblers.

Thaler and Ziemba (1998) goes on to give forms of efficiency in the wagering markets as being weak and strong in nature. Weak form of efficiency exists when no abnormal profits or returns can be obtained by using only price information, and no bets should have positive expected value. Strong form efficiency on the other hand asserts that,

all bets should have equal expected value.

Clearly, efficiency and inefficiency in the wagering market have been a major concern. Economists have thus researched into its effects on the returns of gamblers.

2.2 Empirical Review

Empirical studies regarding the optimality of a bet coupled with models of predictions for football in the wagering markets abounds. The questions respectively of what constitutes optimality of a bet, and, how accurate are models for prediction for football matches have been investigated, and results obtained from some leagues, champions leagues data as well as odds data across the world for some time now.

Hirotsu and Wright (2003) through the approach of maximum likelihood estimation (MLE) evaluated the characteristics of teams by empirically using 1999—2000 English Premier League (EPL) real data. The study through the analysis of the data revealed that, the offensive and defensive strengths, preference for playing at home and relative success against opposition teams were better fashioned in that regard. By means of chi—square tests via-a-vis the real data, the duo failed a rejection of the Poisson assumption made in the model. The duos', with the assistance of the given data graphically (in Cartesian plane) illustrated the offensive and defensive strengths of teams with a subsequent conclusion that "no team has a particular overall propensity of success both for gaining and for the avoidance of losing possession. That is, there is no team with absolute home advantage in possession" (italicised mine). This is against the backdrop of similar work by the duo in (2002) that described and laid an explanation of a Markov model of football match, where representation of stochastic transition rates of scoring and conceding goals were not only key variables of interest, but, also the rate of gaining and losing possession were addressed using explanatory variables of home advantage, offensive strengths and defensive strengths of teams. Stochastic rates of transition arise due to the change of ball possession and goal scoring of opposing teams at an estimated point in time. Other works

that seek to empirically analyse real data of football scores tow the lines similar to Hirotsu and Wright (2003) and (2002), in the form and fashion of non—parametric tests that is spiced up in graphical analysis framework. Maher (1982) through the Poisson regression model and using maximum likelihood estimation (MLE) approach, did an analysis of English football league divisions from 1971-1972, and estimated the offensive strength potency at scoring goals as well as the defensive strength laxity against conceding goals for each teams. The conclusions arrived at served as the basis for the work of Dixon and Robinson (1998), which incorporated a scoring rate for goals and empirically used an English league and cup football data for three consecutive seasons from 1993—1996 to formulate and obtained a statistical model for predictions.

Using data for five consecutive seasons from 1992—1997 football association league fixtures and home win, draw and away win probabilities, Crowder et al through the refinements of Dixon and Coles (1997)'s independent Poisson model, modelled a 92 football teams for 1992-1997 using the English Football Association (EFA) League. The focus, however, of their paper concentrated on probabilities of teams' win, draws and loses because of the key reliance of such probabilities of bettors' at placing a bet. In like manner, a study to determined whether or not Manchester United indeed deserved to have been crowned winners of the 1995/1996 seasons was conducted by Lee (1997); this he did through the modification of Maher's (1982) model to derive the probabilities for each match and simulated over a number of times to calculate the points awarded to each team. Reep and Benjamin (1968) modelled the number and type of passing moves within a game empirically.

The Poisson distribution has been widely accepted as a suitable model for predicting probabilities of football games, but, without mild deficiency. A simplifying assumption often used in modelling process through this distribution is that of independent between goals scored by home and away teams. Maher (1982) model for example, used two independent Poisson variables where the relevant parameters are constructed as the product of the strength in the attack for one team and the weakness in defence for the

other. An improved and efficient model of football prediction over Maher (1982) and Dixon and Coles (1997) was a birth process model of goal times by Dixon and Robinson (1998). By employing data of over 4000 soccer matches from the English Premier League, evidence was adduced that, the rate of scoring goals changes over the course of a match. This rate tends to increase over the game but is also influenced by the current score. And further applied the model at finding optimal spread bets in the wagering markets.

However, regarding the negative binomial distribution and how useful and accurate it has been over the period, not much work abounds compared to the Poisson distribution counterpart. However, Pollard (1985) empirically makes a point for the Negative binomial distribution as a better fit for the number of goals scored by a home and away teams. This He did by analysing and parsimoniously fitting goals scored for 581 English football games. Reep et al. (1971) examined the fit of the negative binomial distribution to scores from football matches and other goal scoring games. The study however made a conclusion that "chance dominates the game", and thus making it difficult to properly predict outcomes for matches within the model relative to the inherent "noise" in the used data. Hill (1973) makes a claim that, the game of football is either all skill or all chance by justifying Maher (1982) argument and approach through the calculation of the correlation between expert opinions and the final league tables. He further concluded that, even though chance was involved, there was also significant amount of talent that affects the game from the blast of the whistle to the final outcome of the match.

In relation to optimal betting strategies, many writers and researchers adopted a theoretical approach at illustrating optimality strategies for number of bets. Gottlieb (1985, An optimal betting strategy for repeated games), Thorp (1969, Optimal Gambling Systems for Favourable Games), Savage and Dubins (1960, Optimal Gambling System), Breiman (1961, Optimal Gambling System for Favourable game) and others develop optimal betting systems. Dixon and Coles (1997) applied the model of bivariate Poisson distribution and further used for seek for profitable betting strategy for spread betting. In this respect, the ratio of the probabilities of goal score and that of odds should be

positive to cause bettors' to place a bet.

Regarding the efficient and inefficient nature of the wagering markets, data of over 11,000 games from 1985 through to 2003 were employed to test for the potential market inefficiency in the betting and gambling markets. Specifically, the research found out why, by default, deliberately or dictated by the market; bookmakers rationally price out certain strategies at the expense of other equally profitable strategies. This inefficiency is as a result of the statistical overpricing of favourites with home teams being underpriced statistically. (Sinkey and Logan, 2012).

The survey has so far brings to fore the complexities in statistical analysis regarding sports scores and the intricacies involved at placing a bet for optimality of a fund. While the Kelly criterion is very much a recommended strategy for adoption by the bettors' to leap over ruin and ensure optimal funds to meet budget constraints; other issues of market efficiency and inefficiency coupled with the logarithmic function to adopt for use is very much discussed for consideration at placing a bet. Again, various models for football predictions has been proposed for consideration to place a bet, but, without some deficiencies as the Poisson and negative binomial approaches have been criticised from within the circles of statistical analysis of its effectiveness at predicting the probabilities of win, draw or lose regarding games played. Clearly, most of the documented literatures on predictions takes its roots model basics from the idea of Poisson and Negative binomial distributions because of the rates of scores and events involve in the game of football.

CHAPTER 3

METHODOLOGY

This chapter is devoted to data description, theoretical and model framework and methods of estimation for the study. The chapter is in six main sections, with related sub-divisions where necessary. The first section looks at the description of the available data and model formulation for football scores, whiles the remaining sections in summary look at the estimation procedures for the study, and, optimality determination and risk analysis respectively.

3.1 Data Description and Model Formulation

Data has been collected over a season period in the English Premier League (EPL) for reason of easy of its availability, the most bet-on league in Ghana and for consistency purpose. This is because, some three bottom teams get relegated to the lower division in each season, which means that the full compliment of 'same 20 teams' for two or more seasons would naturally not be available; and if available, better baseline would not have been set in this regard. The data collected relates to the 2013/2014 season, for reason that, it serves as the most recent past data for good analysis to be made of teams and further provide a good guide for bettors at placing a bets on these teams.

Specifically, the data collected is the whole season scores for teams regarding both home and away play. This is the entire 380 home and away game scores between teams. Additionally, fouls against a teams, shots on target, corner profiles of teams, yellow cards that cautions team players and red cards that send players off the field data relating to the period in question have all been collected for insightful analysis to be made in that

regard. Season league table is provided for checks to be made of the proposed model. Further work however have been applied to get the data in the form and shape as shown in the appendix section (relating to odds). The data source in respect of the entire season scores was the website *www.footstats.co.uk/index.cfm?false=game*. Accordingly, the basic model proposed in this paper makes use of these data.

However, we explore optimal betting strategy later using sample assigned odds for the same matches from a bookmaker in Ghana's wagering markets. *Mybet.com sport betting company* as is the largest bookmaker in Ghana's betting markets and provides a reason for obtaining odds from their outfit. A typical set of bookmakers' odds for a particular match might be (2:9, 11:5, 6:4) also sometimes written as $(\frac{2}{9}, \frac{11}{5}, \frac{6}{4})$ respectively for home win, draw and away win often referred to as ratio (fractional) odds. We however omit the draw odds for our analysis subsequently for reason that the bettor's fortune is largely tied to teams' win or lose for events of similar kind.

For the above example of odds given, a stake of 9 units in a home win would yield a profit of 2 units if that outcome occurred, with similar interpretations for the draw and away wins respectively. Further, odds θ_1:θ_2 as those given can be transformed into probability P by using the formula given by Dixon and Cole (1997, pp.276) is

$$P = \frac{\theta_2}{\theta_1 + \theta_2} \tag{3.1}$$

Using (3.1) the given set of probabilities for the above given odds are (0.82, 0.31, 0.40) which sum up to 1.53 is a standard phenomenon in the betting and gambling markets. This is thus consistent with earlier discussion that odds of bookmakers provides an implicit level of probabilities for the consideration of bettors.

Further, fractional odds as in the form above can easily be converted to decimals odds for their convenience (in handling them) and for easy analysis purposes. Decimal odds also tell how much profit one is likely to make from a placed bet implicitly. To convert fractional (ratio) odds to decimal ones, we need to take the given decimal (ratio) odds

and add 1 to incorporate the returned stake which we summarise as been

$$\frac{L}{R} + 1 \tag{3.2}$$

where $\frac{L}{R}$ indicates the numerator L and denominator R (conventionally represented) as components of a fractional odd.

3.1.1 Theoretical Framework of Model Specification

The game of betting and gambling has gradually hatched a bloomed market for itself in Ghana; and this is evidenced by the large sports bet markets under the supervision of various bookmakers of different trademarks, and the huge patronage of service by patrons and lovers of association football. This is in spite of the complexities, coupled with the twists and turns of events and happenings in the market.

Football and few other games alike has many admirers, and particularly for football, because of the fun and thrilling experience it provides as well as the opportunity for profiting from its outcome through betting and gambling. However, the prediction accuracy of bettors which serve as the basis for placing profitable bets has not come as an easy task for adoption. And many models of football scores for prediction have long being proposed, including that of Maher (1982), Dixon and Coles (1997), Dixon and Robinson (1998), Crowder et al (2002) have all sought to analyse the dynamic of the football games though through different procedures. These different procedures adopted by these statisticians come with its own strengths and weakness with many offered modifications appropriately suggested for improvement.

In the game of football, teams have their inherent distinct qualities relative to various departments the game offers. These differentials in team's qualities invariably and mostly go to determine the outcomes of games. This explains why when a good team plays a weaker team, there is high certainty of win for the good team. Such preposition does not however undermine the crucial role played by the twin clichés of most sports analysts

(chance and skills), though Reep and Benjamin (1968) argue that, "skill rather than chance dominates the game".

Statistically, the game of football is viewed as a random event because of possible outcomes of win, draw or loss, with associated probabilities of their occurrences in any event of interest. These outcomes are often driven by away and home play (influence) of teams vis-à-vis the number of goals scored in a match for a 90-minutes time span, and possibly an extra (injury) time play. Scores thus become the focus of modelling because of the rich information obtained and the possible formalization of its values. Indeed, this is the starting point of Maher—Poisson approach. Dixon and Coles (1997) offer suggestions for model building that is not alien to Maher—Poisson model which is enumerated here to help create optimal betting strategy through a well specified model.

- The model should take into account different abilities of both teams in a match.
- Evidence shows that home support influences teams' performances, and ought to be a considered part of the model.
- Recent performances of teams' provide somewhat sufficient measure of strengths and abilities for direction.
- Separate measures should be made for goals scoring (attack potency) and goal conceding (defence lapse) of teams in an event.
- In making a summarize view of a team's performance by recent results, account should be taken of the ability of the teams that they have played against.

Evidently, these aspects of the game are composite in nature and therefore must accordingly be evaluated inseparably. In this wise, statistical model that encapsulates these characteristics of teams is worthy of consideration. Fundamentally, the number of goals scored by both home and away teams in any encounter of such is independent and can therefore be viewed as a bivariate Poisson distribution with describe means being

the respective attack and defence qualities of the involved teams. This is the underlying assumption of Maher's seminal prediction model which further received modification by Dixon and Cole (1997) for time dependence as the game evolves.

The reason for considering and modelling scores as independent Poisson distribution is because of its time—bound, random and discrete nature of football scores, and, thus fits perfectly into the chosen distribution. Also, these scores are team—specific due to the distinct inherent strengths of teams, thus making it independent. Secondly, the attack and scoring potency of teams' is at large dependent on the ball possession capability of teams. Maher (1982) posits that, compared to the possession of a team in a match to the likelihood that an attack results in a goal, the latter is minimal. Hence, if the probability p is fixed for a time and attack rates are independent, Poisson distribution arises and becomes appropriate for consideration.

Maher (1982) and Dixon and Cole (1997) model starts with the above framework and subsequently considers the goals scored by teams as being independent. And further through these ideas seek to determine the probability of result between teams. They starts the model building by considering a football match between two teams at a time, t.

Typical of a football match, let the teams be a home team i and an away team j at time t, and respectively represent the scores of these teams as x_{it} and y_{jt} at time t. For the variability of team qualities relative to the attack (strengths or potency) and defence (weaknesses) of teams as well as home advantage factor, Maher (1982), Dixon and Coles (1997) represented the attack strengths of teams i and j at time t as α_{it} and α_{jt}, and respective defence rates as β_{it} and β_{jt}. By the independence assumption between teams' scores, the model appears as

$$P(x_{it}, y_{jt}) = \{exp(-\eta\alpha_{it}\beta_{jt})(\eta\alpha_{it}\beta_{jt})^{x_{it}}/x_{it!}\}\{exp(-\alpha_{jt}\beta_{it})(\alpha_{jt}\beta_{it})^{y_{jt}}/y_{jt!}\} \tag{3.3}$$

where η is the home advantage factor. α_{it}, $\beta_{jt} > 0$. That is, $x_{it} \sim$ Pois $(\eta\alpha_{it}\beta_{jt})$, and $y_{jt} \sim$

Pois $(\alpha_{jt}\beta_{it})$, with x_{it} and y_{jt} being independent.

It is easy to recognise from the above equation that, more goals will be scored by team i against team j when the parameters η (home advantage), α_{it} (attack rate for team i at time t), and β_{jt} (defence rate for team j at time t) increases.

