

Winners and Losers in Soccer World Cup: A Study of Recent History and how to bet if you must*

Fragiskos Archontakis

Institute of Innovation and Knowledge Management (INGENIO)

Universidad Politécnica de Valencia-CSIC

Edificio i4, Camino de Vera s/n

E-46022 Valencia, Spain

Tel.: (+34) 963 877 048

Fax.: (+34) 963 877 991

Email: fraguis@ingenio.upv.es

This Version: June 2004

* This paper has benefited from comments by seminar participants of IASE's 4th International Conference on Sports, Tourism and Culture: 31 May-2 June 2004 held in Athens, Greece. All remaining errors are mine.

ABSTRACT

Football betting in Europe has seen a rapid growth in the last two decades. However, the betting market seems to be rather inert in becoming an efficient market in a similar fashion as inefficiencies have been appearing in the international financial markets. A typical fixed odds set provided by the bookmaker for the result of a soccer match would be: (odds for home team win, odds for draw, odds for away team win). The present work differentiates itself from the others in the relevant literature in the mere fact that the only outcome of probabilistic interest is chosen to be the draw, that is, the most difficult to predict-see Pope and Peel (1989).

The FIFA World Cup is considered to be the most important soccer tournament between national teams from all over the world and is taking place every four years. The data used in the present study come from a 20-year span of World Cup Final Tournaments. Typically the odds include the bookmaker's in-built "take" margin, usually in the range of 11-15%. In order for the gambler to make money out of betting he has to be able to determine the true probabilities of a soccer game better than the bookmaker in order to overcome the bookmakers' profit margin. The estimate for the probability of a draw in a World Cup Final Tournament is found to be 29,76%, which is in agreement with similar results; see for example Dixon and Coles (1997). A simple mathematical sequence, known as the Fibonacci sequence, is used in order to define a consistent betting strategy. It will be shown that for fixed odds given for a draw equal to the value 2.618, the betting rule proposed is giving always at least a unit of profit. Nevertheless, the average for a fixed odds is greater than 3.0, thus we consider the odds also as a random variable and the model is implemented by a Monte Carlo simulation.

JEL classification: L83, C15.

1. Introduction and Literature Review

Football betting in Europe has seen a rapid growth in the last two decades. In the UK there are situated some of the largest betting companies. It is difficult to have a good estimate of the size of the betting market. Indicatively, some betting firms currently operate in 1500-2000 locations in the UK, while there are tens of thousands the employees working in licensed betting offices. Already, a decade ago sports betting was one of the fastest growing areas in the UK's betting industry, see Jackson (1994). More recently, it was found out that "football betting is the fastest-growing form of gambling in the UK", see the Mintel Intelligence Report (2001). Thus, the huge size of the football gaming industry is evident, especially in the last years when Internet has allowed people betting from home using their credit cards, reducing the moving costs and even gambling from another country overcoming home gambling legislations.

However, the betting market seems to be rather inert in becoming an efficient market in a similar fashion as inefficiencies have been appearing in the international financial markets, see Fama (1970) and Osborne (2001). Indeed, there are some evidence that the betting market is far from being efficient. For instance, it is well known that there is a favourite bias, see Thaler and Ziemba (1998) for a collection of examples in various betting markets and Sauer (1998) for an excellent analysis of the economics of wagering markets. A natural question is whether one can take advantage of these inefficiencies and make money out of the betting market by means of an outsider betting strategy. This question has been raised several times in the past and one plausible explanation is given by Pope and Peel (1989). They examine the efficiency of the fixed odds betting market and find that although there is some evidence of ex post inefficiency there did not appear to be profitable betting strategies in hand that could have been implemented ex ante during the sample period. Dixon and Coles

(1997) propose a Poisson parametric model, similarly motivated by the same possibility of potential inefficiencies present in the football betting market. Their model is shown to have a positive return when used as the basis of a betting strategy.

