Modelling prices of in-play football betting markets

Erik Hoog

Handledare: Magnus Wiktorsson
Examinator: Erik Lindstrom

October 2013

Abstract

Odds in sports betting can be seen as a price on a financial contract. This Master thesis is a study of live odds in football matches. Movements and transactions in live odds from 1114 matches are investigated using a high frequency data set. The prices are then modeled using a Poisson distribution, and used to create a trading strategy.

Acknowledgement

In the summer of 2012 I found some excel sheets with historical football odds on the internet which I download and played with in Matlab. For a sport nerd about to write his master thesis in mathematical statistics it opened the door to a fantastic world of both sports and math. A big thanks to my supervisor Magnus Wiktorsson and to my friend Johannes Siven for his tireless reviewing. Last but not least, thank you Stina.

Contents

1 Introduction

The global gambling market is an enormous business and just in 2009, internet gambling generated gross gaming revenue of 21.7 billion dollar worldwide[1]. One of the most popular forms of gambling is sport betting. Betting on a result can add excitement while watching a game, or even be a way of making money. The most common is to bet on 1, X or 2 representing a home team win, a draw and an away team win. Nowadays it is also possible to bet on game details as the number of corners in a match.

Live betting is a bet on an outcome when a match already started. It requires fast communication between bookmaker and gambler since odds need to be updated continuously. Game conditions can change dramatically in just a few seconds with a goal or a red card and that type of information should of course be available for both the one providing and the one betting on an odds. The odds also depend on time left in the match, since the likelihood for different results increase or decrease towards the end. In a match with no goals the likelihood for a 0-0 match, i.e. a draw, is increasing with time left which the odds of course should be compensated for.

These movements in odds and transactions from live betting make up a huge bank of data interesting not only for a sport nerd but also for financial analysts. The entrances of betting markets have made the live odds even more complex and interesting, that the odds are here decided by gamblers and not updated by a bookmaker. This paper is a study of a high frequency data set with live betting odds from 1114 football matches.

The betting markets are new phenomena and will therefore be explained. Basic concepts in a financial market like spread and arbitrage relation will then be studied in the betting market. How people bet before a match and when do they bet during the match? After these investigations a fair way of deciding live odds for 1, X and 2 is suggested. These fair odds will be compared with the actual odds from the data set. Can the deterministic model really price in a reasonable way? Finally using that the model hopefully prices more fairly, i.e. predicts better, than gamblers a profitable betting strategy will be created. To summarize, the live odds from 1114 football matches is in this paper analyzed and treated as any other financial market.

1.1 Literature

The odds in football matches are a well-known and thoroughly studied subject. There are everything from plane match statistics to advanced prediction models made as engineering master thesis using quantification of qualitative data, see for example [5]and [6]. But the live betting markets are quite new[2], and therefore it does not exist much literature on the subject. Since there are few or none academic papers on modelling live prices in betting markets, there are not many references to similar work in this paper.

[1] http://www.indiangaming.org/info/alerts/Spectrum-Internet-Paper.pdf
[2] http://corporate.betfair.com/about-us/betfair-group.aspx

1.2 Betting markets

Odds decide how much you potentially get back for each betted unit. There are several ways of denoting these odds:

Decimal	*Fractional*	*USA*	*Price*
1.5	1/2	-200	0.67
3	2/1	+200	0.33

.

The tabular above shows different notations for odds that gives net earnings of 50 percent, i.e. when betting 10 units you get 15 back.

The odds for a home team to win a match can be seen as a price on a derivative, where the underlying is the final score. Note that it is impossible to buy or sell the underlying, in contrast derivatives on a stock market. An odds of 1.2 corresponds to the price $(1/1.2) = 0.83$ to win 1 unit, so a low odds corresponds to a high price. If the game is fair, the prices for the outcomes 1, X, 2 correspond to their probabilities, and add up to one

$$P(1) + P(X) + P(2) = 1, Price(1) + Price(2) + Price(X) = 1$$

The first internet site with online odds was up and running in 1986[3], and live odds on sports came in 1994. The odds were updated manually in contrast to today where odds change rapidly in split seconds. If the odds were not continuously update it would be possible to exploit that some scenarios and results would be more or less probably with time left in a match.

Ordinary bookmaking companies provide prices that on average a bit too high, ie.

$$Price(1) + Price(X) + Price(2) > 1$$

The prices can be adjusted, depending on how people play, to limit potential losses in a single game. They may still lose money on a single match but will normally profit from their customers in the long run.

There are several Internet betting markets, and more and more people are abandoning traditional bookmakers for better prices. In 1999 Betfair was founded and in 2001 the Betfair betting market was launched and it is the biggest one today [4], present in more than 100 countries and over 950000 customers[5]. So what are the differences when betting on a market instead of using a bookmaker? First it is a matter of whom you are betting against. When placing a bet on bookmaker's website the bookmaker is your counterpart, winning if you lose and vice versa. On a betting market you are buying and selling contracts on a market open for individuals. The betting market works almost exactly as any other market with financial derivatives. The underlying when you buy or sell probabilities in a football match is the result in the match. A bet can be seen as any financial derivative with a fixed maturity, namely when the match is over. The first and biggest betting market is Betfair,

[3]https://svenskaspel.se/?pageid=/sport/oddsetlangen/information
[4] http://www.spelbors.se/allt-om-spelborser/vilka-spelborser-finns-pa-natet/
[5]http://corporate.betfair.com/about-us/betfair-group.aspx

The Betfair live market is a limit order market. A limit order is an order to buy or sell a contract to a specific price or better. It is only executed if the market reaches the limit price, i.e. if someone chose to buy the offered contract.

