Optimal Pricing in the Online Betting Market

Maurizio Montone*

April 2015

Abstract

I find that the optimal price of a bet for a risk-averse bookmaker is a function of elasticity of demand, operating costs, and the number of outcomes of the betting event. Such a price, however, is dynamically adjusted if the flow of bets follows an unexpected pattern, thus generating arbitrage opportunities for investors. An empirical analysis of the online betting market supports these predictions. I show that (1) bookmakers with greater market power earn higher expected returns; (2) the inclusion of an additional outcome prompts an increase in markup; and (3) arbitrage opportunities arise on a daily basis, and increase in frequency at the end of the betting period. The results suggest that bookmakers' attitude towards risk is more important than asymmetric information or insider trading to explain betting market prices.

*Erasmus University Rotterdam, Erasmus School of Economics, 3000DR Rotterdam, The Netherlands. I would like to thank Franklin Allen, Gennaro Bernile, Jeffrey Butler, Marco Pagano, and Facundo Piguillem, for many precious insights and suggestions. I also thank participants at the International Atlantic Economic Society (IAES) Best Undergraduate Paper Competition 2008 Final in Montreal for many helpful comments.

1 Introduction

The betting market is widely recognized as an ideal framework for economists to analyze market efficiency (Thaler and Ziemba, 1988; Gandar et al., 1988; Camerer, 1989; Golec and Tamarkin, 1991; Brown and Sauer, 1993a, 1993b; Woodland and Woodland, 1994; Dare and MacDonald, 1996; Gray and Gray, 1997; Sauer, 1998). Since the advent of online trading, the market has changed dramatically (see e.g. Mainelli and Dibb, 2004). The literature, however, has kept a surprisingly narrow perspective (Sauer, 2005), either considering profit maximization problems in isolation (Kuypers, 2000; Strumpf, 2003; Levitt, 2004), or searching for pricing errors (Forrest et al., 2005; Forrest and Simmons, 2008; Vlastakis et al., 2009), without studying equilibrium prices. This paper tries to tackle this issue, both theoretically and empirically, with an asset pricing model of the optimal markup in the online betting market.

I find that the optimal price of a bet for a risk-averse bookmaker is a function of the elasticity of demand, operating costs, and the number of outcomes of the betting event. Such a price, however, is dynamically adjusted if the flow of bets follows an unexpected pattern, thus generating arbitrage opportunities for investors. An empirical analysis of the online betting market supports these predictions. Using survey data on bookmaker characteristics and a unique dataset on real-time prices, I show that (1) bookmakers with greater market power earn higher expected returns; (2) the inclusion of an additional outcome prompts an increase in markup; and (3) arbitrage opportunities arise on a daily basis, with an instantaneous gross rate of return of 1.14% per operation, and increase in frequency at the end of the betting period.

The results suggest that bookmaker preferences are more important than asymmetric information or insider trading to explain betting market prices. This is due to two reasons. First, asymmetric information should be more severe at the beginning of the betting period (Gandar et al., 1998), as the information set typically improves over time, especially for sporting events. Second, if some traders could observe a private signal (Shin, 1991, 1992), they would have an incentive to bet on that outcome only, which implies that all bookmakers should adjust odds in the same direction. I show that these two predictions are counter-factual, as the odds adjustments take place at the end of the betting period, and in a variety of directions.

Even in the absence of asymmetric information, one could argue that the odds adjustment may be simply driven by a price discovery process, i.e. changes in probability estimates rather than markup. If so, the dispersion in bookmakers' estimates should decrease as the betting event approaches, rather than increase, as prices converge to the correct value. I show, however, that this prediction is counter-factual too. In fact, the dispersion in bookmakers'odds actually increases over the betting period, reaching its peak on the day of the betting event.

The paper also sheds new light on the findings of Shin (1993), who argues that a positive relationship between the optimal markup and the number of outcomes represents compensation for insider trading, in a competitive market with risk-neutral bookmakers. I show that this relationship also holds for risk-averse bookmakers, even in the absence of asymmetric information, because a higher number of outcomes inflates the variance of both conditional and unconditional profits. Consistent with this idea, I show that bookmakers exhibit risk-averse behavior and earn higher expected returns on three-outcome bets than they do on two-outcome bets.

A bet represents an elementary Arrow-Debreu security, which makes betting markets rather similar to financial markets in many respects (Jaffe and Winkler, 1976). In its simplest form, it requires the specification of an event made up of a finite number of mutually exclusive outcomes and the participation of two counterparties: a gambler, who takes a long position; and a bookmaker, who takes a short position. The contract works as follows. The gambler picks a particular outcome and deposits an amount of money, called *stake*, with the bookmaker. If the outcome is realized, the bookmaker pays the stake back to the gambler, multiplied by a coefficient called *odd*. Otherwise, no further transaction occurs and the bookmaker keeps the stake.

Research shows that bookmakers are unlikely to trade at a loss (Kuypers, 2000; Levitt, 2004), because they typically use odd compilers, i.e. experts who engage in a sophisticated analysis of past data and provide unbiased estimates of the probabilities of outcomes (Gandar et al., 1988; Forrest et al., 2005). I build on these findings in my theoretical framework, and propose a betting market model in which bookmakers are as least as well informed as gamblers. Therefore, I rule out the instance of insider trading such as e.g. match fixing. For all practical purposes, however, this assumption is irrelevant: many online bookmakers set wagering limits, which makes it impossible for perfectly informed inside traders to realize arbitrarily large profits. I also rule out the possibility that bookmakers take advantage of gambler biases, if any (Camerer, 1989; Brown and Sauer, 1993a; Dare and MacDonald, 1996). In fact, as long as they can also take long positions, they can arbitrage other bookmakers in case they set the wrong prices.

The model predictions can be summed up as follows. For a risk-averse bookmaker, the optimal price of a bet is a function of operating costs, which I assume to be a fixed proportion of revenues, and the following three discriminating factors: (1) the elasticity of the bookmaker's demand, which is a function of gambler preferences and the bookmaker's market share; (2) the number of outcomes, as more outcomes imply more volatile profits and therefore a higher markup; and (3) the residual length of the betting period, as bookmakers have an incentive to dynamically adjust their odds over time to reduce the conditional variance of their profits.

In the second part of the paper, I test these predictions in the online betting market. This is an instructive setting to do research for at least two reasons. First, it is international and, with only a few exceptions, has no formal boundaries. Then, it should be more efficient than old-fashioned local betting shops. Second, this market as a whole is still relatively unexplored. To the best of my knowledge, this is the first paper that looks into the optimal pricing in this market, or the business practices of its participants.

In order to test the first two model predictions, I use a survey of 82 online bookmakers from the website of bookmaker statistics "Top100bookmakers" (www.top100bookmakers.com), complemented by information from the official bookmaker websites, collected on August 31, 2012. For each bookmaker, I observe the markup applied to bets on six major sports, including baseball, basketball, football, hockey, soccer, and tennis, as well as a variety of operating and structural characteristics of the business. The operating features include promotional services offered, safety of transactions, customer service, and wager restrictions. The structural features are age, Alexa daily reach[1], number of languages supported, and number of currencies accepted.

The number of languages and currencies, in particular, play a key role in my identification strategy. In fact, these two variables are likely to affect betting in different ways. A gambler from e.g. the euro area is typically unable to place a bet with a bookmaker who only accepts deposits in dollars, or has to pay a fee to do so. Similarly, a gambler is unlikely to open an account with a bookmaker whose language she does not understand. Nonetheless, there is an important difference between the two cases. In the first case, it is either impossible or inefficient for the gambler to place the bet in her own currency. In the second case, the gambler may still place the bet provided she knows some basic English, which is the common language of all online bookmakers. As a consequence, currency areas can be thought of as *de facto* market segments. The number of languages, instead, may constitute a proxy for the bookmaker's market share. In fact, supporting more languages allows the

[1]Defined as the number of daily visits to the bookmaker's website estimated by Alexa Internet Inc.

bookmaker to reach non-English speakers as well, which increases the size of her clientele within a given currency segment.

In order to test these predictions, I look into bookmakers' overall expected returns, defined as the equal-weighted average of the expected returns on all six major sports. The mean is 5.83%, which exactly coincides with the median, and ranges from 2.37% to 10.35% across bookmakers. Consistent with the first model prediction, I find that a 50% increase in the number of languages supported is associated with an increase in overall expected returns by 16.5 basis points per bet (t-stat 2.10). This finding is consistent with the idea that bookmakers exploit market frictions, such as e.g. geographic separation, to discriminate prices (Bruce and Johnson, 2001; Strumpf, 2003). On the operating side, I find that one additional promotion offered is associated with a decrease in expected returns by 48 basis points (t-stat -2.97), and a 20% increase in customer service quality is associated with a decrease in expected returns by 28 basis points (t-stat -1.93). Therefore, there is a negative relationship between markup and services, which suggests that services may be a tool used by bookmakers with low market power in order to increase their demand.

These regressions, however, have three issues. First, I only consider overall returns, rather than distinguish between sports categories. Second, most variables are highly correlated, which may inflate standard errors and then bias t-stats downwards. Third, the sample size is relatively small, with 82 observations. For these reasons, I repeat the tests in a panel setting, in which I consider the markup set by each bookmaker on each of the six major sports in the sample. The number of observations then increases to nearly 450. This setting also allows me to control – where possible – for sports fixed effects, which picks up differences in popularity across sports, and bookmaker fixed effects, which captures firm-level unobserved characteristics, and cluster standard errors at the bookmaker level.

The estimates are similar in sign, magnitude, and statistical significance. I find that a 50% increase in the number of languages supported is associated with an increase in expected returns by 20 basis points (t-stat 2.91). These results are consistent with the idea that bookmakers with greater market power face less elastic demand, and therefore earn higher expected returns.

Next, I introduce a new variable in the analysis, defined as the difference between the number of payment methods accepted for deposits and the number of payment methods available for withdrawals, which I call "net options". Fairness would require that these numbers be equal. Surprisingly, however, this is not the case: the difference between these two numbers is large and highly significant (5.23, t-stat 8.84), which suggests that bookmakers tend to make it easier to deposit money than to withdraw it. Therefore, I take this variable as a proxy for the bookmaker's trustworthiness. Consistent with this interpretation, I find that a unit increase in the number of net options is associated with a decrease in expected returns by 4 basis points, , even though slightly outside of the rejection region (t-stat -1.49). Therefore, there is mild evidence that a lower level of trustworthiness is (slightly) penalized by investors through lower expected returns.

The model predicts that risk-averse bookmakers should set a higher markup for events with a greater number of outcomes, as they imply a higher volatility of profits. On the contrary, a risk-neutral bookmaker would not require compensation for that. I test this implication as follows. The available markup estimates refer to six major sports. Four of them, i.e. baseball, basketball, football, and tennis, typically represent two-outcome bets, the outcome being the victory of either side. In fact, even though a tie is theoretically possible, it is extremely rare due to the high-scoring nature of these sports. Hence, most bookmakers do not accept bets on it. On the contrary, soccer and hockey are sports with much lower scores, which makes ties very common. As such, they typically constitute three-outcome bets. Then, I define the three-outcome premium as the difference in average markup between three-outcome and two-outcome bets.

Consistent with the risk-aversion hypothesis, I find that the premium is positive and highly significant both in the time series regression (1.40%, t-stat 8.87) and in the panel setting (1.37%, t-stat 8.56), in which I control for bookmaker fixed-effects.

In the last part of the paper, I test the third model prediction and try to shed some light on the phenomenon of arbitrages in the online betting market. The model predicts that risk-averse bookmakers have an incentive to dynamically adjust their odds to reduce the conditional variance of profits. In turn, this mechanism should generate arbitrage opportunities for investors.

