UNIVERSITEIT VAN AMSTERDAM

Master's Thesis

Arbitrage opportunities in sports betting markets

Academic supervisor

Rafael Matta

Author

Len van der Sluijs

Abstract

In this paper I will provide the data and results related to the theoretical model developed in Len van der Sluijs (2013). By using a unique hand-collected database I am able to make inferences on what factors affect the duration and profitability of arbitrage opportunities in sports betting markets. From these results a link can be made to the influences of competition on arbitrage opportunities within the sports betting market. In line with the findings of the model developed in Len van der Sluijs (2013) the results show that an increase in competition is positively related with the duration of arbitrage opportunities in sports betting markets.

May 1th 2015

University of Amsterdam

Table of contents

''By posting relatively high quotes a bookmaker will attract more betting volume on his website. This might enable bettors to construct a combination of bets that seem to provide a risk free profit i.e. an arbitrage bet. I expect that in the future there will still be some arbitrage opportunities, but nevertheless no big money can be made since the bookmakers will impose limits or cancel your account. Arbitrageurs also affect Betfair, but to a lesser extent. We do not ban arbitrageurs, but they do harm us because of the many orders being processed. This is why rules apply for bettors who bet a given number of times within a certain period.''

— Quote out of an interview that I had with Andrew Black, Founder and former CEO of Betfair (The largest betting exchange in the world)

1. Introduction

Over the years many papers have used the betting market to test the efficient market hypothesis (Golec, 1991) (Hausch, 1981) (Pope, 1989) (Vlastakis, 2009). According to Fama (1970) a market is efficient if it is not possible to consistently generate returns in excess of the market on a risk adjusted basis based on historical prices, publicly available or insider information. One way to evaluate the efficiency of a market is to see whether exploiting the ''inefficiency'' could lead to risk adjusted and cost covering returns, i.e. if arbitrage opportunities exist. Previous studies have demonstrated these inefficiencies in the betting market. The reason of this inefficiency could be that bookmakers can strategically post unfair quotes in order to exploit the biases of bettors (Levitt, 2004). When the quotes posted by bookmakers diverge too far from the ''fair'' quotes considerable differences can be found amongst bookmakers. If the variance amongst posted quotes is large enough arbitrage opportunities arise for the unbiased bettors and arbitrageurs.

Although previous studies have addressed the inefficiencies in sports betting markets the question remains whether competition among bookmakers could preclude bookmakers from exploiting the biases of bettors, forcing quotes to reflect the ''fair'' odds. In order to fill this gap in the existing literature I developed a model of competition in which I showed how the existence of arbitrageurs gives rise to persistent arbitrage opportunities in equilibrium (Len van der Sluijs, 2013)[1]. In this paper I will provide the data backing up the theory developed in Len van der Sluijs (2013). By compiling a unique hand-collected dataset over a period of thirty days I was able to follow the details of 590 arbitrage opportunities over time. With the use of a statistical program this dataset allowed me to make inferences on what factors affect the duration and profitability of these arbitrage opportunities. In line with the findings of the model presented in Len van der Sluijs (2013) the results show that an increase in competition is positively related with the duration of arbitrage opportunities in sports betting markets. In

[1] The model can be found in appendix A and the proof of the model can be found in appendix B, C, D, E and F.

contrast with the efficient market hypothesis I also show that the inefficiencies in the betting market do not disappear quickly and can be exploited for long periods of time.

The paper is structured as follows. First a review of the findings of previous literature is presented. Then the mechanics of an arbitrage opportunity in the sports betting market will be discussed. In the third part of the paper the research methodology is presented after which the results will be discussed in the fourth part of the paper. Finally a summary of the findings, along with recommendations for future research, will be presented in the conclusion.

2. Literature review

A number of academic papers on sports betting markets have addressed arbitrage opportunities. Hausch and Ziemba (1990) were among the first to demonstrate the inefficiencies in these markets. In their article, cross-track betting is studied, which was a new form of wagering at the time. Cross-track betting made it possible for bettors to place bets at their (home) track for races that were run at other tracks. Because each track had it's own betting pool, differences in payoffs were not uncommon. This made cross-track betting an interesting market to investigate, because if the differences in payoffs amongst tracks are big enough, arbitrage opportunities might occur. The variance in these payoffs has to be large enough to cover the costs associated with placing the bets and leave room for the bettor to ensure a profit. Hausch and Ziemba (1990) have shown that, in their sample of races with the possibility of cross track betting, differences can be large enough to ensure a profit. The same is presented by Edelman & O'Brien (2004) in their sample of Australian thoroughbred races. In both articles the authors developed an arbitrage model that demonstrates the inefficiencies of the market and in the article written by Hausch and Ziemba (1990) an optimal capital growth model is used to exploit these inefficiencies.

The problem with the arbitrage opportunities presented in these papers is that they can only be identified ex post. This is due to the nature of the betting system, which is called pari-mutual betting. In pari-mutual betting bets are placed in pools from which the bookmakers take their cut and payoffs are determined by the final betting volume placed at each outcome. This makes it impossible to construct a combination of bets that yield a risk free profit ex ante. So for a bettor to successfully implement the model developed by Hausch and Ziemba (1990), he has to learn the odds of each track in the last few minutes of betting, run the model and finally have employees place the optimal stakes. The bettor would have to take on the risk of changing odds in the last few minutes. The effect of changing odds in the last 2 minutes are examined by Hausch, Ziemba & Rubinstein (1981). They found that the expected returns were indeed affected by last minute changes in odds, but in most cases the arbitrage bets would stay profitable. Although considerable evidence is presented in favour of the market being weakly inefficient, the model developed by Hausch and Ziemba (1990) contradicts these findings. Information asymmetry is given as a possible explanation for the differences in payoffs amongst tracks (Hausch & Ziemba, 1990).

In contrast to pari-mutual betting, bettors do have the possibility to achieve risk free arbitrage opportunities by exploiting the differences in quotes in the traditional bookmaker system. Here payoffs are known ex ante, so if bets are placed simultaneously the risk of changing odds affecting profit is eliminated. Multiple articles on sports betting markets cover arbitrage opportunities in the traditional bookmaker system (Vlastakis et al., (2009), Verbeek (2012) and Franck et al., (2012)).

Pope and Peel (1989) examined odds posted by four different bookmakers in the UK during the 1981-1982 season. In their sample, only one arbitrage opportunity was discovered and this opportunity offered a guaranteed return of two percent (Pope & Peel 1989). Pope & Dixon (2004) studied the odds in the UK football betting market in the seasons 1993-1994, 1994-1995 and 1995-1996. In their sample they used the fixed odds posted by three bookmakers. In this sample no combination of bets could yield a positive return. The possibility that bookmakers became more efficient forecasters was brought forward as an explanation for these findings (Pope & Dixon, 2004).

A paper by Vlastakis et al. (2009) contradicts these findings. In this article, a dataset is compiled with closing odds posted by 6 European bookmakers of whom 5 operate online. In this dataset 12,420 matches were examined and in 0.5 percent of all matches a bettor could construct a combination bet yielding a guaranteed profit. If only the online bookmakers were examined the percentage would drop to 0.1 percent. Although the quantity of the arbitrage opportunities that were found might not seem surprising, the magnitude of the returns that could be obtained by combined betting is. Of all arbitrage opportunities found, more than 50 percent would have allowed bettors to achieve returns over 12 percent. A maximum return of more than two hundred percent was found and the average return of the arbitrage possibilities in their data set was 21.78 percent. A possible explanation for this might be that no corrections were made for outliers in their data set, which might occur when bookmakers post the wrong quote by accident (Vlastakis et al., 2009).

On top of the extraordinary returns found by Vlastakis et al. (2009), the quantity of arbitrage opportunities found may also be far less than the actual number of arbitrage bets a bettor could construct. This is caused by the limited number of bookmakers that were included in their data set and to the fact that they only incorporated closing odds. Vlastakis et al. (2009) give a number of explanations for their findings. They argue that a bettor trying to exploit these opportunities does not have to entail a loss for the websites involved, because the websites will try to balance their own books (Vlastakis et al.,2009). This finding is backed by Levitt (2004) & Kuypers (2000), who both mentioned that bookmakers might state a different quote than the ‘’fair’’ quote with the purpose of exploiting betting biases. This would imply that if differences exist between the preferences of bettors active on each website, differences may also occur in the books of these bookmakers (Levitt, 2004) (Kuypers, 2000). Another explanation for the existence of these opportunities given by Vlastakis et al. (2009) is that it might also be the case that bookmakers post odds that attract attention from bettors and arbitrageurs for advertisement reasons. The bookmaker could control the risks involved by imposing limits on bets that can be used for arbitrage (Verbeek, 2012), (Vlastakis et al., 2009).

A paper written by Verbeek (2012) includes no less than 67 bookmakers, whereas Vlastakis et al. (2009) only looked at 6 bookmakers. In his paper, Verbeek (2012) examines 7,225 matches and because an arbitrage bet can only be made with a minimum of 2 bookmakers he deducts the matches for which only one bookmaker offers a wager. The total amount of matches offered by at least two bookmakers comes down to 6,495 and the average number of bookmakers that offered wagers on these matches was 34%. The matches considered were played in January 2011 and January 2012. Verbeek (2012) found an astonishing amount of arbitrage opportunities. In his data set an arbitrage bet could be constructed in 25.8% of all matches. This comes down to a total of 1,789 matches with an average return of 2.32%. It becomes clear that the difference in the amount of arbitrage opportunities found is due to the number of bookmakers included in their datasets. This paper also contradicts the conclusion of Pope & Dixon (2004), who mentioned the improved efficiency of bookmakers as a reason that they did not find any arbitrage opportunities in their data set. The true explanation can probably be found in the number of bookmakers included (Verbeek, 2012).

In 2000 a new form of betting emerged, namely exchange betting. In exchange betting a person would no longer be paid out by the bookmaker in the case of a win, instead the person will be matched with a bettor that is situated on the other side of the contract. This new form of betting also allows bettors to bet against a given outcome, similar to short selling in equity markets. Exchange betting also has implications for the manner in which odds are determined. The odds are no longer determined by a bookmaker, but by a double auction process that matches supply and demand (Franck et al., 2012).

A paper by Franck et al. (2012) examines inter-market arbitrage by compiling a data set using odds posted by 10 bookmakers and one bet exchange (Betfair). Inter-market arbitrage would involve a combined bet at a traditional bookmaker and a betting exchange. They estimate that the existence of betting exchanges has increased the number of arbitrage opportunities. In a sample of 11,933 matches they found 102 intra-market arbitrage opportunities with an average return of 0.9% and 2287 inter-market arbitrage opportunities with an average return of 1.2%.

A number of reasons are brought forward for the large amount of inter-market arbitrage opportunities found in their data set. Their first argument is the manner in which quotes are determined. Previous studies have shown that bookmakers and bettors have different expectations of the outcome of sporting events (Levitt, 2004). This could cause systematic differences in odds between betting exchanges and bookmakers. A second argument is that bookmakers charge higher commission costs than betting exchanges, because bookmakers bear the risk for the bets placed at their website. A third argument is that it is no longer necessary to place bets on all possible outcomes. One could now simply bet with a traditional bookmaker and sell the same bet at a more favourable price at the betting exchange (Franck et al., 2012).

2.1 Steven Levitt

A paper written by Steven Levitt (2004) deserves some extra attention since the underlying idea and the profit function brought forward in his paper are the basis of the model derived in my previous paper. Levitt (2004) postulates that bookmakers simply post a price for a given match and if that price is not equal to the market-clearing price, the bookmaker might be subjected to considerable risks. Well-informed bettors might thus be able to benefit from this incorrect price, which results in losses for the bookmaker.