Imagine a match between Liverpool and Manchester United with Liverpool as the home team. If you believe that Liverpool will win the match, there are at least two ways to win money on the 1, X and 2 market. You can either bet on Liverpool, i.e. buy a contract that pays out if the result is a 1, or you can offer/sell a contracts on that pays out for a X or a 2. This way you will earn the amount for which you sold the contract, if Liverpool doesn't loses or draws.

How do Betfair and other betting markets make money? Betfair does not risk losing money on a specific result as an ordinary betting company would. Instead they take a commission on each winning, a percentage of the winnings. The percentage varies between 2-5%", depending on the volumes you bet. The amount is calculated from your net earnings on a match.

2 The Data

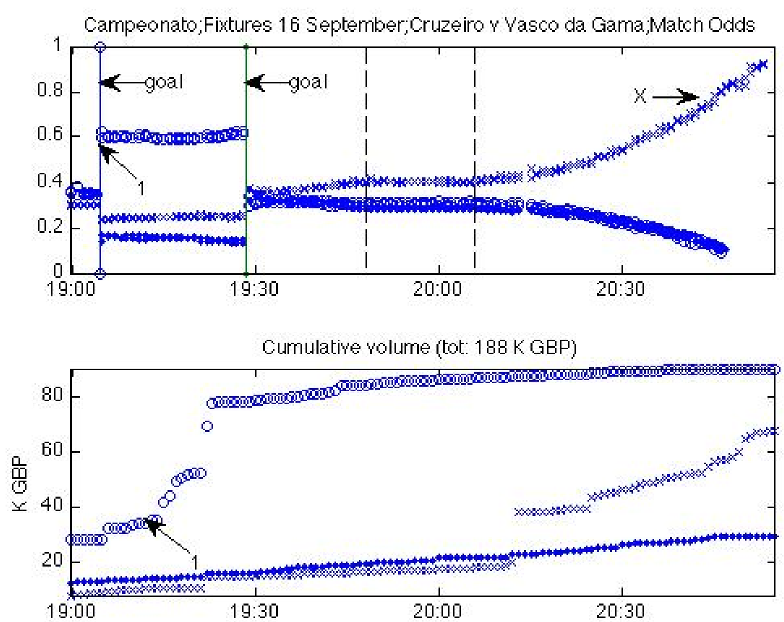

Figure 1: Presentation of a random match.

There is lots of information about football matches available, both pre match and live data. Everything from easily processed data like goal times to more hard evaluated data of injured players. The high frequency data set used in this thesis consists of live prices from 1114 football matches, parsed down from Betfair.com. Every 15:th second the bid and ask prices for 1, X and 2 were stored along with the score, match time and betting volume. A marker for kick off, half time and final whistle were also saved along with data for when the market is closed, for example during a penalty kick. The data was saved in matrices via Matlab making it easy to work with. Figure 1 plots all the data available for one match.

To conclude, for each match the following data was gathered:

Title	Names of teams playing and type of gamed played
Kick off time	Real time for kick off
Half time	Real time for the end of the first half
Half time end	Real time for when the second half begin
Final whistle	Real time for the end of second half
Gathered every 15th second	
Score	The number of goals scored by each team
Bid and ask prices	Bid and ask prices for 1,X and 2
Market status	Boolean vector that representing an opened or closed market

2.1 Market properties

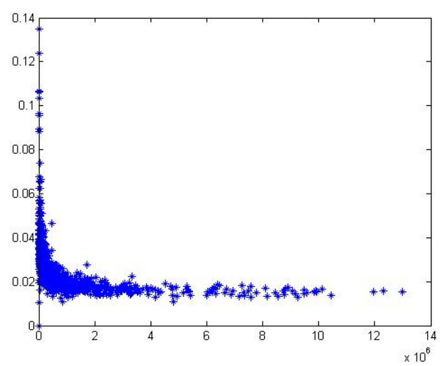

Figure 2: The volumes (k GBP) ploted against the spreads

Prices on financial markets are decided by supply and demand. The current highest price demanded on an underlying product is called the bid price, and the lowest price offered, the ask price. The size of the spread between these prices depends on multiple factors, the most important one- the cash flow in the market. Figure 3 and 2 shows how the spread in the Betfair market decreases with the amount betted/cash flow, which is natural and typical for any financial markets.

When do people live bet? Figure 4 display the amount betted in each minute. Even though half time would be a natural point to make a bet without stress, the plot reveals no visible increase. There is though a stable increase in

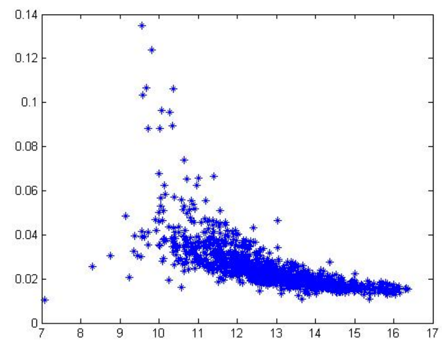

Figure 3: The logarithm of the Volume (k GBP) ploted against the spread

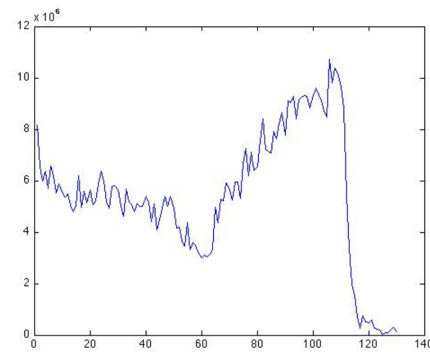

Figure 4: The amount betted in each minute.

volume during the second half, showing that people prefers to bet with lower risk.