Note that there are three possible reasons as to why bookmakers may want to change their prices (see e.g. Gandar et al., 1988): (1) release of new information after betting begins; (2) prediction errors on the part of the aggregate betting public and/or bookmakers; (3) the randomness of order flow. In the model I propose, bookmakers produce unbiased predictions. This is a realistic setting for the online betting market, because even if there were heterogeneity in predictive skills across bookmakers, the low-skilled price setters would be able to readily mimic the correct prices set by the high-skilled ones in real time. For this reason, the adjustment is unlikely to be driven by insider trading (Gandar et al., 1998), and initial prices are likely to be aligned. Note that in such a setting, the release of new information would move all prices in the same direction, and the same would happen if gamblers were prone to sentiment. In light of this, as long as there exists a subset of informed bookmakers, the only likely reason that moves prices in the online betting market is random order flow.

In order to test this prediction, I use a a time-series of arbitrage opportunities available in the online betting market for sporting events from January 1$^{\text{st}}$ to April 30$^{\text{th}}$, 2008. The source is the website "Infobetting" (www.infobetting.org), which provides two main services: (1) comparison of real-time odds posted by a large sample of online bookmakers for a variety of major sports, and (2) notification of arbitrage opportunities for the observed betting events. The data have been personally collected by the author in real time twice a day, at 2pm and 8pm respectively, for a total of 242 observations. The sample period captures the most important part of the season for most of the sports considered, which in turn should imply enough betting activity to test the odds adjustment hypothesis. The choice of the timing, on the other hand, reflects the fact that most sporting events take place in the afternoon between 2pm and 8pm (Brown and Sauer, 1993b), hence odds adjustments should be wider within that time frame.

Interestingly, I observe at least one arbitrage opportunity per session. On average, I find that the number of such opportunities increases by 7 units per session on weekends (t-stat 6.40), and by 4 units on match days (t-stat 3.37). These effects are significantly stronger for the afternoon sessions, which increases the number of opportunities by 6 units on weekends (t-stat 3.85), and by another 6 units on match days (t-stat 3.46). These findings reflect three facts: (1) most gamblers are likely to place their bets over the weekend, when they have more spare time; (2) the odds adjustment is more likely to happen on match days, as the betting period comes to an end; (3) most matches take place before 8pm, so most of the action in the market from both bookmakers and gamblers happens before then.

As a last test, I analyze which characteristics make bookmakers more likely to generate arbitrage opportunities. I find that a 50% increase in the number of languages supported is associated with an increase in arbitrage opportunities created by 4 units (t-stat 1.76), and a 50% increase in the number of currencies accepted is associated with an increase in arbitrage opportunities created by 6 units (t-stat 2.81). These results reflect the fact that the bookmaker's presence in more linguistic and currency areas makes the odds accessible to a larger number of gamblers, which then makes it more likely to create an arbitrage opportunity. I also find that a 50% increase in the bookmaker's age is associated with an increase in the probability of producing at least one arbitrage opportunity by 9.17% (t-stat 3.10), which indicates that the bookmakers that have stronger reputation can also afford to adjust their odds more aggressively.

Overall, the results are consistent with the idea that bookmakers are risk-averse. As such, they are consistent with Fingleton and Waldron (1999), and stand in contrast with the hypothesis of bookmakers' risk-neutral preferences (Shin 1991, 1992, 1993), and risk-seeking preferences (Strumpf, 2003). The evidence on arbitrage opportunities I provide is similar to Vlastakis, Dotsis and Markellos (2009), but I show that this is a phenomenon that is not limited to a few online bookmakers, but rather it is a widespread feature of the market. Importantly, the presence of such opportunities is not necessarily inconsistent with market efficiency (see e.g. Sauer, 1998), and has a theoretical foundation in the management of profit volatility.

The paper proceeds as follows. Section 2 introduces the model. Section 3 presents the data. Section 4 discusses the empirical results. Section 5 concludes.

2 A betting market model

2.1 Fundamentals of betting

Betting markets are similar to financial markets in many aspects (Jaffe and Winkler, 1976), as a bet represents an elementary Arrow-Debreu security. The contract requires the specification of an event made up of n mutually exclusive outcomes, and the participation of two types of agents, a gambler and a bookmaker. The contract works as follows. At time 0, the gambler takes a financial position on one of the n outcomes by transferring an amount of money S, called *stake*, to the bookmaker. At time 1, the event unfolds. If the realized outcome coincides with the one specified in the contract, the bookmaker will pay the gambler an amount of money equal to S multiplied by a pre-specified coefficient q, called the *odd*, which then represents a gross rate of return. Otherwise, no transaction occurs and the bookmaker retains S. Then, the bookmaker's expected profit can be expressed as follows:

$$E(\pi_b) = (1 - p)S + p(S - qS) \tag{1}$$

where p is the probability that the outcome occurs. Equation (1) can be rearranged more compactly as:

$$E(\pi_b) = S(pq - 1) \tag{2}$$

Note that there are three main variables of interest. The stake S, which is the gambler's control variable; the odd q, which is the bookmaker's control variable; and the probability p, which is exogenous. Since this is a zero-sum game, the gambler's profit is symmetric to the bookmaker's, so that:

$$E(\pi_b) + E(\pi_g) = 0 \tag{3}$$

An odd is said to be *fair* if it makes the expected profit equal to zero on both sides. Solving out from (2):

$$q^F = \frac{1}{p} \tag{4}$$

Therefore, a bookmaker has a positive expected profit if she sets the odds such that $q < q^F$. In general, the odd can be expressed as:

$$q = \alpha q^F \tag{5}$$

where $\alpha \in (0, 1)$ if the bookmaker can apply a markup and make (2) positive. Therefore, alpha and the markup are inversely related. Combining (4) and (5), alpha can be expressed as:

$$\alpha = pq \tag{6}$$

which then allows to rearrange equation (2) as:

$$E(\pi_b) = S(\alpha - 1) \tag{7}$$

If the bookmaker accepts bets on all n outcomes of the betting event, (6) can be summed over all i's and yield:

$$\sum_{i=1}^{n} p_i = \alpha \sum_{i=1}^{n} \frac{1}{q_i} \tag{8}$$

and since $\sum_{i=1}^{n} p_i = 1$, alpha can be expressed as a function of the odds:

$$\alpha = \left(\sum_{i=1}^{n} \frac{1}{q_i} \right)^{-1} \tag{9}$$

which makes it possible to observe the bookmaker's markup even without knowing the probabilities of the outcomes.

Note that (9) also has an interpretation in terms of returns. In fact, since the bookmaker's revenues are equal to S, and the expected costs are αS, then the gross expected return for the bookmaker is:

$$E(1 + r_b) = \frac{1}{\alpha} \tag{10}$$

which in betting market jargon is known as the "overround" or "vigorish" (Levitt, 2004, Forrest and Simmons, 2008). The question remains, however, as to why a gambler would accept a bet with a negative expected return. There three possible answers: she may (1) exhibit risk-seeking preferences (Weitzman, 1965); (2) hold biased beliefs (see e.g. Ziemba and Hausch, 1987); or (3) draw some utility from gambling (Conslik, 1993). The present work, however, does not take a stance on this issue, but rather focuses on the activity and preferences of bookmakers: as long as they are informed about the probability of outcomes and can take both long and short positions, the beliefs of gamblers do not play a central role in this setting. I show this below.

2.2 Bookmakers' risk-return trade-off

In this section, I analyze how a bookmaker's risk-return trade-off is affected by gamblers' beliefs, operating costs, and business expansion.

2.2.1 Gamblers' beliefs

If there is more than one bookmaker in the market, profits may not be equal across states even if gamblers' probability beliefs are correct. In fact, unless the bookmaker is a monopolist, her clientele represents a random sample from the gamblers' population and her revenues x_b will be less than the market revenues X. Then, the proportion of bets the bookmaker receives on outcome i will be p_{ib}, with possibly $p_{ib} \neq p_i$ for some i.

This fact affects the bookmaker's risk-return trade-off. Expected profits are unaffected, both conditionally:

$$E(\pi_b|P_b) = x_b - \sum_{i=1}^{n} p_i(p_{ib}x_b)q_i = x_b - x_b \sum_{i=1}^{n} p_i p_{ib} \frac{\alpha_b}{p_i} = x_b(1-\alpha_b) \tag{11}$$

and unconditionally:

$$E(\pi_b) = x_b - \sum_{i=1}^{n} p_i(p_i x_b)q_i = x_b - x_b \sum_{i=1}^{n} p_i p_i \frac{\alpha_b}{p_i} = x_b(1-\alpha_b) \tag{12}$$

where $P_b = (p_{1b}, \ldots, p_{nb})$ is the vector of all p_{ib} realizations. On the contrary, the volatility of profits is affected, both conditionally:

$$var(\pi_b|P_b) = \sum_{i=1}^{n} p_i \left((\pi_{ib}|P_b) - E(\pi_b|P_b)\right)^2 = (\alpha_b x_b)^2 \sum_{i=1}^{n} \frac{(p_i - p_{ib})^2}{p_i} \tag{13}$$

and unconditionally:

$$var(\pi_b) = (\alpha_b x_b)^2 \sum_{i=1}^{n} \frac{E\left((p_i - p_{ib})^2\right)}{p_i} = (\alpha_b x_b)^2 \sum_{i=1}^{n} \frac{\sigma_{ib}^2}{p_i} = (\alpha_b x_b)^2 (n-1) \frac{X - x_b}{X - 1} \tag{14}$$

where $var(\pi_b) \equiv E\left(var(\pi_b|P_b)\right)$, and I use the fact that p_{ib}, the sample proportion, has mean equal to p_i and variance equal to:

$$\sigma_{ib}^2 = \frac{p_i(1-p_i)}{x_b} \frac{X - x_b}{X - 1} \tag{15}$$

as the population of bets has finite size X.

If gambler beliefs are biased, the expected proportion of bets on outcome i for the bookmaker becomes:

$$E(\hat{p}_{ib}) \equiv \hat{p}_i = p_i + \delta_i \tag{16}$$

which again does not affect expected profits:

$$E(\pi_b) = x_b - \sum_{i=1}^{n} p_i(\hat{p}_i x_b)q_i = x_b - x_b \sum_{i=1}^{n} p_i \hat{p}_i \frac{\alpha_b}{p_i} = x_b(1-\alpha_b) \tag{17}$$

but does affect the variance, as it implies one extra-term with respect to (14):

$$var(\pi_b) = (\alpha_b x_b)^2 \sum_{i=1}^{n} \frac{var(\hat{p}_{ib}) + \delta_i^2}{p_i} \tag{18}$$

which is due to the fact that the bookmaker cannot exploit the gamblers' bias by incorporating it in the odds, or she would be arbitraged by other informed traders.

2.2.2 Operating costs

Operating costs include the bookmaker's license fees as well as labor and equipment costs. In this setting, I express them as a fraction t of her revenues. The effect of such costs is again asymmetric, but in this case expected profits are affected:

$$E(\pi_b) = x_b(1-t) - \sum_{i=1}^{n} p_i(p_i x_b)q_i = x_b(1-t) - x_b \sum_{i=1}^{n} p_i p_i \frac{\alpha_b}{p_i} = x_b(1-\alpha_b-t) \tag{19}$$

while the volatility of profits is not:

$$var(\pi_b) = E\left(x_b(1-t) - (p_i x_b)q_i - x_b(1-\alpha_b-t)\right)^2 = E\left(x_b - (p_i x_b)q_i) - x_b(1-\alpha_b)\right)^2 \tag{20}$$

as the terms in t cancel out.

