The Revelation of Insider Trading using the Spread: A Comparison with the Sports Betting Market

Christophe Barraud[1] and Philippe Gillet[2]

Abstract

This article compares methodologies which can be used to detect insider trading on financial markets and frauds on sports betting industry. There is a lot of evidence coming from the literature showing the similarities on financial markets and sports betting market. Furthermore, some authors highlight a spread enlargement depending on the proportion of uninformed investors during insider trading on the financial markets. Considering that, we used the bid-ask spread to see if such a phenomenon exists on sport betting industry. We used a sample of 10 litigious tennis games compared with a control-sample to examine spread's behavior in the betting industry. We find out that, as on financial markets, spreads of litigious games increase on gambling markets before the beginning of the match, when the proportion of uninformed bettors is low. So, we can make two conclusions: on the one hand, the spread behavior could be used to detect frauds on gambling markets. On the other hand, we confirm spread's behavior during insider trading on financial markets, which is comforting the strong hypothesis of the market efficiency theory.

Keywords:

Market efficiency, betting industry, spread, insider trading, strong hypothesis.

Introduction

The recent opening of online sports betting in Southern Europe and particularly in France, coinciding with 2010 Soccer World Cup, gives a new visibility to this activity. It becomes difficult in these countries to consult a website without being solicited by advertisements for

[1] University of Paris-Dauphine, Paris, France
[2] University of Paris-Sud, Paris, France

Corresponding author :
Philippe Gillet, Email : Philippe.gillet@u-psud.fr

online gambling sites. However, this activity has long existed in English speaking countries, including Britain, where the art of bookmaking is integral part of British culture. In fact, both, amounts wagered and transactions are important. In 2008, the turnover generated by English bookmakers was about 10 billion pounds. Opening markets in Europe should see this trend increasing exponentially. Figure 1 shows the monthly variation in trades' number of the biggest betting company, Betfair, compared on this statistic with financial markets. Betfair generates fewer trades than the CME but more than the LSE or NBOT. Its activity is comparable to an average-sized market as shown by the study of Roman (2007)[1].

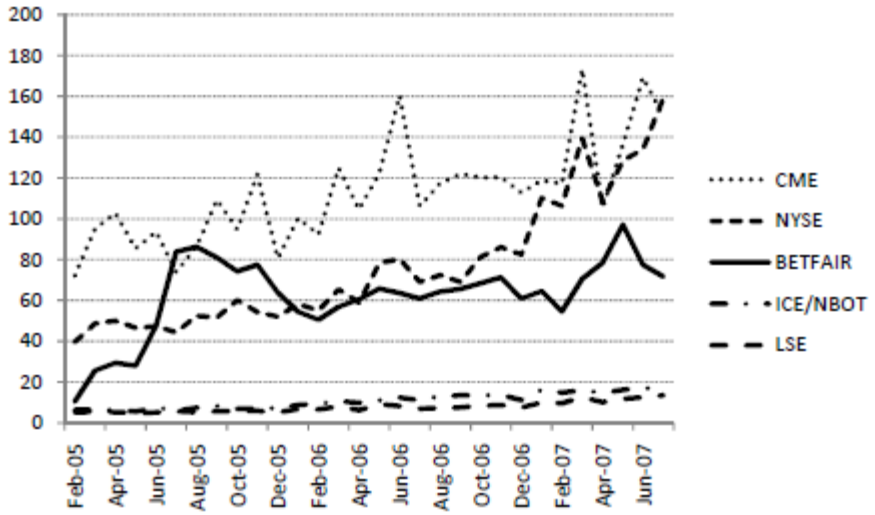

Figure 1. Number of trades executed each month (in millions).

The sports betting market may look distant from the financial markets. However, Barraud (2012) shows that it provides an experimental framework close enough in terms of organization, kind of participants and rules, to test the theory of informational efficiency. The existence of bets on sporting events creates opportunities for frauds, similar to those that can be found in the financial markets. It is very tempting to influence an athlete (or a team) to lose intentionally, and thus to collect gains without having taken risk. In doing so, the bettor has inside information: he knows the result before the other bettors and can make profits.

The aim of this paper is to use existing similarities between financial markets and sports betting market to compare the detection of insider trading on financial markets and frauds in the sports betting market through the study of spread's behavior. In doing so, the notions of strong form efficiency of financial markets on the one hand and fairness in sports betting market on the other hand will be discussed and compared.