The present work differentiates itself from the others in the relevant literature in the fact that the only random outcome of interest is chosen to be the draw, that is, the most difficult to predict. Indeed, Pope and Peel (1989) have found that in comparing the odds provided by four betting firms, the standard deviations for the draw probabilities were consistently lower when compared with the other two outcomes, i.e. a win of either team. They conclude that "...this behaviour could simply reflect a general inability to predict draw outcomes with any degree of reliability...".

The purpose of the present paper is to bridge the gap that seems to exist between the theory, i.e. the numerous theoretical papers regarding football betting markets, and the practice, that is the relatively scarce literature regarding simple applied techniques on gambling. To this end, our aim is to find a good estimate of the draw-distribution of football matches and propose a way of taking advantage of this information. Furthermore, we agree with the results of previous researchers regarding the games distribution and reconsider a variation of the, so-called, martingale casino strategy as a method of betting. A simple mathematical sequence, known as the Fibonacci sequence, is used in order to define a consistent betting strategy. It will be shown that for fixed odds given for a draw equal to 2.618, the betting rule proposed is giving always at least a unit of profit. Nevertheless, fixed odds vary from game to game and their average is found to be approximately 3.0. Thus, we decide to consider the odds also as a random variable and the model is implemented by a Monte Carlo simulation.

The rest of the paper is organised as follows. Section 2 explains how fixed odds work and briefly describes the data. Section 3 provides the estimation of draw and some by-products regarding the robustness of the parameter across time. The betting rule is discussed and analysed in Section 4. Finally, Section 5 concludes.

2. Analysis of Fixed Odds and Data Set

There is a wide variety of information that can be extracted by a single football match. The final score, the half-time score, the total number of goals scored, the players who scored, in which minute were the goals scored, etc. In order to keep things simple and, above all, to distinguish the present work from the relevant bibliography, we choose to look into the probability of a game resulting in a draw. Hence our probability space will have only two outcomes (draw, not draw) and is defined as follows:

$$p = \text{Prob}[\text{game } x \text{ is a draw}] \text{ and } q = 1-p = \text{Prob}[\text{game } x \text{ is not a draw}].$$

A typical fixed odds set provided by the bookmaker for the result of a soccer match would be: (home team win, draw, away team win) = (5/4; 21/10; 9/5) in fractional odds, or (2.25; 3.10; 2.80) in decimal odds. In a state monopoly case (Greece is an example) there is only one set of quoted odds. In the UK, though, there is a rich variety of betting-offices. The website *www.oddschecker.com* is providing updated online information regarding most of quoted odds offered in the UK betting market. For example, on 16th September 2003 the Champions League match of AC Milan versus Ajax Amsterdam was taking place in Milan, Italy. In Table 1 it is easy to find that the best odds for AC Milan to win are 4/7, the best odds for Ajax to win are 11/2 and the best odds for a draw are 13/5.

Let us now consider again the set of bookmakers' odds for the particular match above, chosen to be (2.25; 3.10; 2.80). In a situation of maximum uncertainty one would expect the true odds against each of three outcomes to be 2/1 or (3.0; 3.0; 3.0). However, the bookmaker's quoted odds might well be (2.7; 2.7; 2.7) since he has

allowed for an in-built profit margin. The bookmaker will define a set of odds which follows from his a priori subjective probability appointed regarding the match's result. For odds o_1/o_2 by using the formula $p = o_2/(o_1 + o_2)$ one can find the probability appointed by the bookmakers to the particular result. Hence, for our example, the corresponding set of bookmaker's announced probabilities is (0.444; 0.323; 0.357). Adding up the probabilities it turns out to be that their sum is approximately 1.124, i.e. greater than unity (that one might have expected). This is because the odds include the bookmaker's in-built "take" margin. In this particular example the bookmaker's profit is 12.4%. For calculating the true probabilities appointed by the bookmaker, one has to scale the bookmakers' announced probabilities by the factor 1.124 to find that the true probability for a draw is 0.287, i.e. the set is (0.395; 0.287; 0.318) which sum up to unity and are considerably lower than the announced probabilities. In order for the gambler to make money out of betting he has to be able to determine the true (underlying) probabilities of a game more accurately than the bookmaker so that he can overcome the bookmaker's profit margin.