2.2.3 Business expansion

A bookmaker may increase her demand using tools other than price. In particular, she can expand her business either in her own market, for instance by offering promotions or better customer services, or in different market segments. Following an increase in demand, her expected profits increase:

$$\frac{\partial E(\pi_b)}{\partial x_b} = 1 - \alpha_b(1 - |\epsilon_{\alpha_b,x_b}|) > 0 \tag{21}$$

for any value of ϵ_{α_b,x_b}, the elasticity of supply, and the effect is increasing in the bookmaker's market power, i.e. decreasing in alpha. The effect on the variance of profits, instead, is ambiguous:

$$\frac{\partial var(\pi_b)}{\partial x_b} = \alpha_b^2 \frac{n-1}{X-1}\left(-2x_b + X\left(1 + |\epsilon_{X,x_b}|\frac{x_b-1}{X-1}\right) - 2|\epsilon_{\alpha_b,x_b}|(X-x_b)\right) \tag{22}$$

where ϵ_{X,x_b} represents the percentage change in the pool of bets (X) that the bookmaker faces if the increase in demand (x_b) comes from a new market, and has a positive effect on the variance of profits, in that it implies a decrease in the bookmaker's size with respect to the whole market. On the contrary, ϵ_{X,x_b} is zero if the bookmaker decides to expand within the same market, and the increase in demand prompts a decrease in profit volatility, provided that the bookmaker's market power is above the following threshold:

$$|\epsilon_{\alpha_b,x_b}| > \frac{X - 2x_b}{2X - 2x_b} \tag{23}$$

Hence, the benefit of expansion is larger for the most successful bookmakers, who can pursue higher expected profits and lower variance.

2.3 Optimal pricing

Next, I consider the problem of setting the optimal odd for a risk-neutral and a risk-averse bookmaker, in the presence of operating costs.

2.3.1 Risk-neutral preferences

If the bookmaker is risk-neutral, she maximizes her expected profits with respect to alpha and obtains:

$$\alpha^* = \frac{1-t}{1+\frac{1}{\epsilon_{x_b,\alpha_b}}} \tag{24}$$

where $\epsilon_{x_b,\alpha_b} > 0$ represents the elasticity of demand for the bookmaker's services. Then, in a perfectly competitive market, infinitely elastic demand for the firm implies that the optimal alpha is entirely determined by operating costs. The comparative statics yield the following two results:

$$\frac{\partial \alpha_b^*}{\partial t} < 0 \tag{25}$$

$$\frac{\partial \alpha_b^*}{\partial \epsilon_{x_b,\alpha_b}} > 0 \tag{26}$$

which indicate that the optimal alpha decreases with operating costs, and increases with the elasticity of demand. In particular, note that alpha will be less than one if (1) operating costs are non-zero, i.e. $t \in (0, 1)$, and (2) for finite values of the elasticity, which is the case for an oligopolistic market.

In this setting, the variance of profits does not matter to the bookmaker.

2.3.2 Risk-averse preferences

If the bookmaker is risk-averse, she will seek an optimal trade-off between risk and return. Here, I consider the case of mean-variance preferences. The bookmaker then solves:

$$\max_{\alpha} E(u) = E(\pi_b) - \frac{\gamma_b}{2} var(\pi_b) \tag{27}$$

and the first-order condition is the following non-linear expression for alpha:

$$\epsilon_{x_b,\alpha_b} \frac{1-\alpha_b - t_b}{\alpha_b} - 1 - \gamma_b \alpha_b \frac{n-1}{X-1} \left(X \left(1 - \frac{\epsilon_{x_b,\alpha_b}}{2} \right) - x_b(1+\epsilon_{x_b,\alpha_b}) \right) = 0 \tag{28}$$

The comparative statics yield three results. The first two are qualitatively analogous to (25) and (26) from the risk-neutral case. The third one, instead, is as follows:

$$\frac{\partial \alpha_b^*}{\partial n} < 0 \tag{29}$$

and indicates that the optimal alpha decreases with the number of outcomes of a bet. In fact, an increase in this number increases profit volatility without affecting expected profits.

2.4 Book imbalance

At the end of the betting period, after all bets are made, the bookmaker's state profits can be expressed as:

$$\pi_{ib} = X - (p_{ib}X)q_i = X - Xp_{ib}\frac{\alpha_b}{p_i} = X\left(1 - \alpha_b \frac{p_{ib}}{p_i}\right) \tag{30}$$

for all i. Note that the markup here may serve as an insurance mechanism for the bookmaker. In fact, even if $p_{ib} > p_i$ for some i, all of her state profits may still be positive provided the following condition holds for all outcomes:

$$\alpha_b < \frac{p_i}{\hat{p}_i} \tag{31}$$

However, the profits may still not be equal in all states of nature. In this case, the bookmaker's book is said to be unbalanced. A risk-neutral bookmaker would not be bothered in this instance, but a risk-averse bookmaker would. In particular, she has the possibility to adjust the conditional variance of profits leaving expected returns unchanged. The intuition is as follows. Suppose the realization of vector P_b is far off from the vector of true probabilities P, but the betting period is not over yet[2]. In this case, the bookmaker may stimulate additional flows of bets so as to rebalance her book (Gandar et al., 1988). In fact, she may increase the markup on the outcomes that have received a greater proportion of bets, while simultaneously decreasing the flow for the other outcomes, thus leaving the average markup on the event still at the same level but decreasing profit volatility.

The purpose of such adjustments is to make gamblers' expected profits temporarily unequal across outcomes, thus generating a new investment pattern that can make p_{ib} closer to p_i for all i. To see how this is done, consider a two-outcome bet where the bookmaker's current state profits are $\pi_1 > 0$ and $\pi_2 < 0$. The bookmaker might then set a lower odd (i.e. higher markup) for outcome one and a higher odd (i.e. lower markup) for outcome two. This change would immediately make the second outcome more attractive to gamblers, thus generating more investments on that side of the book. Symmetrically, the first outcome would become less attractive, thus generating a lower amount of additional bets. Once the book is rebalanced, the bookmaker can switch the markup back to unprofitable levels for both outcomes. This is why the adjustment is dynamic, as it can be enforced during the betting period.

One important issue is how to determine the optimal odds adjustments. This can be thought of as a trial and error process, in which the bookmaker can increase an odd even above its fair value (i.e. $q_i > q_i^F$) for a short period of time. In fact, she does not lose money on that outcome as long as the additional stakes for the overvalued odd help her rebalance her book. At the end of the betting period, the state profits for the bookmaker will look as follows:

$$\pi_{1b} = \sum_{t=1}^{T} (x_{1b}(t) + x_{2b}(t)) - \sum_{t=1}^{T} (q_{1b}(t)x_{1b}(t)) \tag{32}$$

$$\pi_{2b} = \sum_{t=1}^{T} (x_{1b}(t) + x_{2b}(t)) - \sum_{t=1}^{T} (q_{2b}(t)x_{2b}(t)) \tag{33}$$

[2]The betting period for a given event is the time span between the moment a bookmaker starts accepting bets and the moment she stops doing so. Usually, the closing moment is just a few hours or even minutes before the event takes place.

where each t represents the time interval between two consecutive odds adjustments. For each t then, the bookmaker will post a different set of odds and receive a different flow of bets. From (32) and (33), it is clear that a risk-averse bookmaker should be primarily concerned with making the two state-contingent costs as close as possible in order to reduce the volatility of profits across states. To this end, she will not mind setting an overvalued odd for a limited period of time t, as long as total costs are equal and less than her revenues at the end of the betting period. The bookmaker's average return on outcome i will be:

$$1 + \bar{r}_b \cong \frac{1}{\bar{\alpha}_{ib}} \frac{p_i}{p_{ib}} \tag{34}$$

where $\bar{\alpha}_{ib}$ is the average markup (weighted by time intervals) applied on odd i during the betting period, which is approximately equal to the optimal overround:

$$\frac{1}{\bar{\alpha}_{ib}} \frac{p_i}{p_{ib}} \cong \frac{1}{\alpha_b^*} \tag{35}$$

Hence, if a bookmaker faces an over-investment in outcome i, i.e. $p_{ib} > p_i$, she will temporarily set the average markup on that outcome below its equilibrium value, i.e. $\bar{\alpha}_{ib} < \alpha_b^*$. In case of under-investment, instead, $\bar{\alpha}_{ib} > \alpha_b^*$.

Note that the same strategy can be used to increase profit volatility, which is the case in which the bookmaker takes a position on a given outcome herself (Strumpf, 2003). Whether the odds adjustment takes place and why, then, is an empirical question, and I try to shed some light on this issue below.

2.5 Arbitrages

If the odds adjustment is wide enough over time, it may give rise not only to positive expected returns but also to arbitrage opportunities. In fact, all it takes is that different bookmakers simultaneously set complementary odds above the fair value. To see this, consider a two-outcome bet in a market with two bookmakers A and B. Suppose their state profits are currently $\pi_{1A} > 0$, $\pi_{2A} < 0$ and $\pi_{1B} < 0$, $\pi_{2B} > 0$ respectively. In order to rebalance their books, the bookmakers may independently decide to increase the odds for the states with negative profits (q_{2A} and q_{1B}) and simultaneously decrease the other odds (q_{1A} and q_{2B}). If the increase is large enough, an arbitrage opportunity arises.

This is easy to prove. Let w_0 be the arbitrageur's investment. Since an arbitrage implies making a positive and riskless profit in all states of nature, she has to simultaneously invest in all complementary outcomes of a given event. Let λ_i be the fraction of w_0 she invests in outcome i, so that her capital outlay for that outcome is $\lambda_i w_0$. The arbitrage operation implies:

$$(\lambda_i w_0) q_i^d > w_0 \tag{36}$$

for all i, where q_i^d is the dominant odd for outcome i, i.e. the highest odd available in the market for that outcome. Equivalently, (36) can be rearranged as:

$$\lambda_i > \frac{1}{q_i^d} \tag{37}$$

Summing over all i's, and using the fact that $\sum_{i=1}^{n} \lambda_i = 1$:

$$\sum_{i=1}^{n} \frac{1}{q_i^d} < 1 \tag{38}$$

which represents the arbitrage condition, i.e. the inequality that must be satisfied by all odds on a set of complementary outcomes in order for an arbitrage to be profitable. The arbitrageur's revenues can be expressed as:

$$w_1 = w_0 \left(\sum_{i=1}^{n} \frac{1}{q_i^d} \right)^{-1} \tag{39}$$

which implies the following riskless return:

$$1 + r_f = \left(\sum_{i=1}^{n} \frac{1}{q_i^d} \right)^{-1} \tag{40}$$

and the following optimal investment in each outcome:

$$\lambda_i^* w_0 = \frac{w_1}{q_i^d} \tag{41}$$

The striking feature of such arbitrage opportunities is that they are Pareto-efficient. In fact, they help bookmakers rebalance their books, and simultaneously earn arbitrageurs risk-free money.

2.6 Model predictions

The model predictions can be summed up as follows. First, the price of a bet is determined by (1) operating costs, which I assume to be proportionally equal across bookmakers, and (2) the elasticity of the bookmaker's demand with respect to alpha, which is a function of gambler preferences and market share. Second, risk-averse bookmakers (1) charge a higher markup on events made up of a greater number of outcomes, and (2) have an incentive to dynamically adjust their odds in order to reduce the conditional variance of their profits. In the next section, I take these predictions to the data.

3 The data

In order to test the model predictions, I perform an analysis of the online betting market. This is a relatively new framework, as it took place with the advent of the internet and has been growing steadily since the early 2000s (see e.g. Mainelli and Dibb, 2004). It also represents an instructive setting to do research for at least two reasons. First, it is international and – with only a few exceptions – has no formal boundaries. Then, it should be more efficient than old-fashioned local betting shops. Secondly, this market as a whole is still relatively unexplored. To the best of my knowledge, there is no formal study that looks into the optimal pricing of this market, or the business practices of its participants.

I use three sets of data in my analysis, which include operating and structural characteristics of a sample of online bookmakers, and a unique dataset of arbitrage opportunities available in the online betting market. I discuss each of them in turn.

3.1 Bookmakers' characteristics and expected returns

The first dataset is a survey of 82 online bookmakers from the website "Top100bookmakers" (www.top100bookmakers.com), complemented by information collected from the official bookmaker websites[3], collected on August 31st, 2012. For each bookmaker, I collect a variety of operating and structural features of the business, and the markup applied to bets on six major sports, including baseball, basketball, football, hockey, soccer, and tennis.