Techniques development to identify insider trading is particularly interesting for both practitioners and theorists of financial markets. For practitioners, and particularly supervisors, they allow them to spot insider trading and to repress them, if they are strictly prohibited. Using the spread, organizers of sports betting may also easily identify suspicious games. For theorists, the identification of those operations allows to ensure on the one hand the reality of the strong form efficiency and on the other hand whether the use of privileged information is beneficial or not to investors. Nevertheless, testing such kind of market efficiency is tricky for searchers: when insider trades are prohibited, they are not all identified, except those suppressed by the supervisory authorities. When tests are made on the basis of these crimes, they are inevitably biased, we can easily think that the reactions of returns and trading volumes for insider trading which have not been identified by supervisors are different from those that have been identified, it is also perhaps why the first operations were identified and the others did not. This explanation is consistent with Kyle's theoretical model (1985).

This article is split into three parts. In the first part, we will present the notion of spread and specifications of sports betting. The second part will focus on the methodology and description of our sample. The results and key findings will form the third part.

The spread

Tests to detect insider trading are divided into three main groups: tests on abnormal returns, tests on abnormal volumes and tests on the spread. Tests on abnormal returns are both the best known and numerous. They are based on the well-known methodology of event study initiated by Fama, Fisher, Jensen and Roll. In their work, Meulbroeck (1992) and Guivarc'h (1997) use this methodology and do not validate the hypothesis of strong efficiency form, showing that using privileged information allow insiders to obtain abnormal returns. Insider trading can also be detected by observing, with a methodology quite similar, the existence of abnormal trading volumes. Fishe and Robe (2004) analyze transactions made by informed agents dealing on the Nyse, Nasdaq and Amex, and observe that transaction volumes increase on the day the insiders execute their orders. Finally, a more original detection of insider trading is based on the observation of the spread. The underlying assumption is while an insider is trading, spread may behave abnormally. Chung and Charoenwong (1998) show that in the case of a price driven market, market-makers widen the bid-ask spreads of stocks handled by the informed traders. Similarly, in a order driven market, Guivarc'h (1996) shows

that during an insider trading, short limited donors are placing orders in a less aggressive way, leading to a wider bid-ask spread.

Two types of spread exist: the bid-ask spread also called the quoted spread and the effective spread. The quoted spread is defined as the difference between asking and biding price and is usually calculated in relative terms (% of assets), to allow comparison between stocks of which the price levels differs, as follows:

$$fa_t = \frac{Ask_t - Bid_t}{\left[\frac{Askt + Bidt}{2}\right]} \quad (1)$$

The bid-ask spread assumes that transactions are always carried out on the best present limits in the order book, however, several authors including Huang and Stoll (1996) show that the best limits shown in the book do not correspond systematically to the price of transaction. This is the case on the Nasdaq, where some transactions are within the range when it is wide. In addition to the quoted spread, the effective spread takes into account the actual price of a transaction and is calculated as follows:

$$fe_t = 2 * \frac{\left|\frac{Ask_t + Bid_t}{2} - P_t\right|}{\left[\frac{Ask_t + Bid_t}{2}\right]} \quad (2)$$

In this article we will therefore study the spread's behavior as it may appear in the world of sports betting, we will seek to determine whether the conduct of the spread identified in insider trading on financial markets is reflected in the match-fixing in the sports betting market or not. Thus, if the spread's behavior is the same in both worlds and in the same scenarios, the use of the spread to detect insider trading on financial markets or fraud in the sports betting market should be strengthened.

The Betting Exchange

The betting exchange is a type of bets which allows the bettor to play not against a bookmaker, but directly with other players. Bettors have the option to bet against the

execution of an event and to position themselves as sellers, in this case, they are called *Layers*. Thus, if we take the example of a horse race with five runners and we want to bet on the defeat of one of them, it is not necessary to bet on the winning of the other four, as would be the case in a conventional system, now, it's possible to bet on the defeat of the horse who seems less able to win. This system simply allows any punter to take the place of the bookmaker. In this system, the odds will not be permanently fixed and will evolve according to changing market conditions, ie by financial flows corresponding to the players' expectations. Like traditional financial markets, bettors can place multiple order types.

Betfair

Our sample is issue from Betfair, which is the largest online betting exchange in the world and the only one providing total access to data. Generally in the world of sports betting and more particularly on Betfair, competition mechanisms are close to those of financial markets and even surpasses them in terms of perfection:

- Betfair is the biggest company of sports betting, its popularity is the guarantee of a large number of punters, large volumes and thus liquidity.

- Information is free and is simultaneously revealed to all punters, for example in case of tennis matches, people have to watch TV to see if a player surpasses another or not.

- Everyone is able to see all orders, buying or selling. There is no hidden order[2].

- Bet can be placed at any time before the start of an event or while the event unfolds, we call this period the *Live Betting*. So there is free entry and free exit to the market.

- There is no delayed time when you want to bet before the start of the event, and 10 seconds during the Live Betting.

- Agents are price takers, the total amounts wagered are important and only one person can't influence prices.

- Generally[3], the remuneration of Betfair equals 5% of net profits. So there is no transaction cost in the strict sense since only the earnings are taxed.