Similarly, there will be *more* edge of goals over the other end when the factor $\alpha_{jt}\beta_{it}$ increases. This model helps to assess the scores of teams involve in a game. The equation further says that, given two respective teams as indexed home team i and away team j, the probability of a team win or loss can easily be estimated relative to the (known) values of the parameter estimates as in equation. (3.3)

The arduous task however is for the estimation of the parameters as in (3.3), especially for n number of teams. However, it is easy to infer accordingly that, for n teams, the attack parameters are heuristically $(\alpha_{1t}, \alpha_{2t}, \alpha_{3t}, ..., ..., \alpha_{nt})$, with defence parameters being $(\beta_{1t}, \beta_{2t}, \beta_{3t}, ..., ..., \beta_{nt})$ and home advantage parameter η (fixed) to be determined. Maximum Likelihood Estimation (MLE) approach is thus adopted to estimate these parameters for n number of teams.

For the avoidance of over parameterization, a constraint is imposed on the model in the estimation of the parameters. The constraint is thus given as

$$n^{-1}\sum_{i=1}^{n}\alpha_i = 1 \tag{3.4}$$

That is, the sum of the attack parameters is equal to the number of teams in the season, thus making the model parameters quite unique and plausible for consideration.

For the English Premier League (EPL) of 20 teams in a season, with our focus of estimating teams scoring rates as in the λ_{aorb}, from fixed home advantage factor η, home and away attack and defence parameters of opposing teams. There is 20 distinct scoring intensities of teams to be estimated.

Setting up the pseudolikelihood function with teams indexed $k = 1, 2, 3, ..., ..., ...N$, with

corresponding scores (x_k, y_k), this takes the form

$$L(\alpha_i, \beta_i, \eta; i = 1, 2, ..., ...n) = \prod_{k=1}^{N} e^{-\alpha_{i(k)}\beta_{j(k)}\eta}(\eta\alpha_{i(k)}\beta_{j(k)})^{x_k} e^{-\alpha_{j(k)}\beta_{i(k)}}(\alpha_{j(k)}\beta_{i(k)})^{y_k} \quad (3.5)$$

where $i(k)$ and $j(k)$ respectively represent the home and away teams locked up in match k. According to Dixon and Coles (1997), a structural limitation of (3.5) stems from the static nature of teams performance as evidenced by their respective attack and defence capabilities α_i and β_i of the model.

Evidently, the model above makes good use of scores (important though) and home advantage factor without recourse to other features of the game of football. In fact, as earlier revealed in chapter two of this paper, many sports prediction model tow the line of scores modelling by using attack and defence rates which are easily derived from the data of goals scored and conceded, without key cognizance of other characteristics of interest on the game which invariably determine football scores either playing at home or away. That notwithstanding, many researchers have analysed separate variables and their influences on games including that of weather, pitch suitability, red cards, shots on targets and others of such which are fundamental to the scores of teams in football games. This limitation of previous models stems from the lack of *joint representations* of many interest variables for score determination and the associated difficulty in estimating these values, and a source of major motivation for the exploration of other key variables, with key emphasis on estimating the home and away scoring rates of teams.

3.1.2 Model Formulation for Football Scores

This paper in line with Maher (1982) and Dixon and Coles (1997) prediction model formulates a Poisson regression model that extends to incorporates a key variable not in the above model, and further seek to determine how significant these variables are, in the model for prediction purposes through statistical tests. However, the focus will be to estimating the scoring rates of teams (both home and away) which are linear combinations

of these variables from the formulated Poisson regression Model (PRM).

Suppose there are two teams indexed k and l playing a game i; conventionally, let's represent the goals scored by these two teams respectively by X_{ki} and Y_{li} at time t. Further, let λ_h and λ_a respectively represents mean *home scores* and mean *away scores* of the above indexed teams at time t on the tacit assumption that, the goals scored are independent Poisson distribution. It easily follows that, a match between teams k and l is *bivariate Poisson* random variable.

If the mean λ_h is made to reflects the home attack strength, defence lapse of the away team, home edge and *shots on target* of the home team. And the mean λ_a also made to reflects the attack potency of the away team, defence lapse of the home team, and *shots on targets* for the away team.

In essence, each team has its own specific characteristics determining its expected scores per match. As earlier explained, and like the model by Dixon and Cole and Maher, the probability of win, draw and loss of a match in this instance is given by

$$P(X_{ki}, Y_{li}) = \{exp(-\lambda_h)\lambda_h^x/x!\}\{exp(-\lambda_a)\lambda_a^y/y!\} \tag{3.6}$$

That is, $X_{ki} \sim$ Pois (λ_h) and $Y_{li} \sim$ Pois (λ_a) with $(X_{ki}, Y_{li}) \sim BP(\lambda_h, \lambda_a)$ where BP is bivariate Poisson. However, to estimate the probabilities as in equation (3.6), we need the estimates of λ_h and λ_a.

The described respective mean scores of the teams involved as reflecting its characteristics is a Poisson and must be positive. Therefore, the logarithms of the means can be considered as a linear combination of its factors (characteristics). In this wise, respectively, the following equation holds

$$Log(\lambda_h) = \mu + home_{adv} + att_{hi} + def_{ai} + shot_{hi} \tag{3.7}$$

Symbolically, the above equation can be rewritten as

$$Log(\lambda_h) = \mu + \eta + \alpha_{hi} + \beta_{ai} + \gamma_{hi} \tag{3.8}$$

Likewise for the away team, the following equation hold for the mean score

$$Log(\lambda_a) = \mu + att_{ai} + def_{hi} + shot_{ai} \tag{3.9}$$

Similarly, equation 3.9 can symbolically be written as

$$Log(\lambda_a) = \mu + \alpha_{ai} + \beta_{hi} + \gamma_{ai} \tag{3.10}$$

for $i = 1, 2, 3...., ..., n$ and indicative of a game played and n is the number of teams, with α_{hi} and β_{hi} being the attack strength and defence ability of team k home team in game i, while α_{ai} and β_{ai} encapsulates the corresponding offensive (attacking) and defensive performances of team l (away team), η is a constant parameter denoting a home advantage effect for home team, with γ_{hi} and γ_{ai} as the respective shots on targets for teams k and l respectively. The described models as stated in (3.8) and (3.10) are known as the Poisson regression model which we describe momentarily.

Indeed, this is an extension of Maher and Dixon and Coles work that seeks to incorporates *shots on targets* of competing teams to assess the probability of win or lost of teams.

3.1.3 Model Inference

The model of equations (3.8) and (3.10) relates to expected goals for number of teams of both home and away teams (vis-a-vis the intensities of scores) which are linear combinations of its factors. The resulting values of these models (equations) help to directly estimate the probabilities of (3.6).

Also, it is pretty obvious that, cognizance of the number of teams in the English premier

league (the league for whose data is used), model estimates for each team is to be provided (i.e the scoring rates of teams).

As suggested by Dixon and Cole and other researchers, a constraint is imposed on the model to avoid it been over-parameterised, and also for purpose of model identifiability. Thus, in line with Maher's work, the mean sum of the parameters is made to sum up to 1.

$$n^{-1}\sum_{k=1}^{n}\alpha_k = 1$$

Same can be made for the β_s and γ_s

That is, the above restriction ensures that, the number of parameters are uniquely assigned to each teams in the premiership.

As stated already, these parameters can be estimated through the maximum likelihood approach. Setting up the maximum likelihood for the model, we have

$$L(\alpha_i, \beta_i, \gamma_i, \eta, \tau; i = 1, 2, ..., ...n) = \prod_{k=1}^{N} \lambda_h^x e^{-\lambda_h} \lambda_a^y e^{-\lambda_a} \tag{3.11}$$

Where $\lambda_h = \mu + home_{adv} + att_{hi} + def_{ai} + shot_{hi}$ and $\lambda_a = \mu + att_{ai} + def_{hi} + shot_{ai}$

Very noticeably, the estimation of the parameters (from which scoring rates of the teams are generated) is not straight forward because of the lack-of-closed-form nature of the function, and optimisation techniques are resorted to in this instance to aid the estimation of the parameter values. This is because, the probability surface for maximum-likelihood Poisson regression is always concave, making Newton–Raphson or other gradient-based methods appropriate estimation techniques.

3.1.4 Goodness of Fit and Model Diagnostic Tests

The questions relating to the above proposed model that easily comes to mind can be stated as:

i. what is the probability that the factors determining the mean score of either home

or away team are significant.

ii. If it does, how strong is the influence.

Tests for statistical significance is used to address the former to assess whether or not the influence of these parameters on mean score for a particular at point in time is a mere coincidence or indeed a relationship, while the later question is addressed by measures of association. This ensures the reliability of the model and very much so for the consideration of bettors.

The formulated models above are implemented with described data at the *'data description'* stage. And the above questions are address through tests of the significance of the parameters in the models and through null deviance. The former is easily and informally done by direct comparison between significance levels. Also, Wald's test of significance can be adopted for this all important exercise.

The assumption of independent Poisson earlier made regarding the models above are tested with the available data to determine the relevance of the model and to avoid spurious effect of the model, if any. This is discussed in section (4.1)

Regarding the goodness of model fit relative to the available data and statistical packages output, null deviance statistical measure easily comes to mind in the circles of statistical analysis.

3.1.5 Separate Home Advantage Factor

The influence of home advantage factor in the game of football is very much documented, Dixon and Coles (1997) and Maher(1982) for the least, make references to it.

It is easily determined by the ratio (dividing) of the number of points amassed at home by the total of earned point of the season, with the mindset that, the home advantage factor if really influenced play for the home-team should reflect in the 'statistics' of the game-*ceteris paribus*. That is, a home advantage factor must work magic for the home team beyond individual team strength of the involved teams, in which case it must reflect

in the number of fouls earned or committed, more shots on targets, cards against opposing teams both yellow and red as well as number of corners earned in the match. Clearly, these variables must reflect to depict and make a case for home teams cognizance of the euphoria created in favour of the home 'fans support'.

In the stance of this argument, a model is fitted to assess which of the above variable is key for home score determination for the study of bettors in placing profitable bets. Also, the model is compared with an away play model to determine the best model fitted by these variables.

The model thus fits a Poisson regression to home and away scores as the dependent variable with the above variables as the explanatory variables.

$$Log(\lambda_h) = Inter + FoulF + FoulA + RedCA + CornersP + YellowCA + ShotsPER \tag{3.12}$$

The value of λ_h in equation (3.12) should help in the estimation of the probability of a 'home play game' on the assumption that, a home advantage effect will result in one more goal margin in any game played. That is, cognizance of the influence of home play, we assumed any game should result in the edge of at least a goal margin over the visiting team-ceteris paribus. The opposite could possibly happen.

Further, on the assumption that goal score follows a Poisson, the probability of scoring a home or any away goal is thus given by

$$P(X_k = x) = \frac{e^{-\lambda_h}\lambda_h^x}{x!} \tag{3.13}$$

It is pretty obvious from above that, the value of λ_h directly helps to estimate the probability of a home win relative to a given score.

Further, we generate values for the home advantage factor or effect and seek to know and assess how home advantage influences performance in playing at home vis-a-vis the available data set.

3.1.6 Deviance Statistics

In the GLM framework, it is customary to use a quantity known as deviance to formally assess model adequacy and to compare models. In statistical analysis, deviance statistics are identical to those obtained using likelihood ratio test. And further describes the quality of fit for a model that is often used for statistical hypothesis testing. In fact, it is a generalization of the idea of using the sum of squares of residuals in ordinary least squares to cases where model-fitting is achieved through maximum likelihood.
The deviance for a model M_1, based on a dataset y is defined to be

$$D(y) = -2(log(p(y|\theta_1)) - log(p(y|\theta_s))) \tag{3.14}$$

The factor θ_1 denotes the fitted values of the parameters in the model M_1, with θ_s denoting the fitted parameters for the "full model".
In the class of generalized linear model (GLM), where it has a similar role to residual variance from Anova in linear models (RSS). Suppose in the framework of GLM, we have two nested models M_1 and M_2, such that, M_1 contains the parameters in M_2, and k additional parameter(s). Then, under the null hypothesis that M_2 is true model, the difference between the deviances for the models follow approximately chi-squared distribution with k-degree of freedom.
In essence, deviance values help to determine the level of dispersion (under or over) within a model for correction or otherwise. The relationship between deviance and its degree of freedom of a given model provides a basis for testing models level of dispersion. That is, the value of the fraction *deviance / degree of freedom (df)* is the basis of knowing, and for decision making regarding whether the underlying assumption of variance being equal to the mean is satisfied (ideal situation) or variance being greater than mean (over dispersion) relative to the Poisson distribution is satisfied or violated.
Specifically, if the fraction *Deviance value / df* >1, there is presence of overdispersion for

which action is needed to remedy the situation, however, if the *Deviance / df* <1, there is underdispersion.

The negative binomial and the zero-inflated poisson models are adopted to remedy the over or under dispersion difficulties in models like the Poisson.

3.1.7 Statistical Package Usage

The statistical packages used for this study is the R package and spreadsheet analysis. The reason for their use are because while the former is generally acceptable and know for its user friendly, the latter is accustomed to easy generation of its results as well as providing better view of its working environment.

In the case of the former, inbuilt packages such as *sandwich, gplot, ggplots2, MASS* and others will be used for fitting and estimation of the Poisson regression parameters and graphical plots.

Data is initially keyed-in into excel environment, imported into SPSS (16.0) for further data structuring to be made and then sent into the R console for the generation of the relevant statistical measures and values.

3.2 Relevant Descriptions

3.2.1 Generalized Linear Model(GLM)

The Poisson Regression Model(PRM) shortly to be discussed in this paper is a classic statistical model that is amenable to the family of generalized linear models.

Generalized Linear Models (GLMs) relates to class of statistical models that link responses to linear combinations of predictor variables, that includes commonly encountered types of dependent variables and error structures in special cases.

In addition to regression models for continuous dependent variables, models for rates and proportions, binary, ordinal and multinomial variables and counts can be handled as

GLMs.

3.2.2 The Poisson Distribution

3.2.3 Definition

The number of accidents in a year, the number of patients visiting a hospital, the number of fire outbreaks, including goals scored per match by teams and others including Feller (1957) example of the number of flying-bomb hits in his classic text on Probability are better accounted for by the Poisson distribution. This is because, the distribution is frequently used to model the random number of occurrences of some 'rare' event over a given time period. The probability distribution is defined by a single parameter often describe by (lambda) such that:

$$P_r(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{3.15}$$

for $x = 0, 1, 2, 3....$, and $\lambda > 0$, with the unique feature that, the mean and variance of the distribution are equal. That is,

$$\mathbf{E(x) = Var(x) = \lambda} \tag{3.16}$$

The parameter λ in the above equation indicates the rate at which the events take place per unit of time.

Because, the mean is equal to the variance, it follows that any factor affecting one will invariably affect the other. Thus, the usual assumption of homoscedasticity would not be appropriate for count or Poisson data. In fact condition (3.16) is a property of the Poisson Model. In the event that, the condition in equation (3.16) do not hold, for which Var(x)> E(x), then there is over dispersion in the model for which the negative binomial or the Poisson zero inflated models are again adopted for rectification. Underdispersion

is also corrected via these models (Karlis and Ntzoufras, 2003).