3.1.1 Operating characteristics

The bookmakers' operating characteristics are reported in Table 1. I consider four categories: promotional services, safety of transactions, customer service, and wager restrictions. Panel A reports the descriptive statistics. The promotions a bookmaker may choose to offer are bonus schemes for prospective new customers, additional deposits, referrals, and loyalty status. Most of the bookmakers in the sample (76%) offer promotions from at least one of these four categories, and both the mean and the median number of promotion categories offered are exactly equal to one. I define the promotion index as the number of promotions offered by the bookmaker.

Safety variables include the possession of the Secure Sockets Layer (SSL) encryption technology certificate, and the number of fair gaming body subscriptions the bookmaker has. The SSL certificate is very important, as it guarantees all gamblers' private data and financial transactions are secure. In fact, 93% of the bookmakers in the sample have it. The fair gaming body membership is also important, as it indicates which independent panel will regulate a dispute should it arise. Surprisingly, however, 62% of the sample has no membership with any fair gaming body. I combine the provision of SSL encryption and the number of fair gaming body subscriptions into one variable, which I define as the safety index.

Customer service variables include the indication of office hours for a help desk, and the number of available means of communications to contact the bookmaker, including e-mail, telephone, chat, and ordinary mail. Interestingly, not all bookmakers provide office hours – only 76%. The average number of available means is quite large, instead, 3.34 out of 4. I combine these two variables into one, which I define as the customer service index.

Wagering restriction is a dummy variable that takes on value one if the bookmaker enforces

[3] The bookmakers included are: 10Bet, 12Bet, 188Bet, 24hPoker Sports, 5Dimes, 888 Sport, AllYouBet, Bet-at-Home, Bet3000, Bet365, Bet770, BetAdria, Betboo, BetClic, Betdaq, Betfair, Betfred, Betinternet, BetOnline, Betoto, BetPhoenix, Bets10, Betsafe, Betsson, BetVictor, Betway, Blue Square, Bodog, Bovada, Boylesports, Bwin, Canbet, Centrebet, ComeOn, Coral, Dafabet, Digibet, Dobet, DOXXbet, Efbet, Expekt, FortunaWin, Gamebookers, GoalBet, GoldBet, IASbet, Intertops, Interwetten, Jetbull, Ladbrokes, Leon Bets, Meridianbet, MyBet, NordicBet, Noxwin, Offsidebet, Paddy Power, PartyBets, Pinnacle Sports, PlanetOfBets, Redbet, Sbobet, Skybet, Sportbet, Sportingbet, Sportingbet Australia, Sports Interaction, Stan James, Tempobet, The Greek, Tipico, Titan Bet, Tobet, TopBet, Totesport, Triobet, Unibet, WagerWeb, Whitebet, William Hill, World Bet Exchange, YouWin.

a limit to either the maximum wager or the maximum winning. These two limits actually serve the same purpose of setting a restriction on the maximum potential outflow for the bookmaker. Since they vary across types of bets, I prefer to use a dummy variable rather than a numerical one. Maximum winnings are restricted in approximately half of the sample (51%).

Panel B presents the correlation matrix for the operating characteristics. Some interesting patterns arise. The number of promotions has a negative and highly significant correlation with the SSL certificate possession (-0.46, p-value < 0.01). This suggests that bookmakers may tend to compensate for a lower level of safety by offering more promotions. The proportion of bookmakers that offer at least one promotion has a positive and significant correlation with the available number of means of communication (0.22, p-value < 0.05). This makes sense, as the bookmakers who engage in promotional investments should also be easier to contact. Interestingly, none of the customer service variables is correlated with any of the safety variables. Maximum winning restrictions are positively and significantly correlated with the number of fair gaming body subscriptions (0.21, p-value < 0.10), the subscription to at least one fair gaming body (0.31, p-value < 0.01), the safety index (0.22, p-value < 0.10), the number of means of communication (0.29, p-value < 0.01), and the customer service index (0.22, p-value < 0.05). Therefore, such restrictions are compensated by higher safety and better customer service.

3.1.2 Structural characteristics

The bookmaker structural characteristics are reported in Table 2. I consider four variables: Alexa daily reach, age, number of languages supported, and number of currencies accepted. Panel A presents the descriptive statistics. Alexa daily reach is the proportion of Alexa toolbar users (expressed in per million terms) that visited the bookmaker website on a daily basis in the 6-month period from March 1st to August 31st, 2012. The mean daily reach is 139.41 millions, whereas the median is only 26.60. The distribution is skewed to the right, as the range goes from 0.30 to 3,345.00. This measure can be interpreted as a proxy for the bookmaker's size.

The bookmaker's age, defined as the number of years of activity, can be thought of as a proxy for the bookmaker's reputation. The mean is approximately 16 years, whereas the median is only 10. The distribution is again skewed to the right, which reflects the fact that the range is from as little as 1 to as much as 126 years.

The number of languages supported by the bookmaker's website represents the number of different linguistic areas that the bookmaker operates in, independently of the currency. On average, bookmakers support 7 languages, while the median is 5. Very few bookmakers (18%) are monolingual, i.e. only support one language, which is hardly surprising in an international market. The maximum number of languages supported is 23. The number of currencies accepted by the bookmaker for payments follows a similar pattern. The average number is approximately 7, the median is 5, and the range is between 1 and 28. Again, few bookmakers only support one currency (29%). Note that many currency areas may be reached by supporting only a few languages.

These two variables, however, are likely to affect betting in different ways. A gambler from e.g. the euro area is unable to place a bet with a bookmaker who only accepts deposits in dollars, or can do so by paying a fee. Similarly, a gambler is unlikely to open an account with a bookmaker whose language she does not understand. Nonetheless, there is an important difference between the two cases. In the former, it is impossible or inefficient for the gambler to place the bet in his own currency. In the latter, the gambler may still place the bet provided she knows some basic English, which is the common language of all bookmakers. Therefore, currency areas can be thought of as *de facto* market segments. On the other hand, the number of languages may constitute a proxy for the bookmaker's market share *within* a given segment. In fact, supporting more languages allows the bookmaker to reach non-English speakers as well.

Panel B presents the correlation matrix for the structural characteristics. The Alexa daily reach has a positive and highly significant correlation with age (0.31, p-value < 0.01). This is expected, as older and better known bookmakers should attract more gamblers. Alexa also has a positive and highly significant correlation with the number of languages supported (0.39, p-value < 0.01) and currencies accepted (0.47 p-value < 0.01). Therefore, international expansion seems increase the customer base. Age is not correlated with the number of languages supported, but it is positively and significantly correlated with the number of currencies (0.25, p-value < 0.05). This suggests that better known bookmakers are more likely to expand their business to new market segments, rather than increase their market share within a segment. The number of languages and the number of currencies are positively and highly correlated (0.51, p-value < 0.01).

Finally, Table 3 reports the distribution of headquarter location in the sample. Among the top five countries, four are considered as "tax havens", with the predominance of Malta and Gibraltar, which feature 26 and 11 observations respectively. In third place is the UK, with 8 bookmaker headquarters, followed by Costa Rica and Curacao, both with 7 observations. Overall, the proportion of bookmakers located in a tax haven is rather high (85%).

3.1.3 Expected returns

Equation (10) shows that a lower alpha, i.e. a higher markup, implies higher expected returns for the bookmaker. Table 4 presents the sample statistics for expected returns, expressed in logs, on the set of six major sports. The expected return on a major sport event is 5.83%, and exactly coincides with the median. The distribution is symmetric and ranges from 2.37% to 10.35%. These numbers imply an average alpha equal to 0.9449 (i.e. 0.0551 below the fair level), ranging from 0.9062 to 0.9768. As can be noticed from the number of observations, not all bookmakers accept bets on all major sports. For instance, 79 bookmakers in the sample accept bets on soccer, but only 69 accept bets on football. The major sports that provide the lowest and highest expected returns are baseball (4.62%) and soccer (6.79%) respectively.

The model predicts that a higher elasticity of demand implies a lower optimal markup, and therefore lower expected returns. For this reason, it is instructive to look at the third Column of Table 5, which reports the estimated number of fans for each major sport. The estimate is from the website "Mostpopularsports" (www.mostpopularsports.net) and is based on Alexa daily reach[4]. Since many people may follow more than one sport, the same person may be included in the calculation more than once, hence the overall estimated number of fans (7.8 billion) exceeds the world's population. Consistent with the model's prediction, the ranking of sports based on fans numbers almost coincides with the ranking on returns, with the only exception of baseball.

3.2 Arbitrage opportunities

The second dataset is a time-series of arbitrage opportunities available in the online betting market for sport events from January 1st to April 30th, 2008. The source is the website "Infobetting (www.infobetting.org), which provides two main services: (1) comparison of real-time odds posted by a large sample of online bookmakers[5] for a variety of major sports, and (2) notification of

[4]Other sources provide similar estimates

[5]The number of bookmakers covered has varied over time. In the sample period, the bookmakers involved were the following 90: 10Bet, 188Bet.com, 5dimes, Acttab, Admiral, AstraBet, Bet at home, Bet on Bet, Bet1128, Bet24, Bet365, Betandgame, Betandwin, Betclic, Betcris, Betdaq, Betdirect, Betfair, Betinternet, Betklass, Betonmarkets, Betsense, Betshop, Betsson, Better.it, Betway, Bill hurley, Bluesquare, Bodog, Boyle, Canbet, Cashmans, Casino

arbitrage opportunities for the observed betting events. The data have been personally collected by the author in real time twice a day, at 2pm and 8pm respectively, for a total of 242 observations. The sample period captures the most important part of the season for most of the sports considered, which in turn should imply enough betting activity to test the odds adjustment hypothesis. The choice of the timing reflects the fact that most sporting events take place in the afternoon between 2pm and 8pm (Brown and Sauer, 1993b), hence book imbalance risk should be greater within that time frame.

Table 5 presents the dataset, with a breakdown in three categories: the number of arbitrage opportunities available, the characteristics of the arbitrage opportunities with the highest returns, and the first two moments of the distribution of arbitrage returns. All returns are instantaneous and refer to single operations. Panel A presents the descriptive statistics. The average number of arbitrage opportunities observed in a session is approximately 13 and ranges from a minimum of 1 to a maximum of 36. Hence, in all 242 sessions I observe at least one arbitrage opportunity. The rising of arbitrages thus seems to constitute an empirical regularity.

In each session, I identify the dominant arbitrage opportunity as that which yields the largest gross rate of return[6]. The mean return on such opportunities is 5.02% per operation and ranges between as little as 0.07% and as much as 62.87%. Next, I observe the number of bookmakers whose odds generated the dominant arbitrage opportunity. The average number of bookmakers involved is approximately 3, and ranges between 2 and 9. The upper extreme is particularly interesting, as it indicates that as many as 9 bookmakers may simultaneously post dominant odds. Then, I check whether the dominant arbitrage opportunity is typically a two-outcome or a three-outcome bet[7], and find that three-outcome bets account for 69% of the cases. This is consistent with the idea that the conditional volatility of bookmaker profits increases with the number of outcomes, which in turns should constitute a stronger incentive to move odds.

The mean return of all arbitrage opportunities observed is 1.14%, which is a very large number

Venezia, CentreBet, City Index, Davidson S.B., Easybets, Eurobet, Expekt, Fonbet, Fortuna,Gamebookers, Gameday, Globet.it, Goldbet, Gwbet, Iasbet, Interscommessa, Intertops, Isibet, Ladbrokes, Manny Bernstein, Mansion, Match Point, Mediabet, My Sportsbook, Nike, Oddset, Olympic Sports, Partybets, Pianeta Scommesse, Pinnacle Sports, Premierbet, Reno 2000, Scommesseitalia, Sean Graham, Snai, Sportfanatik, Sportingbet, Sports Interaction, Sportwetten, Stan James, Starprice, Sts, Superbet, Supporterbet, Toals, Totalbet, Tote Sport, Totosi.it, Ukbetting, Unibet, Victor Chandler, Vierklee, Vip.com, Wagerweb, William hill, Worldbet, Yabet.