The FIFA World Cup (F.I.F.A. = Fédération Internationale de Football Association) is considered to be the most important soccer tournament between national teams from all over the world and is taking place every four years. The data used in the present study come from a 20-year span of soccer World Cup Final Tournaments (WCFT). Data consists of 336 full-time WCFT match results. All data come from the official World Cup's web-site: <http://fifaworldcup.yahoo.com/>

It is found that the last six WCFTs have been won by five different countries, that is: Argentina, Brazil (twice), France, Germany, Italy. Notably, these five countries were also the runner-ups for the years considered. Thus, one can safely claim that the real winners consist of Brazil. Their overall game (90' full-time) score record is found to be (27 wins, 4 draws, 4 defeats), i.e. (77.0:11.5:11.5). This mere fact might be an

explanation of why bookmakers appear to express a favourite bias when quoting their odds.

Table 2 provides a summary report of all data used and indicates as a fraction the amount of games drawn per number of total games played. It is seen for example that in the Group Matches stage there is a 28.75% probability of draw found, while in the Final Competition stage of the tournament, where the importance of each game is the highest (and the fear for losing higher as well), is circa 32.29%. Furthermore, it is interesting that in Italy 1990 there were very few draws in the Group stage. However, in the Final stage there was a surprising 50%, which brought the overall tournament close to the total average.

3. Estimation of the probability of draw

The underlying distribution of the data y_1, y_2, \dots, y_T is the Bernoulli. The Bernoulli distribution is a discrete distribution having two possible outcomes labelled by $y = 0$ and $y = 1$ in which $y = 1$ ("success") occurs with probability p and $y = 0$ ("failure") occurs with probability $q=1-p$, where $0 < p < 1$.

The Bernoulli probability distribution function of each observation is $f(y) = (1-p)^{1-y} p^y$ and the log-likelihood function, to be maximized, $l(p|y) = \sum[(1-y_i)\log(1-p) + y_i\log(p)]$.

The first order conditions are generated by the equations $d l(p;y)/dp = 0$, with solution the maximum likelihood estimator (MLE) of p to be:

$$\hat{p} = \sum y_i / T = \bar{y} .$$

From the Theorem regarding the asymptotic normality of the MLE (see, for example, Rice (1995) for more details) we derive that for large samples $\hat{p} \approx N(p, p(1-p)/T)$, i.e.

the MLE of p follows a Gaussian distribution with mean p and variance $p(1-p)/T$.

Thus, an approximate 95% confidence interval (C.I.) for \hat{p} would be:

$$\mathfrak{S}=[\hat{p} \pm 2 \sqrt{p(1-p)/T}].$$

From Table 2 we have that the total number of draws is 100, while the total number of games is $T=336$ observations. Hence $\hat{p} = \sum y_i/T = \bar{y}$, and as a result $\hat{p} = 100/336 = 0,2976$. The estimate for the probability of a draw in a World Cup Final Tournament is found to be 29,76%, in agreement with similar results; see for example Dixon and Coles (1997, page 267) and Stefani (1983, p.322).

An obvious observation derived from Figure 1 is that there seems to be some kind of structural change observed in the USA 1994 WCFT, in the sense that there is a considerably lower amount of draws compared with the other tournaments. The only plausible explanation we could come up regarding this incident was the rule passed in 1992 that does not allow the goalkeeper to catch the ball with his hand when receiving a pass by a team-mate:

1992: Backpass ruling: Law XII - Fouls and Misconduct

“On any occasion when a player deliberately kicks the ball to his own goalkeeper, the goalkeeper is not permitted to touch it with his hands. If, however, the goalkeeper does touch the ball with his hands, he shall be penalised by the award of an indirect free-kick to be taken by the opposing team from the place where the infringement occurred,...” [Source: <http://www.fifa.com/en/game/historylaws.html>]

However, it is evident from Figure 1 that in the next tournament there was an immediate and considerable correction to this effect and the estimate of a draw went back to its normal levels and very close to its overall average. In effect, Figure 1 suggests that the cumulative draw probability for the WCFTs games (except USA 1994) is approximately within the bands of [0.30, 0.35], that is, roughly 1/3 of the WCFTs games is a draw.