Levitt then describes three scenarios in which a price setting mechanism would allow bookmakers to achieve profits:

In the first scenario the bookmaker is able to set a price that will balance his books in advance. Here the bookmaker does not need to be able to predict the outcome of any sporting event, he simply needs to have a good prediction of betting behaviour. The bookmaker will ensure a profit by collecting the commission regardless of the outcome.

In the second scenario the bookmaker is consistently better than bettors in forecasting results of sporting events. In this case the bookmaker is able to set a price that ‘’equalizes the probability that a bet placed on either side of a wager is a winner’’ (Levitt, 2004). The bookmaker will not be able to balance the volume on any particular game, but will collect the commission on average. In this situation however, a more skilled bettor will be able to win from the bookmaker on average, contrary to the first situation.

In the third scenario the bookmaker is not only more sophisticated in forecasting the outcome of sporting events, but is also good in anticipating betting behaviour. Here, the bookmaker will consistently set a price, different to the market-clearing price, with the purpose of exploiting betting biases. The bookmaker will now be able to skew the odds against the team in favour. However the bookmaker should not skew the odds too much since skilled bettors might be able to recognize the correct price and then will be able to ensure profits when prices diverge too far from the correct price.

Levitt analyses a data set of 20,000 wagers placed by 285 bettors on the American National Football League with one bookmaker. He examines his data set in a unique way since he looks at the prices and quantities of the bets placed instead of simply looking at the prices of each bet. By doing so he can see whether the bookmaker is either trying to balance his books or accepts risk on a single match. In addition, Levitt is able to monitor the betting behaviour of the bettors since he has access to the bets placed by each of the 285 bettors in his data set. This enables him to find out if some bettors are able to beat the bookmaker on a regular basis.

From this analysis, Levitt (2004) is able to draw the following conclusions. His first finding is that the examined bookmaker does not try to balance his books, since at least two thirds of the bets are on one side of the wager in 50 % of all games examined. His second finding is that bookmakers appear to exploit betting biases by skewing the odds against the team in favour. Finally, the examined bettors seem to be unable to beat the bookmaker on a regular basis. Although his data set consists of wagers posted by only one bookmaker, he expects that the results can be generalized to the entire market.

In addition to the analysis of his dataset, Levitt (2004) tries to show how the traditional bookmaker system works by developing a profit function.

The profit function of a bookmaker is stated as follows:

$$E(Bookmaker\ profit) = [(1-P)F + P(1-F)](1+V) - [(1-P)(1-F) + PF].$$

Where $P =$ the probability that the favourite wins, $F =$ the fraction of dollars bet on the favourite, V = the bookmaker's commission (the vig) which is only paid on losing bets.

From this profit function it follows that the terms in the left set of brackets produce the case in which the bookmaker wins. This amount is multiplied by $1+V$, which is the bookmaker's commission. The terms in the right set of brackets define if the bookmaker loses and has to pay the winning bets placed by the bettors.

Taking all this into consideration, the profit function can be simplified to:

$$E(Bookmaker\ profit) = (2+V)(F+P-2PF) - 1.$$

This makes clear that the bookmaker's profit is simplified to $V/2$ in the cases that the bookmaker makes sure that either the probability that the two teams win or the money bet on each team is equal. In these cases the bookmaker would eliminate his risks and simply collect the commission.

''In reality the fraction of money bet on the favourite is a function of the probability that the favourite actually wins, i.e. $F = F(P), \frac{\delta F}{\delta P} > 0$'' (Levitt, 2004).

When the derivative is taken of the bookmaker's profit function with respect to P, the formula for the optimal P is obtained.

$$[1 - 2F(P)] + \frac{(1-2P)\delta F}{\delta P} = 0.$$

The benefit the bookmaker would obtain by skewing the odds if bettors do not respond to price changes is represented by the term in brackets. The right side of the formula shows the effect on earnings of bettors who switch to the team with better than fair odds.

By skewing the odds of the favourite the bookmaker can offer less than fair odds and still attract more than 50% of the bets on the favourite. This will allow the bookmaker to obtain profits higher than simply having $P = 0.5$.

The bookmaker should not have P too far away from $P = 0.5$ since it will become more costly for him if bettors switch from the favourite to the underdog. Besides this a price too far away from $P = 0.5$ would allow skilled bettors to exploit the skewed odds and achieve positive expected profits. When determining the distance from $P = 0.5$ it should be taken into consideration that due to the vig a bettor has to win 52.4% of all bets to make a profit (Levitt, 2004).

2.2 Ben R. Marshall

In addition to the paper written by Steven Levitt (2004) the research of Ben R. Marshall (2009) also deserves some extra attention since his research is closely related to the research being conducted in this thesis. In his paper ''How quickly is temporary market inefficiency removed'' Ben Marshall (2009) investigates data from a subscription based program that scans the quotes offered by 50 bookmakers for arbitrage opportunities using proprietary software, similar to the program used in this paper. The data examined in his paper is from the period January 2003 to December 2005. After removing arbitrage opportunities that were for the exact same bet with at least one of the same bookmakers 509,679 arbitrage opportunities remained in the dataset. By collecting the specifics of each arbitrage bet and information of the bookmakers involved in these arbitrage opportunities he was able to show statistically what the determinants were for the profit and duration of these arbitrage opportunities. He finds a median arbitrage profit of 1.5% with a median duration of fifteen minutes. Furthermore he finds that 75% of all arbitrage opportunities disappear within fifty minutes and that the longest a single arbitrage bet could be exploited was slightly under a day. Although I use a similar dataset the results found in this thesis, with respect to the duration of the arbitrage opportunities, differ substantially. The longest a single arbitrage bet could be exploited in my dataset is approximately twenty-eight days.

3. The mechanics of an arbitrage bet

Before presenting the data it is of importance to understand the mechanics behind an arbitrage bet. An arbitrage opportunity only arises when the variance between the quotes posted by the bookmakers is large enough. The arbitrageur has to place bets on each outcome of the event in order to obtain a risk free profit. It is obvious that when bettors were to bet on all outcomes at a single bookmaker they would incur a loss. So in order to exploit an arbitrage opportunity an arbitrageur has to find a set of quotes that makes it possible to obtain a risk free profit no matter what the outcome is. As soon as the arbitrageur has found quotes from different bookmakers for the same sport event that vary enough, the arbitrageur can calculate the stakes and check if he has truly found an arbitrage opportunity. This process is described below.

The arbitrageur first has to check whether the combination of bets will yield a risk free profit. This can be done by calculating the amount the arbitrageur has to wager in order to obtain one dollar. If the amount the arbitrageur has to wager exceeds one dollar no arbitrage opportunity exists. If the amount that needs to be wagered is less than one dollar the arbitrageur can obtain the difference as a risk free profit. In order to present this theoretically I need to define the following variables:

TAW = Total Amount Wagered.

X_i = The quote posted by a bookmaker on outcome i.

n = The number of possible outcomes of the sport event.

The total amount wagered is defined as follows:

$$TAW = \sum_{i=1}^{n} \frac{1}{X_i} \tag{1}$$

The total amount wagered has to be less than 1 in order for there to be an arbitrage opportunity. If this is the case the profit then can be calculated as follows:

$$\text{Profit} = \frac{1 - \sum_{i=1}^{n} (\frac{1}{X_i})}{\sum_{i=1}^{n} (\frac{1}{X_i})} \qquad \text{Where } \sum_{i=1}^{n} (\frac{1}{X_i}) < 1 \tag{2}$$

The stakes to place on each possible outcome of the event are determined as follows:

$$\text{The stake on outcome } i = \frac{(\frac{1}{X_i})}{(\frac{1}{X_1}) + (\frac{1}{X_2}) + \cdots + (\frac{1}{X_n})} \tag{3}$$

The formulas presented above are obtained from a paper written by Ben Marshall (Marshall, 2009). These are the formulas that are used most often in order to calculate the stakes that need to be placed on each outcome to exploit an arbitrage opportunity.

I will now demonstrate how this process works in practice by examining a randomly selected arbitrage opportunity. I will first show that no arbitrage opportunity exists when a bettor were to place bets on all outcomes with one bookmaker. Secondly, the stakes and the profits will be determined for the same match by combining quotes of different bookmakers. Please note that combining quotes from different bookmakers only results in a risk free profit if the variance between the quotes posted by these bookmakers is large enough.

The sport event considered in this example is a soccer match in the women's Toppserien of Norway. The match is scheduled to take place on the 30th of April at 18:00 pm Central European Summer Time (CEST). The opposing teams are Røa and Kolbotn. Let's first consider the quotes posted by one bookmaker, namely Mybet.

Table 1. Røa – Kolbotn 01-05-2015, 19.00pm

	Over 2.5 goals	Under 2.5 goals
Mybet	1.65	2.30

Table 1: In this table one can find the quotes posted by Mybet for the soccer match between Røa and Kolbotn. In the first row the outcomes of the match are illustrated. In the next row one can find the bookmaker and the corresponding quotes. As shown in the table the amount that a bettor would place will be multiplied by 1.65 if more than 2.5 goals are scored in this match. In the case that less than 2.5 goals are scored the bettors amount will be multiplied with 2.3.

When a bettor places bets on each outcome at Mybet the total amount wagered would come down to:

$$TAW = \sum_{i=1}^{n} X_i = \frac{1}{1.65} + \frac{1}{2.30} \approx 1.0408$$

Since 1.0408 is larger than 1 no arbitrage opportunity exists when a bettor places bets on both outcomes of the event.

Now consider the quotes posted by two bookmakers, namely Mybet and Cashpoint.

Table 2. Røa – Kolbotn 01-05-2015, 19.00pm

	Over 2.5 goals	Under 2.5 goals
Mybet	**1.65**	2.30
Cashpoint	1.35	**2.95**

Table 2: In this table one can find the quotes posted by Mybet and Cashpoint for the soccer match between Røa and Kolbotn. In the first row the outcomes of the match are illustrated. In the next two rows one can find the bookmakers and the corresponding quotes. The quotes that are used to construct the arbitrage bet are in bolt.

The arbitrageurs will now only use the most favourable quotes for each outcome. The total amount wagered would then come down to:

$$TAW = \sum_{i=1}^{n} X_i = \frac{1}{1.65} + \frac{1}{2.95} \approx 0{,}945$$

Since 0.945 is less than 1 an arbitrage opportunity arises by combining the most favourable quotes of each bookmaker. By doing so the arbitrageur will ensure a profit no matter what the outcome of the sport event is. The profit is calculated as follows.

$$\text{PROFIT} = \frac{1 - \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i} = \frac{1 - (\frac{1}{1.65} + \frac{1}{2.95})}{\frac{1}{1.65} + \frac{1}{2.95}} \approx 0.0582 \approx 5{,}82\%$$

By betting with both bookmakers the arbitrageur is able to achieve a risk free return of 5,82%. In order to obtain this return the arbitrageur has to calculate the stakes to bet on each outcome. The stakes are determined as follows:

$$\text{Stake to bet on over 2.5 goals} = \frac{\frac{1}{1.65}}{\frac{1}{1.65} + \frac{1}{2.95}} \approx 0.6413 \approx 64{,}13\%$$

$$\text{Stake to bet on under 2.5 goals} = \frac{\frac{1}{2.95}}{\frac{1}{1.65} + \frac{1}{2.95}} \approx 0.35869 \approx 35{,}87\%$$

By placing 64,13% of the total betting amount on over 2.5 goals and 35,87% on under 2.5 goals the arbitrageur is able to obtain a risk free profit of 5,82% no matter what the outcome of the soccer match is. In table 3 below a summary of the process is presented.