[6] Unfortunately I cannot directly observe the presence of fees or other operating costs, then all returns are expressed in gross terms.

[7] All the betting events reported by Infobetting have either two or three outcomes.

for an instantaneous rate of return. The standard deviation is 1.42%, which makes for a large coefficient of variation (125%). I also calculate the trimmed mean and standard deviation by leaving the dominant returns out of the calculation. The estimates change to 0.80% and 0.84% respectively, which still generates a large coefficient of variation (105%). Hence, arbitrage returns seem to be rather volatile.

Panel B presents the correlation matrix for all of the above variables, along with a new set of dummy variables that capture the timing of arbitrages. I define a "weekend" variable as a dummy that takes on value one if the observation is recorded between Friday and Sunday, and 0 otherwise. This variable should capture gamblers' market participation. In fact, sport betting is likely to constitute a leisure activity by most gamblers, hence it is more likely to be pursued during weekends when free time is greater. I define a "match day" variable as a dummy that takes on value one if the observation is recorded on Wednesdays, Saturdays and Sundays and zero otherwise, as most sporting events take place on those three days. This variable should then capture the days in which book imbalance risk should be greatest, which in turn should imply wider and more frequent odds adjustments. I define a "Midday" variable as a dummy that takes on value one if the observation was recorded in the afternoon session (2pm), and 0 in the evening session (8pm).

The number of arbitrage opportunities available is highly correlated with all the time dummies. The correlation with the weekend dummy is positive and highly significant (0.32, p-value < 0.01). This makes sense, as book imbalance and the consequent odds adjustment are likely to manifest when most bets are placed. The correlation with the match day dummy is also positive and highly significant (0.19, p-value < 0.01). This is consistent with the idea that the incentive to adjust odds is strongest when books are about to close, which by definition happens on the match day. The correlation with the midday dummy is positive and highly significant (0.40, p-value < 0.01). This reflects the fact that most sport events take place during the day, hence most books are closed by the evening. As a consequence, there is greater odds adjustment in the afternoon, which generates more arbitrage opportunities.

The number of arbitrages also has a positive and significant correlation with the dominant arbitrage return (0.28, p-value < 0.01), the trimmed mean arbitrage return (0.22, p-value < 0.01), the trimmed standard deviation of returns (0.15, p-value < 0.05), and the number of bookmakers involved (0.16, p-value < 0.05). These estimates reflect the fact that when book imbalance risk is greatest, the distribution of odds becomes more dispersed. As a consequence, more bookmakers have an incentive to adjust their odds on different events and outcomes, which in turns implies more arbitrage opportunities and higher returns. Consistent with this interpretation, the number of bookmakers involved with a dominant arbitrage is positively correlated with the weekend dummy (0.11, p-value < 0.10), the match day dummy (0.18, p-value < 0.01), and the midday dummy (0.12, p-value < 0.10).

Three-outcome arbitrage opportunities are highly correlated with the match day dummy (0.14, p-value < 0.01). This is consistent with the idea that greater profit volatility constitutes a stronger incentive for bookmakers to adjust their odds, especially near book closure.

Finally, the weekend dummy is negatively correlated with the mean arbitrage return (-0.17, p-value < 0.01), the trimmed mean return (-0.12, p-value < 0.10), and the standard deviation of returns (-0.12, p-value < 0.10). Therefore, the return distribution at the weekend is steeper and shifted to the left. This is consistent with the hypothesis of late market participation, as the presence of more gamblers implies that arbitrage opportunities are exploited faster and disappear sooner. Arbitrage returns, then, are correspondingly lower and less volatile.

4 Empirical results

4.1 The determinants of expected returns

Following the model's guidance, a bookmaker's market power should be the key determinant of expected returns. On the other hand, the markup may be related to the bookmaker's operating characteristics in an ambiguous way. If a higher market share is due to the provision of better services, such as customer care or safety, then these characteristics should be positively related to the markup. Alternatively, bookmakers with a lower market share may try to increase their demand by providing better services.

I test these predictions in Table 6. The dependent variable is the bookmaker's expected returns, expressed in logs. In column (1) I include the structural characteristics, all expressed in logs because of their highly skewed distribution. Consistent with the conjecture, the coefficient of the number of languages supported by the bookmaker is positive and significant (0.0033, t-stat 2.10), which indicates that a 50% increase in such a number is associated with an increase in expected returns by 16.5 basis points per bet. The other coefficients, instead, are not significant with t-stats well below one.

In column (2) I test the explanatory power of operating characteristics. I find that the coefficient of the promotions index is negative and highly significant (-0.0048, t-stat -2.97), which indicates that each additional promotion offered is associated with a decrease in expected returns by 48 basis points. The coefficient of the customer service index is also negative, even though only marginally significant (-0.0028, t-stat -1.93), which indicates that a 20% increase in customer service quality is associated with a decrease in expected returns by 28 basis points. Therefore, there is a negative association between markup and services, which suggests that services may be a tool used by bookmakers with low market power in order to increase their demand.

Next, I introduce a variable defined as the difference between the number of payment methods accepted for deposits and the number of payment methods available for withdrawals. Fairness would require that these numbers be equal. Surprisingly, however, this is not the case: the difference between these two numbers is large and highly significant (5.23, t-stat 8.84), which suggests that bookmakers tend to make it easier to deposit money than to withdraw it. Therefore, I take this variable as a proxy for the bookmaker's trustworthiness, and call it "net options".

In columns (3) and (4) I re-estimate the first two regressions with the addition of net options. The coefficient of this new variable is negative but outside the rejection region, both in the model of structural characteristics (-0.0004, t-stat -1.30) and in that with operating characteristics (-0.0003, t-stat -1.05). The other coefficients are virtually unchanged from the other columns.

These regressions, however, have three issues. First, I only consider overall expected returns, with no distinction between sports categories. Second, most variables are highly correlated, which may inflate standard errors and then bias t-stats downwards. Third, the sample size is relatively small, with 81 and 82 observations respectively for the structural and the operating model. For this reason, I re-estimate these equations in a panel setting, in which I consider the markup set by each bookmaker on the six major sport events listed above. The number of observations then increases to 453 and 459 respectively for the structural and the operating model. This setting allows me to control – where possible – for sports fixed effects, which picks up differences in popularity among sports fans, and bookmaker fixed effects, which captures firm-level unobserved characteristics. Also, I cluster standard errors at the bookmaker level.

The results are in Columns (5) and (6), where I control for sports fixed effects. In Column (5), a 50% increase in a bookmaker's age is associated with an increase in expected returns by 20 basis points (t-stat 2.92). In Column (6), an additional promotion is associated with a decrease in expected returns by 48 basis points (t-stat -2.62), and a 20% increase in customer service quality is associated with a decrease in expected returns by 32 basis points (t-stat -1.98). The findings are again consistent with the idea that bookmakers with a less elastic demand earn higher expected returns, and bookmakers with lower market power tend to provide better services. Finally, I find that a unit increase in the number of net options is associated with a decrease in expected returns by 4 basis points, even though the coefficient lies slightly outside of the rejection region (t-stat -1.49). Therefore, there is mild evidence that a lower level of trustworthiness is penalized by investors through lower expected returns.

4.2 The three-outcome premium

Equation (30) shows that risk-averse bookmakers should set a higher markup for events with a greater number of outcomes, as they imply a higher volatility of profits. On the contrary, a risk-neutral bookmaker would not require such compensation. Below, I test this implication.

My empirical strategy is as follows. The available markup estimates refer to six major sports. Four of them, i.e. baseball, basketball, football, and tennis, typically represent two-outcome bets, the outcome being the victory of either side. In fact, even though a tie is theoretically possible, it is an extremely rare outcome due to the high-scoring nature of these sports. Hence, most bookmakers do not accept bets on it. On the contrary, soccer and hockey are sports with much lower scores, which makes ties very common. Hence, they typically constitute three-outcome bets. Then, I define the three-outcome premium as the difference in average expected returns, expressed in logs, between three-outcome and two-outcome bets.

An intuitive approach would be to perform a simple t-test on the premium. Unfortunately, however, this may not be enough. In fact, the two-outcome sports in the sample (soccer and hockey) are also the sports with the highest estimated number of fans worldwide. Hence, the test result may be driven by the difference in popularity. Then I complement the test with a time series regression analysis of the premium on the bookmaker's structural and operating features, and a panel regression in which I control for bookmaker fixed effects.

Table 7 presents the regression output. Consistent with the risk-aversion hypothesis, the premium in Column (1) is positive and highly significant (1.40%, t-stat 8.87). In Column (2), I introduce the set of bookmakers' structural characteristics. None of the variables has any explanatory power over the premium, which is still positive and highly significant (1.68%, t-stat 3.22). In Column (3), I introduce the set of operating characteristics. Again, none of the coefficients is significant. The premium, however, despite being unaffected in its magnitude, is now no longer significant (1.44%, t-stat 1.63).

In columns (4) to (6), I switch to the panel setting. The dependent variable is expected returns, expressed in logs, on the six major sports, and the main variable of interest is a dummy variable that takes on value one for three-outcome bets. In column (4) I control for bookmaker fixed effects, and find that the three-outcome premium is still positive and highly significant (1.37%, t-stat 8.56). In columns (5) and (6) I replace bookmaker fixed effects with structural and operating characteristics, and find that the premium is positive and highly significant both in the structural model (1.35%, t-stat 9.36) and in the operating model (1.39%, t-stat 9.43). Therefore, the results suggest that the premium is not driven by bookmaker-specific features.

In order to show that the observed premium is indeed driven by risk-aversion, below I provide evidence that bookmakers do indeed exhibit risk-averse behavior in their odds adjustments.

4.3 Odds adjustments

The model predicts that risk-averse bookmakers have an incentive to dynamically adjust their odds to reduce the conditional variance of profits. In turn, this mechanism should generate arbitrage opportunities for investors. I test this hypothesis below. First, I look into the pattern of arbitrage opportunities over time, and then I analyze its relationship with bookmaker characteristics.

4.3.1 Arbitrage opportunities over time

The dependent variable is the number of arbitrage opportunities available in the online betting market in a given session, either afternoon (2pm) or evening (8pm). Table 8 presents the results. In Column (1) I use the weekend dummy as the only explanatory variable. As expected, its coefficient is positive and highly significant (t-stat 5.26) and indicates that arbitrage opportunities over the weekend increase by approximately 5 units. In column (2), I introduce the midday dummy and its interaction with the weekend dummy. The coefficient of the weekend dummy is still positive and highly significant (t-stat 8.32), the coefficient of the midday dummy is positive and significant

(t-stat 2.08), and the coefficient of the interaction term is positive and highly significant (t-stat 5.63). The results indicate that in a given session there are almost 9 more arbitrage opportunities during the weekend, 2 more opportunities in the afternoon, and 9 more opportunities during the weekend afternoons. This pattern captures the fact that most sporting events take place before 8pm, and market participation is likely to be highest over the weekend, so the odds adjustment follows the same pattern.

In Column (3), I replace the weekend dummy with the match day dummy. The results are rather similar. The coefficient of the midday dummy is still positive and significant (t-stat 2.05), the coefficient of the match day dummy is positive and significant (t-stat 6.07), and the coefficient of the interaction term is positive and highly significant (t-stat 5.10). The results indicate that arbitrage opportunities increase by nearly 7 units on match days, increase by a further 2 units in the afternoon, and increase by an additional 8 units during match day afternoons. This is again consistent with the fact that most sport events take place before 8pm.