4. Betting Rule

The betting rule we propose has similarities with the so-called martingale strategy, used in the casino's roulette. It is "naïve", in the sense that one bets only for a draw but this makes it also simple to apply. The betting strategy suggested, coined the Fibonacci betting rule, is as follows: assuming that a draw will eventually come in a series of games we apply the strategy of betting continuously for a draw with amounts defined by the Fibonacci sequence. The whole point is to find out what amount of bet one should place each time in a betting sequence. To this end we need to define the Fibonacci sequence.

A Fibonacci sequence is defined as the element 1, followed by another 1, and each element thereafter is the sum of the previous two elements. For example, the first few elements of a Fibonacci sequence are: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ...

The mathematical sequence is produced by the recursive formula:

$$a_{n+1} = a_n + a_{n-1}, \text{ where } a_1 = 1 \text{ and } a_2 = 1.$$

Its characteristic equation is $g(\lambda) = \lambda^2 - \lambda - 1 = 0$ with characteristic roots equal to $\lambda_{1,2} = (1 \pm \sqrt{5})/2$. The positive root $(1 + \sqrt{5})/2$ is also known as the "golden ratio" and is approximately equal to $\phi = 1.618$, since for consecutive Fibonacci terms it is known that their ratio $a_{n+1}/a_n \rightarrow \phi$, when $n \rightarrow \infty$. The n^{th} Fibonacci term is given by Binet's Formula: $a_n = [((1 + \sqrt{5})/2)^n - ((1 - \sqrt{5})/2)^n] / \sqrt{5}$.

The gambler has a probability p of winning on any one of a sequence of bets. If he places a bet of a_n monetary units, assuming that the given odds are a fixed number b , then the amount won will be ba_n . The probability of winning for the first time at the x^{th} bet is given by $p(1-p)^{x-1}$, i.e. it follows the geometric distribution. Its mean is $1/p$ and

its variance is $(1-p)/p^2$. The betting rule is: we start with a unit bet a_1 and then follow it up with another unit bet a_2 . From the 3rd bet onwards, the n^{th} bet a_n is the sum of the last previous ones being terms of a Fibonacci sequence. Table 3 summarizes the Fibonacci betting rule. In Table 3 we denote bets as a_n , the sums of invested bets, i.e. the total cost, as S_n , the revenue as R_n and the profit as $P_n = R_n - S_n$.

We wish to have (at least) a unit of net gain in each series of bets, hence the following equation should hold:

$$a_n b - (a_{n+2} - 1) \equiv 1. \text{ It also holds } a_{n+2} = a_{n+1} + a_n \text{ thus we derive:}$$

$$a_n b - (a_{n+1} + a_n - 1) = 1, \text{ or } b = (a_{n+1} + a_n) / a_n = 1 + a_{n+1} / a_n \Rightarrow$$

$$b \rightarrow 1 + \phi = 2.618, \text{ when } n \rightarrow \infty.$$

It was shown that assuming that for given fixed draw-odds b higher or equal to 2.618, the betting rule proposed is giving at least a unit of profit. Nevertheless, the average for a draw in a pooled data-set of 102 fixed odds was found to be approximately 3.25, thus it is worth trying to apply the Fibonacci betting rule. Note that it seems bookmakers apply different odds depending not only on the game itself but also on the country and the league it belongs to. For example, soccer games in *Serie B*, Italy's second league are expected to provide draws more often than the ones in the *Premiership*, England's first league. However, there is a serious drawback in applying the Fibonacci betting rule. Notice that for $p=0.3$ the mean of the geometric distribution is given by $1/p=10/3$, while its variance is $(1-p)/p^2=70/9$, which implies that often one has to be quite patient. This is better seen by trying to answer the following question: How large must the gambler's initial capital be in order to sustain this betting system through the x^{th} bet given that he lost all previous $x-1$ bets?