Before looking in to live prices, let us look at markets before matches. The market for 1, X and 2 open long before a match so when a match just started the prices are quite stable, especially for mathes with big cash flows.

How about the prices for outturns in the sample space 1, X, 2, are they on average fair or unbiased? Figure 5, 7 and 6 are histograms over the payout-frequency for 1, X and 2 sorted by the price of the contracts. The x axis represents price intervals, $0 - 0.1, 0.1 - 0.2, ..,$ and the y axis the observed win frequency for that interval. For prices of 0.5 to be fair they should on average payout 50% of the time. The staples should produce a linear curve representing that the probability to win increase with the price, which is clearly seen in the pictures. The small divergences can be explained by the interval partition, leaving some intervals rather small. It is no big generalization to say that the time zero prices are unbiased.

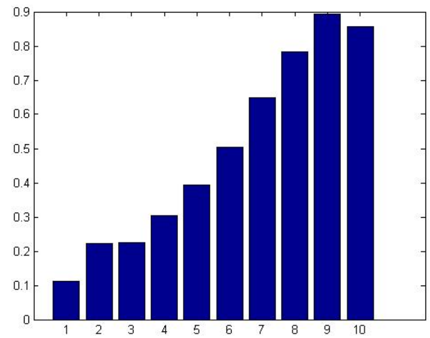

Figure 5: The payouts for 1

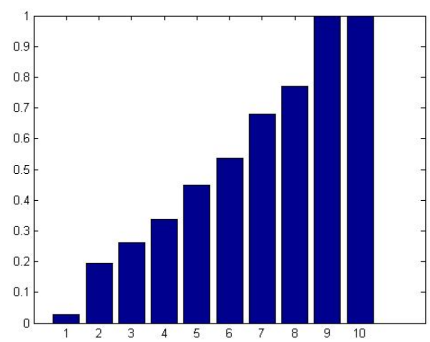

Figure 6: The payouts for 2

The probability for the outturns 1, X, 2 follows the relation:

$$P(1) + P(X) + P(2) = 1.$$

How about the prices? Imagine that the ask prices $price(1) + price(X) + price(2) < 1$, then there is an arbitrage situation. For example, if the ask prices are $price(1) = 0.4, price(X) = 0.3$ and $price(2) = 0.2$, then buy all the three contracts . Independently of outturn you would retrieve one unit and only have spent 0.9 on the portfolio.

The same arbitrage possibility holds if the bid prices

$$price(1) + price(X) + price(2) > 1$$

, except that you then sell the three contracts.

In the 1114 matches these arbitrage possibilities never occur. Plot 8 show the sum of the bid and ask prices for one match, revealing that the sum of the ask prices never goes under 1 and the bid prices never goes over.

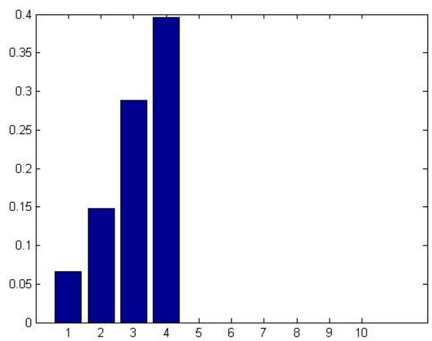

Figure 7: The payouts for X

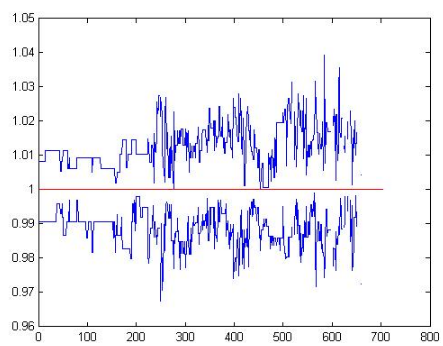

Figure 8: The sum of the bid and ask prices for 1, X and 2 during a match, showing that the ask prices never goes under 1 and bid prices never goes over.

Many betting markets and bookmakers buy and sell derivatives on the same underlying, the outcome in a specific football game. So even if the arbitrage relations are respected in the Betfair market, one can combine and by a portfolio from different markets/bookmakers. There are even companies that are on the lookout for such events, [6].

3 The model

When creating a model for live prices there are some main properties to strive for. If the model later will be used calculate prices in real time, it has to be sufficiently fast, especially if it will be incorporated in a trading strategy. The model should also be unambiguous meaning that some input parameters should

[6] www.oddsportal.com/sure-bets/

create unique prices for that condition. The pricing model has to manage all possible situations, not leaving some special conditions that are impossible to price. Some of the dynamics affecting the result is very hard to qualitative analyze and model, so using data that are easily interpreted is also desirable.

3.1 Poisson pricing model

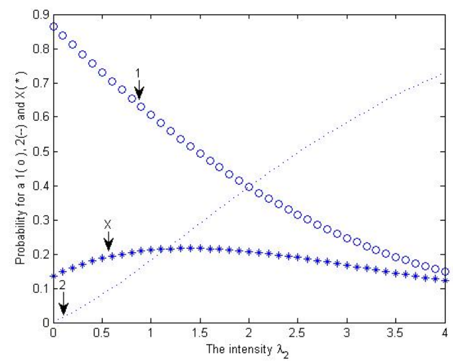

Figure 9: The probability for 1, X and 2 for $\lambda_1 = 2$ and $\lambda_2 \in [0, 4]$