In Column (4) I let all of the above variables coexist within the same model. The coefficients of the weekend and the match day dummies are still positive and highly significant (t-stats 6.40 and 3.37 respectively). Interestingly, then, the market participation and the book closure proxies keep their separate effects despite their correlation, which is consistent with their different interpretations. In particular, on weekends there are 7 more arbitrage opportunities per session, and on match days there are 4 more. The interactions of the weekend and the match day dummies with the midday dummy are again positive and highly significant (t-stats 3.85 and 3.46 respectively). The results indicate that on weekend afternoons there are 6 more arbitrage opportunities per session, and on match day afternoons there are 5 more. However, the midday dummy alone now loses its explanatory power (t-stat 0.58). This is not surprising, as it indicates that apart from match days and the weekend there is no significant difference in the number of arbitrages available between the half-day and the end of day sessions. In fact, if sports events are far, the incentive for bookmakers to move odds is lower.

In Column (5), I re-estimate the equation with the following set of controls: the standard deviation of arbitrage returns as a measure of market volatility, which should proxy for the incentive of bookmakers to adjust odds[8]; a dummy variable that takes on value one if the dominant arbitrage opportunity is a three-outcome bet; the number of bookmakers who posted the odds for the

[8] Using the trimmed standard deviation yields the same results.

dominant arbitrage opportunity; the average number of languages supported by such bookmakers. Interestingly, all the coefficients of the variables of interest are virtually unchanged from Column (4). The coefficient of the standard deviation of arbitrage returns is positive and highly significant (2.92), while the coefficients of the other controls are not significant at any conventional level.

Overall, the pattern of arbitrage opportunities is strongly supportive of the odds adjustments hypothesis, and seems to constitute a structural feature of the online betting market. These findings suggest that bookmakers do manage their profit volatility, which rules out the hypothesis of risk-neutral preferences, but may still be consistent with risk-seeking behavior (Strumpf, 2003). However, the only type of preferences that consistently explains all of the findings I provide is the risk-aversion hypothesis.

Note that these results are inconsistent with an asymmetric information story, according to which bookmakers may be at a disadvantage against some better informed traders. This is due to two reasons. First, asymmetric information should be more severe far from the betting event, as the information set typically improves towards the end of the betting period, especially for sporting events[9]. As a consequence, the odds adjustment should happen more often on non-match days. Second, informed traders should bet on one side of the book, rather than all two or three. In fact, if they observe a signal on a given outcome, they have an incentive to bet on that outcome – not on the others. Therefore, odds should move in the same direction for all bookmakers with no arbitrage opportunities. Both implications are counter-factual.

The results are also unlikely to be driven by changes in probability estimates rather than markup. In fact, as the betting event approaches the information set of all market participants improves. Then, if anything, the dispersion in estimates should decrease rather than increase as the betting period shrinks. This is the opposite of what I find.

4.3.2 Arbitrages and bookmakers' characteristics

In this last section, I look into the relationship between the number of arbitrage opportunities generated by a bookmaker, and her operating and structural characteristics. Since 59% of the bookmakers in the sample produce no arbitrage opportunity, I first estimate a Tobit regression. The results are in Table 9, Columns (1) to (3). In Column (1), the only regressor is the three-outcome

[9]Think, for instance, of information such as team news, injuries, and tactics, among other things, that are typically released in the run-up to a match.

premium. I find that a 1% increase in the premium is associated with a marginally significant decrease in the number of arbitrage opportunities created by 6 units (t-stat -1.75). This result reflects two facts. First, a higher markup works as an insurance mechanism for the bookmaker (see equation 31), as it makes the book less likely to become unbalanced. Second, a higher premium implies lower odds on three-outcome bets. Since most arbitrage opportunities are three-outcome bets (69%), a higher premium makes it less likely to post a dominant odd.

In Column (2) I include the bookmakers' structural features, along with the number of net options. I find that a 50% increase in the number of languages supported is associated with a marginally significant increase in arbitrage opportunities created by approximately 4 units (t-stat 1.76), and a 50% increase in the number of currencies accepted is associated with a highly significant increase in arbitrage opportunities created by 6 units (t-stat 2.81). These results reflect the fact that the presence in more linguistic and currency areas makes the odds accessible to a larger number of gamblers, which makes it more likely to create an arbitrage opportunity. The coefficient of the three-outcome premium is negative and highly significant (t-stat -2.66), which indicates that a 1% increase in the premium is associated with a decrease in arbitrage opportunities created by approximately 8 units. The other coefficients, instead, are not significant at any conventional level.

In Column (3), I replace the structural characteristics with the operating ones. I find that a one point increase in the safety index is associated with a highly significant increase in the number of arbitrage opportunities created by 12 units (t-stat 3.35). This finding indicates that the safest bookmakers can afford to adjust their odds more aggressively. The coefficient of the three-outcome premium is still negative, but slightly outside of the rejection region (t-stat -1.60). The other coefficients are also positive, even though not significant.

Finally, I analyze the relationship between the probability of producing at least one arbitrage opportunity and bookmaker characteristics. To this purpose, I estimate a linear probability model. The results are in Columns (4) to (6). In Column (4), the only regressor is the three-outcome premium. The coefficient is negative but not significant (-0.59, t-stat -0.15). In Column (5) I introduce the structural characteristics, along with net options. I find that a 50% increase in the bookmaker's age is associated with a highly significant increase in the probability of producing at least one arbitrage opportunity by 9.17% (t-stat 3.10), which indicates that the bookmakers that have a stronger reputation can also afford to adjust their odds more aggressively. A 50% increase in the number of languages supported is associated with a significant increase in the probability of producing at least one arbitrage opportunity by 9.99% (t-stat 3.82), which is consistent with the results from column (2). The other coefficients, instead, are not significant. In Column (6) I replace the structural characteristics with the operating ones. I find that a one point increase in the safety index is associated with a highly significant increase in the probability of producing an arbitrage opportunity by 15.25% (t-stat 3.15), which is consistent with the results from column (3). I also find that a one point increase in the customer service index is associated with a marginally significant increase in the probability of producing an arbitrage opportunity by 11.03%, which suggests that bookmakers that serve their customers better can also afford to adjust their odds more aggressively.

5 Conclusion

I find that the optimal price of a bet for a risk-averse bookmaker is a function of the following three factors: (1) the elasticity of the bookmaker's demand, which exogenously depends on gamblers' preferences and endogenously on the bookmaker's market share; (2) the number of outcomes, because a bet with a higher number of outcomes implies more volatile profits and thus calls for a higher markup; and (3) the residual length of the betting period, as it is optimal for bookmakers to dynamically adjust their odds over time to reduce the conditional variance of their profits.

An empirical analysis of the online betting market supports these predictions. I show that (1) bookmakers with greater market power earn higher expected returns; (2) the inclusion of an additional outcome prompts an increase in markup; and (3) arbitrage opportunities arise on a daily basis, with an instantaneous gross rate of return of 1.14% per operation, and increase in frequency at the end of the betting period. To my knowledge, this is the first paper to study the optimal pricing and the phenomenon of arbitrages in the online betting market as a whole, both theoretically and empirically.

The results suggest that bookmaker preferences are more important than other competing explanations, such as asymmetric information or insider trading, to understand betting market prices. This is due to two reasons. First, asymmetric information should be more severe at the beginning of the betting period, as the information set typically improves over time, especially for sporting events. Second, if traders could observe a private signal, they would place their bet on that outcome only, which implies that bookmakers should adjust odds in the same direction. I show that these two predictions are counter-factual, as the odds adjustments take place near the betting event, and in a variety of directions.

I also rule out the possibility that the observed behavior might simply be a price discovery process. If so, the odds adjustment should reflect changes in probability estimates rather than markup. However, that implies that the dispersion in bookmakers' estimates should decrease as the betting event approaches, rather than increase, as prices should converge to the correct value. I show that this prediction is counter-factual too. In fact, the dispersion in bookmakers' odds actually increases over the betting period, reaching its peak on the day in which the event takes place.

The paper also sheds new light on the findings of Shin (1993), who argues that a risk-neutral bookmaker demands a premium for bets with a higher number of outcomes as a compensation for insider trading risk. I show that if bookmakers are risk-averse, rather than risk-neutral, they demand such a premium even in the absence of insider trading, because a higher number of outcomes inflates the variance of both conditional and unconditional profits. Consistent with this idea, I show that bookmakers exhibit risk-averse behavior and earn higher expected returns on three-outcome bets than they do on two-outcome bets.

The evidence on bookmakers' risk-aversion is consistent with Fingleton and Waldron (1999), and stands in contrast with the hypothesis of bookmakers' risk-neutral preferences (Shin 1991, 1992, 1993), and risk-seeking preferences (Strumpf, 2003). I also show that the phenomenon of arbitrage opportunities in the online betting market is frequent and widespread, rather than sporadic and limited to a few bookmakers (Vlastakis, Dotsis and Markellos, 2009). The emergence of such opportunities, however, is not inconsistent with market efficiency (see e.g. Sauer, 1998), and has a theoretical foundation in the bookmakers' goal to decrease conditional profits.

References

Ali, Mukhtar M., 1977: *"Probability and utility estimates for racetrack gamblers"*, Journal of Political Economy, 85, 803-815.

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt, 1982: *"Racetrack betting and informed behavior"*, Journal of Financial Economics, 10, 187-194.

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt, 1984: *"Market efficiency in racetrack betting"*, Journal of Business, 57, 65-75.

Asch, Peter, and Richard E. Quandt, 1986: *"Racetrack betting: The professors' guide to strategies"*, Auburn House, Dover, Mass.

Asch, Peter, and Richard E. Quandt, 1987: *"Efficiency and profitability in exotic bets"*, Economica, 59, 278-298.

Avery, Christopher, and Judith Chevalier, 1999: *"Identifying investor sentiment from price paths: The case of football betting"*, Journal of Business, 72, 493-521.

Brown, William O., and Raymond D. Sauer, 1993a: *"Does the basketball market believe in the hot hand? Comment"*, American Economic Review, 83, 1377-1386.

Brown, William O., and Raymond D. Sauer, 1993b: *"Fundamentals or noise? Evidence from the professional basketball betting market"*, Journal of Finance, 48, 1193-1209.

Bruce, Alistair C., and Johnnie E. V. Johnson, 2001: *"Efficiency characteristics of a market for state-contingent claims"*, Applied Economics, 33, 1751-1754.

Camerer, Colin F., 1989: *"Does the basketball market believe in the hot hand?"*, American Economic Review, 79, 1257-1261.

Dare, William H., and S. Scott MacDonald, 1996: *"A generalized model for testing the home and favorite team advantage in point spread markets"*, Journal of Financial Economics, 40, 295-318.

Fingleton, John, and Patrick Waldron, 1999: *"Optimal determination of bookmakers' betting odds: Theory and tests"*, Trinity Economic Paper Series, Technical Paper No.96/9, Trinity College, Dublin.

Forrest, David, John Goddard, and Robert Simmons, 2005: *"Odds-setters as forecasters: The case of English football"*, International Journal of Forecasting, 21, 551-564.

Forrest, David, and Robert Simmons, 2008: *"Sentiment in the betting market on Spanish football"*, Applied Economics, 2008, 40, 119-126.

Gandar, John M., William H. Dare, Craig R. Brown, and Richard A. Zuber, 1998: *"Informed traders and price variations in the betting market for professional basketball games"*, Journal of Finance, 53, 385-401.

Gandar, John M., Richard A. Zuber, Thomas OBrien, and Benjamin Russo, 1988: *"Testing rationality in the point spread betting market"*, Journal of Finance, 43, 995-1008.

Golec, Joseph, and Maurry Tamarkin, 1991: *"The degree of inefficiency in the football betting market"*, Journal of Financial Economics, 30, 311-323.

Gray, Philip K., and Stephen F. Gray, 1997: *"Testing market efficiency: Evidence from the NFL Sports Betting Market"*. Journal of Finance, 52, 1725-1737.

Hausch, Donald B., William T. Ziemba, and Mark Rubinstein, 1981: *"Efficiency of the market for racetrack betting"*. Management Science, 27, 1435-1452.