Table 3. Røa – Kolbotn 01-05-2015, 19.00pm

	Over 2.5 goals	Under 2.5 goals
Mybet	**1.65**	2.30
Bet365	1.57	2.35
Tipico	1.60	2.20
Cashpoint	1.35	**2.95**
Stake	0.6413*1000 = 641.30	0.3587*1000 = 358.7
Revenue	641.30*1.65 = 1058.145	358.7*2.95 = 1058.165
Profit	1058.145-(641.3+358.7)= 58.145	1058.165-(641.3+358.7)=58.165
ROI	58.145/1000=5.81%	58.165/1000=5.82%

Table 3: In this table one can find the quotes posted by Mybet, Bet365, Tipico and Cashpoint for the soccer match between Røa and Kolbotn. In the first row the outcomes of the match are illustrated. In the next 4 rows one can find the bookmakers and the quotes that they offered for these bets. The quotes which are used to construct the arbitrage bet are in bolt. In the last 4 rows the stakes for each bet are calculated and their respective returns are shown.

4. Data

4.1 Data source

The data has been hand collected from a program called Rebelbetting. Rebelbetting is a Swedish subscription paid software program that uses proprietary software to identify arbitrage opportunities across bookmakers in the sports betting industry. Full membership is available in four different subscriptions: 39 euro's for a one week trial; 129 euro's for one month; 499 euro's for six months and 799 euro's for one year. Rebelbetting was founded in 2007 by ClarobetAB and currently has users from over 120 countries.

In order to obtain the data a one-month membership was purchased, granting me access to all arbitrage opportunities detected by the software for 32 days. The program can be downloaded from the website www.rebelbetting.com and can be used after entering a username and password. The program shows all arbitrage opportunities detected by the software that are currently available in the market. By clicking on a specific arbitrage opportunity the details of the arbitrage bet are revealed. The program also offers a feature that calculates the exact stakes that need to be placed on a specific bet in order to exploit the arbitrage opportunity. If the arbitrageur decides that the opportunity is worth exploiting the program can open the exact pages of the involved bookmakers on request such that the arbitrageur can immediately place his bets within the scope of the program. Rebelbetting supports nine types of sports, which are monitored from 52 bookmakers. A list of the sports and bookmakers is provided in appendix G and H.

4.2. Data collection process

The data has been hand collected twice a day over a period of thirty days. At any moment in time twenty arbitrage opportunities were monitored. The goal was to constantly have a variety of sports and potential profits in the dataset. On the 27th of January 2015 at 12.00pm twenty arbitrage opportunities were selected. From the moment the arbitrage opportunities entered the dataset they were tracked in order to see what happened with these arbitrage opportunities over time. That same day at 18.00pm the previously entered arbitrage opportunities were looked up in the program to see whether they still existed and whether the bookmakers had revised their quotes. Some arbitrage opportunities vanished by that time and were subsequently dropped out of the dataset. The remainder of the arbitrage opportunities were still available and hence could still be exploited. The details of these arbitrage opportunities were then entered in the dataset in order to examine how these arbitrage opportunities change over time. After dropping out the arbitrage opportunities that vanished and tracking the remainder of the previously entered twenty arbitrage opportunities the data set was filled with new arbitrage opportunities until there were twenty exploitable arbitrage opportunities in the dataset again.

For a period of thirty days these steps were followed to obtain the full dataset. Every day at 12.00pm and 18.00pm the twenty arbitrage opportunities that were still exploitable in the dataset were tracked until they left the dataset. The arbitrage opportunities could either leave the dataset when the opportunity vanished or when the specific event had taken place. The number of arbitrage opportunities that had left the dataset were replaced by new arbitrage opportunities in order to keep the amount of arbitrage opportunities being tracked at twenty.

Monitoring the arbitrage opportunities and updating the dataset took about three hours per day (an hour and a half per time).

4.3 Variables

During the collection period the specifics of each arbitrage opportunity were entered and monitored in the dataset. By doing so the following list of variables were created: date of entry, time of entry, time until last betting opportunity (in minutes), type of sport, type of bet, league, teams/players, bookmakers, quotes, profit (in percentages, i.e. 8% or 0.5%), age (in minutes), # of times, change of profit, better option, change with better option and only available with different bookmaker. A precise description of each variable can be found in appendix I.

After the collection period a number of variables were created without using the program. The following variables were created after the collection period: country/countries of teams/players, Same Country Dummy (1= all from same country), Book Dummy (1= 3 bookmakers & 0= 2bookmakers), Herfindahl (normalized Herfindahl index), Top 5 dummy (1=all bookmakers are in the top 5, i.e. they appear most in my dataset, 0= at least one bookmaker is not in top 5), IBAS Dummy (1= all are a member of IBAS & 0= at least one is not a member of IBAS), Weak Dummy (1=all bookmakers have a license from a weakly regulated country & 0= at least one bookmaker has a license from a strongly regulated country), Member UK Dummy (1= all are a member of the UK gambling commission & 0= at least one bookmaker is not a member of the UK gambling commission), Sport Dummy (1= American football; 2= basketball; 3= hockey; 4=rugby; 5=soccer; 6=tennis), Public Dummy (1= all are public & 0= at least one bookmaker is not public). A precise description of each variable can be found in appendix J.

4.4 Caveats of the dataset

Since the dataset was only updated every day at 12.00pm and 18.00pm it is less complete than when it would've been updated continuously. This was the plan originally, however, creating a program that would monitor the arbitrage opportunities continuously turned out to be harder than expected. By monitoring the arbitrage opportunities manually a smaller dataset has been collected due to time constraints. By automating this process the dataset could have been much larger. The effect on the precision of the age variable is ambiguous. On one hand automating the process could have been better since in some occasions the arbitrage opportunity vanished before the first time they were updated. However, the duration variable stated in Rebelbetting was not always accurate. It seemed as if the software lost the arbitrage opportunity for a second in some cases and found it immediately after. This would have resulted in a false duration variable since the software would show a duration shorter than the actual duration. By collecting all the information manually I was able to calculate the exact age by hand for each arbitrage opportunity.

5. Methodology

In order to assess how market competition affects the pricing of quotes on sporting events three regressions are performed using the data presented in the previous section. The variables of interest are the duration and profit of the arbitrage opportunities. By looking at the effects of the explanatory variables on the duration and profit of arbitrage opportunities it might be possible to make inferences on how market competition influences the price setting mechanism of the bookmakers. I would expect to see that the variables that could be used as proxies for increased market competition will have a positive effect on duration.

5.1 Cox proportional hazard model

The first statistical method used to determine what factors drive the duration of arbitrage opportunities is the Cox proportional hazard model. The Cox model is used to explore the relationship between the survival of a particular event and several explanatory variables. In this case the particular event occurs when the arbitrage opportunity ceases to exist. In my dataset all arbitrage opportunities vanish at some point in time so no censoring of the data was required. Since the variable of interest is the duration of the arbitrage opportunities, the decision was made to use a duration model, of which the Cox proportion hazards model is one. It is important to note that the dependent variable in a Cox model is the hazard function at a given point in time, with the hazard function being the risk that the arbitrage opportunity ceases to exist at time t. By using this model it is possible to estimate how each variable affects the risk of the arbitrage opportunity disappearing. Before running the Cox regression the failure event has to be specified, which in this case is specified as the moment that the arbitrage opportunity is no longer exploitable. By including the explanatory variables presented in the previous section the hazard function can be expressed as follows.

$$h(t) = h_0(t)\exp\{\beta_1\, Book\ Dummy + \beta_2\, Sport\ Dummy + \beta_3\, Same\ Country\ Dummy$$
$$+ g\ (t)(\gamma_1\, Herfindahl + \gamma_2\, Profit + \gamma_3\, Top\, 5\ Dummy + \gamma_4\, IBAS\ Dummy$$
$$+ \gamma_5\, Public\ Dummy + \gamma_6\, Weak\ Dummy + \gamma_7\, Member\ UK\ Dummy)\}$$

The cox model provides estimates of $\beta_{1,} \dots, \beta_k$ and $\gamma_1, \dots, \gamma_k$, but does not provide a direct estimate of the baseline hazard $h_0(t)$, where the baseline hazard is comparable to the constant in a standard regression model. The regression coefficients $\beta_{1,} \dots, \beta_k$ are time invariant covariates, whereas the regression coefficients $\gamma_1, \dots, \gamma_k$ are time varying covariates. These coefficients represent the proportional change in the risk of the arbitrage opportunity vanishing, while taking into account the changes in the explanatory variables.

5.1.1 Interpretation of the hazard ratio

The results of the Cox proportional hazard model are interpreted differently than those of standard regressions since the estimation results are shown in terms of hazard ratios. A hazard ratio can be interpreted as follows: if the hazard ratio for book dummy equals 0.4, with book dummy indicating whether there are three or two bookmakers involved in the arbitrage bet (1= three bookmakers), one could say that when three bookmakers are involved in the arbitrage bet the hazard is sixty percent smaller than when two bookmakers are involved. In other words, when there are three bookmakers needed to construct the arbitrage bet, the arbitrage opportunity is sixty percent less likely to disappear compared to an arbitrage bet that only needs two bookmakers. In the case that the hazard ratio for book dummy would be 1.3 this would mean that the hazard is thirty percent larger for arbitrage opportunities with three bookmakers compared to arbitrage opportunities with two bookmakers, or is thirty percent more likely to vanish.

5.1.2 Expectations

Given that I expect that increased competition increases the prevalence and duration of arbitrage opportunities, since competition could be a motivation to increase quotes on certain sport events causing further price dispersion among the quotes offered by bookmakers on each outcome, the expectations of the hazard ratios for each variable can be formed.

Since book dummy indicates whether there are two bookmakers or three bookmakers involved in the arbitrage bet, a hazard ratio below one can be expected. As the number of bookmakers involved in an arbitrage bet increases, competition will increase as well.

For the sports dummy hazard ratios below one are expected for the more popular sports and hazard ratios above one for less popular sports. There will be more betting volume on the popular sports, thus one could expect that the competition is stronger for these sports.

The same country dummy variable indicates whether all teams/players are from the same country. I would expect to see a hazard ratio larger than one since the bettors are more likely to know what the fair quotes should be, given that there is more history between both teams. When bettors are better able to detect when a quote is too high or too low compared to the fair quote this imposes a risk for the bookmaker to skew the odds by too much.

For the potential arbitrage profit a hazard ratio above one is expected since a higher profit could entail a larger loss for one of the bookmakers involved in the arbitrage bet. This makes it more likely for bookmakers to revise their quotes in order to get the potential profit down.

The top 5 dummy variable indicates that all bookmakers involved in the arbitrage bet are among the top 5 bookmakers involved in arbitrage bets in the dataset. Because of this I would expect that these bookmakers are competitive companies that are trying to offer the best quote. Assuming that competition increases the duration of arbitrage opportunities, a hazard ratio smaller than one is expected.

The IBAS dummy variable indicates that all bookmakers involved in the arbitrage bet are a member of the independent betting adjudication service (IBAS), which handles disputes between bettors and bookmakers. It is likely that these bookmakers are bigger companies that have already obtained some market share. Because these bookmakers most probably already have a relatively large client base they are less likely to offer quotes with which long lasting arbitrage opportunities can be exploited. They might still offer quotes different than the fair quote in order to exploit betting biases, but will most likely revise their quotes as soon as they notice that their quotes can be used in arbitrage bets. Therefore a hazard ratio smaller than one is anticipated.

The same reasoning applies for the variables member UK dummy and public dummy, which indicate whether all bookmakers involved in the arbitrage bet are regulated by the UK gambling commission and whether all bookmakers involved are listed on the stock exchange.

The weak dummy variable indicates whether all bookmakers involved hold a license from a weakly regulated country. These bookmakers are most likely smaller companies and are therefore more likely to increase their quotes in order to obtain their share of the market. Because of this a hazard ratio smaller than one is expected for such companies.

The Herfindahl variable is probably the best proxy for market competition out of all the variables. The higher the Herfindahl index the more concentrated the market is, thus the lower the competition. Because of this I would expect a hazard ratio above one.