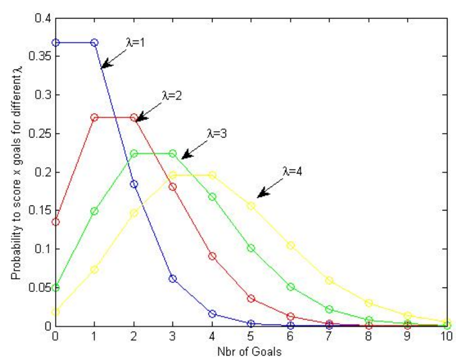

Figure 10: The probability to score x goals plotted for different λ

Imagine a football match between $team_1$ and $team_2$. It is common to model the result of a match as the realization of two independent Poisson distributed random variables, with parameters λ_1 and λ_2, respectively. Then the number of goals scored by $team_1$ has the probability mass function

$$\mathbb{P}(\text{team}_1 \text{ scores i goals}) = e^{-\lambda_1} \frac{\lambda_1^i}{i!}, \text{ for } i \in \mathbb{N}.$$

The probability mass function only depends on the intensity λ_1. If the intensity is defined as number of goals over 90 minutes, figure 10 displays the likelihood for x goals for different intensities.

Since two teams are assumed to score independently of one another, it is easy to calculate the probability for a specific result. The probability for the outcome $i-j$ (i.e. $team_1$ scores i goals and $team_2$ scores j goals) is given by:

$$\mathbb{P}(\text{result } i-j) = e^{-\lambda_1-\lambda_2}\frac{\lambda_1^i\lambda_2^j}{i!j!}$$

The probability for a match winner, $\mathbb{P}(\text{team}_1 Win)$ is the sum of the probabilities for all combinations where team_1 has scored more goals than team_2,

$$\mathbb{P}(\text{team}_1\text{ win}) = \sum_{i=1}^{n}\sum_{j=0}^{i-1}\mathbb{P}(\text{team}_1\text{ score i goals})\mathbb{P}(\text{team}_2\text{ score j goals}),$$

and the probability for a draw,

$$\mathbb{P}(\text{Draw}) = \sum_{i=0}^{n}\mathbb{P}(\text{team}_1\text{ and team}_2\text{ score i goals}) = e^{-\lambda_1\lambda_2}\frac{(\lambda_1\lambda_2)^i}{i!}.$$

The probability for 1, X and 2 varies and depends only on the two intensities. Figure 9 displays the three probabilities for a fixed λ_1 and a varying λ_2.

So the fair prices for 1, 2 and X can be calculated using a function with only the two intensities as input.

$$f(\lambda_1,\lambda_2) \to (\text{Fairprice1}, \text{Fairprice2}, \text{FairpriceX}) = (\mathbb{P}(\text{team}_1\text{ win}), \mathbb{P}(\text{team}_2\text{ win}), \mathbb{P}(\text{draw}))$$

This is a well known method to calculate the fair price for 1, X and 2, which Maher suggested already in 1982, see [2], [3] or [1] . In the function f above, $[0,\mathbb{N}_0]^2 \to [0,1]^2$, for every pair of intensities, λ_1 and λ_2 the prices for 1, X and 2 exist and are unique [7].

3.2 Poisson live pricing model

Let team_1 and team_2 score like two independent Poisson processes s_1 and s_2 with for now known scoring intensities λ_1 and λ_2 per 90 minutes.

The number of goals scored at minute $t \in [0,90]$, $s_1(t)$ and $s_2(t)$ are then Poisson distributed with expected value,

$$E[s_k(t)] = \frac{t\lambda_k}{90}, t\in[0,90], k=1,2,$$

The fair price for the outcome 1, $(s_1(90)) > s_2(90)$, at match minute t is :

$$price(t,x,y) = \mathbb{P}(s_1(90) > s_2(90)|s_1(t)=x, s_2(t)=y),\ \text{for } t\in[0,90],\ \text{and } x,y=0,1,2,\dots.$$

[7] http://mathoverflow.net/questions/138768/does-px-1x-2-and-px-1-x-2-where-x-1-and-x-2-are-independent-and-po

The probability is conditional on the current score leading to that the probability for a team to score decreases with time left even if the intensity is constant, see figure 11. Here wee see that the probability to score 1 goal is increasing in the beginning since $\lambda = 2$ and the team does not score. The distribution is without "memory" , but at a certain point there is not enough time left so the probability to score 1 goal is decreasing even if $\lambda = 2$.

The probability that $s_1(90) > s_2(90)$ can be calculated in the same manner as in the previous section with the respect to the current score and the time decreasing probability for goals.

$$\mathbb{P}(s_1(90) > s_2(90) | s_1(t) = x, s_2(t) = y), \text{ for } t \in [0, 90], \text{ and } x, y = 0, 1, 2, \ldots.$$

The function providing the fair live prices is still unique and exists for every pair of constant intesities λ_1 and λ_2:

$$f(\lambda_1, \lambda_2, t) \to (Fairprice1(t), Fairprice2(t), FairpriceX(t)), \lambda_1, \lambda_2 \geq 0, t \in [0, 90].$$

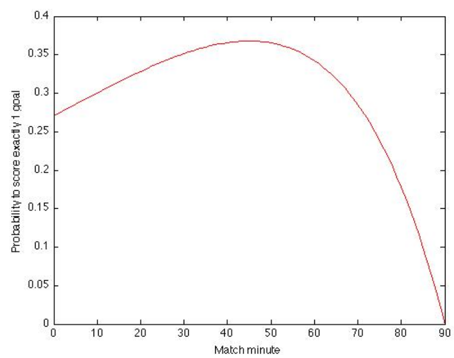

Figure 11: The probability for a team with intensity $\lambda_1 = 2$ to score exactly 1 goal throughout a match given that they have scored 0 goals so far

3.3 Intensity calibration

Do teams really score independently of each other? Intuitively the motivation to score depend on the opponent current score, but a simple control of the 1114 matches, reveals a correlation of -0.0297 and a p-value of 0.3249. The correlation is not significant, so the motivation to score may increase but not the actual intensity. Using that team$_1$ and team$_2$ score with the stochastic variables $X_1 \in Po(\lambda_1)$ and $X_2 \in Po(\lambda_2)$, how can the constant intensities be estimated? One opportunity is to look at historic data and for example let λ_1 be estimated by the average number of goals team$_1$ scored in their last 10 games. This method is logic but not valid, it does not use the dynamics between two specific teams and requires historic data.