- Betfair has introduced specific rules that were intended to avoid any possibility of making profit without risk. The bets are immediately suspended when an unexpected event is likely to significantly alter the price structure. For example, during a soccer game, when the referee rules out a player, bets are interrupted.

- As a financial market, there are minimal differences in prices (ticks) function of the odd value. The odd is therefore not continuous.

The advantage of using tennis

In this article we will look at the case of tennis games and more particularly to men's professional tennis matches because:

- Tennis a great opportunity to study insider trading because one player can tip the outcome of the match, which is obviously not the case with sports team.

- The odds of individual sports are easier to analyze than those of team sports as outlined in Avery and Chevalier (1999) or Forrest and Simmons (2002).

- Tennis has the advantage to set two competitors who often play fairly. Information about them is clear and easy to find when, for example, for horse racing, there are lots of different starters and jockeys, which multiplies opportunities and complicates choices.

- In a sport like tennis, draws do not exist, therefore bettors have four possibilities: Bet on the winning of Player 1 (Back), bet on the defeat of Player 1 (Lay), bet on the victory of Player 2 (Back) and bet on the defeat of Player 2 (Lay). We easily notice that in the lack of arbitrage opportunity, the four possibilities became two: bet on the victory of one player or the defeat of his opponent back to the same.

Presentation of our data and tests

Figures come from the site http://www.fracsoft.com. We have a spread for each player every 20 seconds which means 180 quotations per hour and player. Unfortunately, it is impossible to obtain spreads at the time of the transaction, which will limit our investigations[3].

We retained the quoted and effective spread listed as follows:

$$fa_{intz} = \frac{\left(Ask_{intz} - Bid_{intz}\right)}{\left[\frac{Ask_{intz} + Bid_{intz}}{2}\right]} \qquad fe_{intz} = 2 \times \frac{\left|\frac{Ask_{intz} + Bid_{intz}}{2} - P_{intz}\right|}{\left[\frac{Ask_{intz} + Bid_{intz}}{2}\right]} \qquad (3)$$

- fa_{itnz}: the quoted spread listed at time t for the player i[4], the period n[5] and the match z[6].
- fe_{itnz}: the effective spread listed at time t for the player i, the period n and the match z.
- Ask_{intz}: the lowest asking price displayed on the book at time t for the player i, the period n and the match z.
- Bid_{intz}: the highest bidding price displayed on the book at time t for the player i, the period n and the match z.
- P_{intz}: the last price traded[7] at the time t, for the player i, the period n, and the match z.

The methodology is split into three phases:

- First, we isolated ten matches[8] that have been investigated or seriously suspected[9] and quoted in the community of sports betting as possible match-fixing. Matches with fraud are fortunately rare and in this population, only ten matches meet the following criteria:
    - The liquidity of bets is sufficient and data are available.
    - They have always been won by the presumed outsider[10].
    - They were held in the early rounds of tournaments (1st and 2nd round), which gives a comparable basis (beyond in tournaments, the liquidity improved significantly).

[3] This situation makes impossible the establishment of standardized tests such as the George, Kaul and Nimalendran (1991) to study the evolution of the spread's components.

The concept of citation in the community of sports betting can be defined as follows: there are numerous internet forums and newspapers dealing with sports news and sports betting on the same areas. In the case of the published list of suspicious matches, several of them have been associated with various abnormalities (odds' changes not consistent with the scores' evolution, players' attitude, betting volumes). Details on some of these games, their progresses, their outcome and the origin of the fraud presumption are detailed in Appendix 1.

- Then, we formed a reference sample of 25 games[11] in which no alleged insider dealing has been detected and which will be our benchmark. These 25 games have been selected randomly from a larger sample of 100 matches for which:

    - The liquidity of bets is sufficient and data are available.
    - They have always been won by the outsider presumed.
    - They were held in the early rounds of the tournament (1st and 2nd).

The games' list included in the reference sample is described in Appendix 2 and these were never named as suspects.

- Finally, the observation period was divided into two sub-periods:

    - From two hours before the match to the beginning of the match.
    - From the beginning of the match to the end of the match.

The spreads listed more than two hours before the match were dropped for consistency, and all the data have been adjusted so that it does not take into account the spreads without counterpart.

The following tables list the spreads' number selected for both our reference sample and our suspicious matches.