Consider the random variable S_n , the amount of capital that the gambler needs in order to sustain the Fibonacci betting strategy. We will answer the question above by calculating the probability density function of S_n . From Table 3 and the comment

on the geometric distribution, we deduce that the random variable S_n has probability density given by $Pr[S_n = a_{n+2} - 1] = p(1-p)^{n-1}$, with $p=0.3$. Hence, the expected value of S_n is given by $E[S_n] = \sum (a_{n+2} - 1) p(1-p)^{n-1}$, for all $n=1,2,3,\dots$. It follows that $E[S_n] \rightarrow \infty$, when $n \rightarrow \infty$. That is, no finite amount of money is sufficient to sustain the Fibonacci betting system. However, there is a slight catch in this calculation: the WCFTs have a finite number of 64 games and there are always more than ten draws in this series of games, i.e. on average in every six games there is a draw appearing.

In order to investigate in more detail the expected gains of the profit distribution we consider the bookmaker's odds b to follow a Gamma random variable G (with mean μ and variance σ^2) and the model is implemented by a Monte Carlo simulation experiment. The "odds-data" are taken from two well-known betting firms. They consist on the bookmakers' odds for a pooled sample of 102 games from major national championships (England, France, Germany, Italy and Spain), international friendlies between national teams and Champions League games. The empirical distribution followed (see Figure 2) has empirical mean $\mu = 3.25$ and standard deviation $\sigma = 0.42$. Using the above information and via the method of moments we have fit a gamma probability distribution $G(t; \alpha, \beta) = \beta^\alpha t^{\alpha-1} \exp(-\beta t) / \Gamma(\alpha)$ to the data. From the empirical mean and variance we derive its estimated parameters $\alpha = 60$ and $\beta = 18.5$. Note that by using the Gamma distribution one actually violates the assumption on the bookmaker's odds $b \geq 2.618$.

The Monte Carlo simulation makes clear the point that although the mean of the Profit distribution is positive and larger than unity, it is accompanied by a very large standard deviation. In fact, the Profit distribution's density seems to be centered in an area relatively near to zero, as is also seen in Table 4 and in Figure 3.

Table 4 suggests that the mean of the Profit distribution can be safely considered to be statistically equal to zero. Thus, it is highly unlikely that one will end up always in

positive. For instance, in the case where the simulated odds violate the assumption $b \geq 2.618$ there may be a long series of $N-1$ non-draws followed by a draw with low odds, i.e. with $b < 2.618$, which inevitably will give to $P_N = R_N - S_N$ a negative value. Thus, naturally the Fibonacci betting rule fails to provide a positive gain with certainty.

5. Conclusions

This paper considers the unexplored fact in the wagering literature that the event of a draw between two soccer teams is random and difficult to model. This is especially true for bookmakers when setting the football games' odds. By using a 20-year span dataset of Soccer World Cup data we find out that the probability of draw is relatively stable and found to be equal to 29,76%. An interesting by-product of this result is the fact that there seemed to be some kind of structural change observed in the USA 1994 Tournament, in the sense that there is a notably lower amount of draws compared with the other tournaments due to the 1992 rule that does not allow the goalkeeper to catch the ball with his hand when receiving a pass by a team-mate. However, in the next tournament there was a considerable correction to this effect and the estimate of a draw went back to its normal levels.