The time zero prices from Betfair were earlier shown to be relatively unbiased. Since the prices are fair they correspond to probabilities on different outturns. Therefore they make excellent input to model the intensities. Using that the teams score as described in the last section, a desirable function is;

$$g(price_1(t_0), price_2(t_0)) \to (\lambda_1, \lambda_2)$$

where

$$f(\lambda_1, \lambda_2) = (price_1(t_0), price_2(t_0).$$

Since the intensity is constant through the match, λ_1 and λ_2 contains all information needed to calculate the fair prices throughout a match.

The function f is invertible, so these functions can be implemented numerically in Matlab. The intensities are modeled so that the probabilities for team 1 and 2 to win equal the prices at Betfair just before the game starts which is done through brute force search in a grid of Poisson intensities. The grid is created by calculating probabilities P(1) and P(2) for every combination of intensities $\lambda_1 = 0, 0.01...3$ and $\lambda_2 = 0, 0.01...3$ Lets say that $ObservedPrice_{t0}(1) = 0.3$ and $ObservedPrice_{t0}(2) = 0.4$. A search through the grid tells which intensities that gives the matching probabilities, ie. when $abs(P(1) - 0.3) < 0.01$ and $abs(P(2) - 0.2) < 0.01$. The grid is only calculated once before hand which makes the look up very fast and Price(X) is calculated by $Price(X) = 1 - Price(1) - Price(2)$.

When implementing the Poisson technique numerically, a value for n have to be chosen, when adding probabilities. The probability for a team to score more than 10 goals is almost zero with realistic intensities. So without generalizing too much n=10 have been used in this implementation.

4 Model evaluation

There are several important questions when evaluating the Poisson live model. Are the price movements realistic? And what about the constant scoring intensity, what are the effects on prices? The perfect scenario would be for it to predict results better than the market, and then use it to make a profitable trading algorithm. To compare the movements with the observed prices some case studies are presented below.

4.1 Case studies

Figures 12 and 13 visualize the Betfair prices together with the model prices for two 0-0 games. The Poisson model handles the time decay in prices well, in the sense that they follow the market. The price for X goes against 1 as time passes, representing the likelihood for a draw, 0-0 game, goes up for each second with no goals. The models prices are of course fixed during half time, but the players do not seem to change their minds particularly either. Even with one goal the model seems to agree with the players, see figure 14. It handles the crucial price jumps after the goal very good; it is almost exactly as the observed prices. The prices after goals can be troublesome though. In a match with many goals it becomes obvious, like in 15. The model values the goals differently than the players.

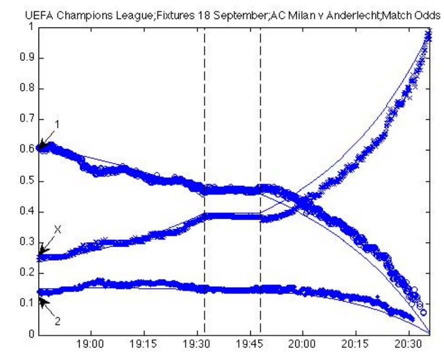

Figure 12: Example with no goals

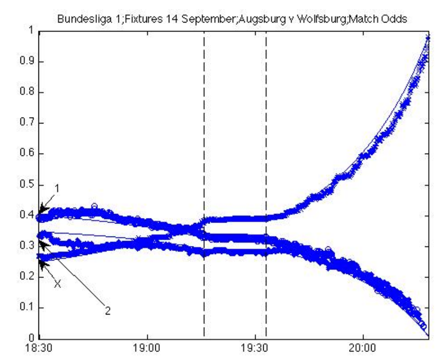

Figure 13: Another example with no goals

The Poisson model only uses time past and the current score as price changing parameters, leaving some factors unaccounted for. Figure 16 shows that the model doesn't adjust for a red card as the players on Betfair do.

4.2 The price jumps after goals

The most spectacular and crucial events when pricing outcomes in a football game are the price jumps after goals. Depending the current score and time of a goal, it can be decisive or almost meaningless. A 4-0 goal in the 89 th minute will not affect the prices for 1,X or 2 very much. But an outcome-changing goal

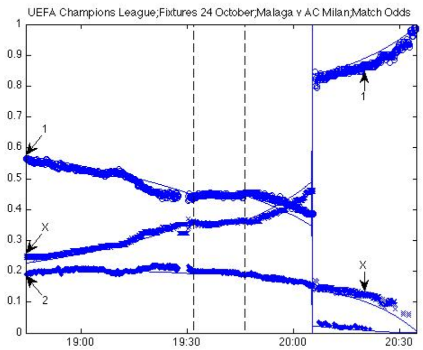

Figure 14: Example with one goal

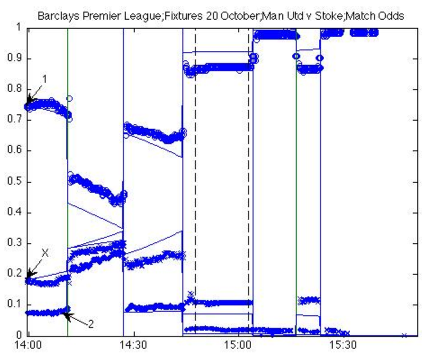

Figure 15: Example with many goals

late in the match leads to very dramatic changes. How do the model value goals compared to gamblers? Figure 17 reveals that the model is over shooting for both negative and positive big jumps, compared to observed prices. By plotting the jumps, without goals in the last fifteen minutes, proves that most of the big jumps are because of late goals, see figure 18. For smaller jumps they seem to agree more.