Table 1. Number of spreads selected for our reference sample for each player

Number of matches	Number of spreads Period 1	Number of spreads Period 2
25	8967	7484

Table 2. Number of spreads selected for each suspicious match for each player

Match title	Number of spreads Period 1	Number of spreads Period 2
Suspect1	171	351
Suspect2	360	516
Suspect 3	360	405
Suspect4	360	314
Suspect5	360	256
Suspect6	360	267
Suspect7	360	194
Suspect8	360	3720[12]
Suspect9	360	357
Suspect10	178	155

In our sample of suspicious matches, the average spreads for each period, each match and each player (favourite and outsider) was calculated as follows:

$$Fa_{inz} = \frac{\sum_{t=1}^{T} fa_{intz}}{T_{inz}} \qquad Fe_{inz} = \frac{\sum_{t=1}^{T} fe_{intz}}{T_{inz}} \tag{4}$$

- Fa_{inz}: the average quoted spreads displayed for player i, the period n and the match z.
- Fe_{inz}: the average effective spreads displayed for player i, the period n and the match z.
- fa_{intz}: the quoted spread at the time t, of player i, the period n and the match z.
- fe_{intz}: the effective spread quoted at time t, player i, the period n and the match z.
- T_{inz}: the spreads' number listed of player i, the period n and the match z.

For the reference sample, we ordered all spreads all matches combined (excluding insiders) and we then divided by the total number of spreads: therefore, we have determined for each sub-period, average reference spreads.

$$Fa_{in} = \frac{\sum_{z=1}^{25} \sum_{t=1}^{T} fa_{intz}}{N_{in}} \qquad Fe_{in} = \frac{\sum_{z=1}^{25} \sum_{t=1}^{T} fe_{intz}}{N_{in}} \tag{5}$$

- Fa_{in}: the average reference quoted spreads for the period n and the player i.
- Fe_{in}: the average reference effective spreads for the period n and the player i.
- N_{in}: the set of all spreads all matches combined for the period n and the player i.

We then compared averages for each period, each type of player (favourite and outsider) and for each suspicious match to our overall average to see what the trends that emerged. For this, 11 samples were made for each period, the first consisting of all the listed spreads for each player (presumed favourite and outsider) all games combined, resulting only of our reference sample, and ten other grouping quoted and effective spreads traded for every player, every match suspects being distinct. In addition it was interesting to observe whether or not the spreads followed a normal distribution. This hypothesis was rejected after checking the normality test of Jarque-Bera (1987). The test was applied below to each of the 11 samples studied.

Table 3. Test results of Jarque-Bera for each sample for the presumed **favourite**

Quoted spreads										
P.value associated with the Jarque-Bera (**favourite)**										
Period 1	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.0000	0.0000	0.0000	< 0.0001	< 0.0001	0.0000	< 0.0001	< 0.0001	0.00000	0.00000	< 0.0001
Period 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Effective spreads										
P.value associated with the Jarque-Bera (**favourite)**										
Period 1	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.0000	0.0000	0.0000	0.0267	< 0.0001	0.0000	< 0.0001	< 0.0001	< 0.0001	0.0000	< 0.0001
Period 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	< 0.0001	0.0000	0.0000	0.0000	0.0000

Table 4. Test results of Jarque-Bera for each sample for the presumed **outsider**

Quoted spreads										
P.value associated with the Jarque-Bera test (**outsider)**										
Period 1	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.0000	< 0.0001	< 0.0001	0.0000	< 0.0001	0.0000	0.0021	0.0000	< 0.0001	< 0.0001	0.0000
Period 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.0000	0.0000	< 0.0001	0.0000	< 0.0001	0.0000	< 0.0001	< 0.0001	0.0000	< 0.0001	< 0.0001
Effective spreads										
P.value associated with the Jarque-Bera test (**outsider)**										
Period 1	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.0000	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0000	< 0.0001	0.0000	< 0.0001	0.0000	0.0000
Period 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.0000	0.0000	0.0000	0.0000	< 0.0001	0.0000	< 0.0001	< 0.0001	0.0000	< 0.0001	< 0.0001

Except for one sample, the test of Jarque-Bera consistently rejected the hypothesis of normality at the threshold of 1%. In fact, we then used the nonparametric test of Mann and Whitney (1947), insofar as it is a rank test to compare the means of two independent samples not having the same size, from a different distribution and does not match the requirements of normality for parametric tests. We compared for each period, the sample of our benchmark to each of the other samples to see if the mean was significantly different. Then, to give more weight to our analysis, we also calculated the standard deviation of spreads, which is defined as follows:

$$\sigma a_{inz} = \sqrt{\frac{1}{T_{inz}} \times \sum_{t=1}^{T} (fa_{intz} - Fa_{inz})^2} \qquad \sigma e_{inz} = \sqrt{\frac{1}{T_{inz}} \times \sum_{t=1}^{T} (fe_{intz} - Fe_{inz})^2}$$

- σa_{inz}: the standard deviation of the quoted spreads of the player i, in period n and the match z.
- σe_{inz} represents the standard deviation of the effective spreads of the player i, in period n and the match z.