The betting strategy suggested is the Fibonacci betting rule: assuming that a draw will eventually come in a series of games we apply the "naïve" strategy of betting continuously for a draw with amounts defined by the Fibonacci sequence. It is shown that for fixed odds given for a draw equal to 2.618, the betting rule proposed is giving always at least a unit of profit. Nevertheless, the average for a fixed odds is often greater than 3.0, thus we consider the odds also as a random variable and the model is

implemented by a Monte Carlo simulation. The simulation experiment unveils that the Fibonacci betting rule is likely to be giving a positive gain but with great uncertainty. Added to this, the possibility of “gambler’s ruin” in order to sustain the Fibonacci betting system indicates that in an uncertain soccer world one should keep on searching for an improved long-run betting strategy.

References

[1] Dixon, M. J. and Coles, S. G. (1997). “Modelling Association Football Scores and Inefficiencies in the Football Betting Market”, *Applied Statistics* 46: 265-280.

[2] Fama, E. F. (1970). “Efficient Capital Markets: A Review of Theory and Empirical Work”, *Journal of Finance* 40: 793-805.

[3] Forrest, D. and Simmons, R. (2001). “Globalisation and Efficiency in the Fixed-odds Soccer Betting Market”, *Mimeo*. University of Salford, Salford, UK.

[4] Jackson, D. (1994). “Index Betting on Sports”, *The Statistician* 43: 309-315.

[5] Mintel Intelligence Report (2001). “Online Betting”, Mintel International Group Ltd, London.

[6] Osborne, E. (2001). “Efficient Markets? Don’t bet on it”, *Journal of Sports Economics* 2: 50-61.

- [7] Pope, P.F. and Peel, D.A. (1989). "Information, Prices and Efficiency in a Fixed-Odds Betting Market", *Economica* 56: 323-41.
- [8] Rice, John A. (1995). *Mathematical Statistics and Data Analysis*, 2nd Edition, Duxbury Press, Belmont, California.
- [9] Sauer, R.D. (1998). "The Economics of Wagering Markets", *Journal of Economic Literature* 36: 2021-2060.
- [10] Stefani, R. T. (1983). "Observed Betting Tendencies and Suggested Betting Strategies for European Football Pools", *The Statistician* 32: 319-329.
- [11] Thaler R.H. and Ziemba W.T. (1998). "Parimutuel Betting Markets: Racetracks and Lotteries", *Journal of Economic Perspectives* 2: 161-174.

Tables & Figures

Table 1 Major UK boomakers' odds for the Champions League soccer match between AC Milan vs Ajax Amsterdam (16.09.2003) provided by oddschecker.com

 Updated at 16-09- 2003 10:02 refresh prices 																		
8/15	8/15	8/15	8/15	1/2	8/15	1/2	4/7	8/15	8/15	8/15	1/2	1/2	5/9	4/7	8/15	1/2	6/11	AC Milan
12/5	12/5	9/4	5/2	12/5	9/4	13/5	9/4	12/5	12/5	12/5	12/5	12/5	12/5	5/2	12/5	12/5	12/5	Draw
5	5	5	9/2	5	5	9/2	9/2	5	5	5	5	11/2	21/4	9/2	5	5	26/5	Ajax

The odds we show come directly from the online bookmakers. Whilst every effort is made to ensure that the odds are correct, it is your responsibility to check before you place a bet.

Table 2 Summary report of number of draws appeared in each Group and Final Stages per Tournament Year. Total sample observations are $T = 336$.

World Cup's Year	Spain 1982	México 1986	Italy 1990	USA 1994	France 1998	Korea- Japan 2002
Group Matches						
Group 1	5/6	3/6	0/6	1/6	2/6	3/6
Group 2	0/6	2/6	1/6	2/6	4/6	1/6
Group 3	1/6	1/6	0/6	3/6	2/6	1/6
Group 4	2/6	1/6	1/6	0/6	2/6	1/6
Group 5	3/6	2/6	1/6	2/6	4/6	2/6
Group 6	1/6	2/6	5/6	0/6	1/6	3/6
Group 7					1/6	1/6
Group 8					0/6	2/6
Total Groups	12/36	11/36	8/36	8/36	16/48	14/48
Final Competition						
Round of 16	1/3	1/8	4/8	2/8	2/8	3/8
Quarter-Finals (except 1982)	2/3	3/4	2/4	1/4	1/4	2/4
	0/3					
	1/3					
Semi-Finals & Finals	1/4	1/4	2/4	1/4	1/4	0/4
Total Final Competition	5/16	5/16	8/16	4/16	4/16	5/16
Total Tournament	17/52	16/52	16/52	12/52	20/64	19/64