4.3 Is the intensity really constant

In the Poisson model, the probability for a team to score only depends on time past and the scoring intensity remains the same throughout a game. How does that cope with reality? Plot number 20 show the distribution of all goal times. The high staple to the right represents all overtime goals. Without them, the relative frequency intensity seems to constantly increase throughout the match, see plot 19.

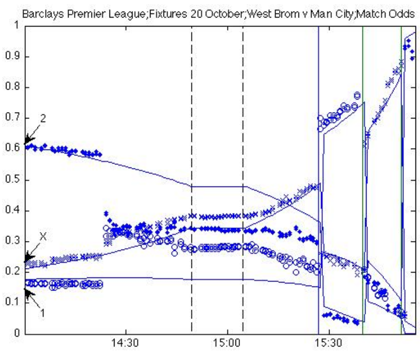

Figure 16: Example with prices not being adequate due to a red card

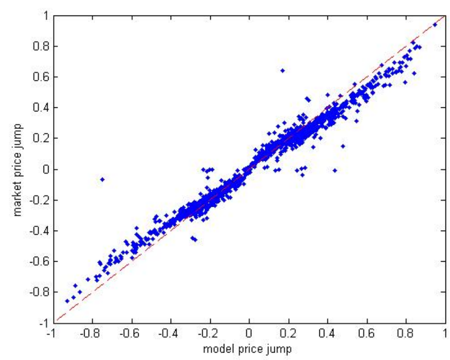

Figure 17: The size of the observed price jump for 1 after goals plotted against the size of the price jump by the Poisson model

With the assumption that the goals are Poisson distributed, the times between them are exponentially distributed. To investigate the intensities further figure 22 and 21 clearly shows that there is another goal intensity when the score is 0-0. It is very intuitively that the teams use the first few minutes to warm up. The likelihood estimates for the scoring intensities are a goal every 39.68 minutes when the score is 0-0 and once every 31.95 minutes for other score states. The likelihood estimates includes not only the time between goals but also the observation that there was no goal between the last goal in the match and final whistle. Using these intensities to calculate probabilities for 0-0 matches we get:

	Expected time to first goal 39.68 min	*Expected time to following goals 31.95 min*
Estimated probability for a 0-0 match	5.28%	9.36%

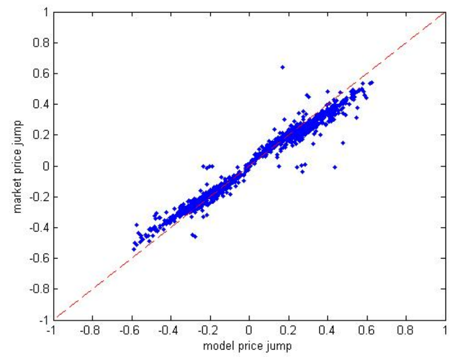

Figure 18: The size of the observed price jumps without goals made in the last 15 minutes.

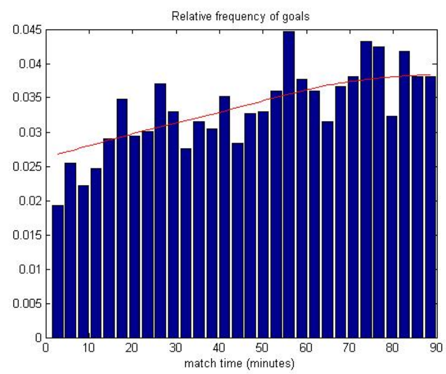

Figure 19: Goaltimes without overtime goals smooth down

The actual observed part of 0-0 matches is 8.06% which is between the two intensities above.

Another hypothesis is that if the score difference is only one the moral to score would be higher, resulting in more goals for that state. Figure 23 reveals however no clear sign of this.

4.4 Pricing in live matches and an analogy with Black and Scholes

When creating a pricing model for a financial market it is natural to glimpse at the Black and Scholes market. They described a market where the underlying, the stock, can be modelled as a Brownian motion. This led to that derivatives, like vanilla options, could be priced deterministically.

Given the spot price on a stock together with the volatility, the Black and

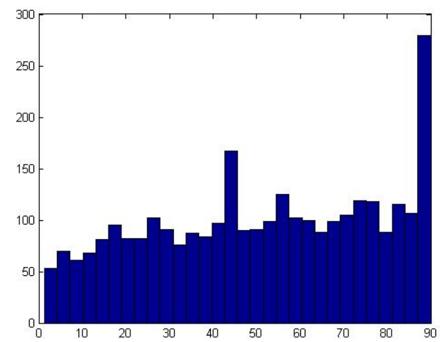

Figure 20: Goaltimes with all goals after match minute 90 in one bar.