Finally, the standard deviation of spreads of our reference sample may be defined as follows:

$$\sigma a_{in} = \sqrt{\frac{1}{N_{in}} \times \sum_{z=1}^{25} \sum_{t=1}^{T} (fa_{intz} - Fa_{inz})^2} \qquad \sigma e_{in} = \sqrt{\frac{1}{N_{in}} \times \sum_{z=1}^{25} \sum_{t=1}^{T} (fe_{intz} - Fe_{inz})^2}$$

- σa_{in}: the standard deviation of quoted spreads on the player i, in period n match all confused.
- σe_{in}: the standard deviation of effective spreads of player i, in period n match all confused.

Results and comments

Descriptive analysis of the reference sample

The evolution of average spreads of our reference sample, standard deviation and frequency of volumes traded bets (outsider + favori) are grouped in the tables below:

Table 5. Average, standard deviation of quoted and effective spreads, and frequency of bets traded volumes for the presumed **favourite**

Quoted spreads			
Period	Average	Standard deviation	Frequency of volumes traded bets
1	0,939%	1,965%	18,551%
2	12,839%	27,386%	81,449%
Effective spreads			
Period	Average	Standard deviation	Frequency of volumes traded bets
1	0,887%	1,964%	18,551%
2	11,660%	27,004%	81,449%

Table 6. Average, standard deviation of quoted and effective spreads, and frequency of bets traded volumes for the presumed **outsider**

Quoted spreads			
Period	Average	Standard Deviation	Frequency of volumes traded bets
1	2.691%	2.059%	2.967%
2	19.458%	32.733%	97.033%
Effective spreads			
Period	Average	Standard Deviation	Frequency of volumes traded bets
1	2.206%	2.350%	2.967%
2	17.226%	32.111%	97.033%

Before going into details of the calculations, it should be emphasized that the results obtained using the quoted or effective spreads are close because it's not possible to treat within the spread unless the spread is wide and if two orders are executed at the same price simultaneously. Similarly, the format of our data (high frequency) ensures that the last traded price is necessarily very close to the quoted spread.

Concerning the average, standard deviation of quoted and effective spreads and the frequency volumes traded bets, we observe that they increase significantly between the period 1 and 2. Between 2 hours before the match and the beginning of the match, the punters have almost all the same information and are expected to converge on the estimate of a price, which results in small changes in odds and therefore small fluctuations of the spread. Spread broadening during the period 2 and frequency increasing of volumes can be explained by market structure. During the period 2, an important flow of information will reach the market (field performance, form factor, point winning...). Based on these information, bettors will be required to assign new probabilities of occurrence in their scenarios and thus to adjust their positions from new orders, leading to a stronger variation of odds, a broadening of the spread and a significant increase of volumes of bets as shown in tables 5 and 6. Upon arrival of new information, the adjustment is not instantaneous, the range will tend to widen and then to narrow.

At the end of the match, the volatility of odds is much greater and the lack of liquidity sometimes destabilizing on the odd's player who is close to defeat (in our study, it is always the presumed favourite player, since only matches for which the outsider wins were retained). As time passes, the more an opponent will detach from each other. This results in a lack of liquidity in the player close to defeat. Symmetrically, the punters have less restraint to bet on an event whose probability of occurrence is high, and expose themselves to very small losses by placing on the sell side. Liquidity improved considerably, the spread is often equals to one or two ticks and volumes traded bets are much more important. It is the *end of match* effect.

Table 7. Example of transactions at the end of a game (Schwank-Andreev Match3)[13]

In Play 5 | Time to race end: -01:22:21 | Time: 09:30:44

Selection	BACK		(108.23%)	(98.84%)		LAY
Igor Andreev	4.7 £4	7.0 £10	7.2 £73	16.0 £9	85.0 £2	100.0 £3
Eduardo Schwank	1.03 £816	1.04 £1 606	1.06 £117	1.08 £49	1.1 £452	1.12 £7

For example, at this point, the quoted spreads for Andreev stood at 75,86% which represents more than 6 times the global average of the presumed favourite, when the quoted spreads for Schwank was 0,94% or nearly 20 times less than the overall average of the presumed outsider. This explains why in this case, from the period 1 to 2, the quoted spread is multiplied by more than 13 times for the presumed favourite who bows and only by a more than 7 for the presumed outsider who wins. *In fine*, three major trends seem to emerge in the reference sample:

- The average quoted and effective spreads widen over time.
- The standard deviation of quoted and effective spreads increases over time.
- The volumes traded bets are concentrated on the period 2.

Figures 2 and 3 illustrate the observations we made concerning the mean and standard deviation of quoted and effective spreads.