3.2.4 Derivation of Poisson Distribution

The Poisson distribution as a discrete distribution is derived as a limiting case or form of the binomial distribution. When the number of success is very large in a Bernoulli trials of specified Probability in each trial. From above, the approximate probability that one of these rare events take place in a short time period of duration $\triangle\, t$ is

$$\lambda.\ \triangle\, t \tag{3.17}$$

The Poisson distribution function arises as the limiting case of the binomial distribution if we divide a unit time period $(0, 1)$ into n equal parts where n is a large positive integer, so that each of these subintervals has length $\triangle\, t = \frac{1}{n}$.

The approximate probability of an event occurring in any one of the subintervals has length $\lambda.\ \triangle\, t = \frac{\lambda}{n}$. Since the interval is very short we will answer that the probability of two or more events occurring in any subinterval is zero. So, in each sub-interval we have either one event occurring in or no event occurring.

Now considering each of the n short interval as a Bernoulli trial where 'success' means that exactly one event occurs in the subinterval Then there are n trials, each with associated probability of success and failure of $p = \dfrac{\lambda}{n}$ and $q = 1-p = \dfrac{n-\lambda}{n}$. Let x denote the number of success in the n trials (i.e the number of events that occur in (0,1)). Approximately, the probability of x events occurring in the unit time period (i.e x success in n trials using the binomial distribution as)

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \tag{3.18}$$

This approximation should get better as n increases, since as the subintervals become smaller the probability of two or more events occurring will approach zero. Rewriting the

probability function as:

$$P(X = x) = \frac{n!}{x!(n-x)!}p^x q^{n-x} \tag{3.19}$$

$$= \frac{n(n-1)(n-2)...(n-x+1)(n-x)!}{x!(n-x)!}(\frac{\lambda}{n})^x(\frac{n-\lambda}{n})^{n-x} \tag{3.20}$$

Factorising n in each term for nth number and cancelling $(n-x)!$ terms out, we get

$$= n^x \frac{(1-\frac{0}{n})(1-\frac{1}{n})(1-\frac{2}{n}).....(1-\frac{-x-1}{n})}{x!}(\frac{\lambda}{n})^x(\frac{n-\lambda}{n})^{n-x} \tag{3.21}$$

Further cancelling n^x in equation (3.21), we get

$$= \frac{(1-\frac{0}{n})(1-\frac{1}{n})(1-\frac{2}{n}).....(1-\frac{-x-1}{n})}{x!}\lambda^x(\frac{n-\lambda}{n})^{n-x} \tag{3.22}$$

Taking the limit of equation (3.22) as n tends to infinity, the first term in the above expression approaches 1 as n increases, since x and λ are fixed, and there are x factors in both the numerator and the denominator. That is, equation (3.22) become

$$= \frac{\lambda^x}{x!}\frac{\lim_{n\to\infty}(1-\frac{0}{n})\lim_{n\to\infty}(1-\frac{1}{n})\lim_{n\to\infty}(1-\frac{2}{n})....\lim_{n\to\infty}(1-\frac{-x-1}{n})}{x!}\lim_{n\to\infty}(\frac{n-\lambda}{n})^{n-x} \tag{3.23}$$

It is pretty easy to know that, equation (3.23) becomes

$$\frac{\lambda^x}{x!}(1.1.1.1....1)\lim_{n\to\infty}(\frac{n-\lambda}{n})^{n-x} = \frac{\lambda^x}{x!}\lim_{n\to\infty}(\frac{n-\lambda}{n})^{n-x} \tag{3.24}$$

However, from fundamental calculus, the limit of the second term is:

$$\lim_{n\to\infty}(1-\frac{\lambda}{n}) = e^{-\lambda}. \tag{3.25}$$

Hence, we have the probability function of the Poisson distribution with parameter λ:

$$P(X=x)=\lim_{n\to\infty}[\frac{n(n-1)(n-2)...(n-x+1)}{(n-\lambda)^x}(1-\frac{\lambda}{n})^n\frac{\lambda^x}{x!}]=\frac{e^{-\lambda}\lambda^x}{x!} \tag{3.26}$$

This is the derivation of the Poisson distribution function from which the log link function specifies the Poisson regression model which we discuss presently.

Further, the distribution of a random variable x with Poisson (λ) distribution as given by equation (3.15) is re-written as

$$P_r(X=x)=\frac{e^{-\lambda}\lambda^x}{x!} \tag{3.27}$$

$$=exp\{x\log\lambda-\lambda-\log x!\} \tag{3.28}$$

$$=exp\{\frac{x\log\lambda-\lambda}{1}-\log x!\} \tag{3.29}$$

So the canonical parameter is $\log\lambda$. In this case, the dispersion parameter is 1. It is for this reason that R output as will be shown in beneath some output tables give the message (*Dispersion Parameter for Poisson family taken to be 1*)in the summary output when Poisson distribution is fitted in R. Also, this explains why Poisson distribution is a one-parameter exponential family distribution.

The above derivation of the Poisson from the limiting form of the binomial distribution can also be obtained from the idea of stochastic processes under conditions of occurrences in time interval being very small and proportional to the length of the interval and negligible in a sense, with a disjoint time interval being mutually independent.

3.2.5 The Poisson Regression

The Poisson regression is used to model count data in the form and nature of the above model, where the logarithm of the expected value is seen as the linear combination of unknown parameters to be estimated. It further assumes that, the dependent variable(s)

λ_h and λ_a has a Poisson distribution (which has been describe above). Poisson regression models are generalized linear models with the logarithm as the link function and a Poisson error, with the Poisson probability distribution as the probability distribution of the response. For example, a generalized linear model with link log is given

$$\log(\mu_i) = x^{'}\beta \tag{3.30}$$

From equation (3.30), the regression coefficient β_j represent the expected change in the log of the mean per a unit change in the predictor x_j. Exponentiating equation (3.30), the multiplicative result is

$$\mu_i = \exp(x_i^{'}\beta) \tag{3.31}$$

The coefficient in equation (3.31) thus represent a multiplicative effect of the $j-th$ predictor on the mean of the model, which accounts for the unit change predictor variables. In essence, if $x \in \mathbb{R}^n$ is a vector of independent variables, then the model takes the form $\log(E(Y|x)) = \alpha + x^{'}\beta$, where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^n$. More compactly, this can be written as $\log(E(Y|x)) = x^{'}\beta$ such that there are $(n+1)$ dimensional vector consisting of n independent variables concatenated to a vector of ones. In this scenario, θ is simply concatenated to a vector of β.

The advantage of using the log link lies in the observation that, with count data the effects of predictors are often multiplicative than additive.

3.3 Estimation Techniques

3.3.1 Maximum Likelihood Estimation (MLE)

The idea of maximum likelihood estimation originally postulated by R.A Fisher in the $1920s$, states that, the desired probability distribution is the one making the the observed data "most probable". That is, parameter vector is sought after for the maximization of

the likelihood function.

The log-linear Poisson model discussed in sections (3.2.4) and (3.2.5) is a GLM with specified Poisson error and link log, and maximum likelihood estimation (MLE) approach is thus resulted to in getting parameter estimates.

MLE is a statistical technique largely known for estimating model parameters. Basically, this statistical technique seeks to answer the question: what model parameters are most likely to characterise a given set of data? As the name connotes, MLE proceeds to maximize a likelihood function, which in turn maximizes the agreement between the model and the available or given data.

More intuitively, MLE derives the parameters for a probability density functions (PDF) of a particular or known distribution. The likelihood function is obtained by considering the PDF as function of distribution's parameters. The joint likelihood of the full data set is the product of these functions. Indeed, this product is generally very small, and in this instance, the likelihood function is normally replaced by a log-likelihood function. Both the latter and the former yields the same result, but the latter is just more tractable and elegant.

More succinctly, suppose $x_1, x_2, x_3, ..., ..., x_n$ are independent and identically distributed (iid) random variables (rvs) of a given PDF (in the continuous case) and probability mass function (PMF), in the discrete case. Then it follows that, the rvs. have a joint function

$$f_\theta(x_1, x_2, x_3, ..., x_n) = f(x_1, x_2, x_3, ..., x_n|\theta). \tag{3.32}$$

Given the observed values $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$, the likelihood function of θ is the function $lik(\theta) = f(x_1, x_2, x_3, ..., x_n|\theta)$ considered as a function of θ.

The MLE of θ is that value of θ that maximises $lik(\theta)$-the value that makes the observed data the very *most Probable*. Given that the X_t, $t = 1, 2, 3, ..., n$ are iid, then equation

(3.32) simplifies to

$$lik(\theta) = \prod_{t=1}^{n} f(x_t|\theta) \tag{3.33}$$

For simplicity and mathematical elegance, logarithmic is applied to equation (3.33) so as to log-maximized the function because of the difficulties involved in evaluating the product of equation (3.33). Applying logarithm to equation (3.33),

$$lik(\theta) = \sum_{t=1}^{n} \log(f(x_t|\theta)) \tag{3.34}$$

Equation (3.34) is a known form of *log-likelihood function* technique for estimating parameter estimates. Furthermore, equation (3.34) is maximized by differentiating the function wrt. (with respect to) each of the variable of interest and setting the result to zero (0).

This underscore the point that, parameter estimates for a known distribution can easily be estimated vis-a-vis the MLE approach. For clarity we illustrate the above concept from the perspective of two known distributions (Poisson and Normal distributions). The reason for these distribution lies in the fact that, while the former is discrete in nature and fits into in our scheme of things for this paper, the latter is continuous; and these provides a basis for evaluation, if any.

The method of maximum likelihood provides estimators that have both reasonable intuitive basis and desirable statistical properties.

That is, for $x_1, x_2, ...xn$, independent and identically distributed (iid) Poisson rvs, with each random variable having the pdf

$$P(X = x) = \frac{\lambda^x e^{\lambda}}{x!} \tag{3.35}$$

Then the joint product of the marginal frequency, the log-likelihood is

$$L(\lambda) = \sum_{i=1}^{n}(x_i \log \lambda - \lambda - \log x_i!) \tag{3.36}$$

$$= \log \lambda \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \log x_i! \tag{3.37}$$

The maximum value of λ is thus found by differentiating and equating to 0 for the FOC (first order condition);

$$L^{'}(\lambda) = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0 \tag{3.38}$$

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.39}$$

In the case of the standard normal distribution, is $x_1, x_2, ...x_n$ are iid. rvs, then the joint density function is

$$f(x_1, x_2, ...x_3|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{(2\pi)}} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2) \tag{3.40}$$

$$l(\mu, \sigma) = -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{3.41}$$

$$\frac{\delta l}{\delta \mu} = \frac{1}{\sigma^2}\sum_{i=1} n(x_i - \mu) \tag{3.42}$$

$$\frac{\delta l}{\delta \sigma} = \frac{-\mu}{\sigma} + \sigma^{-3}\sum_{i=1} n(x_i - \mu)^2 \tag{3.43}$$

From here, once again, these derivatives are set to 0 to help estimate the values of μ and σ

3.3.2 Properties of Maximum Likelihood Estimation

- Simple and easy to use.
- MLE provides standard errors, statistical tests and other results useful for inference.

However, MLE require strong assumptions to be made about the structure of the data.

3.3.3 Akaike Information Criterion

The Akaike information criterion (AIC) is used to assess the relative quality of a statistical model for a given set of data. And thus provide a good basis for the check of the models above. For the maximised values of the likelihood function L of the number model parameters, the AIC value can be determine as

$$AIC = 2k - 2ln(L) \tag{3.44}$$

Where k is the number of predictors of the model.
AIC rewards goodness of fit as determined by the maximum likelihood function. Aikake (1974) opines that, given two models, the preferred one is the model with minimum AIC value.
It is known as its properties for its prediction goodness, and asymptotically efficient model selection criterion for adoption.

3.4 The Kellly Criterion Revisted

A scientist working for Bell labs, John Larry Kelly, Jr. (1956) in a self-published article entitled "A new interpretation of information rate" demonstrated that, in order to achieve maximum growth of wealth, at every bet a gambler should optimize the expected value of the logarithm of his capital. He pivoted his postulation under the assumptions that, the gambler's capital is infinitely divisible and all profits are reinvested. Thus, the *Kelly criterion* is known to maximises the expected value of the capital of the bettor.
Essentially, the criterion ensures optimization of the bettor's fund relative to amount wagered for a series of placed bets. That is, the criterion makes the bettor to take a second pulse as to what amount to bet even in the face of 'obvious' favour of a game to

avoid becoming shock of huge lose to being brought to ruin.

We derive the criterion momentarily and explore other key features at obtaining the optimality of series of bets as the gambler seeks upsurge in income through gambling and betting on sports results.

Consider a sequence of games with probabilities p of winning and $q = 1 - p$ of losing with an allocated bankroll $GHCB$ to be spread. If a decimal odds θ is offered as a reward for the gambler vis-a-vis correct prediction of the outcome of a game; then on the assumption that, the bettor wins a number of games and losses a number of games also, and cognizance of the fraction of bankroll wagered on every game, the Kelly criterion is derived mathematically as:

$$\frac{p\theta - q}{\theta} \tag{3.45}$$

where $\theta \neq 0$

3.4.1 Derivation of the Kelly Criterion

For a fraction $f \in [0, 1]$, with bankroll B of the gambler and an associated odds θ as the received reward for correct prediction, as the stated assumptions of the criterion, if the gambler bets fB and wins, the new bankroll of the gambler from B is

$$B_1 = B + \theta fB = (1 + f\theta)B \tag{3.46}$$

where f, B and θ assume the usual meanings as describe above.

In the second game, the gambler would bet fB_1 which is equivalent to

$$fB_1 = f(1 + f\theta)B \tag{3.47}$$

On the assumption that, the gambler wins again, then it easily follows that, the new

bankroll of the bettor becomes

$$B_2 = (1 + f\theta)B_1 = (1 + f\theta)^2 B$$

Further, if the gambler sadly losses the third bet, the bankroll B becomes

$$B_3 = (1 - f)B_2 = (1 + f\theta)^2(1 - f)B$$

That is, substituting B_2 from above and writing the expression in organized form.
Evidently, after n games, if the gambler has won w games and lost l games, the bankroll becomes

$$B_n = (1 + f\theta)^w(1 - f)^l B \tag{3.48}$$

In this regard, the gain in the bankroll of the gambler relative to the number of wins and loses is

$$Gain_n = (1 + f\theta)^w(1 - f)^l \tag{3.49}$$

From equation (3.48), it is pretty obvious to observe that, the bankroll is growing or shrinking exponentially. But, as earlier explained, the Kelly criterion maximizes the geometric mean of the wealth of the gambler.
Thus, the geometric mean G is the limit as n approaches infinity of the n^{th} root of the gain

$$G = lim_{n\to\infty}((1 + f\theta)^{w/n}(1 - f)^{l/n}) \tag{3.50}$$

Replacing w/n as p (probability of win) and l/n as $q = 1 - p$ (probability of a lost), equation (3.50) become

$$G = (1 + f\theta)^p(1 - f)^q \tag{3.51}$$

For this value of G, the value of the gambler's bankroll after n games is

$$B_n = G^n B$$

This is the value of the bankroll that would occur if exactly pn games are won and exactly qn games are lost after a series of games or number of games, all things being equal (*ceteris paribus*).