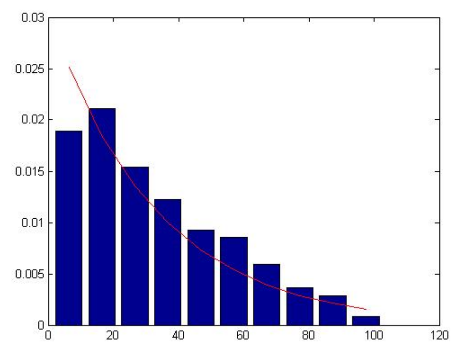

Figure 21: Distribution of minutes to the first goal, together with the corresponding exponential curve

Scholes formula will give you the fair price for a vanilla option with strike price K, time to maturity t and with a interest rate r. For example the price of a European call option is given by:

$$C(S,T) = N(d_1)S - N(d_2)Ke^{-r(T-t)}$$

When pricing derivatives in the Black and Scholes framework one talk about implied volatility. Implied volatility is the volatility that when calculating the derivative using the pricing model gives the current market value. The Poisson model for live prices in football matches does not have a clear volatility parameter. Volatility is a measure of variations of the underlying. The underlying in a football match is the score, which in the Poisson model is decided by the scoring intensity. So the analogy from the Black and Scholes volatility is a function of the two teams goal intensities.

When pricing derivatives in the Black and Scholes model with strike prices

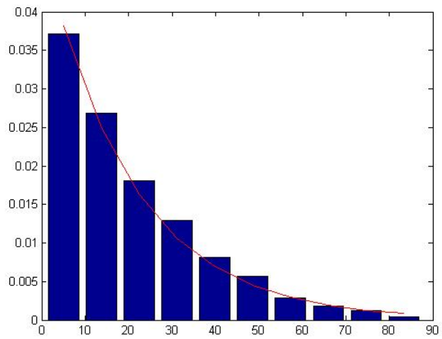

Figure 22: Distribution of minutes to the next goal except the first one, together with the corresponding exponential curve

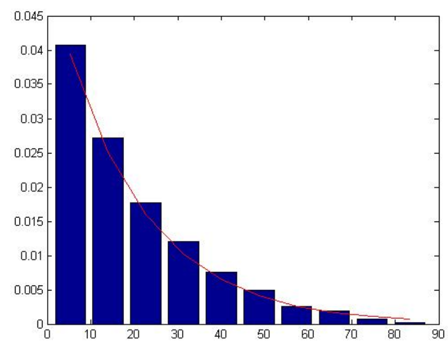

Figure 23: Distribution of minutes to the next goal when the goal difference is one, together with the corresponding exponential curve

far from the spot price, the market and model price differ a lot, and creates what is known as a volatility smile. The analogy for the Poisson model is the price jumps after goals, where it have been seen that big jumps are harder to handle.

The no arbitrage assumption in the Black and Scholes market is also respected in at least the Betfair market.

Another interesting remark about the live odds prices is that the un-scaled prices equals their probabilities, $P = Q$. It is both convenient but also a bit confusing in long thinking strains about results.

In the Black and Scholes market the underlying is modelled as a geometric Brownian motion. Maybe it would be possible to model the scoring intensities throughout a game in a similar manner, instead of letting them be constant.

4.5 Pricing differences

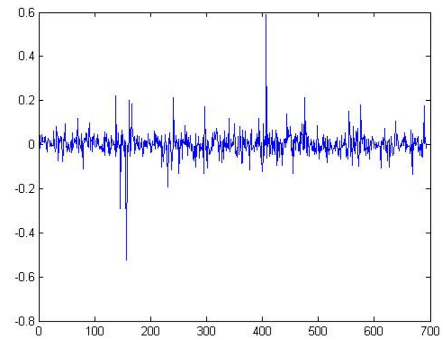

Figure 24: The difference between observed price and the Poisson price given for random points in random matches

The case studies are an indication that the Poisson model is good. But how does it differ from the observed prices? Taking the absolute differences between observed prices and Poisson prices gives a mean difference of 0.03 or 10%. Figure 24 represent differences for some random points. The big spikes, from very different pricing, can be due to a red card or different reaction times after a goal. It is not entirely fair to compare movements like this. Slightly different reaction times, can give a big diff for a time point even if the prices have almost identically movements. A better way is to look at how well the Poisson model predicts results compared to the gamblers/market.

5 Prediction and trading strategy

What if the Poisson model gives fair prices and therefore predicts results better than the players. The utopian scenario would be to live profit from buying or selling contracts when the market prices are off. To implement and test a strategy like this there are several technical difficulties to overcome.

A transaction on a betting market takes some seconds to execute. This removes the advantage of faster information about a goal, a red card or anything that is game changing. Whenever a game-changing event occurs all unfilled orders are erased and the market is rebooted or reset. The time that the market is closed differs from time to time. The data set in this paper tells, for every time point, if the market were opened or closed. The data price data are for every 15th second. Preferably this should have been more frequent especially after a goal so that the freezing and resetting of the data could be followed more intensively.

A trading strategy can be constructed in many ways, the easiest way is to buy or sell a mispriced contract and keep it until it expires. Figure 25 is the

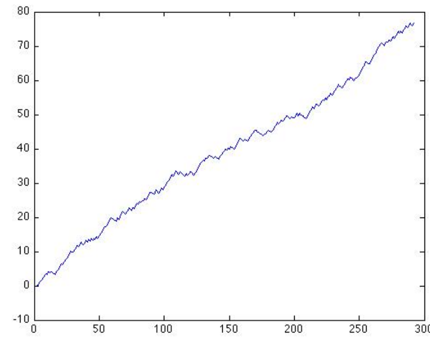

Figure 25: Profit and loss curve when buying one contract per match, not regarding closed markets or transaction times