Table 3 Betting rule based on a Fibonacci sequence

Index n	Bets a_n	Sum S_n	Revenue R_n	Profit P_n
1	1	1	b	$b-1$
2	1	2	b	$b-2$
3	2	4	$2b$	$2b-4$
4	3	7	$3b$	$3b-7$
5	5	12	$5b$	$5b-12$
6	8	20	$8b$	$8b-20$
7	13	33	$13b$	$13b-33$
\vdots	\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots	\vdots
n	a_n	$a_{n+2} - 1$	$a_n b$	$a_n b - (a_{n+2} - 1)$

Table 4 Simulation results of betting experiment

Number of replications	Simulated mean of Profit distribution	Simulated std. deviation of Profit distribution
10	7.25	10.27
20	4.79	9.53
30	3.66	6.22
60	5.69	12.22
100	8.19	45.52
200	41.19	305.93
500	34.82	561.44
1000	17.58	158.13
5000	49.47	1786.10
10000	47.11	2482.20
100000	1054.10	280200.0

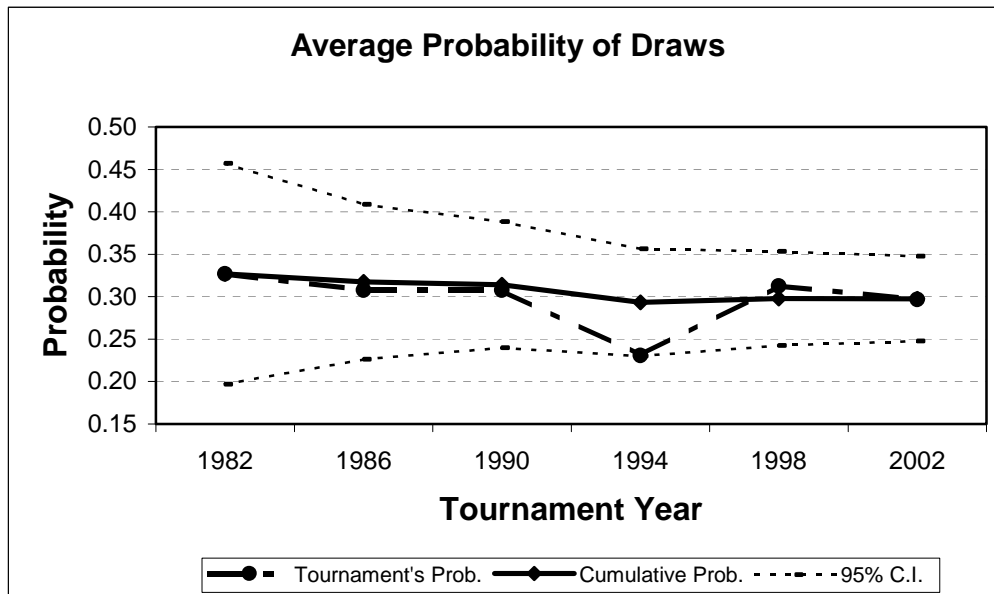


Figure 1 Average probability of draws per tournament and cumulative probability along with its 95% confidence intervals