Jaffe, Jeffrey F., and Robert L. Winkler, 1976: *"Optimal speculation against an efficient market"*, Journal of Finance 31, 49-61.

Kuypers, Tim, 2000: *"Information efficiency: an empirical study of a fixed odds betting market"*, Applied Economics, 32, 1353-63.

Levitt, Steven, 2004: *"Why are gambling markets organized so differently from financial markets?"*, Economic Journal, 114, 223-46.

Mainelli, Michael, anad Sam Dibb, 2004: *"Betting on the future: Online gambling goes mainstream financial"*, Center for the Study of Financial Innovation, CSFI/New York CSFI, 68, December 2004.

Sauer, Raymond D., 1998: *"The economics of wagering markets"*, Journal of Economic Literature, 36, 2021-64.

Sauer, Raymond D., 2005: *“The state of research on markets for sports betting and suggested future directions”*, Clemson University, Working paper.

Shin, Hyun Song, 1991: *“Optimal betting odds against insider traders”*, Economic Journal, 101, 1179-1185.

Shin, Hyun Song, 1992: *“Prices of state-contingent claims with insider traders, and the favourite-longshot bias”*, Economic Journal, 102, 426-435.

Shin, Hyun Song, 1993: *“Measuring the incidence of insider trading in a market for state-contingent claims”*, Economic Journal, 103, 1141-1153.

Snyder, Wayne W., 1978: *“Horse racing: Testing the efficient markets model”*, Journal of Finance, 33, 1109-1118.

Strumpf, Koleman, 2003: *“Illegal sports bookmakers”*, University of North Carolina at Chapel Hill, Working Paper.

Thaler, Richard H., and William T. Ziemba, 1988: *“Parimutuel betting markets: Racetracks and lotteries”*, Journal of Economic Perspectives, 2, 161-174.

Vlastakis, Nikolaos, George Dotsis and Raphael N. Markellos, 2009: *“How efficient is the European football betting market? Evidence from arbitrage and trading strategies”*, Journal of Forecasting, 28, 426-444.

Weitzman, Martin, 1965: *“Utility Analysis and Group Behavior: An Empirical Analysis”*, Journal of Political Economy, 73, 18-26.

Woodland, Linda M., and Bill M. Woodland, 1994: *“Market efficiency and the favorite-longshot bias: the baseball betting market”*, Journal of Finance, 49, 269-279.

Ziemba, William T., and Donald B. Hausch, 1987: *“Dr. Z’s Beat the Racetrack”*, William Morrow, New York.

Table 1. Bookmakers' operating characteristics

Operating characteristics of a sample of 82 bookmakers from the online betting market. The variables include: a dummy variable that takes on value one if the bookmaker offers at least one promotion; a promotions index, defined as the sum of promotions offered by a bookmaker; a dummy variable that takes on value one if the bookmaker has an SSL encryption certificate; a dummy variable that takes on value one if the bookmaker has at least one subscription to fair gaming bodies; the number of subscriptions to fair gaming bodies; a safety index, defined as the sum of the SSL encryption dummy and the number of fair gaming body subscriptions; a dummy variable that takes on value one if the bookmaker holds office hours for customer service; the number of means of communication through which the bookmaker can be reached; a customer service index, defined as the sum of the office hours dummy and the number of means of communication; a dummy variable that takes on value one if the bookmaker enforces either maximum wagering or maximum winning limits. Dummy variables are denoted by (d). Panel A reports the descriptive statistics, Panel B the correlation matrix. The data sources are a 2012 survey from the website "Top100bookmakers" (www.top100bookmakers.com) and the bookmakers' official websites.

Panel A: Descriptive Statistics

Variable	Observations	Mean	St. Deviation	Median	Min	Max
Promotions (d)	82	0.76	0.43	1.00	0.00	1.00
Promotions Index	82	1.02	0.80	1.00	0.00	3.00
SSL Encryption (d)	82	0.93	0.26	1.00	0.00	1.00
Fair Gaming (d)	82	0.38	0.49	0.00	0.00	1.00
Fair Gaming	82	0.67	1.04	0.00	0.00	4.00
Safety Index	82	1.60	1.12	1.00	0.00	5.00
CS Hours (d)	82	0.76	0.43	1.00	0.00	1.00
CS Means	82	3.34	0.74	3.00	1.00	4.00
CS Index	82	4.10	0.88	4.00	1.00	5.00
Max. Winning (d)	82	0.51	0.50	1.00	0.00	1.00

Panel B: Correlation Matrix

Variable	P	PI	SSL	FG(d)	FG	SI	CSH	CSM	CSI	MW
Promotions (d)	1									
Promotions Index	0.73^{a}	1								
SSL Encryption (d)	-0.46^{a}	-0.16	1							
Fair Gaming (d)	0.01	0.03	0.22^{b}	1						
Fair Gaming	-0.01	0.01	0.18	0.83^{a}	1					
Safety Index	-0.11	-0.02	0.40^{a}	0.82^{a}	0.97^{a}	1				
CS Hours (d)	-0.02	-0.06	-0.16	-0.03	-0.04	-0.08	1			
CS Means	0.13	0.22^{b}	0.00	0.05	-0.01	-0.01	0.07	1		
CS Index	0.10	0.16	-0.08	0.03	-0.03	-0.05	0.55^{a}	0.87^{a}	1	
Max. Winning (d)	0.06	0.07	0.10	0.31^{a}	0.21^{c}	0.22^{c}	-0.04	0.29^{a}	0.22^{b}	1

(a) $p < 0.01$, (b) $p < 0.05$, (c) $p < 0.10$

Table 2. Bookmakers' structural characteristics

Structural characteristics of a sample of 82 bookmakers from the online betting market. The variables include: the bookmaker's website daily reach from Alexa Internet Inc., defined in millions; the bookmaker's age, defined as the number of years of activity; the number of languages supported by the bookmaker's website; a dummy variable that takes on value one if the bookmakers website supports one language only; the number of currencies accepted for payments by the bookmaker; a dummy variable that takes on value one if the bookmaker only accepts one currency for payments. Dummy variables are denoted by (d). Panel A reports the descriptive statistics, Panel B the correlation matrix. The data sources are a 2012 survey from the website "Top100bookmakers" (www.top100bookmakers.com) and the bookmakers' official websites.

Panel A: Descriptive Statistics

Variable	Observations	Mean	St. Deviation	Median	Min	Max
Alexa (mln)	81	139.41	454.45	26.60	0.30	3345.00
Age	82	15.83	20.84	10.00	1.00	126.00
Languages	82	7.01	6.21	5.00	1.00	23.00
Monolingual (d)	82	0.18	0.39	0.00	0.00	1.00
Currencies	82	6.66	5.86	5.00	1.00	28.00
Monocurrency (d)	82	0.29	0.46	0.00	0.00	1.00

Panel B: Correlation Matrix

Variable	Alexa	Age	Languages	Monolingual	Currencies	Monocurrency
Alexa (mln)	1					
Age	0.31^a	1				
Languages	0.39^a	0.17	1			
Monolingual (d)	-0.11	0.16	-0.46^a	1		
Currencies	0.47^a	0.25^b	0.51^a	-0.29^a	1	
Monocurrency (d)	-0.16	-0.03	-0.24^b	0.32^a	-0.62^a	1

(a) $p < 0.01$, (b) $p < 0.05$, (c) $p < 0.10$

Table 3. Bookmakers' headquarters location

Location of the headquarters for a sample of 82 bookmakers from the online betting market. Panel A includes countries commonly regarded as tax havens, Panel B reports the other countries. The data sources are a 2012 survey from the website "Top100bookmakers" (www.top100bookmakers.com) and the bookmakers' official websites.

Panel A: Tax Havens

Country	Frequency	Percent
Malta	26	37.14
Gibraltar	11	15.71
Costa Rica	7	10.00
Curacao	7	10.00
Isle of Man	6	8.57
Antigua and Barbuda	5	7.14
Kahnawake	3	4.29
Alderdey	2	2.86
Ireland	1	1.43
Panama	1	1.43
Philippines	1	1.43
Total	70	100.00

Panel B: Other Countries

Country	Frequency	Percent
UK	8	66.67
Australia	2	16.67
Costa Rica	2	16.67
Total	12	100.00

Table 4. Bookmakers' expected returns

Expected returns, defined as the "overround", for a sample of 82 bookmakers from the online betting market. The returns are expressed in logs and calculated separately for each of six major sports, including baseball, basketball, football, hockey, soccer, and tennis, and then aggregated in an overall equal-weighted average. The variables reported for each sport are: the number of bookmakers in the sample that accept bets on that particular sport; the estimated number of worldwide fans from Alexa Internet Inc.; and a set of descriptive statistics, including mean, standard deviation, median, and range of bookmaker returns. The data sources are a 2012 survey from the website "Top100bookmakers" (www.top100bookmakers.com) and the bookmakers' official websites.

Variable	Observations	Fans	Mean	St. Deviation	Median	Min	Max
Overall	82	7.8 bln	5.83%	1.25%	5.83%	2.37%	10.35%
Baseball	76	0.5 bln	4.62%	1.62%	4.26%	1.69%	9.44%
Basketball	78	0.4 bln	5.32%	1.64%	4.97%	2.57%	10.17%
Football	69	0.4 bln	4.98%	1.44%	4.69%	2.47%	10.80%
Hockey	79	2.0 bln	6.63%	2.26%	6.95%	2.76%	11.60%
Soccer	79	3.5 bln	6.79%	1.83%	6.77%	1.88%	11.15%
Tennis	78	1.0 bln	6.36%	1.58%	6.39%	2.18%	10.80%

Table 5. Arbitrage opportunities in the online betting market

Descriptive statistics of arbitrage opportunities available in the online betting market from January 1st to April 30th, 2008. Each day has two observation sessions, at 2pm and 8pm respectively, for an overall number of 242 sessions. The variables included for each session are: the number of arbitrage opportunities available; the maximum arbitrage return, defined as the rate of return on the most profitable (dominant) arbitrage opportunity; the number of bookmakers that posted the odds for the dominant arbitrage opportunity; a dummy variable that takes on value one if the dominant arbitrage opportunity is a three-outcome bet; the mean and standard deviation of returns on all arbitrage opportunities available, both simple and trimmed, where the latter leaves dominant arbitrage returns out of the calculation. Dummy variables are denoted by (d). Panel A includes descriptive statistics. Panel B reports the correlation matrix, and includes three indicators of timing: a dummy variable that takes on value one if the observation is recorded between Friday and Sunday ("Weekend"); a dummy variable that takes on value one if the observation is recorded on Wednesdays, Saturdays and Sundays ("Match Day"); and a variable that takes on value one for 2pm sessions ("Midday"). The data source is the website of real-time odds comparison "Infobetting" (www.infobetting.org).