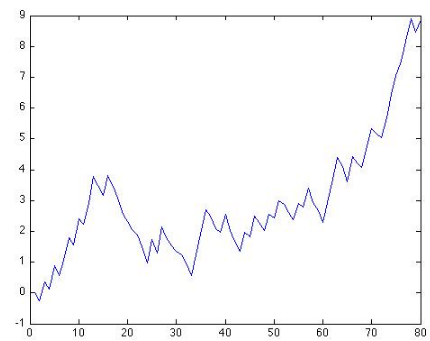

Figure 26: Profit and lost curve when buying one contract per match using closed market but transaction time

profit and lost curve when buying contracts that differs more than 0.3 units in price from the Poisson model. In this case the algorithm buys contracts even in a split even when the market is closed. Most of the transactions are just after a goal before the gamblers have had a chance to stabilize their prices and the market is built up by very few contracts. When taking away the possibility to buy during a closed market, figure 26 still shows a clearly positive PoL curve. But when also demanding a transaction delay on the affairs, all the profitable potential buys during goals are gone, see figure 27.

It could be a good idea to investigate the profitability in the market after goals, since it earlier where shown that the Poisson prices differed from the observed. Again it is a delicate question for how long the transaction time should be simulated. Since the time intervals between information are 15 seconds it is not given that a buy or sell signal from the algorithm can be executed in the

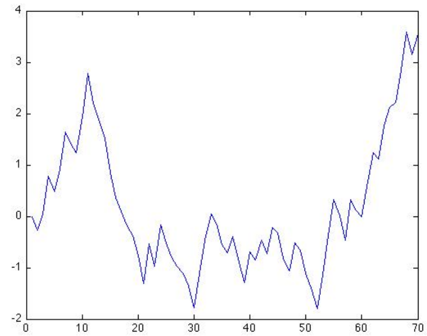

Figure 27: Profit and lost curve when buying one contract per match, with concern for closed markets and transaction times

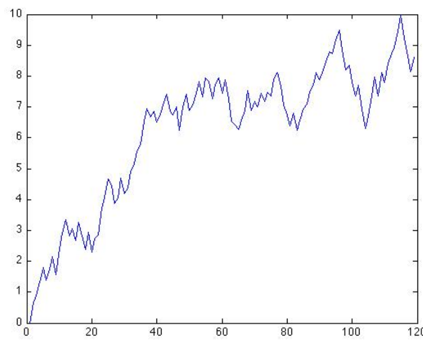

Figure 28: Profit and lost curve when buying contracts after goals, regarding closed market but not transaction time

same time point. Figure 28 represent a scenario when there is no delay from the buy/sell signal to the actual trade. But if the transaction takes one time point the clear rising curve is gone, see figure 29. This shows that the profitable scenario in figure 28 is because of fluctuations in the market after goals. It is possible that this is a potential market but with the data I possess it is hard to be certain. It can also be a hint that the market is quite effective.

6 Conclusions

The live betting markets are still a new phenomenon and the lack of academic research of live prices makes even fundamental investigations interesting. This paper opens up for further research both improving the current model and with more data extend it to more markets than just to bet on 1, X and 2. The Poisson

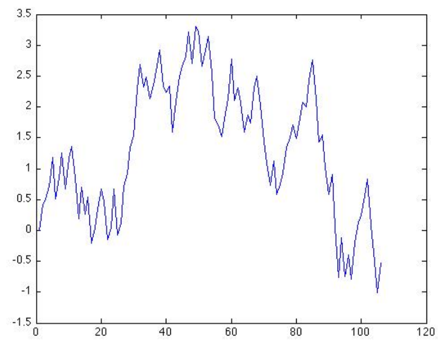

Figure 29: Profit and lost curve when buying contracts after goals, regarding closed market and transaction time

pricing model presented in this paper works well as a benchmark model even with some obvious flaws like the incoherence for a red card. It does not predict results better than the market, or if it does it is too little to be profitable. Perhaps an even better model could be created if the intensities could be re calibrated depending game dynamics The obvious difference in intensities when the score is 0-0 is also something that could be modeled separately. There are other studies of made with a modified Poisson models for example see, [1] or [3] and maybe that is something that can be done with live prices as well.

If more data had been gathered during the match, like ball possession or corners, even more opportunities would have opened. The possibility to bet on exact result, like studied in [4], is also something that would be interesting to do with live data and see how well the Poisson model performs. The betting markets also provide markets on other sample spaces like the next team to score, and maybe it could be used to hedge bets on 1, X and 2. The attempt to make a trading algorithm left some unanswered questions, and it would have been interesting with more frequent data around the goals.

For obvious reasons, bookmakers keep it a secret how they update their live prices. But it is not unbelievable that they use some kind of Poisson model. Since it is a rather fast pricing method it can surely be used, at least with some modifications.

7 References

References

[1] Mark J. Dixon and Stuart G. Coles. *Modelling Association Football Scores and Inefficiencies in the Football Betting Market* Applied Statistics Volume 46 Issue 2 (1997).

[2] M.J. Maher. *Modelling Association Football scores* Statistica Neerlandica, 36, (1982).

[3] Dimitris Karlis and Ioannis Ntzoufras *Analysis of sports data by using bivariate Poisson models* The Statistician,52, Part 3, (2003)

[4] Michael Cain, David Law and David Peel *The favourite-longshot bias and market efficiency in uk football betting* Scotish Journal of Political Economy vol 47 No1 (2000).

[5] Niko Marttinen *Creating a Profitable Betting Strategy for Football by Using Statistical Modelling* M.Sc. Department of Statistics, Trinity College, Dublin (2001).

[6] Rasmus B. Olesen *Assessing the number of goals in soccer matches* M.Sc. Department of Computer Science, Aalborg University, (2008)