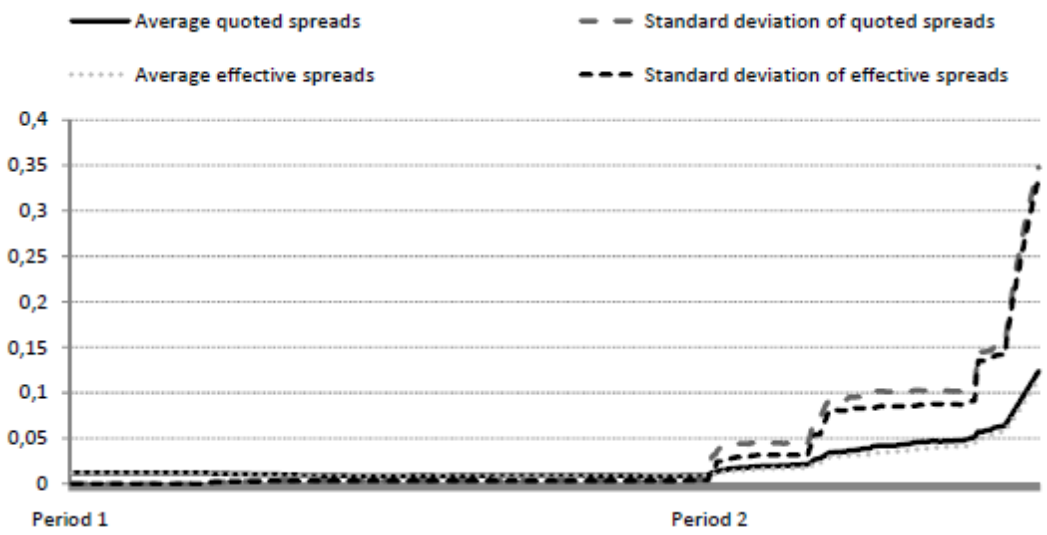

Figure 2. Evolution of the average, standard deviation of quoted and effective spreads for the presumed **favourite** for a game chosen randomly from the Reference sample (Match8).

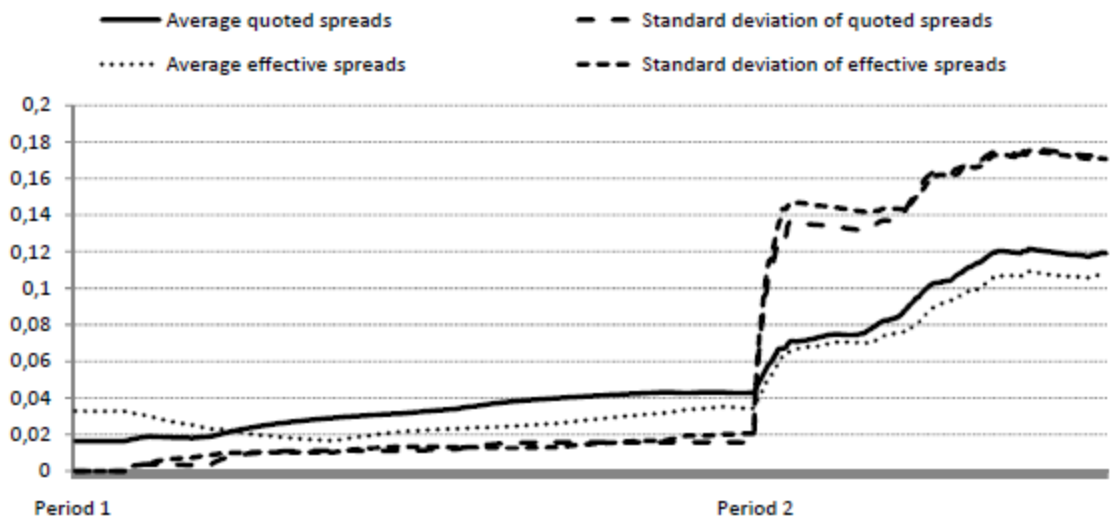

Figure 3. Evolution of the average, standard deviation of quoted and effective spreads for the presumed **outsider** for a game chosen randomly from the Reference sample (Match8).

Descriptive analysis of suspicious matches

Tables 8 to 13 reproduce the results found for the suspicious matches:

Table 8. Average quoted and effective spreads for each sample and the presumed **favourite**

Average of quoted spreads										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.939%	2.477%	2.188%	1.149%	0.957%	0.867%	1.065%	0.833%	1.873%	0.844%	0.802%
12.839%	10.149%	3.040%	6.386%	7.353%	4.944%	9.707%	17.595%	2.386%	10.158%	9.724%
Average of effective spreads										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
0.887%	2.256%	2.146%	0.951%	0.983%	0.862%	1.056%	3.884%	1.887%	0.767%	0.749%
11.660%	7.633%	3.206%	7.121%	7.513%	5.021%	9.953%	18.101%	2.374%	9.436%	9.471%

Table 9. Average quoted and effective spreads for each sample and the presumed **outsider**