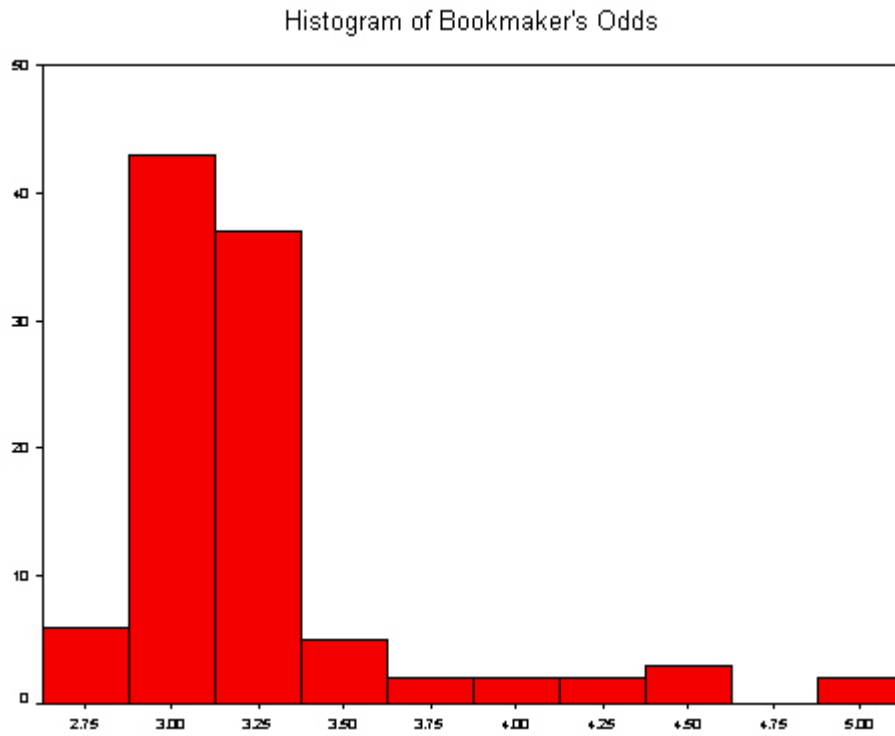


Figure 2 Bookmaker's Odds for a pooled sample of 102 observations

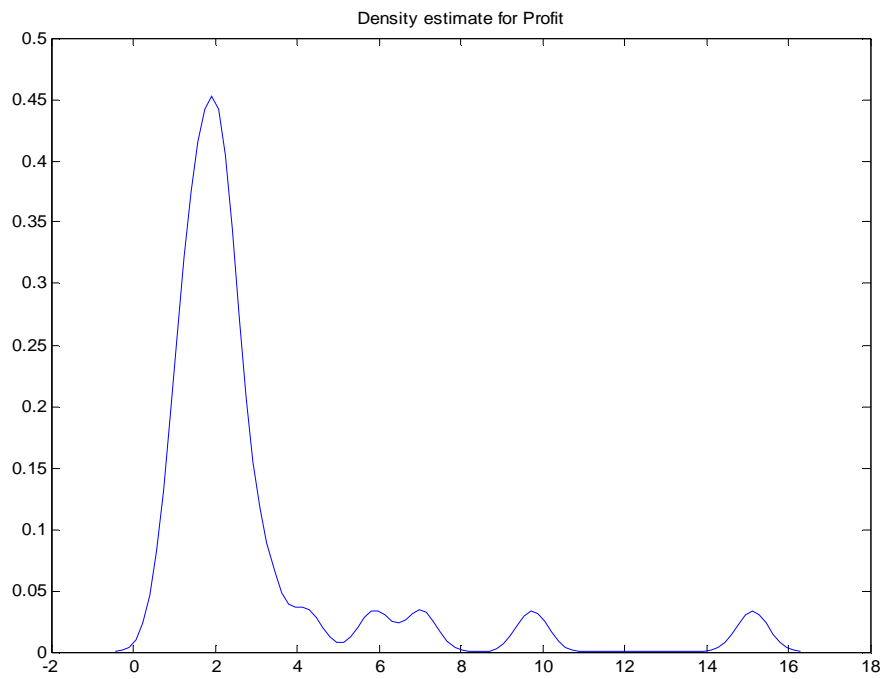


Figure 3 Example of a probability density function estimate for the Profit distribution by using the Fibonacci betting rule ($n=30$)