Intuitively, the geometric mean arises because if n games are played with a probability p of winning a game and a probability q of losing a game, the expected number of wins is pn and the expected number of loses is qn.

At this point, the fraction f to maximize G the gain in a given bankroll is to differentiate equation (3.51) with respect to (w.r.t.) f.

Taking the logarithm of equation (3.51), which appears as

$$\log G = p \log(1 + f\theta) + q \log(1 - f) \tag{3.52}$$

Differentiating equation (3.52) wrt. to f, we have

$$\frac{dG}{df} = 0 \tag{3.53}$$

Differentiating and equating the result to 0, we have

$$\frac{dG}{df} = \frac{p\theta}{1 + f\theta} - \frac{q}{1 - f} = 0 \tag{3.54}$$

$$\frac{p\theta}{1 + f\theta} = \frac{q}{1 - f} \tag{3.55}$$

$$p\theta(1 - f) = q(1 + f\theta) \tag{3.56}$$

$$p\theta - p\theta f = q + qf\theta \tag{3.57}$$

Simplifying we have

$$p\theta - q = q\theta f + p\theta f \tag{3.58}$$

We know however from elementary algebra and Probability that, $q = 1 - p$ since $q + p = 1$.

The equation just above can be rewritten as

$$p\theta - q = (1-p)\theta f + p\theta f \tag{3.59}$$

Expanding the RHS of the expression (3.59), we have

$$p\theta - q = f\theta - f\theta p + p\theta f \tag{3.60}$$

This reduces to

$$\Rightarrow p\theta - q = f\theta \tag{3.61}$$

Making f the subject, we have the optimized fraction to bet as being

$$f^* = \frac{p\theta - q}{\theta} = \frac{Edge}{Odds} \tag{3.62}$$

Note that, $B_j, j = 1, 2, 3.....n$ indicates the value of the bankroll at time j.

Heuristically, the Kelly criterion is premised on the understanding that, the bettor knows before hand with some degree of certainty that a selection (pick) will be favourable or otherwise to wager a fraction of the bankroll in placing a bet. In other words, probability of winning or losing an event must be known for the Kelly criterion to be of good use for the gambler. Like we have already pointed out, probability of winning or losing a game is a convoluted mix of variables of interest on the game of association football. However, the provided odds from bookmakers relative to an event in question implicitly provides a basis for getting real probabilities from the perspective of the bookmaker.

Odds collected from the bookmakers plus the obtained probabilities from the above model should provide a better basis for calculating the criterion values from which optimality situations are established for the gambler. Specifically, if the probability of win over a sequence of plays is greater than the probability of lose, then it suffices to explain that, the gambler returns can be expected to be positive at least for the next $(n+1)$ game.

3.5 The Risk Concept of Betting and Gambling

The concept and existence of risk in enterprises including gambling and betting regarding the amount of money a bettor is likely to lose, and the possibility and positivity of increasing the return of the bankroll provide a basis for critical analysis.

That is, the game of betting and gambling is widely known for its indeterminate outcome making it risk prone. Risk in layman perspective can be described as hazard, a chance of bad consequence, exposure to misfortune, lose possibility of any kind. Embrechts et al (2005 p.2), define risk as the quantifiable likelihood of loss or less than expected returns. That is, according to Embrechts and his friends, risk is the fear or possibility of losing an underlining asset in an enterprise of any kind. This is unpleasant if it so happens. Most unpleasant however, is when all "available assets or funds" are lost to the point of being brought to bankruptcy. To this end, there comes a lot of risks for the attraction and attention of actuaries and financial analysts. This includes credit risk, market risk, operational risk and liquidity risk. In fact these are non—exhaustive lists as many other forms of risks exist for analysis.

However, in this paper we restrict risk to the gambler. To the gambler, risk is the risk of experiencing gambler's ruin, an actuarial concept which states that, the gambler will eventually lose entire bankroll while playing against an adversary. Equally, the risk of the bookmaker— the mandated body to sanction betting and gambling, can also be determined. To the bookmaker, the risk of experiencing losses emanating from claims of gamblers' placed bets, resulting in less-than-expected returns which results in insolvency. In actuarial circles, losses are captured as a function of loss frequency (i.e. the number of losses) and loss severity (i.e. the size or quantum of loss). Essentially, both the gambler and the house are not immune to the unfortunate effects of being at risk. Like the gambler, the bookie's risk of losses over different time periods can be quantified to assess the point bankruptcy.

Investors are risk averse; that is, given the same expected return, they will choose the investment for which that return is more certain and higher than that for which is improbable and lesser.

3.5.1 The Gambler's Ruin Theory

3.5.2 Definition

We now seeks to determine the risks of success or failure of a gambler who goes to a casino with a specified bankroll initially with the intent of seeking upsurge in income or at worst go home with same revenue as before.

That is, on a typical betday the bettor is faced with two possibilities at close of day:

- The gambler achieve the goal of winning the desire amount of money.
- The gambler ends up with no money.

In the case of the second scenario, the gambler is ruined, thus the name gambler's ruin. That is, the probability of an individual losing sufficient gambling money to the point at which continuity on is no longer considered an option to recoup initial funds and losses suffered. This takes into consideration the probability of winning p, the probability of incurring loses q, and the portions of an individual bankroll that is in play or at risk. Prominently, this is known as the probability of ruin (PoR) or risk of ruin (RoR).

3.5.3 Derivation of Gambler's Ruin Theory

The gambler's ruin is derived by the idea of Markov property (which we review later). Supposing that the gambler become cautious on a bet day and wagers one dollar each time the game is played, with some probability p of winning or $q = 1 - p$ of losing. Let W_n denote the total fortune after the nth wager. The gambler's (as a rational being and risk-averse person) objective is to reach a total fortune of N, without first being brought

to bankruptcy. If the gambler succeeds, then the gambler is said to win the game. In any case, by the filtration of the game the gambler stops playing after winning or getting ruined, whichever happens first.

Nonetheless, while the game rolls on, $\{W_{n:n\geq 0}\}$ forms a simple random walk.

$$W_n = \gamma_1 + \gamma_2 + ... + \gamma_n, W_0 = i \tag{3.63}$$

where $\{\gamma_n\}$ forms an i.i.d sequence of r.vs distributed as $p(\gamma = 1) = p$, $p(\gamma = -1) = q$ and represents the earnings on the successive gambles.

Since the game stops when either $W_n = 0$ or $W_n = N$, let $\tau_i = \min\{n \geq 0 : W_n \epsilon \{0, N\} | W_0 = i\}$ denotes the time at which the game stops when $W_0 = i$. If $W_{\tau i} = N$, then the gambler wins, if $W_{\tau i} = 0$ then the gambler is ruined. By letting $P_i = P(W_{\tau i=N})$ denote the probability that the gambler wins when $W_0 = i$. Clearly, $P_0 = 0$ and $P_N = 1$ by definition. and Computing P_i, when $1 \leq i \leq N - 1$, and conditioning on the outcome of the first wager, when $\gamma_1 = 1$ or $\gamma_1 = -1$ this yields

$$p_i = pP_{i+1} + qPi - 1 \tag{3.64}$$

If $\gamma_1 = 1$, the gambler's total fortune increases to $W_1 = i+1$, and by the Markov property the gambler will now win with probability P_{i+1}. Similarly, if $\gamma_1 = -1$, then the gambler fortune decreases to $W_1 = i - 1$ and by the idea of Markov property the gambler will win now with probability P_{i-1}. The probabilities corresponding to these outcomes are p and q which yields equation (3.64). From the usual $p + q = 1$, the scenario above can be re-written as

$$pP_i + qP_i = pP_{i+1} + qP_{i-1} \tag{3.65}$$

Dividing (3.65) through by p, factorising and simplifying, we have

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}) \tag{3.66}$$

From equation(3.66), and in particular

$$P_2 - P_1 = \frac{q}{p}(P_1 - P_0) = (\frac{q}{p})P_1 \tag{3.67}$$

since $P_0 = 0$ so that

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = (\frac{q}{p})^2 P_1 \tag{3.68}$$

Generalising the results of equations (3.67) and (4.2), such that $0 < i < N$

$$P_{i+1} - P_i = (\frac{q}{p})^i P_1 \tag{3.69}$$

In essence,

$$P_{i+1} - P_1 = \sum_{k=1}^{i}(P_{k+1} - P_K) \tag{3.70}$$

$$P_{i+1} - P_1 = \sum_{k=1}^{i}(\frac{q}{p})^k P_1 \tag{3.71}$$

which results that

$$P_{i+1} = P_1 + P_1 \sum_{k=1}^{i}(\frac{q}{p})^k = P_1 \sum_{k=0}^{i}(\frac{q}{p})^k \tag{3.72}$$

$$= \begin{cases} P_1 \dfrac{1 - (\frac{q}{p})^{i+1}}{1 - \frac{q}{p}} & \text{if } p \neq q; \\ P_1(i+1) & \text{if } p = q = 0.5; \end{cases} \tag{3.73}$$

Choosing $i + 1 = N$ and using the fact that $P_N = 1$, equation (3.73) results in

$$1 = P_N = \begin{cases} P_1 \dfrac{1 - (\frac{q}{p})^N}{1 - \frac{q}{p}} & \text{if } p \neq q; \\ P_1 N & \text{if } p = q = 0.5; \end{cases} \tag{3.74}$$

from which it can be infer that,

$$P_i = \begin{cases} \dfrac{1-(\frac{q}{p})^i}{1-(\frac{q}{p})^N} & \text{if } p \neq q; \\ \dfrac{i}{N} & \text{if } p = q = 0.5; \end{cases} \tag{3.75}$$

It follows that, $1 - P_i = q_i$ is the probability of ruin.

That is, the probability of ruin denoted q_i is thus given by

$$q_i = \begin{cases} \dfrac{(\frac{q}{p})^N-(\frac{q}{p})^i}{(\frac{q}{p})^N-1} & \text{if } p \neq q; \\ 1-\dfrac{i}{N} & \text{if } p = q = 0.5; \end{cases} \tag{3.76}$$

where i is the initial capital and N is the attained wealth after number of plays with $P_i + q_i = 1$. The probability of successful win plus the probability of getting ruined must sum up to 1.

3.5.4 Implication of The Gambler's Ruin for Infinite Wager

From the above derivation of the gambler's ruin problem as in equations (3.75) and (3.76), the actuarial risk of the gambler getting infinitely rich or ruined is easily . ascertained. That is, from equation (3.75), if $p > 0.5$ then $\dfrac{q}{p} < 1$ and therefore as the expected amount to reach N become large

$$\lim_{N \longrightarrow \infty} P_i = 1 - (\frac{q}{p})^i > 0, P > 0.5 \tag{3.77}$$

However, if $P \leq 0.5$, then $\frac{q}{p} \geq 1$ and thus from equation (3.75),

$$\lim_{N \longrightarrow \infty} P_i = 0, for P \leq 0.5 \tag{3.78}$$

From equations ((3.77)) and ((3.78)), it is pretty easy to see that, if the gambler starts

with the wealth $x_0 = i$ and wishes to continue gambling forever until if possibly ruined, with the intention of earning as much money as possible such that there is no winning value N, the gambler only stops if ruined.

In essence, equation ((3.77)) says that if $P > 0.5$ each game in his favour; then there is positive probability that the gambler will never get ruined but instead will become infinitely rich . Equation (3.78) on the other hand says that, if $P \leq 0.5$ (each game not in his favour), then there is that probability the gambler will get ruined.

Therefore, with an initial stake of say GHC5 and probability of winning $p = 0.6$, the probability of the gambling obtaining a fortune of $N = GHC12$ without going broke can easily be calculated from above formulae. That is, With $i = 5$ $N = 12$ $q = 1 - p = 0.4$ and thus $\frac{q}{p} = \frac{2}{3}$, the probability of obtaining N without going broke is given by $P_{i=5} = \dfrac{1 - (\frac{q}{p})^i}{1 - (\frac{q}{p})^N} = \dfrac{0.8683}{0.9923} = 0.8750$ which further means that about 0.12% the gambler is likely to go broke. And about 0.86%, the gambler will infinitely become rich, and 0.13% he/she will get broke infinitely.

3.5.5 Markov Property

The Markov Property states that, given the present state X_n at time n, the future $\{X_{n+1}, X_{n+2}....\}$ is independent of the present $\{X_0, X_1, ...X_{n-1}\}$. Markov Processes is therefore a random process in which future event is independent of the past given the present.

This property is amenable to the gambler's ruin theory and its derivation. It is also used in assessing the point of exit (stopping time) of the gambler, which this paper highlighted briefly at the introductory stage.

CHAPTER 4

ESTIMATION, ANALYSIS AND DISCUSSION OF RESULTS

This chapter presents the estimation of relevant parameter values and discussion of results. The analysis is carried out based on the data described in chapter three. The chapter is in five sections. Section one presents the results of the model formulation for football scores of both home and away play. Specifically, separate home advantage factor or effect is estimated in additions to tests of significance of the variables of influence of both home and away scores, whiles assessing the model assumption through summary statistics. Section two considers and discusses the probabilities values of both home and away scores based on the estimated values of the mean home and away scores intensities of teams. Interested graphs are provided for pictorial view of the distribution of the data. The third section looks at the Kelly criterion, while section four and five are respectively devoted to estimation of gambler's ruin and other summary measures of the study.

4.1 Model Assumption and Results

As the lambdas vary from one match to another, there remain no direct way to test for the validity of the Poisson assumption. Nonetheless, we can assess in average sense whether the assumption holds. Presently are the summary results (statistics) and histograms to demonstrate the distribution of home and away scores in the premier league for 2013/2014 season.

Table 4.1: Summary Statistics

	hscore	**ascore**
Min.	0.00	0.00
Mean	1.575	1.195
Median	1.00	1.00
1st Quater	0.50	0.00
Max	7.00	6.00
3rd Quater	2.00	2.00
Var	1.9011	1.4380

The mean home score as seen in Table (4.1) is 1.575 with the variance being 1.901. This means, the mean of the home score is almost equal to the variance of it. The away score central tendencies also shows that, the mean is 1.195 and the variance of given in the table as 1.438, these values are not quite vast as they're almost equal. In fact, these give a sense of the probability distribution the scores of the matches follow. The maximum scores in a match for the season was a home based score of 7 goals, with an average of 1.5755 goals in a match at home. Clearly, teams in the EPA in the season being considered on average scored more home goals than away.