Average of quoted spreads										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
2.691%	7.955%	5.096%	2.823%	8.939%	6.725%	3.457%	3.573%	3.182%	3.872%	1.327%
19.458%	1.747%	6.882%	16.394%	30.021%	15.118%	19.265%	31.102%	9.743%	47.621%	30.428%
Average of effective spreads										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
2.206%	6.682%	5.030%	2.850%	8.187%	8.418%	3.240%	3.186%	2.610%	2.401%	1.327%
17.226%	2.038%	7.770%	16.827%	30.033%	24.158%	19.356%	30.892%	9.320%	46.239%	32.144%

Table 10. Standard deviation of quoted and effective spreads for each sample and the presumed **favourite**

Standard deviation of quoted spreads										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
1.965%	10.149%	8.847%	0.669%	0.359%	0.064%	0.656%	0.277%	1.291%	0.273%	0.277%
27.386%	10.187%	4.752%	9.065%	13.332%	11.505%	13.955%	31.949%	3.353%	28.076%	25.561%
Standard deviation of effective spreads										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
1.964%	10.131%	8.866%	0.458%	0.500%	0.065%	0.848%	0.424%	1.389%	0.294%	0.329%
27.004%	10.784%	5.235%	10.956%	13.018%	11.006%	14.073%	32.173%	4.154%	26.536%	24.850%

Table 11. Standard deviation of quoted and effective spreads for each sample and the presumed **outsider**

Standard deviation of quoted spreads										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
2.059%	8.561%	4.730%	1.867%	4.725%	7.136%	1.452%	10.463%	2.026%	2.676%	0.100%
32.733%	1.681%	7.838%	27.805%	49.430%	28.304%	42.561%	55.096%	8.651%	61.364%	51.700%
Standard deviation of effective spreads										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
2.350%	8.453%	5.059%	2.234%	4.566%	13.487%	2.155%	10.504%	1.976%	1.673%	0.100%
32.111%	2.640%	9.534%	27.170%	49.212%	74.102%	42.368%	54.100%	8.880%	61.510%	51.014%

Table 12. Frequency of volumes traded bets for each sample and the presumed **favourite**

Frequency of volumes traded bets										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
18,55%	55,03%	21,24%	28,05%	15,22%	10,73%	18,64%	47,34%	6,39%	11,10%	13,69%
81,45%	44,97%	78,76%	71,95%	84,78%	89,27%	81,36%	52,66%	93,61%	88,90%	86,31%

Table 13. Frequency of volumes traded bets for each sample and the presumed **outsider**

Frequency of volumes traded bets										
Period 1 and 2	Suspicious matches									
Reference sample	1	2	3	4	5	6	7	8	9	10
2.97%	2.12%	5.91%	10.11%	0.49%	0.14%	2.08%	1.68%	1.24%	1.09%	1.02%
97.03%	97.88%	94.09%	89.89%	99.51%	99.86%	97.92%	98.32%	98.76%	98.91%	98.98%

Our results show that in nine cases, both for favourites or outsiders, spreads tend to widen over time. The exception was the suspect match n°1, this difference can be explained by the fact that despite 2 hours before the match Arguello was the presumed outsider, he became the favourite just before the start of the meeting and accepted this status until the end. Regarding volatility, results are similar in terms of volumes traded bets. It seems consistent with our assumptions, namely that they increase over time; the only exception comes from the suspect match n°1, where insiders placed their bets very quickly.

Statistical analysis and comparaison

Tables 8 and 9 show the evolution of the average spreads for alleged insider may resemble or differ from the reference sample, in other words, the frauds are more or less obvious depending on the matches. For example, figures 4 and 5 show the evolution of the average quoted and effective spreads for our reference sample and three suspicious matches that we selected.