Below a table is presented with the expectations for each variable.

Tabel 4: Expectations of Hazard Ratio's

Variables	**Hazard ratio**
Time Invariant Covariates	
Book Dummy	<1
Sport Dummy	<1 or >1
Same County Dummy	>1
Time Varying Covariates	
Herfindahl	<1
Profit	>1
Top 5 Dummy	<1
IBAS Dummy	>1
Public Dummy	>1
Weak Dummy	<1
Member UK Dummy	>1

5.2 Time fixed effects model with duration as the dependent variable

In addition to the Cox proportional hazards model I also determine which factors influence duration by performing a time fixed effects regression. The decision to use a fixed effects regression over a random effects regression is based on the outcome of a Hausmann test. The regression performed is expressed below.

$$Duration = \alpha + \beta_1\ Herfindahl + \beta_2\ Profit + \beta_3\ Public\ Dummy + \beta_4\ IBAS\ Dummy$$
$$+ \beta_5\ Weak\ Dummy + \beta_6\ Member\ UK\ Dummy + \beta_7\ Top\ 5\ Dummy$$

5.2.1 Expectations

The reasoning behind the expectations for the time fixed effects regression is the same as explained above for the cox proportional hazard model. The only difference is that the time fixed effects regression does not estimate a hazard ratio. The size of the hazard ratio, however, does reflect the sign that is expected for each coefficient. Hence, a hazard ratio smaller than one corresponds to a positive sign, whereas a hazard ratio greater than one corresponds to a negative sign. Below a table is presented with the expected signs for each variable.

Tabel 5: Expected Signs of Coefficients

Variables	Sign of Coefficent
Herfindahl	-
Profit	-
Public Dummy	-
IBAS Dummy	-
Weak Dummy	+
Member UK Dummy	-
Top 5 Dummy	+

5.3 Time fixed effects model with profit as the dependent variable

Besides duration it is also interesting to see what factors influence profit. Thereto a time fixed effects regression with profit as the dependent variable was executed. Again the decision to do a time fixed effects regression over a random effects regression was based on the outcome of a Hausmann test. The regression performed is presented below.

$$Profit = \alpha + \beta_1\ Herfindahl + \beta_2\ Duration + \beta_3\ Public\ Dummy + \beta_4\ IBAS\ Dummy$$
$$+ \beta_5\ Weak\ Dummy + \beta_6\ Member\ UK\ Dummy + \beta_7\ Top\ 5\ Dummy$$

5.3.1 Expectations

Taking into account the expectations of the effects of competition on the price setting mechanism of bookmakers, the expectations for the signs of the coefficients for each variable can be formed.

For the Herfindahl variable a positive sign is expected. Although increased competition will increase quotes on certain sport events, causing further price dispersion among the quotes offered by bookmakers on each outcome, the smaller bookmakers will not want the potential arbitrage profit to be too large. Since bigger bookmakers will be exploiting betting biases more aggressively, it can be expected that more market concentration will result in higher arbitrage profits albeit for a small amount of time.

For the duration variable a negative sign is anticipated since the potential arbitrage profit will most likely decrease over time.

The sign of the variable public dummy would most likely be positive. Although these companies will revise their quotes as soon as they notice that arbitrage bets can be constructed when using their quotes, they will still offer quotes different than the fair quote in order to exploit betting biases. Because of this they might still be involved in arbitrage bets, but for a shorter amount of time. I expect the sign to be positive since we anticipate these bookmakers to be more aggressive when it comes to exploiting betting biases, which smaller bookmakers might not be able to do as aggressively since this entails more risk.

The same intuition applies to the variables IBAS dummy and member UK dummy. For the weak dummy variable the opposite is anticipated. Since this variable equals one when all bookmakers hold licenses from weakly regulated countries, it can be expected that these bookmakers are relatively small companies and thus might be more actively increasing quotes in order to obtain market share. However, these bookmakers are most likely not involved in arbitrage opportunities with large profits since this exposes them to bigger risks. So these bookmakers are expected to be involved in arbitrage opportunities with longer durations, but with smaller profits. By being involved in arbitrage bets these bookmakers are able to increase their betting volume on their website, while still managing their risks by imposing limits on bettors who only exploit arbitrage opportunities.

For the top 5 dummy variable a negative sign is anticipated. These bookmakers are involved in most of the arbitrage opportunities in the dataset, which implies that they often offer quotes differently than many of their competitors causing arbitrage opportunities to arise. Because of this I expect that these bookmakers will try to keep the potential arbitrage profit low.

Below a table is presented with the expected sign for each variable.

Table 6: Expected Signs of Coefficients

Variables	Sign of Coefficent
Herfindahl	+
Duration	-
Public Dummy	+
IBAS Dummy	+
Weak Dummy	+
Member UK Dummy	-
Top 5 Dummy	-

6. Results & summary statistics

In this section the results of the regressions presented in the previous section are discussed. Before presenting the results I will present the summary statistics.

6.1 Summary statistics

The summary statistics will be presented as follows. First the duration details per sport are presented. After that the potential arbitrage profit details per sport will be discussed. Some statistics on the bookmakers can be found in appendix K.

Below a table is presented with the duration details per sport.

Table 7: Duration – summary statistics

Duration in minutes						
Sports	*# of arbitrage opportunities*	*Median*	*Mean*	*sd*	*min*	*max*
American Football	3	2280	2690.333	2886.458	31	5760
Basketball	158	73.5	267.2405	465.902	0	3600
Hockey	145	172	470.3034	582.8768	1	2700
Rugby	38	706.5	1903.211	5645.573	0	35220
Soccer	180	1108	1541.561	3238.144	1	40440
Tennis	66	99	293.3636	398.4534	1	1500
Total	590	330	826.5322	2402.345	0	40440
Number of times the arbitrage opportunity was updated in the dataset						
American Football	3	1	3.666667	.4618802	1	9
Basketball	158	1	1.322785	.6702883	1	6
Hockey	145	1	1.62069	.8085392	1	4
Rugby	38	2	3.5	7.717127	1	49
Soccer	180	2	2.95	3.467769	1	42
Tennis	66	1	1.378788	.5755306	1	3
Total	590	1	2.050847	2.901459	1	49

From the table it can be seen that there were 590 arbitrage opportunities followed over time in the dataset. The arbitrage opportunity with the longest duration was a soccer match between Villarreal and Red Bull Salzburg, which lasted for 40,440 minutes (approximately 28 days). The arbitrage opportunity was updated for 42 times. It might seem strange that this opportunity was not updated the most often, but this is because the opportunity was already exploitable for a considerable amount of time when it first entered the dataset. The average duration of all arbitrage opportunities in the dataset is 826 minutes and the median is 330 minutes. This is far larger than the results found in the study by Ben Marshall. He finds a median duration of 15 minutes and the arbitrage opportunity with the longest duration in his dataset lasts slightly under a day. Furthermore he finds that 75% of all arbitrage opportunities is removed within 50 minutes.

Below a table is presented with the details of potential arbitrage profit per sport.

Table 8: Profit – summary statistics

Sports	# of arbitrage opportunities	mean	median	sd	Min	max
Profit in percentages						
American Football	3	.8066667	.48	.5658033	.48	1.46
Basketball	158	1.374937	.98	1.214497	.2	7.01
Hockey	145	1.845724	1.32	1.343438	.12	5.97
Rugby	38	1.197895	.975	.9740336	.12	4.42
Soccer	180	2.066333	1.95	1.152501	.14	8.87
Tennis	66	1.065	.61	1.352508	.11	6.99
Total	590	1.65261	1.3	1280033	.11	8.87
Profit in percentages for each observation						
American Football	11	1.367273	1.46	.5558074	.48	2.06
Basketball	209	1.337512	.98	1.1713	.2	7.01
Hockey	235	1.721021	1.23	1.270493	.12	5.97
Rugby	133	1.478722	1.65	.8304822	.05	4.42
Soccer	531	1.729736	1.52	1.070784	.14	8.87
Tennis	91	.9740659	.61	1.194689	.11	6.99
Total	1210	1.572579	1.3	1.134985	.05	8.87

I included the profit for each observation since the profit can vary over time. As can be seen from the table the profit also declines over time since the average profit is lower in the lower part of the table. This makes sense since I would expect that bookmakers will revise their quotes downwards when they notice that their quotes can be used in an arbitrage bet. The potential arbitrage profits in this dataset are lower than the profits found by Ben Marshall in his paper. He finds a median profit of 1.51% and an average profit of 2.03%. The potential arbitrage profits in his paper range from 0.91 to 11.10%, whereas the profits in this dataset range from 0.05% to 8.87%. However, his dataset consists of 509,679 arbitrage opportunities whereas this dataset only includes 590 different arbitrage opportunities.

6.2 Regression results

The regression results of the Cox proportional hazards model are presented below.

Table 9: Regression Results of Cox Model

Dependent Variable: *Hazard Function*	
<u>*Time Invariant Covariates*</u>	
Book Dummy	0.7930604* (0.1014091)
Hockey	0.6371917*** (0.0786849)
Rugby	0.1025563*** (0.0231817)
Soccer	0.1515207*** (0.0218685)
Tennis	1.213294 (0.1985175)
Same Country Dummy	1.401493*** (0.164164)
<u>*Time Varying Covariates*</u>	
Herfindahl	1.00017 (0.0001523)
Profit	1.000016 (0.0000376)
Top 5 Dummy	1.000058 (0.0000892)
IBAS Dummy	0.9998545 (0.0001034)
Public Dummy	0.9998925 (0.0003456)
Weak Dummy	1.000097 (0.000107)
Member UK Dummy	1.000153 (0.0001064)
Number of Observations	1190

From the table it can be seen that the variable book dummy is significant at the 10% level and has the sign that was expected. The hazard ratio indicates that when three bookmakers are involved in an arbitrage bet it is approximately twenty percent less likely to disappear than when the arbitrage bet could be constructed using two bookmakers. This is in line with the expectations since competition increases when there are more bookmakers needed to construct the arbitrage bet.

Looking at the sports dummy variables it can be seen that all sports are significant with the exception of tennis. Basketball was left out of the regression and all hazard ratios depicted in the table are relative to basketball. American Football was also omitted since there were too few American Football arbitrage opportunities in the dataset. Soccer is arguably the most popular sport to bet on out of the list of sports in the dataset. The hazard ratio for soccer indicates that an arbitrage opportunity for soccer lasts approximately eighty-five percent longer than an arbitrage opportunity for basketball. This is in line with the expectations as it was assumed that there is more betting volume for soccer and therefore stronger competition. Arbitrage opportunities on hockey and rugby matches also seem to last longer than for basketball. The longest expected duration in the dataset is for rugby. This could be explained by the fact that the second longest arbitrage opportunity in the dataset was on a rugby match, showing up 49 times in the dataset (lasting for approx. 25 days). Besides this the dataset was compiled during one of the most popular events of the year for rugby, namely the Six Nations championship.

The hazard ratio of the same country dummy indicates that if the teams/players are from the same country, the arbitrage opportunity is approximately forty percent more likely to cease to exist. This is also in line with the expectations since I expected that bettors might be more able to detect when a quote might be different from the fair quote when the teams are from the same country. These teams have probably played against each other more often than teams from different countries. Therefore the bettors have better information and will have a better feel for what the fair quote should be like. Further, the most likely bettors on these matches are often residents of the country, which makes them more familiar with these teams. When bettors are better able to distinguish skewed quotes from fair quotes this imposes a risk for the bookmaker, which may be a reason why bookmakers offer quotes closer to the fair quotes for these types of matches.

Besides the results discussed above, no statistically significant relationships are found using the Cox proportional hazard model.

The second regression performed is the time fixed effects regression model with duration as the dependent variable. The results for this regression are depicted below.