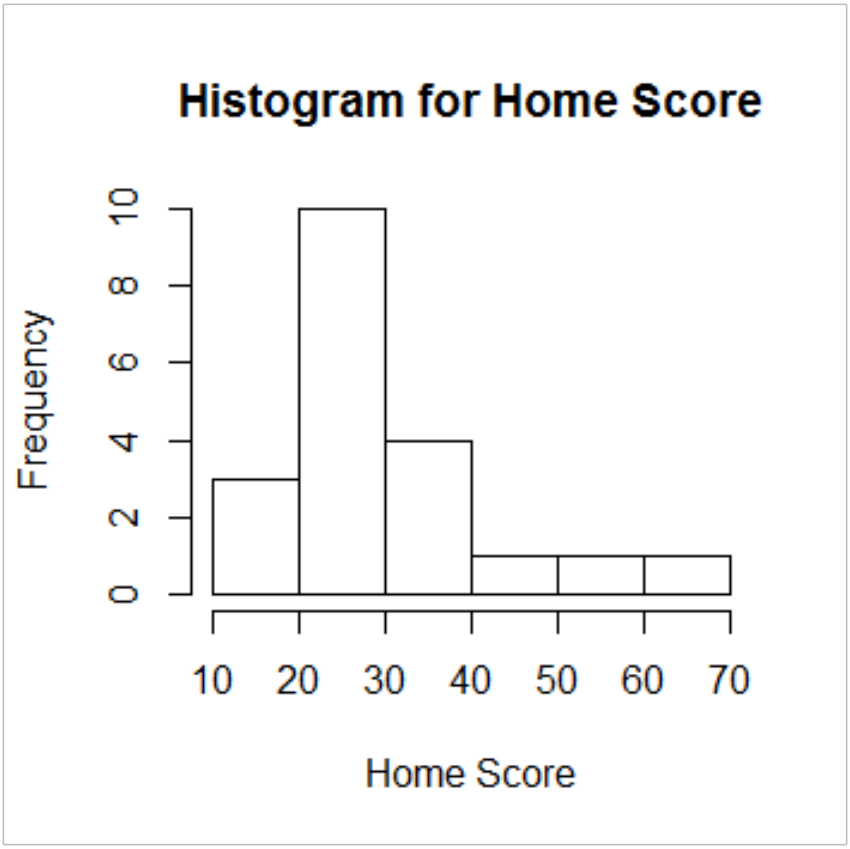


Figure 4.1: Histogram of Home Scores

The figure above give a pictorial representation of the distribution of the home scores for the season. Clearly, more goals tends to be scored at home as depicted by the diagram. We establish this result much clearer momentarily.

In the write up of Dixon and Cole, they concluded from their data set which spanned from 1993-1995 that, Poisson assumption had a nearly perfect fit except for the scores 0-0, 0-1, 1-0 and 1-1. Further, they made an adjustment in their likelihood function, where they included a coefficient to allow for the departure from the independence assumption. It interferes the traditional likelihood function procedure, and thus they are forced to use a so-called "pseudo-likelihood". We are not considering this slight departure from the independence any further in a proper statistical manner due to its complexity in calculations.

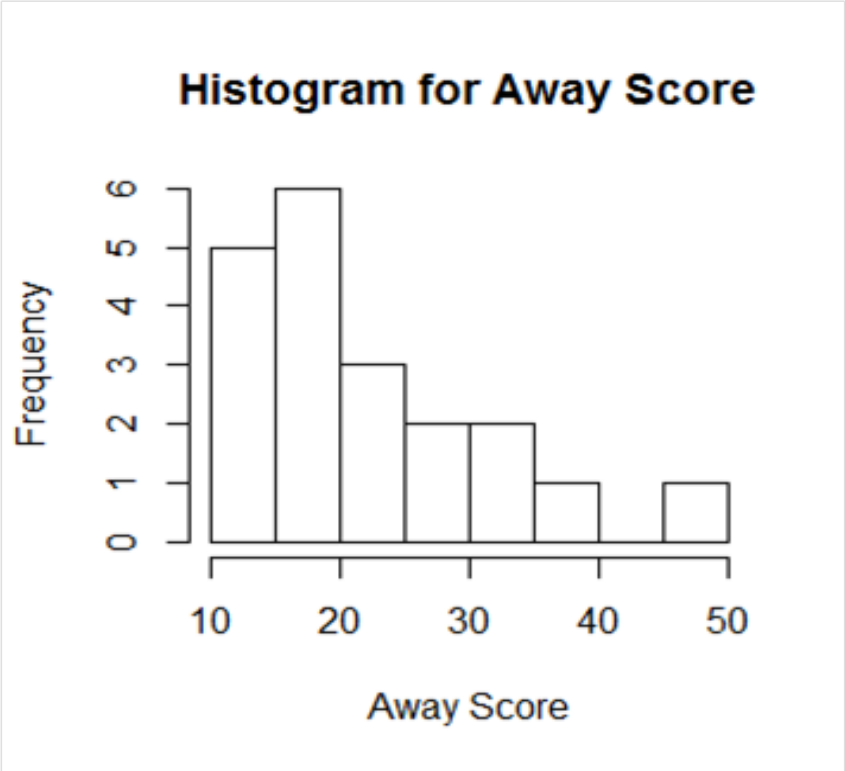


Figure 4.2: Histogram of Away Scores

4.1.1 Basic Model Formulation

Here we establish the validity of the model. Like we have stated previously, the reason for using the Poisson regression stems from the discrete nature of scores results. The R-software output of the Poisson regression model is provided as above. The data for this output is the home and away scores of teams from which teams attack and defensive abilities are taken into account, and other variables of Fouls, red and yellow cards, shots on targets for the entire season of 2013/2014.

The full results or output for the model of other teams is shown in appendices with few shown presently From the table, the estimates of the first 6 teams are given with standard errors of the estimate in brackets. Manchester city was set to 0 for reason of avoiding over parameterization. We see from the table that, Liverpool has higher scoring rate at home followed by Chelsea football club. This is indicated by their home scoring rates of 0.8688 and 0.8032 given under λ_h

Table 4.2: Home estimates of teams' scoring rates

Teams	**Estimate (Std.Error)**	λ_h	**z values**	**pr**$(>\|z\|)$
Liverpool	-0.1406 (0.128)	0.8688	-0.749	0.4541
Chelsea	-.2191 (0.192)	0.8032	-1.142	0.2535
Man United	-0.4733 (0.207)	0.6229	-2.290	0.0220
Everton	-.8150 (0.2312)	0.4426	-3.526	0.00042
Tottenham	-0.7436 (0.226)	0.4754	-3.297	0.00098
Arsenal	-0.6451 (0.218)	0.5246	-2.956	0.003120

Table 4.3: Away estimates of teams' scoring rates

Teams	**Estimate (Std.Error)**	λ_a	**z values**	**pr**$(>\|z\|)$
Liverpool	0.2076 (0.216)	1.2307	0.963	0.3355
Chelsea	-0.3314 (0.248)	0.7179	-1.338	0.1810
Man United	-0.1978 (0.239)	0.8205	-0.829	0.4069
Everton	-0.5725 (0.267)	0.5641	-2.147	0.0318
Tottenham	-0.3677 (0.250)	0.6923	-1.469	0.1419
Arsenal	-0.1082 (0.233)	0.8974	-0.465	0.6421

The non-uniformity in teams estimates suggests that performances are genuinely dynamic. This is seen in the above table by closely examining the values. Manchester city home and away strengths parameters is set to zero as the base parameter. In the closed bracket of column two in both tables are the various standard errors of the estimates. The columns λ_h and λ_a were obtained in both tables by taking the exponent of the estimated values, that is, with Liverpool away estimates of 0.2076 as obtained from the Poisson regression the rate is $e^{0.2076} = 1.2307$ and the same procedures were adopted to get the remaining values. These rates indicate the scoring potencies of teams in both home and away plays. From both tables, it is evidenced that, Liverpool has the highest scoring rates of both home and away strengths of 0.8688 and 1.2307 respectively, with Norwich having low rates of 0.2787 and 0.2820 for both and away respectively as shown on the appendix. Unlike home play, Manchester United has higher away scoring rates next to Liverpool,

which means Manchester United had higher away influence than home. We verify this observation in table 4.7 to ascertain its consistency herein.

These rates of teams performance are used in calculating the marginal and joint probabilities between any two teams using the idea of bivariate Poisson model which we demonstrate shortly;

Table 4.4: Bivariate Poisson estimates of Probabilities of Home and away teams

	0	1	2	3	4	5	6
0	0.1972	0.1616	0.0663	0.0181	0.0037	0.0006	0.0001
1	0.1584	0.1286	0.0576	0.0144	0.0023	0.0005	0.0001
2	0.0636	0.0522	0.0214	0.0059	0.0012	0.0002	0.000
3	0.0170	0.0140	0.0057	0.0016	0.0003	0.0001	0.000
4	0.0034	0.0140	0.0057	0.0003	0.0001	0.000	0.000
5	0.0005	0.0004	0.0002	0.000	0.000	....	...
6	0.0001	0.0001	0.000	....	....	....	...

where the dots .... means 0.0000. The vertical and the horizontal rows of (0,1,2,3,4,5,6) indicate the home and away scores of teams respectively.

Table (4.4) given above provides the marginal and joint probabilities of teams playing home and away with Chelsea home parameter of $\lambda_h = 0.8032$ and Manchester United away parameter of $\lambda_a = 0.8205$. The outputs as displayed in the table grants the gambler the opportunity to calculate the probability of away or home win as well as draw.

By adding the cells probabilities of $(1,0), (2,0), (2,1), (3,0), (3,1)(3,2), (4,3), (5,4), ...$ of such kinds where the possibility of scoring higher at home gives the probability of winning a home play on a team by the gambler. In this scenario, the probability of winning at home as easily seen from the table is given by the summation of the cells with values : $\mathbf{0.1584} + \mathbf{0.0636} + \mathbf{0.0170} + \mathbf{0.0034} + \mathbf{0.0005} + \mathbf{0.0001} + \mathbf{0.0522} + \mathbf{0.0140} + \mathbf{0.0004} + \mathbf{0.0001} + \mathbf{0.0057} + \mathbf{0.0057} + \mathbf{0.0002} + \mathbf{0.0003} = \mathbf{0.3216}$ which translates into 32.16% of winning a home play or placed bet for the gambler all things being equal. Similar calculations could be derived to arriving at the probability of winning an away

play for the bettors consideration. In this respect, we add up the probabilities from the table for which we have $(0,1),(0,2),(1,2),(2,3),(3,4),(4,5),....$ and others of such kinds. Summing the corresponding cells of such scores from the table give **0.1616** + **0.0663** + **0.0181** +**0.0037** + **0.0006** + **0.0001** + **0.0570** + **0.0144** + **0.0023** + **0.0005** + **0.0001** + **0.0059** + **0.0012** + **0.0002** + **0.0003** = **0.3330** which translates into some 33.30% of winning an away game by teams.

In fact, ties probability of two teams playing a game could equally be calculated in same procedure by considering the cells of $(0,0),(1,1),(2,2)$ and others of such kinds. That is, by adding the leading diagonals probabilities of the above table gives the probability of a drawn match.

The probabilities as given in the Table (4.4) was derived by the bivariate Poisson approach. For a score of say (1,0) with a home rate of $\lambda_h = 0.8032$ and an away rate of $\lambda_a = 0.8205$, we have $\frac{0.8032^1 e^{-0.8032}}{1!}\frac{0.8205^0 e^{-0.8205}}{0!} = 0.1584$, all the remaining cells probabilities were obtained in similar fashion.

4.1.2 Model Inference

It is pretty easy to infer that, with a higher scoring rate, there is equally higher probability of win than when the scoring rate is on the low. That is, with a higher away scoring rate of $\lambda_a = 0.8205$ and home scoring rate of $\lambda_h = 0.8032$, the probability of away win is greater than home win in this instance. However, it is quite greater and/or absolute level of probability of win or lose, but nonetheless give an idea of expectant direction for the gambler.

4.1.3 Variables of Influence Significance

This section considers the relevance of variables that influence both home and away scores. Below we provide the results for determining the significance of variables of influence.

Table 4.5: R output of variables of influence on Home scores

Variables	**Estimate (Std. error)**	**Z values**	**pr(>|z|)**
Intercept	1.4892 (1.0878)	1.369	0.1710
FoulF	0.0002 (0.0013)	0.153	0.8780
FoulA	-0.0006 (0.0015)	-0.392	0.6952
RedCA	-0.0064 (0.0252)	-0.257	0.7972
CornersP	0.0066 (0.0016)	4.032	5.5e-05
YellowCA	0.0065 (0.0054)	1.196	0.2316
Shotsperct	0.0230 (0.0174)	1.707	0.0878

Signif. codes: 0 '***' 0.001 '**' 0.01 '**' 0.05 '**' 0.1 '**' 1

(Dispersion parameter for Poisson family taken to be 1)

Null deviance: 82.628 on 19 degrees of freedom

Residual deviance: 27.473 on 13 degrees of freedom

AIC: 145.16

Number of Fisher Scoring iterations: 4

Table 4.6: R output of variables of influence on Away scores

Variables	**Estimate (Std. error)**	**Z values**	**pr(>|z|)**
Intercept	-0.8718 (1.2923)	-0.67	0.4992
FoulF	0.0020 (0.0016)	1.315	0.1884
FoulA	0.0006 (0.0018)	0.809	0.7573
RedCA	0.0207 (0.0289)	-0.717	0.4732
CornersP	0.0048 (0.0020)	2.451	0.0143
YellowCA	0.0093 (0.0064)	1.459	0.1446
Shotsperct	0.0566 (0.0197)	2.867	0.0041

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 70.756 on 19 degrees of freedom

Residual deviance: 18.756 on 13 degrees of freedom

AIC: 130.57

At 95% significance level ($\alpha = 0.5$), we see from the Tables (4.5) and (4.6) and conclude that, corners profile of teams influences teams home and away scores. This explains why most teams take advantage of corners earned to score more goals in games. Corners thus become an important aspect of football games as for scores analysts. We also notice that, shot on targets of teams influence away scores at ($\alpha = 0.05$). On the other hand, Red cards and yellow cards of teams do not influence scores of both home and away. On average, it is known in football circles that, red cards that send players off the pitch do not 'necessarily' contributes to the lose of the affected team or win of the opposition team. Nonetheless, we've in the past seen teams lose games for reason of red card of players which weakens the ability of the team in the department for which the player was, and yellow card of players affecting their output in tackling for fear of second yellow card offence which will results in sending off the player.

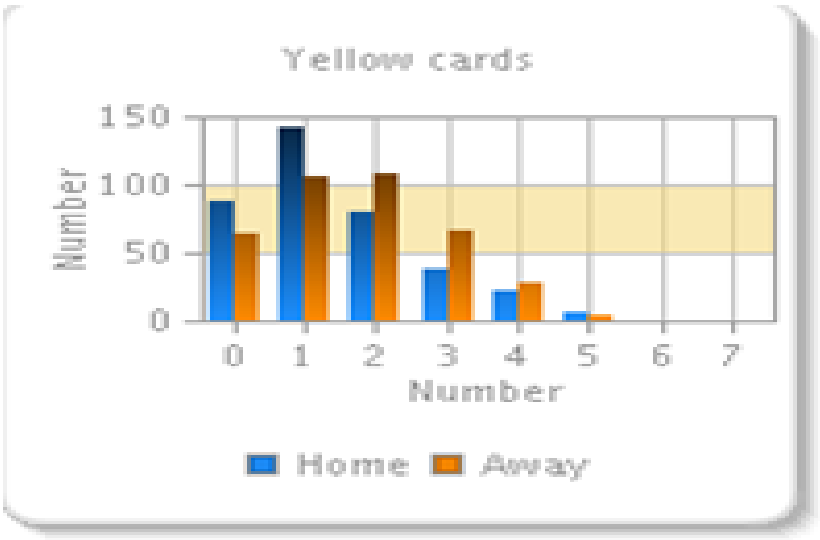


Figure 4.3: Bar chart depicting season's home and away yellow cards

From the figure, we see at glance that, the distribution of the yellow cards are negatively skewed. Further, we see that, there were close to 150 yellow cards for home teams and 100 for away teams of a card. There was however, few cards in the number of 5 for both season's home and away teams.