Panel A: Descriptive Statistics

Variable	Observations	Mean	St. Deviation	Median	Min	Max
Arbitrages	242	12.66	7.20	11.00	1.00	36.00
Max. Return	242	5.02%	7.53%	2.93%	0.07%	62.87%
Bookmakers	242	3.12	1.05	3.00	2.00	9.00
Three-outcome (d)	242	0.69	0.47	1.00	0.00	1.00
Mean Return	242	1.14%	0.95%	0.96%	0.06%	11.19%
Trimmed Mean Return	242	0.80%	0.54%	0.72%	0.05%	6.02%
Returns St. Dev.	242	1.42%	2.08%	0.91%	0.00%	21.41%
Trimmed Returns St. Dev.	242	0.84%	1.13%	0.62%	0.00%	13.83%

Panel B: Correlation Matrix

Variable	A	WE	MA	MI	MR	B	T	M	TM	STD	TSTD
Arbitrages	1										
Weekend (d)	0.32^a	1									
Match Day (d)	0.19^a	(-)	1								
Midday (d)	0.40^a	(-)	(-)	1							
Max. Return	0.28^a	-0.08	0.03	0.13*	1						
Bookmakers	0.16^b	0.11^c	0.18^a	0.12^c	0.07	1					
Three-outcome (d)	0.05	0.09	0.14^b	0.07	0.06	0.52^a	1				
Mean Return	0.10	-0.17^a	-0.05	0.07	0.85^a	-0.04	-0.04	1			
Trim. Mean Return	0.22^a	-0.12*	-0.05	0.15^b	0.63^a	-0.05	-0.09	0.88^a	1		
Return St. Dev.	0.16^b	-0.12^c	-0.01	0.06	0.96^a	0.02	0.03	0.92^a	0.67^a	1	
Return Trim. St Dev.	0.15^b	-0.09	-0.02	0.06	0.83^a	-0.02	-0.03	0.92^a	0.78^a	0.93^a	1

(a) $p < 0.01$, (b) $p < 0.05$, (c) $p < 0.10$

Table 6. The determinants of bookmakers' markup

Regression model of bookmakers' expected returns, defined as the "overround" and expressed in logs, on the following set of structural and operating characteristics: age, expressed in logs, and defined as the number of years of activity of the bookmaker; the bookmaker's website daily reach, as estimated by Alexa Internet Inc., expressed in log-millions; the number of languages, expressed in logs, supported by the bookmaker's website; the number of currencies, expressed in logs, accepted for payments by the bookmaker; a promotions index, defined as the number of promotions offered by the bookmaker; a safety index, defined as the sum of a dummy variable that takes on value one if the bookmaker has an SSL encryption certificate, and the number of subscriptions to fair gaming bodies; a customer service index, defined as the sum of a dummy variable that takes on value one if the bookmaker offers office hours for customer service, and the number of means of communication available to contact the bookmaker; a dummy variable that takes on value one if the bookmaker enforces either maximum wagering or maximum winning limits; and net options, defined as the difference between the number of payment methods accepted for deposits and the number of payment methods available for withdrawals. Columns (1) to (4) present the estimates from OLS regressions, in which the dependent variable is the overall expected return, defined as an equal-weighted average of the expected returns on six major sports: baseball, basketball, football, hockey, soccer, and tennis. Columns (5) and (6) present the estimates of panel regressions, in which the dependent variable is the expected return on each of the above major sports. The model includes sports fixed effects, and standard errors are clustered by bookmaker. The data sources are a 2012 survey from the website "Top100bookmakers" (www.top100bookmakers.com) and the bookmakers' official websites.

	(1)	(2)	(3)	(4)	(5)	(6)
	Esp. Returns	Esp. Returns	Esp. Returns	Esp. Returns	Esp. Returns	Esp. Returns
Age	-0.0002		-0.0007		-0.0008	
	(-0.11)		(-0.38)		(-0.57)	
Alexa	-0.0008		-0.0006		-0.0004	
	(-0.69)		(-0.57)		(-0.42)	
Languages	0.0033**		0.0037**		0.0040***	
	(2.10)		(2.29)		(2.92)	
Currencies	-0.0015		-0.0012		-0.0011	
	(-0.97)		(-0.78)		(-0.83)	
Promotions Index		-0.0048***		-0.0047***		-0.0048**
		(-2.97)		(-2.92)		(-2.62)
Safety Index		0.0005		0.0007		0.0011
		(0.44)		(0.57)		(0.82)
Customer Service Index		-0.0028*		-0.0029*		-0.0032*
		(-1.93)		(-1.99)		(-1.98)
Winning Cap		0.0031		0.0030		0.0041
		(1.05)		(0.98)		(1.18)
Net Options			-0.0004	-0.0003	-0.0004	-0.0002
			(-1.30)	(-1.05)	(-1.49)	(-1.03)
Constant	0.0583***	0.0716***	0.0599***	0.0731***	0.0685***	0.0820***
	(13.39)	(10.18)	(13.28)	(10.20)	(13.89)	(9.65)
Observations	81	82	81	82	453	459
Adj. R-squared	0.02	0.14	0.02	0.14	0.21	0.26

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7. The three-outcome premium

OLS regression of the bookmakers' three-outcome premium, defined as the difference in bookmakers' expected returns between three-outcome and two-outcome bets, and overall expected returns, on the following set of structural and operating characteristics: age, expressed in logs, and defined as the number of years of activity of the bookmaker; the bookmaker's website daily reach, as estimated by Alexa Internet Inc., expressed in log-millions; the number of languages, expressed in logs, supported by the bookmaker's website; the number of currencies, expressed in logs, accepted for payments by the bookmaker; a promotions index, defined as the number of promotions offered by the bookmaker; a safety index, defined as the sum of a dummy variable that takes on value one if the bookmaker has an SSL encryption certificate, and the number of subscriptions to fair gaming bodies; a customer service index, defined as the sum of a dummy variable that takes on value one if the bookmaker offers office hours for customer service, and the number of means of communication available to contact the bookmaker; a dummy variable that takes on value one if the bookmaker enforces either maximum wagering or maximum winning limits; net options, defined as the difference between the number of payment methods accepted for deposits and the number of payment methods available for withdrawals; and a dummy variable that takes on value one for sporting events with three possible outcomes (win, draw, and loss). Columns (1) to (3) present the estimates from OLS regressions, in which the dependent variable is the three-outcome premium, calculated as the average log-return on soccer and hockey betting events, minus the average log-return on baseball, basketball, football, and tennis. Columns (4) to (6) present the estimates of panel regressions, in which the dependent variable is the expected return on each of the above major sports. The model includes bookmaker fixed effects in Column (4), and bookmaker structural and operating characteristics in Columns (5) and (6) respectively. Standard errors are clustered by bookmaker. The data sources are a 2012 survey from the website "Top100bookmakers" (www.top100bookmakers.com) and the bookmakers' official websites.

	(1)	(2)	(3)	(4)	(5)	(6)
	Premium	Premium	Premium	Exp. Returns	Exp. Returns	Exp. Returns
Age		0.0000			-0.0009	
		(0.00)			(-0.62)	
Alexa		-0.0014			-0.0004	
		(-1.07)			(-0.42)	
Languages		-0.0009			0.0040***	
		(-0.46)			(2.92)	
Currencies		0.0022			-0.0012	
		(1.26)			(-0.85)	
Net Options		-0.0001	0.0000		-0.0004	-0.0002
		(-0.22)	(0.07)		(-1.53)	(-1.05)
Promotions Index			0.0007			-0.0049***
			(0.34)			(-2.67)
Safety Index			-0.0019			0.0010
			(-1.30)			(0.80)
Customer Service Index			0.0013			-0.0032**
			(0.74)			(-2.04)
Winning Cap			-0.0049			0.0041
			(-1.28)			(1.16)
Three Outcomes				0.0137***	0.0135***	0.0139***
				(8.56)	(9.36)	(9.43)
Constant	0.0140***	0.0168***	0.0144	0.0412***	0.0542***	0.0677***
	(8.87)	(3.22)	(1.63)	(76.90)	(11.08)	(8.15)
Observations	80	79	80	459	453	459
Adj. R-squared	0.00	-0.03	0.01	0.42	0.14	0.19

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8. The determinants of arbitrages in the online betting market

OLS regression of the number of arbitrage opportunities available in the online betting market from January 1st to April 30th, 2008. Each day has two observation sessions, at 2pm and 8pm respectively, for an overall number 242 sessions. The independent variables are: a dummy variable that takes on value one if the observation is recorded between Friday and Sunday (”Weekend”); a dummy variable that takes on value one if the observation is recorded on Wednesdays, Saturdays and Sundays (“Match Day”); and a variable that takes on value one for 2pm sessions (“Midday”); the standard deviation of returns on all arbitrage opportunities in a given session; a dummy variable that takes on value one if the most profitable (dominant) arbitrage opportunity is a three-outcome bet; the number of bookmakers that posted the odds for the dominant arbitrage opportunity; and the average number of languages supported by the websites of such bookmakers. The data source is the website of real-time odds comparison “Infobetting” (www.infobetting.org).

	(1)	(2)	(3)	(4)	(5)
	Arbitrages	Arbitrages	Arbitrages	Arbitrages	Arbitrages
Weekend	4.6770***	8.9720***		7.3838***	7.5922***
	(5.26)	(8.32)		(6.40)	(6.58)
Midday		2.0571**	2.1594**	0.6106	0.7139
		(2.08)	(2.05)	(0.58)	(0.68)
Midday x Weekend		8.5899***		6.2862***	6.0313***
		(5.63)		(3.85)	(3.69)
Match Day			6.8868***	3.8782***	3.7347***
			(6.07)	(3.37)	(3.24)
Midday x Match Day			8.1867***	5.6253***	5.4254***
			(5.10)	(3.46)	(3.36)
St. Dev. Returns					0.5184***
					(2.92)
Three Outcomes					-0.9917
					(-1.08)
Bookmakers Involved					0.1921
					(0.45)
Languages (average)					-0.0579
					(-0.53)
Constant	10.6857***	11.7143***	12.5362***	10.7170***	10.3715***
	(18.51)	(16.74)	(16.86)	(14.39)	(6.75)
Observations	242	242	242	242	242
Adj. R-squared	0.10	0.34	0.26	0.37	0.38

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9. Arbitrages and bookmakers' characteristics

Regression model of the number of arbitrage opportunities generated by 82 bookmakers from the online betting market (Columns 1-3), and a dummy variable that takes on value one if the bookmaker posted odds that generated at least one arbitrage opportunity (Columns 4-6), in the period from January 1st to April 30th, 2008. The independent variables are: bookmakers three-outcome premium, defined as the difference in bookmaker expected returns between three-outcome and two-outcome bets, and calculated as the average log-return on soccer and hockey betting events, minus the average log-return on baseball, basketball, football, and tennis; age, expressed in logs, and defined as the number of years of activity of the bookmaker; the bookmaker's website daily reach, as estimated by Alexa Internet Inc., expressed in log-millions; the number of languages, expressed in logs, supported by the bookmaker's website; the number of currencies, expressed in logs, accepted for payments by the bookmaker; net options, defined as the difference between the number of payment methods accepted for deposits and the number of payment methods available for withdrawals; a promotions index, defined as the number of promotions offered by the bookmaker; a safety index, defined as the sum of a dummy variable that takes on value one if the bookmaker has an SSL encryption certificate, and the number of subscriptions to fair gaming bodies; a customer service index, defined as the sum of a dummy variable that takes on value one if the bookmaker offers office hours for customer service, and the number of means of communication available to contact the bookmaker; a dummy variable that takes on value one if the bookmaker enforces either maximum wagering or maximum winning limits. Columns (1) to (3) report the estimates of Tobit regressions, Columns (4) to (6) report estimates of linear probability models. The data source is the website of real-time odds comparison "Infobetting" (www.infobetting.org).

	(1)	(2)	(3)	(4)	(5)	(6)
	Arbitrages	Arbitrages	Arbitrages	Arbitrage	Arbitrage	Arbitrage
Three-Outcome Premium	-606.2559*	-772.7425***	-522.7095	-0.5859	-0.6287	1.5175
	(-1.75)	(-2.66)	(-1.60)	(-0.15)	(-0.19)	(0.40)
Age		3.0677			0.1834***	
		(0.67)			(3.10)	
Alexa		3.0946			-0.0141	
		(1.03)			(-0.37)	
Languages		7.3123*			0.0507	
		(1.76)			(0.94)	
Currencies		12.6545***			0.1998***	
		(2.81)			(3.82)	
Net Options		-0.1827	0.3427		-0.0077	-0.0032
		(-0.31)	(0.45)		(-0.82)	(-0.32)
Promotions Index			6.2124			0.0530
			(1.14)			(0.79)
Safety Index			12.2448***			0.1525***
			(3.35)			(3.15)
Customer Service Index			5.4558			0.1103*
			(1.13)			(1.83)
Max Winning			8.8479			0.1360
			(0.79)			(1.07)
Observations	80	79	80	80	79	80
R-squared	0.01	0.14	0.08	-0.01	0.31	0.11

t statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$