Table 10: Regression Results of Fixed Effects Model

Dependent Variable: ***Duration***	
Herfindahl	-2640.5* (1337.6)
Profit	-1201.6*** (183.0)
Public Dummy	-1714.5 (2271.3)
IBAS Dummy	-341.1 (1215.1)
Weak Dummy	2822.0*** (828.0)
Member UK Dummy	-1221.2* (709.1)
Top 5 Dummy	-7959.4*** (808.5)
Constant	7019.2*** (742.5)
Number of Observations	1199
R-squared	0.2039
Adjusted R-squared	-0.576

As can be seen from the table, the Herfindahl variable is statistically significant at the 10% level and has the sign that was expected. As the Herfindahl variable increases, the duration of the arbitrage opportunity decreases. Since the Herfindahl variable measures market concentration, an increase in the Herfindahl variable entails less competition. So as competition increases, i.e. the Herfindahl variable decreases, the duration of the arbitrage opportunity increases. This is exactly in line with the expectations presented in the previous section.

The profit variable is statistically significant at the 1% level and also has the sign that was anticipated. As profit increases the duration of the arbitrage opportunity decreases. This makes sense since it can be expected that quotes involved in arbitrage bets with high potential profits are revised sooner in order to manage the risks of the bookmakers.

The weak dummy variable has a positive sign and is also significant at the 1% level. This confirms what was expected since bookmakers who hold licenses from weakly regulated countries are assumed to be smaller companies which seek to increase their betting volume. These bookmakers will compete with the other bookmakers and will offer quotes at the high end of the spectrum in order to lure bettors to their website, while simultaneously managing their risks by imposing limits on bookmakers who try to exploit the arbitrage opportunity on a regular basis. Since these bookmakers have a smaller client base than the bigger bookmakers, there are less bettors exploiting these arbitrage opportunities than when bigger bookmakers were to offer these quotes. Bigger bookmakers will most likely revise their quotes faster since they already have a lot of betting volume on their website and only post quotes different than the fair quote in order to exploit betting biases.

This is in line with the sign of the variable member UK dummy. I assume that bookmakers that are regulated by the UK Gambling Commission are most likely bigger companies, thus will try to revise their quotes when they notice that these can be used in arbitrage opportunities.

The sign of the top 5 dummy is different than previously anticipated, with the coefficient being statistically significant at the 1% level. The coefficient indicates that when all bookmakers are among the top 5 bookmakers involved in arbitrage opportunities in the dataset, the duration will be shorter than when not all bookmakers are in the top 5. A possible explanation for this could be that these bookmakers revise their quotes faster than the other bookmakers, which might also indicate that they are not trying to attract betting volume but are exploiting betting biases. Another explanation could be that these bookmakers try to attract betting volume but revise their quotes relatively quick to manage their risk.

Besides the results explained above no further statistically significant relationships are found using the time fixed effects regression with duration as the dependent variable.

The third regression performed is the time fixed effects regression with profit as the dependent variable. The results for this regression are presented below.

Table 11: Regression Results of Fixed Effects Model

Dependent Variable: ***Profit***	
Herfindahl	0.631** (0.295)
Duration	-0.0000554*** (0.00000843)
Public Dummy	1.087** (0.486)
IBAS Dummy	0.231 (0.261)
Weak Dummy	-0.324* (0.179)
Member UK Dummy	-0.162 (0.152)
Top 5 Dummy	-0.0382 (0.187)
Constant	1.588*** (0.158)
Number of Observations	1199
R-squared	0.0992
Adjusted R-squared	-0.784

As can be seen from the table the Herfindahl variable is statistically significant at the 5% level and has the sign that was expected. As market concentration increases profit increases as well, which is in line with the reasoning that bigger companies are more aggressively exploiting betting biases. Smaller companies that are trying to attract volume with more attractive quotes will try to keep the potential arbitrage profit small to manage their potential losses. The duration variable is significant at the 5% level and profit decreases as the duration goes up. This makes sense since it is likely that the potential arbitrage profit decreases over time and eventually disappears. Bookmakers will revise their quotes in order to get the arbitrage profit down.

The public dummy variable indicates that when all bookmakers are companies listed on the stock exchange the profit goes up by one percent. This is in line with the reasoning that bigger companies, which I assume that companies listed on the stock exchange are, tend to exploit betting biases more aggressively. This could result in higher arbitrage profits. When the bookmakers notice that their quotes can be used in an arbitrage bet they will revise their quotes, since they are not willing to give their large client base the opportunity to exploit this arbitrage opportunity. So arbitrage opportunities with bigger bookmakers might have higher profits, but these will only be exploitable for a short time.

The sign of the variable weak dummy further confirms this reasoning since bookmakers who hold licenses from weakly regulated countries are assumed to be smaller. These bookmakers try to keep the potential arbitrage profit low, but will leave the arbitrage opportunity exploitable for a longer amount of time in order to attract betting volume.

Besides the results explained above, no other statistically significant relationships are found using the time fixed effects regression model with profit as the dependent variable.

When all the results are taken into consideration it seems that bigger companies tend to have higher potential arbitrage profits, but that these profits are only exploitable for a short period of time. The exact opposite applies to smaller companies. These bookmakers seem to be involved in arbitrage opportunities that last longer, however, the potential arbitrage profit that could be obtained when exploiting the opportunity is smaller.

These results are in line with the theory on the effects of competition on the pricing mechanism of bookmakers. The smaller bookmakers want to compete with the bigger bookmakers and post attractive quotes in order to attract betting volume. These bookmakers do not mind that the arbitrage opportunity is available for a longer period of time since the increase in betting volume will help them grow their client base, make bettors more familiar with the company and allow them to increase advertising revenues on their websites. At the same time they are able to manage their risks by imposing limits on bettors that consistently try to exploit the arbitrage opportunity. The bigger companies do not want to be involved in arbitrage opportunities for long periods of time. They do not want to give their large client base the time to find and exploit these opportunities. Hence, they tend to remove their arbitrage opportunities faster than smaller companies. These bookmakers might still be involved in arbitrage opportunities since they are exploiting betting biases more aggressively and most likely have built expertise in this field. This may result in arbitrage opportunities with high potential profits in some cases. As soon as these bookmakers realize this they will revise their quotes so that the potential arbitrage profit decreases or disappears. For the smaller bookmakers it might be riskier to exploit betting biases since professional bettors could consistently find out when the quotes diverge too far from the fair quote. Furthermore these bookmakers are less likely to have the expertise yet to do this efficiently.

In short, it seems that big bookmakers are involved in arbitrage profits for smaller amounts of time, but could have higher potential arbitrage profits. Smaller bookmakers are involved in arbitrage opportunities for longer amounts of time, but will try to keep the potential arbitrage profit in a range that they are satisfied with.

7. Conclusion

In this paper I have researched the effect of competition on arbitrage opportunities in sports betting markets. By using a unique hand-collected dataset I was able to assess which factors affect the duration and profitability of arbitrage opportunities within these markets. It seems that relatively small companies post high quotes in order to attract more betting volume. These companies are involved in arbitrage opportunities with relatively long durations and low potential arbitrage profits. By keeping the potential arbitrage profit relatively low and by imposing limits on bettors who consistently try to exploit these arbitrage opportunities these bookmakers are able to manage their risk. The bigger bookmakers seem to be involved only in arbitrage opportunities with relatively low durations. An explanation for this could be that these bookmakers are not actively trying to attract betting volume by posting high quotes, but that they are actively exploiting the biases of biased bettors. By doing so they might be involved in arbitrage bets with high potential profits, but it seems that they revise their quotes when they notice that their quotes are used in arbitrage opportunities. The risk of being involved in an arbitrage bet might be higher for these bookmakers since their client base is much larger. In line with the results of the model developed in Len van der Sluijs (2013) I have shown that competition might actually consistently generate arbitrage opportunities. For further research I would recommend to extend the model developed in Len van der Sluijs (2013) such that it can include more than two bookmakers and sport events with more than two outcomes. It might also be interesting to see whether the same empirical results hold for a dataset that is updated continuously and is much larger in size.

Appendix

Appendix A: The model

The model developed here is a simplification of the online betting market. For simplicity's sake only sporting events with two possible outcomes are taken into consideration, but I expect that the results can be generalized to sporting events with more than two possible outcomes.

To examine why arbitrage opportunities exist in the online betting market I will first assume that the entire market is controlled by one bookmaker (a monopoly) and explore what his most profitable strategy would be. Next I will introduce competition and look at the market as if it is controlled by only two bookmakers (a duopoly) and examine what the most profitable strategy would be for either bookmaker assuming that the other also has to set a quote. I will do this by searching for pure Nash equilibria, so mixed strategy equilibria are not considered in this model.

There are four types of players in the model, namely: bookmakers, unbiased bettors, biased bettors and arbitrageurs. I follow Levitt (2004) and specify the quote as a probability, denoted by P. This is done without loss of generality since what is important is the odd $P/(1-P)$. The bookmakers post a quote for every game, which is represented by P. This quote is a probability, given by the bookmaker, that the favourite wins a particular game. The bookmaker will try to choose a quote such that it maximizes his profit. Besides this, the bookmaker collects commissions, which are charged only on losing bets and are represented by V. This assumption is also made without loss of generality since one can think of the quote paid once a bettor wins as net of commission.

The bettors take the quote posted by the bookmaker in to consideration and decide whether they want to bet on the favourite or the underdog. Bettors are endowed with money, which I denote by M_u and M_b for unbiased and biased bettors respectively. The unbiased bettors will always bet on the most favourable quote, which depends on the fair quote Q. This means that whenever the bookmaker posts a quote P such that $P > Q$ the unbiased bettors will bet on the favourite since they can place a bet at a better then fair quote on the favourite. If the bookmaker posts a quote P such that $P < Q$ the unbiased bettors will bet on the underdog since they can place a bet at a better then fair quote on the underdog. In the case that the bookmaker posts a quote P such that $P = Q$ the unbiased bettor is indifferent between betting on the favourite or the underdog. The fraction of money that is bet by the unbiased bettors that goes to the favourite is represented by F_u. The biased bettors are biased towards the favourite. This means that they are willing to bet on the favourite at a less than fair quote. The point at which the biased bettors become indifferent between betting on the favourite and the underdog is represented by Q_b, which is smaller than Q. The fraction of money that is bet by biased bettors that goes to the favourite is represented by F_b.

The arbitrageurs will take advantage of differences in quotes amongst bookmakers as they seek for a risk free profit. After the bookmakers have posted their quotes the arbitrageurs will construct an arbitrage bet whenever the variance between the posted quotes is large enough.

<u>3.1 Monopoly</u>

In this section I will assume that the entire market is controlled by only one bookmaker and will try to find the optimal price setting strategy for the monopolist.

When taking the information provided above into account there are three cases left to consider:

1. $P > Q$.

In this case the biased and unbiased bettors will bet on the favourite, because both will be able to bet on the favourite at a better than fair quote.

2. $Q_b \leq P \leq Q$.

In this case the biased bettors will bet on the favourite and the unbiased bettors will bet on the underdog. The biased bettors will bet on the favourite because the quote posted by the bookmaker is larger than the cut-off level at which biased bettors become indifferent. The unbiased bettors will bet on the underdog since they are able to bet on the underdog at better than fair odds.

3. $P < Q_b$.

In this case the biased and the unbiased bettors will both bet on the underdog. The posted quote by the bookmaker is now below the indifference level of the biased bettors, so biased bettors will bet on the underdog. Unbiased bettors will also bet on the underdog since they always bet on the most favourable quote.

For the model a profit function is used to derive our results from. The profit function is represented below:

$$E(Bookmaker\ profit) = [QM_j(1-F_j)+(1-Q)M_jF_j](1+V)-[QM_jF_jP+(1-Q)(1-F_j)(1-P)M_j]$$

for $j = u,b$. (1)

For each of the three situations described above the bookmaker's profit can be calculated. The profit for each situation is shown below.