Below we also show the distribution of the season's red cards. Unlike yellow cards for

teams, red cards that send off players are few. This indicates that, teams are careful in committing red cards compared to yellow, and perhaps is attributed to being carded twice to necessitates it.

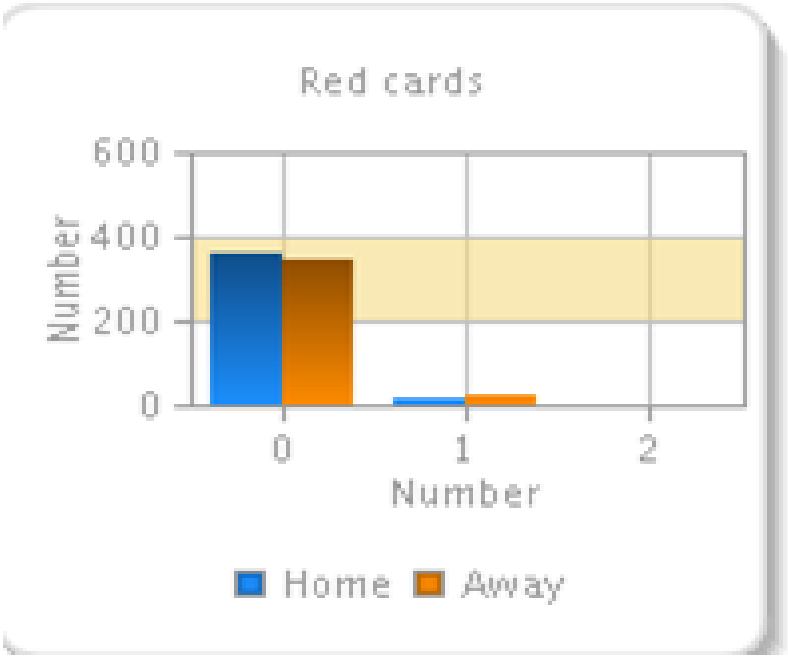


Figure 4.4: Season's depiction of Red cards for home and away teams

4.1.4 Separate Home Advantage Factor

From previous discussion, we have sought to indicate the advantage of playing at home than away. Below are the various estimates for teams' home advantage effect relative to the 2013/2014 season.

From the table, it is pretty obvious to see that, home advantage plays crucial role in the wins of teams. Manchester City (league champions for the season) seems to have taken advantage at playing at home to perhaps earn majority of its total points to emerge victors of the season as evidenced by their home effect of 0.5930. Interesting though, Stoke also made a strong case for themselves in playing at home as evidenced by the home effect of 0.6000 which is higher than Manchester City. This indicates that, Stoke had poor away win records relative to the fact that they placed 9th in spite of their home records. i.e

Stoke earned exactly 60 percent of their total earned points at home.

Table 4.7: Estimates of Teams' Home Advantage

	Teams	**Home Advantage**
1	Manchester City	0.5930
2	Liverpool	0.5714
3	Chelsea	0.5488
4	Arsenal	0.4937
5	Everton	0.4583
6	Tottenham	0.4348
7	Manchester United	0.4219
8	Southampton	0.4286
9	Stoke	0.6000
10	Newcastle	0.4900
11	Crystal Palace	0.4000
12	Swansea	0.3571
13	West Ham	0.3750
14	Sunderland	0.3947
15	Aston Villa	0.4737
16	Hull	0.5676
17	West Brom	0.2500
18	Norwich	0.5455
19	Fulham	0.2813
20	Cardiff	0.4000

West Brom and Fulham unlike Manchester City and Stoke earned only 25 percent and 28 percent of their accumulated points from home as seen from the table by a cursory look. Again, there is relatively high variability in teams performance relative to the home effect estimates of West Brom and Fulham to total points earned, otherwise there seem to be relatively constant home effect.

A close look at the home advantage values further indicates that, most teams earned much of their accumulated at home. In fact a team with 40% of home points is an implicit indication of good strength at home because drawn games (of both home and

away) points plus away points contributes to 60% of the amassed points. Nonetheless, West Brom and Fulham home percentage points remain much to be desired in spite of the general cliche that home play contributes to teams wins.

Further we see from the table that, Manchester United earned only 42.19% of the total season points at home. This confirm earlier observations made in tables 4.2 and 4.3.

These estimates were obtained by dividing the home earned points over total season earned points by each teams. This gives an idea of the percentage of home earned points as against total amassed points that includes ties and away wins.

4.1.5 Model Validation

We validate the efficiency and robust nature of the model above using the 2014/2015 season data. Here, we have 3 new teams in the names of QPR, Burnley and Leceister and the non-availability of Norwich, Fulham and Cardiff for reason of the later 3 being relegated in the 2013/2014 season.

First we calculate the home effects of teams and compared that with the 2013/2014 season to see the consistency in teams' performances.

Table 4.8: Validated Estimates of Teams' Home Advantage

	Teams	**Home Advantage**
1	Chelsea	0.5172
2	Manchester City	0.5316
3	Arsenal	0.4800
4	Manchester United	0.6000
5	Tottenham	0.5156
6	Liverpool	0.4839
7	Southampton	0.5500
8	Swansea	0.4821
9	Stoke	0.5000
10	Crystal Palace	0.3750
11	Everton	0.4468
12	West Ham	0.3830
13	West Brom	0.5456
14	Leceister	0.3947
15	Newcastle	0.5385
16	Sunderland	0.2368
17	Aston Villa	0.4737
18	Hull	0.4286
19	Burnley	0.3636
20	QPR	0.6000

The table 4.8 as shown above indicates within its columns the home effects of teams in the 2014/2015 season. We see from the table that, compared to the 2013/2014 season Manchester city, Chelsea, Arsenal, Hull showed consistency in playing at home by amassing more points accordingly. Then again, Manchester United which had only 42% of home earned points leaped in performance at home and perhaps jumped in the end-of-season position of 7th in 2013/2014 to 4th in 2014/2015 season. Clearly, home effects is thus important in the game of football.

Table 4.9: Away estimates of teams scoring rates (Validated)

Teams	**Estimate (Std.Error)**	λ_a	**z values**	**pr**$(>\|z\|)$
Chelsea	0.3151 (0.253)	1.3703	1.245	0.2132
Man City	0.3151 (0.253)	1.3703	1.2545	0.2132
Man United	-0.2513 (0.291)	0.7778	-0.864	0.3877
Tottenham	-0.0377 (0.275)	0.9630	-0.864	0.3877
Liverpool	-0.2513 (0.291)	0.7778	-0.864	0.3877
Southampton	-0.5233 (0.315)	0.5926	-0.864	0.3877
Swansea	-0.4055(0.304)	0.6667	-1.332	0.1827
Stoke	-0.6568 (0.329)	0.5185	-1.994	0.0461
Crystal Palace	-0.0377 (0.275)	0.9630	-0.137	0.8908
Everton	-0.2513 (0.253)	0.7778	-0.864	0.3877
West Ham	-0.3514 (0.299)	0.7037	-1.173	0.2406
Leceister	-0.2513 (0.2910)	0.7778	-0.864	0.3877
Newcastle United	-0.6568 (0.329)	0.5185	-1.994	0.0461
Aston Villa	-0.7309 (0.338)	0.4815	-2.165	0.0304
Hull	-0.6568 (0.329)	0.5185	-1.994	0.0461
Burnley	-0.6568 (0.329)	0.5185	-1.994	0.0461
QPR	-0.4055 (0.304)	0.6666	-1.332	0.1827

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4.4876e+01 on 19 degrees of freedom

Residual deviance: 4.8850e-15 on 0 degrees of freedom

AIC: 136.05

From table 4.9, we see that, Chelsea and Man City has higher scoring away rates of λ_h 1.3703, implicitly indicating higher probability of win for an away play. Compared to the original estimates in table 4.3, Liverpool had higher away scoring rate for that matter. That notwithstanding, Liverpool still made a point of having higher away scoring rates of $\lambda = 0.7778$, which is similar to Man United scoring rates in spite of the fact that, Liverpool rounded up the season in 6th position as against United's 4th position.

In essence, with Chelsea home scoring rate of $\lambda_h = 0.7250$ and Man United away scoring rate of $\lambda_a = 0.7778$ as indicated in tables 4.9 and 4.10, there is an indication of higher away probability of win than the probability of home win. This is consistent with earlier conclusion reached regarding the calculated probabilities of home win and away win.

Table 4.10: Home estimates of teams scoring rates (Validate Model)

Teams	**Estimate (Std.Error)**	λ_a	**z values**	**pr**$(>\|z\|)$
Chelsea	-0.3216 (0.253)	0.7250	-1.319	0.1873
Man City	0.0953 (0.253)	1.0953	0.436	0.6626
Man United	-0.0513 (0.291)	0.9410	-0.226	0.8209
Tottenham	-0.3930 (0.275)	0.6750	-1.578	0.1146
Liverpool	-0.2877 (0.291)	0.7410	-1.191	0.2336
Southampton	-0.1335 (0.315)	0.8750	-0.577	0.5640
Swansea	-0.4308(0.304)	0.6410	-1.710	0.0873
Stoke	-0.5108 (0.329)	0.6000	-1.978	0.0479
Crystal Palace	-0.7985 (0.275)	0.4500	-2.813	0.0049
Everton	-0.3930 (0.253)	0.6750	-1.578	0.1146
West Ham	-0.5978 (0.299)	0.5500	-2.252	0.0243
Leceister	-0.5978 (0.2910)	0.5500	-2.252	0.0043
Newcastle United	-0.5108 (0.329)	0.6000	-1.978	0.0479
Aston Villa	-0.6445 (0.338)	0.5249	-2.391	0.0168
Hull	-0.8557 (0.329)	0.4250	-2.955	0.0031
Burnley	-1.2040 (0.329)	0.2910	-3.658	0.0003
QPR	-0.7985 (0.304)	0.4500	-2.813	0.0049

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4.7086e+01 on 19 degrees of freedom

Residual deviance: -8.4377e-15 on 0 degrees of freedom

AIC: 141.18

From above tables, determination of the probability of home win, draw and away wins are straight forward for focus teams of Chelsea and Man United as in the main work using

the scoring rates of $\lambda_h = 0.7250$ and $\lambda_a = 0.7778$. Consider the table below

Table 4.11: Bivariate Poisson Estimates of Probabilities (Validated)

	0	1	2	3	4	5	6
0	0.2225	0.1731	0.0673	0.0175	0.0034	0.0005	0.0001
1	0.1613	0.1254	0.0488	0.0127	0.0025	0.0004	0.0000
2	0.0585	0.0455	0.0177	0.0046	0.0009	0.0001	0.000
3	0.0140	0.0101	0.0043	0.0011	0.0002	0.0000	0.000
4	0.0026	0.0011	0.0008	0.0002	0.000	0.000	0.0000
5	0.0004	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000
6	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000

The vertical and the horizontal rows of (0,1,2,3,4,5,6) indicate the home and away scores of teams respectively. Using the idea advanced above, the probability of home win $p(\lambda_h) = 0.2996$ and probability of Away win $p(\lambda_a) = 0.3321$.

Regarding the variables of scores, we see from the table 4.12 that, Corner profile is significant. This indicates that, corner profile of teams influences home scores. This is consistent with 2013/2014 season for teams.

Table 4.12: R output of variables influence for Home scores (2014/2015)

Variables	**Estimate (Std. error)**	**Z values**	**pr(>|z|)**
Intercept	0.9466 (1.328)	0.713	0.4760
FoulF	0.0005 (0.002)	0.344	0.7311
FoulA	0.0002 (0.0017)	0.139	0.7311
RedCA	-0.0032 (0.0276)	-0.117	0.9066
CornersP	0.0060 (0.0029)	1.572	0.1160
YellowCA	-0.0020 (0.0052)	-0.384	0.7010
Shotsontarget	0.0460 (0.0293)	1.572	0.1160

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Poisson family taken to be 1)

Null deviance: 47.086 on 25 degrees of freedom

Residual deviance: 11.782 on 13 degrees of freedom

AIC: 126.96

Number of Fisher Scoring iterations: 4

4.2 Risk And Return Analysis

4.2.1 Kelly Criterion

There is a general cliche in gambling that, "every gambler ought to know the secret to survival by knowing what to throw away and what to keep, when to walk away and when to run," which is implicitly inherent in Kelly criterion and its twin the gambler's ruin.

Earlier in this work, we said that the Kelly criterion relies on known level of winning (probabilities) of games to be wagered in, which also means lose probabilities are assumed to be known beforehand. We recall also that a risk-averse gambler takes a second pulse and refuses to play when $p \leq q$, i.e when the probability of win is less than or equal to the possibility of losing a stake, then it makes no sense to place a wager.

Kelly criterion is therefore known for its optimality feature relative to the capital growth of the gambler. To combine the goals of capital growth and security, an alternative is a *fractional Kelly* criterion, i.e. compute the optimal Kelly investment but invest only a fixed fraction of that amount. Thus security can be gained at the price of growth by reducing the investment fraction.

The table below (generated from excel) provides a helpful role for Kelly criterion values for optimal funds with the underlying intent that the gambler will not be ruined.

Table 4.13: Kelly fractions for given odds

Odds	Probability of win	Probability of lost	Kelly %
2.4	0.42	0.58	N/A
1.2	0.82	0.18	0.67
1.73	0.58	0.42	0.34
7.50	0.13	0.87	N/A
2.25	0.44	0.56	N/A
1.08	0.92	0.08	0.846
1.27	0.56	0.34	0.292
1.44	0.79	0.21	0.644
1.60	0.69	0.31	0.496
1.55	0.63	0.37	0.391
1.62	0.65	0.35	0.434
1.36	0.62	0.38	0.341
1.22	0.73	0.27	0.509
1.67	0.82	0.18	0.712
1.44	0.60	0.40	0.322
2.25	0.69	0.31	0.552
4.50	0.44	0.56	0.316
2.4	0.22	0.78	N/A
8.50	0.48	0.52	N/A
1.15	0.42	0.58	0.18

The column 'Kelly %' gives the estimates of the Kelly fractions for a given amount of money for wagering in a bet. When probability of win is less than that of lose (that is, in a sub-fair situation), the gambler do not wager as it makes no sense in wagering, in which case no percentage of the gambler's bankroll is calculated. N/A means non-available percentage of the fraction to wager because $p \leq q$. For an amount of say GHC 100, with a probability of win of 0.82 as implied in the given odds, the gambler should bet 67 percent of the bankroll which translates into GHC 67 cedis. Even in this situation where the odds are in the favour of the gambler, there is the possibility of the gambler losing the stake for reason well explained regarding the variables of influence. And fractional Kelly is highly

recommended in this regard.