The bookmaker's profit when $P > Q$ is:

$$M_u[(1-Q)(1+V)-QP]+M_b[(1-Q)(1+V)-QP]. \quad (2)$$

The bookmaker's profit when $Q_b \leq P \leq Q$ is:

$$M_u[Q(1+V)-(1-Q)(1-P)]+M_b[(1-Q)(1+V)-QP]. \quad (3)$$

The bookmaker's profit when $P < Q_b$ is:

$$M_u[Q(1+V)-(1-Q)(1-P)]+M_b[Q(1+V)-(1-Q)(1-P)]. \quad (4)$$

Let X be denoted by $X \equiv 2+V-3Q-2QV-P$. Using (2), (3) and (4) we can find the quote range that gives the monopolist the highest profit.

Proposition 1: The monopoly profit function is given by:

(i) if $X > 0$.

(ii) if $X = 0$.

(iii) if $X < 0$.

Proof: see appendix

Proposition 1 gives the monopolist's profit function from which the optimal quote will be chosen. Let's look at each of the functions with Q=1/2. In situation (i) where $P > 1/2$, the condition that (i) is optimal is $P < 1/2$ for $Q = 1/2$. This implies a contradiction with the definition of (i), so (i) can never be optimal. In the third situation where $P < Q_b$, the condition that (iii) is optimal is $P > 1/2$ for $Q = 1/2$. This implies a contradiction with the definition of (iii), so (iii) can never be optimal. In the second situation where $Q_b < P < 1/2$, the condition that (ii) results in an optimum is $P=1/2$ for $Q=1/2$. So solving (ii) is in line with its definition and is the favourite candidate for the optimal quote. Taking all this into consideration, (ii) is the only profit function resulting in a condition for optimality which conforms with its definition. Consequently $Q=1/2$.

From what is discussed above, one finds the optimal quote to be $P=1/2$ when $Q=1/2$.

Corollary 1: If $Q=1/2$, then $P=1/2$.

Proof: see appendix

From this it immediately follows that (ii) is the maximum and that (i) and (iii) will never be the maximum in the case of a monopoly.

So in the case of a monopoly, a profit function belonging to (ii) is the maximum profit:

$$(M_u+M_b)[(V/2)+(1/4)] \tag{5}$$

If I were to include one more bookmaker, the profit for both bookmakers has to be shared if they would both choose $P=1/2$.

This would lead to a profit of: $(M_u+M_b)[(V/2)+(1/4)]/2$. (6)

In principal, this need not be optimal for bookmakers. In other words, bookmakers could have an incentive to deviate to increase profits. However, as I show below in the next sub-section, bookmakers indeed share the profits in the case of a duopoly with only unbiased bettors.

3.2 Duopoly: (only unbiased bettors)

In this section I will assume that the entire market is controlled by two bookmakers. In this market there is only one type of bettor, namely the unbiased bettor. I will try to find the optimal price setting strategy for each bookmaker given that the other bookmaker also has to set a quote.

Proposition 2: With a duopoly and only unbiased bettors the Nash equilibria are:

i. $P_a=P_b=1$

ii. $P_a=P_b=0$

iii. $P_i=1, P_j=0$ for $i,j\in\{a,b\}$ and $i\neq j$

In a situation with competition and only unbiased bettors, the bookmaker's best strategy is not choosing $P = 1/2$ If one bookmaker chooses $P = 1/2$, the other bookmaker can slightly increase or decrease his P and attract all bettors. When the bookmakers are situated at either $P = 1$ or $P = 0$ no one is able to improve his profit at expense of the other. At this point it does not matter whether both choose $P = 0$, $P = 1$ or if one does the opposite of the other.

In situation 1 and 2 no arbitrage opportunities will be available in the market, because both bookmakers will post the same quote. In situation 3 arbitrageurs will be able to exploit the differences in posted odds.

3.2 Duopoly: (biased and unbiased bettors)

In this section I will assume that the entire market is controlled by two bookmakers. In this market there are two types of bettors, namely unbiased bettors and biased bettors. I will try to find the optimal price setting strategy for each bookmaker given that the other bookmaker also has to set a quote.

Proposition 3: In the case of a duopoly with biased and unbiased bettors the unique Nash equilibrium is: $P_a = P_b = 1$.

Proof: see appendix.

In a duopoly with biased and unbiased bettors the optimal choice for each bookmaker is to post a quote equal to 1. Posting a quote equal to 0 is no longer optimal due to the introduction of the biased bettors. If bookmaker a has a quote equal to 0 and bookmaker b has a quote equal to 1, then both bookmakers would share the unbiased bettors and bookmaker b will have all of the biased bettors. Bookmaker a should change his quote to $P_a = 1$ and by doing so both bookmakers will share profits. In every other case the bookmaker who does not post a quote equal to 1 will not attract any of the betting volume. In the case of a duopoly with unbiased and biased bettors no arbitrage opportunities will be available, since in equilibrium both bookmakers will post a quote equal to 1.

3.3 Duopoly: (biased bettors, unbiased bettors and arbitrageurs)

In this section I will assume that the entire market is controlled by two bookmakers. In this market there are three players active, namely unbiased bettors, biased bettors and arbitrageurs. I will try to find the optimal pricing strategy for each bookmaker given that the other bookmaker also has to set a quote.

For an arbitrage opportunity to occur all bets involved in the combined bet need to have a positive expected profit. In the case of a duopoly, the arbitrageur will construct an arbitrage bet by betting on both of the websites. The arbitrageurs have to calculate the stakes that they need to bet on each website. Let α be the stake bet on the favourite on bookmaker a and $1-\alpha$ the stake bet on the underdog on bookmaker b. The quote on the favourite is defined as P with the subscripts indicating which bookmaker offers the quote. This means that $1-P$ is the quote on the underdog. Taking this into consideration the expected profit for the arbitrageurs betting on this match looks as follows:

$$E(profit) = 1/2[\alpha P_a - (1-\alpha)] + 1/2[(1-\alpha)(1-P_b) - \alpha] \text{ .[2]} \tag{6}$$

For an arbitrage opportunity to exist both terms inside of the brackets need to be positive, that is $\alpha P_a - (1-\alpha) \geq 0$ and $(1-\alpha)P_b - \alpha \geq 0$. (7)

These conditions can be simplified to:

$$P_a \geq (1-\alpha)/\alpha \text{ and } (1-\alpha)/\alpha \geq 1/(1-P_b). \tag{8}$$

Since both these conditions need to be met for an arbitrage opportunity to occur I define the condition for arbitrage as follows.

$$P_a \geq (1-\alpha)/\alpha \geq 1/(1-P_b). \tag{9}$$

First, I want to show that bookmakers a and b choosing $P_a = 1$ and $P_b = 0$ is an equilibrium. If we fill in these quotes the arbitrage condition will look as follows.

$$1 \geq (1-\alpha)/(\alpha) \geq 1 \Rightarrow \qquad \alpha = 1/2. \tag{10}$$

Here it becomes clear that in the case of an arbitrage opportunity both bookmakers will share the betting volume of the arbitrageurs. If one of the bookmakers would deviate from these quotes the arbitrage condition will not be satisfied, which means that the arbitrage opportunity does not arise and that the bookmakers will loose the volume being bet by the arbitrageurs. For instance, in the case that $P_a < 1$, then $P_a < 1/(1-P_b)$, which contradicts (9). In the case that $P_b > 0$, then again $P_a < 1/(1-P_b)$, which contradicts (9).

---

[2] One can think the profit function is $1/2[\alpha W_{f,a}] + 1/2[(1-\alpha)W_{u,b}] - 1$. However, since I use probabilities for the quotes we get the same profit function if $W_{f,a} = 1 + P_a$ and $W_{u,b} = 1 + (1-P_b)$. Here $W_{f,a}$ indicates the quote that bookmaker a offers for the favourite and $W_{u,b}$ the quote that bookmaker b offers for the underdog.

When the volume of the bets placed by the arbitrageurs is sufficiently large, then $P_a = 1$ and $P_b = 0$ constitute an equilibrium. I formulate this in the following proposition.

Proposition 4: In the case of a duopoly with unbiased bettors, biased bettors and arbitrageurs the unique Nash equilibrium is $P_i = 1, P_j = 0$ for $i, j \in \{a, b\}$ and $i \neq j$, given that the volume of bets placed by arbitrageurs is sufficiently large.

Proof: see appendix.

When looking at the condition for arbitrage (9) it can easily be seen that no deviations exist that satisfy this condition. In the case that one of the bookmakers deviates, the arbitrage opportunity disappears and the arbitrageurs will not be able to place their bets. This entails that the bookmakers will not have any of the betting volume of the arbitrageurs. From proposition 3 it follows that with the absence of arbitrageurs the only equilibrium is $P_a = P_b = 1$. Since the bookmakers can devise a strategy including arbitrageurs, this will not be an equilibrium in this market. Assuming that the volume being bet by the arbitrageurs is sufficiently large a bookmaker would be better of deviating to $P = 0$, since the profit obtained in this case would be larger than simply sharing the unbiased and biased bettors. Given that the volume being bet by the arbitrageurs is sufficiently large the unique Nash equilibrium in the case of a duopoly with unbiased bettors, biased bettors and arbitrageurs is $P_i = 1, P_j = 0$ for $i, j \in \{a, b\}$ and $i \neq j$.

Appendix B: Proof of proposition 1:

$$(2) = \max\{(2),(3),(4)\}$$

- $(2) \geq 3$

$$M_u[(1-Q)(1+V) - QP] + M_b[(1-Q)(1+V) - QP] \geq M_u[Q(1+V) - (1-Q)(1-P)] +$$

$$M_b[(1-Q)(1+V) - QP] \quad \Leftrightarrow$$

$$M_u[(1-Q)(1+V) - QP] \geq M_u[Q(1+V) - (1-Q)(1-P)] \quad \Leftrightarrow$$

$$(1-Q)(1+V) - QP \geq Q(1+V) - (1-Q)(1-P) \quad \Leftrightarrow$$

$$1 + V - Q - QV - QP \geq Q + QV - 1 + P + Q - QP \quad \Leftrightarrow$$

$$2 + V - 3Q - 2QV - P \geq 0$$

- $(2) \geq (4)$

$$M_u[(1-Q)(1+V)-QP]+M_b[(1-Q)(1+V)-QP] \geq M_u[Q(1+V)-(1-Q)(1-P)]+$$

$$M_b[Q(1+V)-(1-Q)(1-P)] \qquad \Leftrightarrow$$

$$M_u(1+V-Q-QV-QP)+M_b(1+V-Q-QV-QP) \geq M_u(Q+QV-1+P+Q-QP)+$$

$$M_b(Q+QV-1+P+Q-QP) \qquad \Leftrightarrow$$

$$(M_u+M_b)(1+V-Q-QV-QP) \geq (M_u+M_b)(Q+QV-1+P+Q-QP) \quad \Leftrightarrow$$

$$2+V-3Q-2QV-P \geq 0$$

$$(3)=\max\{(2),(3),(4)\}$$

- $(3) \geq (2)$

$$M_u[Q(1+V)-(1-Q)(1-P)]+M_b[(1-Q)(1+V)-QP] \geq M_u[(1-Q)(1+V)-QP]+$$

$$M_b[(1-Q)(1+V)-QP] \qquad \Leftrightarrow$$

$$M_u[Q(1+V)-(1-Q)(1-P) \geq M_u[(1-Q)(1+V)-QP] \qquad \Leftrightarrow$$

$$Q(1+V)-(1-Q)(1-P) \geq (1-Q)(1+V)-QP \qquad \Leftrightarrow$$

$$Q+QV-1+P+Q-QP \geq 1+V-Q-QV-QP \qquad \Leftrightarrow$$

$$-2-V+3Q+2QV+P \geq 0$$

- $(3) \geq (4)$

$$M_u[Q(1+V)-(1-Q)(1-P)]+M_b[(1-Q)(1+V)-QP] \geq M_u[Q(1+V)-(1-Q)(1-P)]+$$

$$M_b[Q(1+V)-(1-Q)(1-P)] \qquad \Leftrightarrow$$

$$(1-Q)(1+V)-QP \geq Q(1+V)-(1-Q)(1-P) \qquad \Leftrightarrow$$

$$1+V-Q-QV-QP \geq Q+QV-1+P+Q-QP \qquad \Leftrightarrow$$

$$2+V-3Q-2QV-P \geq 0$$

$$(4) = \max\{(2),(3),(4)\}$$

- $(4) \geq (2)$

$M_u[Q(1+V)-(1-Q)(1-P)]+M_b[Q(1+V)-(1-Q)(1-P)] \geq M_u[(1-Q)(1+V)-QP]+$

$M_b[(1-Q)(1+V)-QP]$ $\Leftrightarrow$

$M_u(Q+QV-1+P+Q-QP)+M_b(Q+QV-1+P+Q-QP) \geq M_u(1+V-Q-QV-QP)+$

$M_b(1+V-Q-QV-QP)$ $\Leftrightarrow$

$(M_u+M_b)(Q+QV-1+P+Q-QP) \geq (M_u+M_b)(1+V-Q-QV-QP)$ $\Leftrightarrow$

$-2-V+3Q+2QV+P \geq 0$

- $(4) \geq (3)$

$M_u[Q(1+V)-(1-Q)(1-P)+M_b[Q(1+V)-(1-Q)(1-P)] \geq M_u[Q(1+V)-(1-Q)(1-P)+$

$M_b[(1-Q)(1+V)-QP]$ $\Leftrightarrow$

$Q(1+V)-(1-Q)(1-P) \geq (1-Q)(1+V)-QP$ $\Leftrightarrow$

$Q+QV-1+P+Q-QP \geq 1+V-Q-QV-QP$ $\Leftrightarrow$

$-2-V+3Q+2QV+P \geq 0$

Appendix C: Proof of corollary 1:

$$(2) = \max\{(2),(3),(4)\}$$

In order for (2) to be the maximum it needs to be the case that:

$2+V-3Q-2QV-P \geq 0.$

Filling in $Q=1/2$ (the fair quote) gives us:

$2+V-3/2-V-P \geq 0.$ $\Leftrightarrow$

$1/2-P \geq 0.$ $\Leftrightarrow$

$P \leq 1/2.$

But this contradicts the definition of (2).

$$(3)=\max\{(2),(3),(4)\}$$

In order for (3) to be the maximum it needs to be the case that:

$2+V-3Q-2QV-P=0$.

Filling in $Q=1/2$ gives us:

$2+V-3/2-V-P=0. \qquad \Leftrightarrow$

$1/2-P=0. \qquad \Leftrightarrow$

$P=1/2.$

This is in line with the definition of (3).

$$(4)=\max\{(2),(3),(4)\}$$

In order for (4) to be the maximum it needs to be the case that:

$2+V-3Q-2QV-P\leq 0$.

Filling in $Q=1/2$ gives us:

$2+V-3/2-V-P\leq 0. \qquad \Leftrightarrow$

$1/2-P\leq 0. \qquad \Leftrightarrow$

$P\geq 1/2.$

But this contradicts the definition of (4).

Therefore, the optimal profit function must be (3).

Appendix D: Proof of proposition 2:

In the first step I will show that it is impossible to reach an equilibrium in the three situations described. In the second step I will show that each of the remaining cases can be sustained as equilibrium.

Step 1:

1. $0 < P_a, P_b < 1$.

Suppose $P_a = P_b = \overline{P}$. Then a bookmaker can attract all betting volume by charging $\overline{P} + \varepsilon_0$ such that $\overline{P} + \varepsilon_0 < 1$ or $1 - \overline{P} - \varepsilon_1$, such that $1 - \overline{P} - \varepsilon_1 > 0$.

Suppose $P_a > P_b \geq 0.5$. Then bookmaker a can have the same betting volume and increase his profit by charging $P_a - \varepsilon$ such that $P_a - \varepsilon > P_b$.

Suppose $0.5 \geq P_a > P_b$. Then bookmaker a can attract all betting volume by charging $P_a - \varepsilon$ such that $P_a - \varepsilon < P_b$.

Suppose $P_a \geq 0.5 > P_b$ and that $\overline{P} = 1 - P_b$. In the case that $P_a < 1 - P_b$ bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon_0$ such that $P_a + \varepsilon_0 > 1 - P_b$. In the case that $P_a > 1 - P_b$ bookmaker a can have the same betting volume and increase his profit by charging $P_a - \varepsilon_1$ such that $P_a - \varepsilon_1 > 1 - P_b$.

Suppose $P_a \geq 0.5 > P_b$ and that $P_a = 1 - P_b$. Then bookmaker a can attract all betting volume by charging $P_a - \varepsilon_0$ such that $P_a - \varepsilon_0 < P_b$ or charging $P_a + \varepsilon_1$ such that $P_a + \varepsilon_1 > 1 - P_b$.

2. $0 = P_a < P_b < 1$.

In this case bookmaker b could increase his profit at expense of bookmaker a by charging $P_b = 1$ or $P_b = 0$. In these cases the bookmakers would share profits.

3. $0 < P_a < P_b = 1$.

This case is the exact opposite of the previous situation. Here bookmaker a could increase his profit at expense of bookmaker b by charging $P_a = 0$ or $P_a = 1$.

Step 2:

The only sustainable strategies for the two bookmakers are.

1. $P = 0, P' = 0$.
2. $P = 1, P' = 1$.
3. $P = 0, P' = 1$.
4. $P = 1, P' = 0$.

It is easy to check that in all these cases the two bookmakers share the profits and each bookmaker offers the best quote to unbiased bettors to bet either on the favourite or on the underdog. Therefore, any quote deviations by a particular bookmaker will not attract any bettor.

Appendix E: Proof of proposition 3:

The proof for proposition 3 will be provided in two steps. In the first step I will show that it is impossible to reach an equilibrium in the situations described. In the second step I will show that the only sustainable equilibrium is $P_a = P_b = 1$.

Step 1:

The first step is divided in two cases. In the first case both bookmakers charge the same quote and in the second case bookmaker a charges a higher quote than bookmaker b.

Case 1: $P_a = P_b$

Suppose $1 > P_a = P_b \geq 0.5$. Then bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon$ such that $P_a + \varepsilon > P_b$.

Suppose $Q_b \leq P_a = P < 0.5$. Then bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon$ such that $P_a + \varepsilon > 1 - P_b$.

Suppose $P_a = P_b < Q_b$. Then bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon_0$ such that $P_a + \varepsilon_0 > 1 - P_b$ or $P_a - \varepsilon_1$ such that $P_a - \varepsilon_1 < P_b$.

Case 2: $P_a > P_b$

Suppose $0.5 \le P_b < P_a < 1$. Then bookmaker a can have the same betting volume and increase his profit by charging $P_a - \varepsilon$ such that $P_a - \varepsilon > P_b$.

Suppose $P_a \ge 0.5 > P_b \ge Q_b$ and that $\overline{P} = 1 - P_b$. In the case that $P_a < 1 - P_b$ bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon_0$ such that $P_a + \varepsilon_0 > 1 - P_b$ In the case that $P_a > 1 - P_b$ bookmaker a can have the same betting volume and increase his profit by charging $P_a - \varepsilon_1$ such that $P_a - \varepsilon_1 > 1 - P_b$.

Suppose $P_a \ge 0.5 > P_b \ge Q_b$ and that $P_a = 1 - P_b$. Then bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon$ such that $P_a + \varepsilon > 1 - P_b$.

Suppose $P_a \ge 0.5 > Q_b \ge P_b$ and that $\overline{P} = 1 - P_b$. In the case that $P_a < 1 - P_b$ bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon_0$ such that $P_a + \varepsilon_0 > 1 - P_b$ In the case that $P_a > 1 - P_b$ bookmaker a can have the same betting volume and increase his profit by charging $P_a - \varepsilon_1$ such that $P_a - \varepsilon_1 > 1 - P_b$.

Suppose $P_a \ge 0.5 > Q_b \ge P_b$ and that $P_a = 1 - P_b$. Then bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon_0$ such that $P_a + \varepsilon_0 > 1 - P_b$ or charging $P_a - \varepsilon_1$ such that $P_a - \varepsilon_1 < P_b$.

Suppose $0.5 > P_a > P_b \ge Q_b$. Then bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon$ such that $P_a + \varepsilon > 1 - P_b$.

Suppose $0.5 > P_a \ge Q_b > P_b$. Then bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon_0$ such that $P_a + \varepsilon_0 > 1 - P_b$ or charging $P_a - \varepsilon_1$ such that $P_a - \varepsilon_1 < P_b$.

Suppose $0.5 > Q_b \ge P_a > P_b$. Then bookmaker a can attract all of the betting volume by charging $P_a + \varepsilon_0$ such that $P_a + \varepsilon_0 > 1 - P_b$ or charging $P_a - \varepsilon_1$ such that $P_a - \varepsilon_1 < P_b$.

Step 2:

The only situation remaining is $P_a = P_b = 1$ It is easy to see that in this case the bookmakers will share the profits and each bookmaker offers the best quote to both the unbiased bettors and the biased bettors on the favourite. Therefore, any quote deviations by a particular bookmaker will not attract any bettor.

Appendix F: Proof of proposition 4:

The proof for proposition 4 will be provided in two steps. In the first step I will show that it is impossible to reach an equilibrium in the situations described. In the second step I will show that the only sustainable equilibrium is $P_i = 1, P_j = 0$ for $i, j \in \{a, b\}$ and $i \neq j$.

Step 1:

The first step is divided in two cases. In the first case $P_a < 1$. In the second case $1 \geq P_b > 0$.

Case 1: $P_a < 1$.

In the case that $P_a < 1$ there will be no arbitrage opportunities since the condition for arbitrage is not met. This entails that arbitrageurs will not be able to construct an arbitrage bet and that only unbiased bettors and biased bettors need to be considered. From the previous section (proof of proposition 3) it follows that $P_a < 1$ can never be an equilibrium.

Case 2: $1 \geq P_b > 0$.

Suppose that $1 > P_b > 0$. In this case no arbitrage opportunities will be available since the condition for arbitrage is not met. This entails that arbitrageurs will not be able to construct an arbitrage bet and that only unbiased bettors and biased bettors need to be considered. From the previous section (proof of proposition 3) it follows that $1 > P_b > 0$ can never be an equilibrium.

Suppose that $P_b = 1$, then if $P_a < 1$ no arbitrage opportunities will be available since the condition for arbitrage is not met. This entails that arbitrageurs will not be able to construct an arbitrage bet and that only unbiased bettors and biased bettors need to be considered. From the previous section (proof of proposition 3) it follows that $P_b = 1$ and $P_a < 1$ can never be an equilibrium.

Suppose that $P_a = P_b = 1$. In this case no arbitrage opportunities will be available since the condition for arbitrage is not met. This entails that arbitrageurs will not be able to construct an arbitrage bet and that only unbiased bettors and biased bettors need to be considered. From the previous section (proof of proposition 3) it follows that $P_a = P_b = 1$ is an equilibrium in a market with unbiased and biased bettors. In a market with arbitrageurs the bookmakers do have an incentive to deviate since I assume that the volume being bet by the arbitrageurs is sufficiently large. So one of the bookmakers could deviate from $P=1$ to $P=0$ since this would allow the bookmaker to obtain a profit larger than would be the case when both bookmakers share the unbiased and biased bettors, given that the volume being bet by the arbitrageurs is sufficiently large.