A cursory look at the table further indicate that, with a positive probability of winning, the individual bettor should bet higher percentage of the bankroll as against a situation where the win probabilities are low. For an example, with an odds of 1.27 and implied probability of 0.56, the Kelly criterion says the gambler should wager 29.2 percent of the bankroll. Contrasting this with a situation where the probability of win is higher 0.92, the bettor by the Kelly criterion is to wager 84.6 percent. The Kelly percentages were obtained from (3.62) derived in chapter 3 of this write up.

Taking the odds 1.2, probability of win and lose of 0.82 and .018 respectively, the Kelly fraction or percentage using (3.62), we see that $\frac{p\theta - q}{\theta} = \frac{0.82 * 1.2 - 0.18}{1.2} = 0.67$ and similar calculations gave the results for the other columns.

4.2.2 Gambler's Ruin

As the gambler wager the fractions of bankroll for each game based on the Kelly criterion and sometimes half-Kelly, the gambler is faced with the possibilities of getting ruined or getting rich in the long run. We calculate the probabilities of the gambler becoming ruined or otherwise for number of bets.

The natural question to ask is; with a start of say GHC 50, and given levels of win and lose probabilities, what is the possibility that the gambler obtain a certain level of fortune without necessarily going broke or ruined?. Equations (3.73) and (3.75) provide answers to this kind of question. The table below give the ruin probabilities of the gambler relative to the possible risk the gambler faces.

Table 4.14: Calculations and Analysis of Ruin Probabilities

i GHC	N GHC	p	q	p_i	q_i
20	**30**	0.50	0.50	**0.6667**	**0.3333**
20	**35**	0.50	0.5	**0.5714**	**0.4286**
30	35	0.50	0.50	0.8571	0.1429
35	45	0.50	0.50	0.7778	0.2222
40	55	0.50	0.50	0.7272	0.2728
20	30	0.40	0.60	0.0173	0.9827
20	30	0.45	0.55	0.1323	0.8677
30	35	0.45	0.55	0.3661	0.6339
35	**45**	**0.55**	**0.45**	**0.9992**	**0.0008**
5	10	0.45	0.55	0.2683	0.7317
9	10	0.45	0.55	0.7899	0.2101
99	100	0.40	0.60	0.6667	0.3333

Loosely speaking, the bookmaker in any bet day has enough funds to pay winners while retaining the associated vigorish and unlikely to be ruined as compared to the individual bettors. We thus restrict attention to the possibility of the gambler getting ruined and defers that of the bookmaker for any thesis paper in future. From the table 4.14 and previous discussions, the columns labelled i, N, p and q represent the initial capital of the gambler for placed bet, the total fortune to reach, the probability of win and the probability of lose for a particular play respectively. Also, the columns p_i and q_i represent the probabilities of victory (additional earned income) and that of ruin respectively.

A close examination of table 4.14 indicates from the first two rows that, with equal probabilities of play win and lose, the probability of ruin in the case of seeking total fortune of GHC 30 from GHC 20 (ie 30 - 20 = 10 units of income) is 33.33% compared with when the gambler seeks a total fortune of GHC35 from GHC 20 (ie 35- 20 = 15 units of income) with same parameters, in which case, the ruin possibility of the gambler increased by 9.53 (0.4286-0.3333) percentage points. That is, when the gambler expected

fortune to reach from initial capital is higher, the probability of ruin became higher. Also, a close look made from the first-half of the table for further observation confirms this remark. In fact, this is when the probability of win is equal to losing a particular game play. This makes much sense intuitively

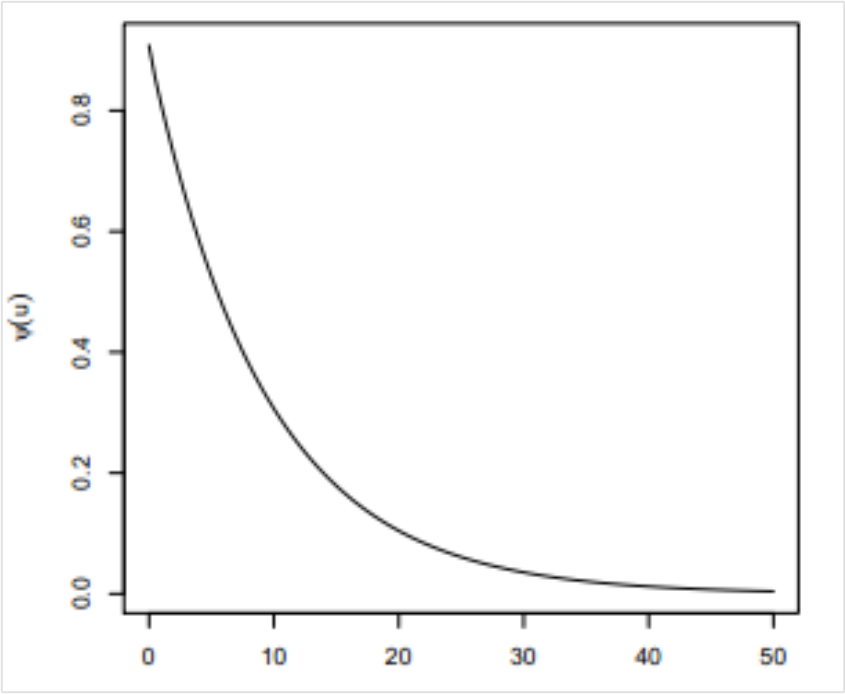
The instance where the probability of win is *not equal* to a lose for a placed bet (ie $p \neq q$), the gambler's risk of ruin (RoR) fluctuates amidst specified levels of probabilities, differential in initial amount and expected fortune. We observe from the first row of the second-half of table 4.14 that, when the $p = 0.40$ and $q = 0.60$ ($p < q$), the ruin probability of the gambler becomes very high (98.27%) with a negligible probability of 1.73% of attaining the goal of $N = 30$ from $i = 20$. In the second row, when the probability of win fell from 60% to 55%, the gambler RoR reduced by 11.5% percentage points from 98.27% to 86.77% cognizance of the same initial amount and expected fortune in both scenarios. Like, our earlier observation, when the expected fortune to reach is GHC 5 (from 30 to 35), with same dynamics of the market as second scenario above, the ROR fell further. This time twice what the probability of win for a placed bet is to 63.39%.

Interesting, with a more positive possibility of winning a placed bet, the probability to reach the expected fortune is very much higher (almost to parity) than the ROR of the gambler. With same dynamics of initial amount i GHC 30, expected fortune N GHC 35 and given probabilities are highlighted in the table, the ROR become very much negligible in which case the gambler expectation become realistic.

This is in consonance with the inherent risk caution provided in the much publicised Kelly criterion discussed extensively in this paper.

Again, in seeking additional GHC 1 to the wealth of the gambler which is minimal compared to the above described scenarios, when the **even** the probability of win is less than that of lose for a placed bet, the RoR of the gambler is low. However, with same conditions again, but higher expected fortune to reach, the gambler RoR increased again affirming earlier conclusion and submission that, the higher expected fortune to reach correlates with the RoR of the gambler.

Figure 4.5: Ruin Probability Graphic



In essence, the table provides some interesting results for consideration of the bettor regarding the eminent risk he/she faces especially with respect to 'expected fortune to reach'. The reason this is quite important stem from the goal of every gambler as wanting to increase the revenue base from bet as large as possible vis-a-vis dynamics of the time but without arbitrage-free risks.

The graphic of figure (4.5), above indicates the that, with a surplus or starting initial capital i, the gambler is likely to be ruined as depicted by the curve.

4.2.3 Optimally Bet Strategy

Against the background of the above arguments, we further investigate whether there is a better gambling strategy for optimal funds for the gambler vis-a-vis betting GHC 1 repeatedly or in units of GHC 10.

For a typical bet day, if the gambler cognizance of the dynamics of the game starts with GHC 20 and with the goal of reaching GHC 200, decides to choosing between

- Play in GHC 1 bets
- Play in GHC 10 bets

We find the probabilities to win in these scenarios.

The probability to win GHC 120 with initial funds of GHC 10 with $p = 0.49$ and $q = 0.51$ is

$$P_{10} = \frac{1 - (\frac{0.51}{0.49})^{10}}{1 - (\frac{0.51}{0.49})^{120}} = 0.0041$$

However, if the gambler bets GHC 20 at each game with same parameters of probabilities of win and lose as above, in which case $N = 6$ and $i = 1$ since we need to make a net total of 5 wins, then the probability to win GHC 120 in GHC 20 bets starting with GHC 20 funds is

$$P_1 = \frac{1 - (\frac{0.51}{0.49})}{1 - (\frac{0.51}{0.49})^6} = 0.1505$$

It is observed clearly that, the chance to win is about 15 in hundred, about four time better than playing in GHC 1 increments. Based on such observation, we indicate against the earlier arguments that, the best strategy for optimal funds is to be bold if the odds of the game are not in your favour and vice-versa. And cautious if the game favours the gambler.

This is interesting at first glance, and 'may' seems inconsistent with earlier conclusions reached regarding the Kelly criterion. A close and deep analysis however points to the fact that, the risk of the gambler is high for number of play in GHC 1 to reach the expected fortune than a bold play to reach the expected fortune. This explains why the Kelly fraction adoption is critical for the security of funds, because, it takes into consideration the dynamics of the game and indicates to the gambler to bet a fraction of the bankroll at any time to increase revenue. We further bear in mind that, the above scenario works better when series of bets are to be placed sequentially.

The understanding of the above will suggest that, if the game is sub-fair ($p < q$), one should bet one's entire fortune. Now if the gambler is to reach a certain target, there is

no reason to bet more than necessary to reach that target. For example, with an initial funds of GHC 65 to reach GHC 110, the bettor can bet GHC 60, but if you want to reach GHC 200, the gambler can bet only GHC 40 keeping some money left to try again if necessary.

CHAPTER 5

SUMMARY, CONCLUSION AND RECOMMENDATIONS

This chapter seek to sum up the work so far and conclusions drawn. Summary of the work and key conclusions for policy formulators, bookmakers and bettors as well as for academic reading are provided below. Policy recommendations from the work are also shared herein for the consideration of bettors and gamblers.

5.1 Summary of Main Results

This research generally sets out to analyse the ruin probabilities of the gambler in placing optimal bets. Specifically, the paper aimed to

a. Formulate prediction model on football scores that estimate probabilities of home win, draw and away win.

b. Determine the influence of some variables as significantly influencing both home and away scores, if any.

c. Examine the Kelly criterion from bookmaker's offered odds.

d. Determine the risk of ruin (RoR) of the gambler, and the optimal betting strategy for adoption.

In respect of objective a., the results in tables (4.3), (4.2) and (4.4) present key outputs for consideration. Table (4.3) for example indicates the home scoring rates of teams, while that of Table (4.2) gives the scoring rates of teams in away play. Thus far, these scoring

rates provides a grave grounds for estimating the probabilities of teams win or lost in both home and away. In fact, Table 4.4 estimates the bivariate probabilities results of teams that predicts home win, draw or away win for teams. In particular reference to Chelsea (home team) and Man United (away team), with respective scoring rates (intensities) of $\lambda_h = 0.8032$ and $\lambda_a = 0.8205$, the probability of home win was found to be 32.16% with probability of away win being 33.30%. That is, the results in Tables (4.3) and (4.2) provide the basis for predicting the win, lose or draw of any two teams vis-a-vis their scoring rates.

For objective (b.), the results are displayed in Tables (4.5) and (4.6). At 95% level of significance, the results indicates that, corner profile of teams influence both home and away scores for teams. Further in Table (4.6), shots on targets was also found to influence teams away scores. These are easily infer from the Tables (4.5) and (4.6) above. This makes intuitive sense, because, more teams have from corners make up from deficits scores to win crucial matches vis-a-vis corners earned. At very best, teams prefer throws to conceding corners, because of the probable effects of corners on scores of teams.

To ensure the security and growth of the wealth of the gambler, the much publicised criterion by famed analyst, Kelly was analysed to determine the fraction of income to bet in any event from offered odds of bookmakers. Analysis of the fraction implicitly inherent in the odds of bookmakers is shown in Table (4.13). At first glance, this will look simplistic, but, more closer look indicates that, over number of plays, the gambler allocation of wealth on successive events become optimally assured. This is the third objective of the paper and the referenced table above presents the related results.

Regarding the optimal strategy of bets for the gambler in avoiding the dire situation of been brought to bankruptcy, the paper sets out to assess the ruin (risk) potential (exposure) of the gambler in the face of unfavourable events as in objective four of this paper. Analysis of different scenarios for the gambler is presented in Table 4.14. The results indicated that, with an equality of win and lose probabilities, the gambler is exposed to ruin when the 'expected fortune to reach' from initial capital is higher,

compared to the situation when the expected fortune to reach is on the lower side. Amidst specified level of win or lose probabilities, the RoR of the gambler fluctuates with key determinants being initial amounts and 'expected fortune to reach.' Further, the paper analysed the strategy to optimal bets in this dilemma by finding that, for a sub-fair ($p < q$) situation, the gambler should bet bold but be cautious in a super-fair ($p > q$). For an infinite number of plays, the paper provides a hint in calculating the probability in this regard.

5.1.1 Conclusions

Cognizance of the results, analysis and discussions of output generated in section four of this paper. The following are the conclusions reached.

- With higher away team scoring intensity, as against home teams, there remain higher probability of away winning.

- Corners profile of teams was found to significantly influence both home and away scores at 95% level of significance. Teams consider corners so important in matches as most teams have through corners come back from deficit in games to win matches hitherto red cards and others failed to work magic for. Reminiscent of this observation is the recent past (2014/2015 Uefa Champion League) exit of Chelsea from the European football association champions league which saw 10 men PSG team come twice from behind through corners to edge Chelsea out.

- Most teams tend to have more points amassed at home than away as evidenced by the estimates of the home advantage effects. Our results indicated that, on the average, 50% of the season earned points are gotten at home compared to away play. Though few teams mildly showed to have strength when playing away and sharing points at home home, possibly.

- The Kelly criterion as established by previous researchers ensure growth and security

of the wealth of the gambler. Its calculated values in the table show that in the long run, the gambler has a back cash to fall on in the unfortunate case of lose of portion of accrued income.

- Much as the ultimate goal (in many respects) is for the gambler to seek upsurge in income, the RoR was found to correlates with higher 'expected fortune to reach' from the initial amount which invariably exposes the gambler to higher risk. Also, when the probability of a placed bet is positive, the RoR of the gambler becomes negligible compared to the opposite situation as was the case considered in table 4.14

- In a sub-fair situation, the gambler is expected to be bold to bet high, but on the low when in a super-fair situation for reason of taking a cautious disposition.

5.1.2 Recommendations

Relative to the work output of this thesis, the following recommendations are offered for bookmakers, gamblers and other stakeholders.

- Gamblers' and bookmakers' must pay due attention to key variables like corners and shots on target of teams in previous encounters to respectively determine the amount to bet and the odds to set for patrons.

- Gambler's must come to understand that, much as home advantage may work magic for teams, the scoring rates of teams in away matches must be carefully considered prior to placed bets.

- The probability of match wins, draw or lose regarding home and away play must be given full regard to placing a bet or setting odds. For bookmakers' though, the set odds has an implied element of probability of win, draw or lose, the gambler must adopt equation (3.1) to determine before hand the likelihood or otherwise of winning or losing a placed bets.

- As an optimal betting strategy, the gambler must bet bold in a sub-fair situation and take a pulse caution in a super-fair situation for optimal funds, all things being equal.

REFERENCES

Bellman, R. and Kalaba, R. E. (1957). Dynamic programming and statistical communication theory. *Transactions of Information Theory, IT*, Vol. 3:917–926.

Breiman, L. (1961). Optimal gambling system for favorable games. *In Proc. 4th Berkeley Symp. Math. Statist. Prob.,*, Vol. 1:65–78.

Browne, S. and Whitt, W. (1996). Portfolio choice and the bayesian kelly criterion. *Advance Applied Probbaility*, 28:1145–1176.

Cain, M., Law, D., and Peel, D. (2000). The favourite-longshot bias and market efficiency in uk football betting. *Scottish Journal of Political Economy*, Vol. 47, No. 1.

Cetinkaya, S., M. P. (1997). Optimal nonmyopic gambling for the generalized kelly criterion. *Naval Research Logistic*, 44 (7):639–654.

Coad, A., Julian, F., Roberts, G. R., and Storey, D. J. (2012). Growth paths and survival chances: An application of gambler's ruin theory. *Journal of Business Venturing*.

Conlisk, J. (1993). The utility of gambling. *Journal of Risk and Uncertainty*, Vol.6:255–607.

Crowder, M., Dixon, M., Ledford, A., and Robinson, M. (2002). Dynamic modelling and prediction of english football league matches for betting. *Journal of the Royal Statistical Society. Series D*, Vol. 51, No. 2:157–168.

Deschamp, B. and Gergaud, O. (2007). Efficiency in betting markets: Evidence from english football. *Journal of Prediction Markets*, Vol. 1:61–73.

Diecidue, E., Schmidt, U., and Wakker, P. P. (2004). The utility of gambling reconsidered. *Journal of Risk and Uncertain*, 29(3):241–259.

Dixon, M. J. and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, Vol. 46:265–280.

Dixon, M. J. and Robinson, M. E. (1998). A birth process model for association football matches. *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 47, No. 3:523–538.

Dubins, L. and Savage, L. J. (1965). *How to Gamble if you Must.* McGraw-Hill.

Dubins, L. E. and Savage, L. J. (1960). Optimal gambling systems. *National Academy of Sciences of the United States of America*, Vol. 46, No. 12:1597–1598.

Ethier, S. N. (1996). A gambling system and a markov chain. *The Annals of Applied Probability*, Vol. 6, No. 4:1248–1259.

Ethier, S. N. (2004). The kelly system maximizes median fortune. *Applied Probability*, Vol. 41 No. 4.

Feller, W. (1968). *An Introduction to Probability theory and its Applications.* John Wiley and Sons, Inc. vol.1.

Ferguson, T. S. (1965). Betting systems which minimize the probability of ruin. *Journal of the Society for Industrial and Applied Mathematics*, Vol. 13, No. 3:795–818.

Finkelstein, M. and Whitley, R. (1981). Optimal strategies for repeated games. *Advances in Applied Probability*, Vol. 13, No. 2:415–428.

Frome, E. L. (1983). The analysis of rates using poisson regression models. *Biometerics*, Vol.39 No.3:665–674.

Glickman, M. E. and Stern, H. S. (1998). A state-space model for national football league scores. *Journal of the American Statistical Association*, Vol. 93, No. 441:25–35.

Gottlieb, G. (1985). An optimal betting strategy for repeated games. *Journal of Applied Probability*, Vol.22:787–795.

Haigh, J. (2000). The kelly criterion and bet comparisons in spread betting. *Journal of the Royal Statistical Society*, Vol. 49, No. 4:531–539.

Harick, G., Goldberg, D. E., Cantu-Paz, E., and Miller, B. L. (1999). The gambler's ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation*, vol. 7:231–253.

Hartvigsen, D. (2009). The action gambler and equal-sized wagering. *Journal of Applied Probability*, Vol.46 (1):35–54.

Harville, D. (1980). Predictions for national football league games via linear-model methodology. *Journal of the American Statistical Association*, Vol. 75, No. 371:516–52.

Hausch, D. B. and Ziemba, W. T. (1995). *Efficiency of Sports and Lottery betting Markets.* North Holland Press.

Hill, I. D. (1974). Association football and statistical inference. *Journal of Applied Statistics*, 23:203–208.

Hirotsu, N. and Wright, M. (2002). Using a markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *The Journal of the Operational Research Society*, Vol. 53 No. 1:88–96.

Hirotsu, N. and Wright, M. (2003). An evaluation of characteristics of teams in association football by using a markov process model. *Journal of the Royal Statistical Society*, Vol. 52, No. 4:591–602.

Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate poisson models. *Journal of the Royal Statistical Society. Series D*, Vol. 52, No. 3:381–393.

Keller, J. (1994). A characterization of the poisson distribution and the probability of winning. *Journal of America Statisti.*, 48:294–299.

Kelly, J. L. J. (1956). A new interpretation of information rate. *Bell System Technology Journal*, Vol.35:917–926.

Knorr-Held, L. (2000). Dynamic rating of sports teams. *Journal of Statistician*, Vol.49:261–276.

Koning, R. (2000). Balance in competition in dutch soccer. *Journal of Statistician*, Vol. 49:419–431.

Kuypers, T. (2000). Information and efficiency: An empirical study of fixed odds betting market. *Applied Economics*, 32:1353–1363.

Latane, H. A. (1959). Criteria for choice among risky ventures. *Journal of Finance*, 30 (5):1213–1227.

Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets. *Journal of Economics*, 114 (495):223–246.

MacLean, L. C., Ziemba, T., and Blazenko, G. (1992). Growth versus security in dynamic investment analysis. *Management Science*, 38:1562–1585.

Maher, M. (1982). Modelling association football scores. *Journal of Royal Statistical Society Neerland*, 36:109–118.

Maslov, S. and Zhang, Y.-C. (1998). Optimal investment strategy for risky assets. *International Journal of Theoretical and Applied Finance*, Vol. 1:377–387.

McSharry, P. E. (2007). Altitude and athletic performance: Statistical analysis using football results. *British Medical Journal*, Vol. 335, No. 7633:1278–1281.

Mohan, C. (1995). The gambler's ruin problem with correlation. *Biometrika*, No.3/4:486–493.

Moroney, M. J. (1956). *Facts from Figures.* Penguin 3rd. edition.

Moya, F. E. (2012). Statistial methodology for profitbale sports gambling. Master's thesis, Simon Fraser University.

Ottaviani, M. and Norman, P. S. (2007). The favourite-longshot bias: An overview of the main results.

Piotrowski, E. and Schroeder, M. (2007). Kelly criterion revisited: optimal bets. *The European Physical Journal B*.

Reep, C. and Benjamin, B. (1968). Skill and chance in association football. *Journal of Royal Statistical Society A*, Vol. 131:581–585.

Reep, C., Pollard, R., and Benjamin, B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society*, Vol. 134, No. 4:623–629.

Ridder, G., Cramer, J. S., and Hopstaken, P. (1994). Prediction and home advantage for australia rules football. *Journal of Applied Statistics*, 19:251–261.

Rosett, R. N. (1965). Gambling and rationality. *Journal of Political Economy*, Vol.73:595–607.

Rotando, R. and Thorp, E. O. (1969). Optimal gambling systems for favourable games. *International Statistical Institute (ISI)*, Vol.37(3):273–293.

Rue, H. and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society*, Vol. 49, No. 3:399–418.

Sinkey, A. S. and Logan, T. D. (2012). Are sports betting markets prediction markets? evidence from a new test. *Journal of sports economics*, Vol.15(1):45–63.

Sudderth, W. D. (1971). A gambling theorem and optimal stopping theory. *Annals of Mathematical Statistics*, 42 No. 5.

Thaler, R. H. and Ziemba, W. (1988). Parimutuel betting markets: Race tracks and lotteries. *Journal of Economic Perspectives*, Vol.2:161–174.

Thorp, E. O. (1997). The kelly criterion in blackjack, sports betting, and the stock market. In *Gambling and Risk Taking*.

Truelove, A. J. (1970). Betting system in favorable games. *Annals of Mathematical Statistics*, Vol.41 No.2:551–566.

Warwick, J. (2007). 91.57 modelling penalty competitions to decide football matches. *The Mathematical Gazette*, Vol. 91, No. 521:342–348.

Whitrow, C. (2007). Algorithms for optimal allocation of bets for many simultaneous events. *Applied Statistics*, Vol. 56 part. 5:607–623.

APPENDIX

5.2 APPENDIX A

5.2.1 Odds of Bookmaker

Table 5.1: Probabilities obtained from given odds

Odds and its transformed probabilit					
Home Odds (A)	Draw Odds(B)	Away Odds(C)	Decimal H odds(D)	Decimal D 0dds(E)	Decim Odds(
7/5	11/5	6/4	2.4	3.2	2.5
2/9	3/1	13/2	1.2	4.00	4.00
8/11	3/2	4/7	1.73	2.50	1.57
13/2	3/1	2/5	7.50	4.00	1.40
5/4	21/10	41/20	2.25	3.10	3.05
1/12	13/2	14/1	1.08	8.50	15.00
4/5	12/5	14/15	1.80	3.4	1.93
4/15	18/5	15/2	1.27	4.6	8.50
4/9	3/1	19/4	1.44	4.00	5.75
3/5	13/5	18/5	1.60	3.6	4.60
11/20	5/1	11/1	1.55	6.00	12.00
8/13	12/5	17/4	1.62	3.4	5.25
4/11	4/1	15/2	1.36	5.00	8.50
2/9	5/1	13/2	1.22	6.00	7.50
4/6	5/2	19/5	1.67	3.50	4.80
4/9	16/5	17/2	1.44	4.20	9.50
5/4	11/5	15/8	2.25	3.20	2.88
7/2	12/5	13/20	4.50	3.40	1.65
11/10	1/5	39/20	2.10	1.20	2.95
7/5	11/5	6/4	2.40	3.2	2.50
15/2	19/4	1/5	8.50	5.75	1.20
3/20/2	5/1	11/1	1.15	6.00	12.00
3/1	5/2	3/4	4.00	3.50	1.75
8/13	12/5	18/5	1.62	3.40	4.60
9/4	9/4	43/20	2.11	3.25	3.15

5.3 Appendix B

5.3.1 Full Results of Model Scoring intensities for both Home and Away Teams

Table 5.2: R output of home scores of teams through MLEs

Teams	**Estimate (Std.Error)**	λ_h	**z value**	**pr(>|z|)**
Liverpool	-0.1406 (0.128)	0.8688	-0.749	0.4541
Chelsea	-.2191 (0.192)	0.8032	-1.142	0.2535
Arsenal	-0.4733 (0.207)	0.6229	-2.290	0.0220
Everton	-.8150 (0.2312)	0.4426	-3.526	0.00042
Tottenham	-0.7436 (0.226)	0.4754	-3.297	0.00098
Man United	-0.6451 (0.218)	0.5246	-2.956	0.003120
Southampton	-0.8920(0.238)	0.4098	-3.756	0.000173
Stoke	-0.9754 (0.245)	0.3770	-3.986	6.71e-05
Newcastle	-1.2205 (0.268)	0.2951	-4.550	5.3e-06
Crystal Palace	-0.6451(0.218)	0.5246	-2.956	0.00312
Swansea	-0.8528 (0.234)	0.4262	-3.641	0.000272
West Ham	-1.0664 (0.253)	0.3442	-4.215	2.5e-05
Sunderland	-1.1151(0.258)	0.3279	-4.328	1.51e-05
Aston Villa	-1.0198 (0.249)	0.3607	-4.101	4.12e-05
Hull	-0.9754 (0.2447)	0.3770	-3.986	6.71e-05
West Brom	-1.2777 (0.2743)	0.2787	-4.659	3.18e-05
Norwich	-0.8920 (0.238)	0.4098	-3.756	0.000173
Fulham	-0.9754 (0.245)	0.3770	-3.986	6.71e-05
Cardiff	-4.733(0.207)	0.6229	-2.290	0.022

— Sig.codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 8.2628e+01 on 19 degrees of freedom

Residual deviance: -1.5099e-14 on 0 degrees of freedom

AIC: 143.68

Number of Fisher Scoring iterations: 3

Table 5.3: R output of Teams Away scores

Teams	**Estimate (Std.Error)**	λ_h	**z values**	**pr(>|z|)**
Liverpool	0.2076 (0.216)	1.2307	0.963	0.3355
Chelsea	-0.3314 (0.248)	0.7179	-1.338	0.1810
Arsenal	-0.1978 (0.239)	0.8205	-0.829	0.4069
Everton	-0.5725 (0.267)	0.5641	-2.147	0.0318
Tottenham	-0.3677 (0.250)	0.6923	-1.469	0.1419
Man United	-0.1082 (0.233)	0.8974	-0.465	0.6421
Southampton	-0.5725 (0.267)	0.5641	-2.147	0.0318
Stoke	-0.6678 (0.275)	0.5128	-2.147	0.0152
Newcastle	-0.6678 (0.275)	0.5128	-2.147	0.0152
Crystal Palace	-0.9555 (0.309)	0.3846	-3.145	0.0017
Swansea	-0.6190 (0.271)	0.5385	-2.287	0.0222
West Ham	-0.9555 (0.309)	0.3846	-3.145	0.0017
Sunderland	-0.6678 (0.275)	0.5128	-2.428	0.0152
Aston Villa	-1.0245 (0.312)	0.3590	-3.288	0.0010
Hull	-0.7732 (0.2850)	0.4615	-2.713	0.0067
West Brom	-0.7732 (0.2850)	0.4615	-2.713	0.0067
Norwich	-1.2657 (0.341)	0.2820	-3.707	0.0027
Fulham	-0.8910 (0.297)	0.4102	-3.001	0.0027
Cardiff	-1.1787 (0.330)	0.3077	-3.570	0.0004

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 7.0756e+01 on 19 degrees of freedom

Residual deviance: -1.5543e-14 on 0 degrees of freedom

AIC: 137.82

Number of Fisher Scoring iterations: 3

5.3.2 R codes for Poisson Regression

Call: glm(formula = ASCORE factor(TEAMS), family = poisson(log), data = matches)

Call: glm(formula=HSCORE factor(TEAMS),family=poisson(log),data=matches)

Call: glm(formula = ASCORE FOULF + FOULA + REDCA + CORNERSP + YELLOWCA + SHORTSPER, family = poisson(log), data = matches)

Call: glm(formula = HSCORE FOULF + FOULA + REDCA + CORNERSP + YELLOWCA + SHORTSPER, family = poisson(log), data = matches)