Step 2:

In order to fulfill the condition for arbitrage it always needs to be the case that $P_a \geq 1/(1-P_b)$. It can easily be seen that this condition can only be met when $P_a = 1$ and $P_b = 0$. If one bookmaker were to deviate, the arbitrage opportunity will disappear and the bookmakers will loose the volume being bet by the arbitrageurs.

Appendix G: A list of the 9 sports supported by Rebelbetting

American Football, Aussie Rules, Baseball, Basketball, Ice Hockey, Rugby League, Rugby Union, Soccer & Tennis.

Appendix H: A list of the 52 bookmakers monitored by Rebelbetting

10Bet, 12Bet, 188bet, 888sport, 5Dimes, Admiral.at, Bet365, Bet-at-home, BetClic, BetCRIS, Betdaq, Betfair, Betfred, Betinternet, Betonline, Betsafe, Betsson, Bookmaker, Boylesports, Bwin, cashpoint, centrebet, Coral, Dafabet, DiamondSportsBook, Expekt, Gamebookers, Interwetten, Intertops, Ladbrokes, Luxbet, Marathonbet, Marathonbet.co.uk, Matchbook, Mybet, Nordicbet, Noxwin, Paddypower, Pinnacle Sports, SBObet, Smarkets, Sportsbet AU, StanJames, Tab AU, Tatts, TheGreek, Tipico, Titanbet, Unibet, William Hill, William Hill AU & Winner.

Appendix I: A list of the variables created during the collection period

- Date of entry
  - The date at which I was either entering the arbitrage opportunity or was updating the details of the arbitrage opportunity.
- Time of entry
  - The time at which I was either entering the arbitrage opportunity or was updating the details of the arbitrage opportunity.
- Time until last betting opportunity
  - The time that remained until the last possible opportunity to place a bet on a specific event. In the case that this was longer than an hour the variable was rounded to the nearest hour. The variable is measured in minutes, i.e. 120 minutes.

- Type of sport
    - The type of sport of a specific arbitrage opportunity.
- Type of bet
    - The combination of types of bets that were required in order to exploit the arbitrage opportunity.
- League
    - The league in which the event took place
- Teams/players
    - The teams or players that were involved in a particular arbitrage opportunity.
- Bookmakers
    - The bookmakers that offered the highest quotes for a specific sport event, i.e. the bookmakers that were involved in the arbitrage opportunity.
- Quotes
    - The quotes offered by the bookmakers with the highest quotes. These are the quotes with which the potential arbitrage profit can be calculated. One can simply read the quote as a multiplication factor of the initial stake placed.
- Profit
    - The potential arbitrage profit. This is the profit, net of fees, that the arbitrageur would obtain after exploiting the arbitrage opportunity.
- Duration
    - The duration of the arbitrage opportunity. In the case that this was longer than an hour the variable was rounded to the nearest hour. The variable is measured in minutes, i.e. 120 minutes.
- # of times
    - The number of times that a specific arbitrage opportunity occurred in the dataset.
- Change of profit
    - The change of the potential arbitrage profit compared to the last update of the potential arbitrage profit of that same arbitrage opportunity.
- Better option
    - In some cases a higher potential arbitrage profit could be obtained by using one or more different bookmakers than the bookmakers that previously had the best

quotes on a specific event. If this was the case a ‘’yes’’ would be entered in the better option variable.

- Change with better option
  - In case that a better option became available the change of the potential arbitrage profit with the better option compared to the last update of the potential arbitrage profit of that same arbitrage opportunity would be entered.
- Only available with different bookmaker
  - In some cases on of the bookmakers that originally offered a quote that was involved in the arbitrage opportunity dropped his quote such that when the arbitrageur would use this revised quote it would not be able to obtain a profit. In these cases a ‘’yes’’ was entered in the ‘’only available with different bookmaker’’ variable and the ‘’bookmakers’’ variable would be updated.

Appendix J: A list of the variables created after the collection period

- Country/countries of teams/players
  - The country/countries of the teams/players that were involved in a specific arbitrage opportunity.
- Same country dummy
  - The dummy equals 1 in the case that all teams or players are from the same country. The dummy would equal 0 when the teams or players are from different countries
- Book dummy
  - In order to construct an arbitrage bet two or three bookmakers are needed. The book dummy would equal 1 in the case that three bookmakers were required and 0 when only two bookmakers were required.
- Herfindahl
  - For each bookmaker the frequency in the dataset was determined. I than calculated the market share for each bookmaker per arbitrage opportunity. With this information the normalized Herfindahl index was calculated as follows: $Herfindahl = \frac{(H-\frac{1}{N})}{1-\frac{1}{N}} \; where \; H = \sum_{i=1}^{N} s_i^2$
- Public dummy
  - For each bookmaker I looked up whether the bookmaker or the parent company of the bookmaker was listed on a stock exchange. In the case that all

bookmakers involved in a specific arbitrage bet were listed on a stock exchange the public dummy would equal 1. If at least one of the bookmakers that was involved in a specific arbitrage bet was not listed on a stock exchange the dummy would equal 0.

- Sport dummy
  - I created a dummy variable for sports where 1 indicates American Football, 2 indicates Basketball, 3 indicates Hockey, 4 indicates Rugby, 5 indicates Soccer and 6 indicates Tennis. American Football was eventually omitted since there were too few observations. Basketball is omitted from the regression results in order to prevent the dummy variable trap. Hence, all other sports are relative to Basketball.
- Top 5
  - Some bookmakers appear more often in the data set than others. These bookmakers are ranked based on the frequency they are involved in arbitrage opportunities. The top 5 dummy equals 1 if all bookmakers involved in a specific arbitrage were in the top 5. The dummy equals 0 if at least one bookmaker involved in a specific arbitrage bet were not in the top 5.
- Member IBAS dummy
  - ‘’The Independent Betting Adjudication Service (IBAS)acts as an impartial adjudicator on disputes that arise between betting/gambling operators and their customers after they have been through the operator's own internal dispute procedures and if a deadlock exists. The IBAS Panel of betting experts apply their specialist knowledge to the facts and adjudicate by reference to the operator's own terms and conditions but do not rule on complex legal issues. As well as offering effective dispute resolution procedures, IBAS also check that operators have complied with the standards set by the Gambling Commission and with the IBAS terms and conditions of registration. IBAS rulings are binding on all parties, without prejudice to any legal proceedings that may be commenced subsequently’’ (IBAS).

    Bookmakers can choose to be a member of the Independent Betting Adjudication Service. The IBAS dummy equals 1 if all bookmakers involved in a specific arbitrage bet are members of the Independent Betting

Adjudication Service. The IBAS dummy equals 0 if at least one of the bookmakers that was involved in a specific arbitrage bet is not a member.

- Member UK gambling commission dummy
  - ‘’The Gambling Commission was setup under the Gambling Act 2005 to regulate commercial gambling in Great Britain in partnership with licensing authorities. The gambling commission is an independent non-departmental public body sponsored by the Department for Culture Media and Sports’’ (UK gambling commission).

  Bookmakers can obtain a license from the UK gambling commission. The member UK gambling commission dummy equals 1 if all bookmakers involved in a specific bet hold a license issued by the UK gambling commission. The dummy equals 0 if at least one of the bookmakers involved in a specific arbitrage bet does not hold a license issued by the UK gambling commission.

- Weak dummy
  - For each bookmaker I looked up the country from which the operating license was issued. The regulation on betting activities within these countries can either be strong or weak. The weak dummy equals 1 if all bookmakers involved in a specific arbitrage bet hold an operating license from a weakly regulated country. The weak dummy equals 0 if at least one of the bookmakers involved in a specific arbitrage bet does not hold a license from a weakly regulated country, i.e. a strongly regulated country. Weakly regulated countries are: Antigua & Berbuda, Costa Rica, Curacao, Gibraltar and Panama. Strongly regulated countries are: Australia, Isle of Man, Malta and the United Kingdom.

Appendix K: Additional bookmaker information

Number of bookmakers in dataset	40
Number of bookmakers that hold a license from a strong regulated country	25
Number of bookmakers that hold a license from a weak regulated	15
Number of bookmakers that are ibas members	17
Numbers of bookmakers that are not ibas members	23
Number of bookmakers that hold a license from the uk gambling commission	28
Number of bookmakers that do not hold a license from the uk gambling commission	12
Number of bookmakers that are listed on the stock exchange	14
Number of bookmakers that are not listed on the stock exchange	26
Number of arbitrage opportunities with two bookmakers	439
Number of arbitrage opportunities with three bookmakers	151
Number of observations with bookmakers that all have a license from a weak regulated country	362
Number of observations with bookmakers that are all ibas members	168
Number of observations with bookmakers that are all members of uk gambling commission	396
Number of observations with bookmakers that are all listed companies	24
Number of observations with bookmakers that are all in the top5	338
Average normalized Herfindahl index of all observations	0.3516699
Median normalized Herfindahl index	0.2935134

Bibliography:

Dixon, M., and P. Pope, 2004, The Value of Statistical Forecasts in the UK Association Football Betting Market, International Journal of Forecasting 1 (4), 697-711.

Edelman, D., and N. O'brien, 2004: Tote Arbitrage and Lock Opportunities in Racetrack Betting. The European Journal of Finance 10 (5), 370-378.

Fama, E., 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, The Journal of Finance 25 (2), 383-417.

Franck, E., E. Verbeek and S. Nüeusch, 2012, Inter-market Arbitrage in Betting, Forthcoming in Econometrica.

Golec, J., & M. Tamarkin, 1991, The Degree of Inefficiency in the Football Betting Market, Journal of Financial Economics 30, 311-323.

Hausch, D., and W. Ziemba, 1990, Arbitrage Strategies for Cross-Track Betting on Major Horse Races, The Journal of Business, Vol. 63, No. 1, Part 1, 61-78.

Hausch, D., W. Ziemba and M. Rubenstein, 1981, Efficiency of the Market for Racetrack Betting, Management Science 27, 1435-1452.

IBAS, http://www.ibas-uk.com/

Kuypers, T., 2000, Information efficiency: An Empirical Study of Fixed Odds Betting, Applied Economics 32, 1353-1363.

Marshall, B., 2009, How Quickly Is Temporary Market Inefficiency Removed?, The Quarterly Review of Economics and Finance 49, 917-930.

Paton, D., and L. V. Williams, 2005, Forecasting Outcomes in Spread Betting Markets: Can Bettors Use 'Quarbs' to Beat the Book?, Journal of Forecasting 24, 139-154.

Pope, P., & D. Peel, 1989, Information, Prices and Efficiency in a Fixed-Odds Betting Market, Economica 56, 323-341.

Rebelbetting, www.rebelbetting.com

Sauer, R. D., 1998, The Economics of Wagering Markets, Journal of Economic Literature, Vol. 36, 2021-2064.

Levitt, S. D., 2004, Why Are Gambling Markets Organized So Differently From Financial Markets?, The Economic Journal 114, 223-246.

Thaler, R., and W. T. Ziemba, 1988, Parimutual Betting Markets: Racetracks and Lotteries. Journal of Economic Perspectives 2, 161-174.

UK Gambling commission, http://www.gamblingcommission.gov.uk/

Verbeek, R., 2012, Arbitrage Opportunities and Biased Betting Behaviour, Working Paper, University of Amsterdam.

Vlastakis, V., G. Dotsis and R. N. Markellos (2009), How Efficient is the European Football Betting Market? Evidence from Arbitrage and Trading Strategies, Journal of Forecasting 28, 426-444.

William Hill Annual Report 2012.

Interview: Andrew Black (founder and former CEO of Betfair, the largest betting exchange in the world)

Interview: Robert Markwick (angel investor Betfair)

Interviews conducted in Malta with multiple employees working for different bookmakers