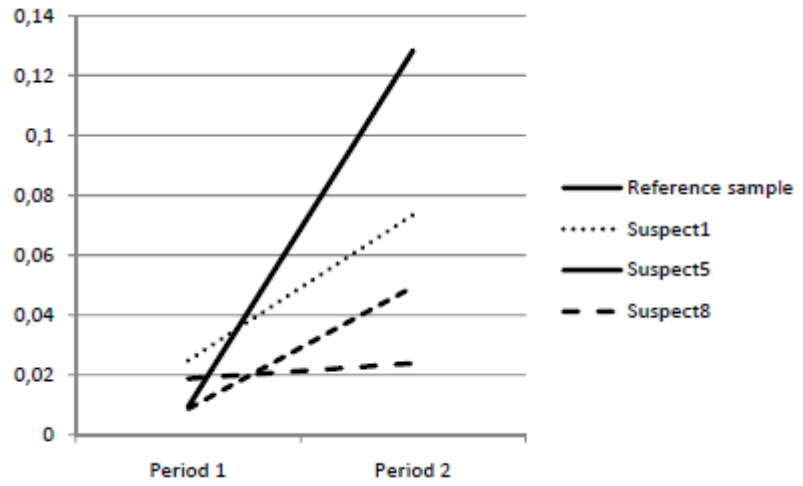

Figure 4. Average quoted spreads for 4 samples and the presumed **favourite.**

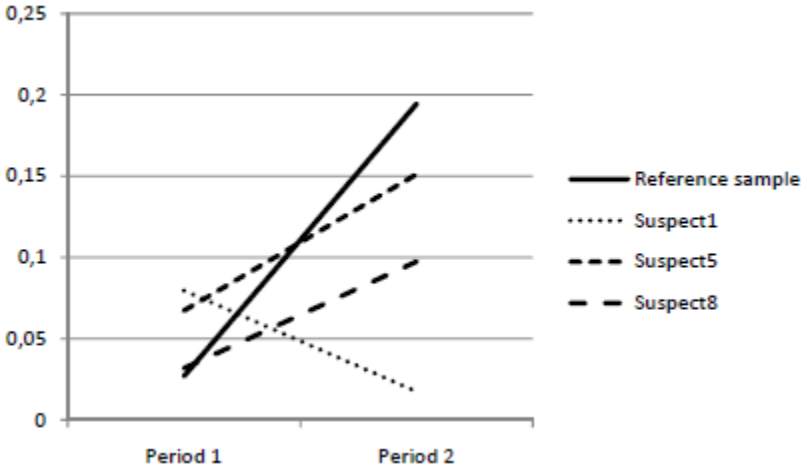

Figure 5. Average quoted spreads for 4 samples and the presumed **outsider.**

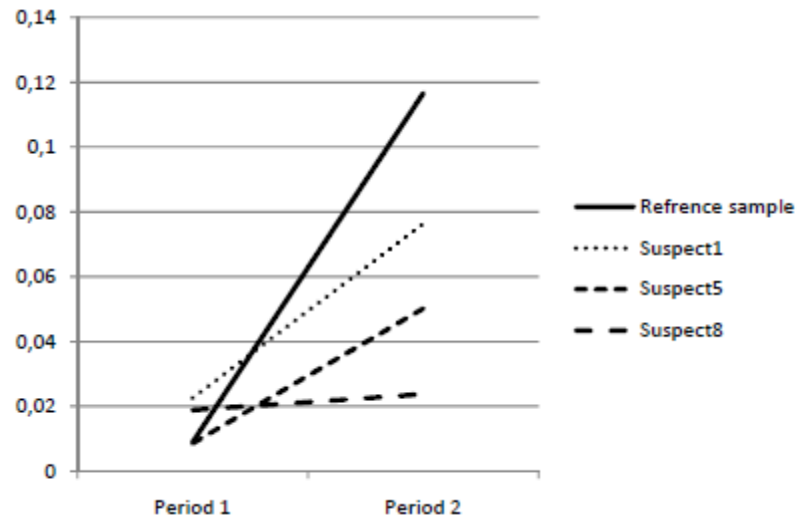

Figure 6. Average effective spreads for 4 samples and the presumed **favourite.**

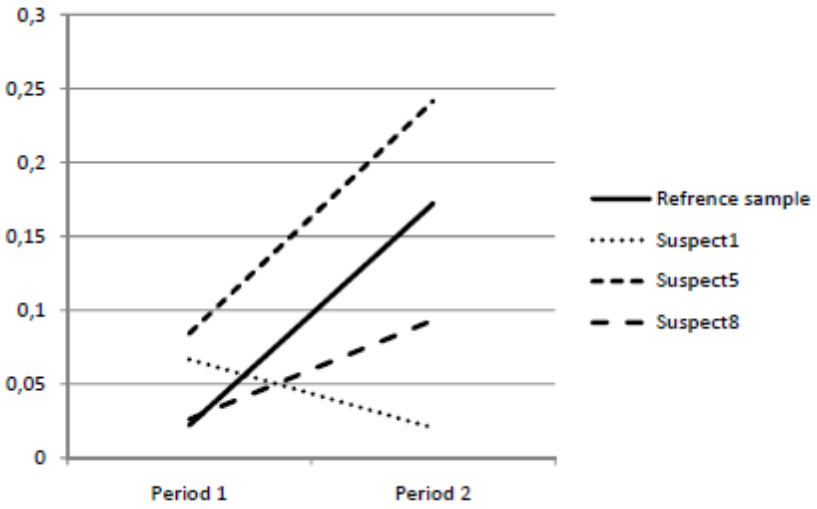

Figure 7. Average effective spreads for 4 samples and the presumed **outsider.**

First, it appears again that the results obtained with the quoted and effective spreads are very close. Some differences exist between the slope of the line of our reference sample and slopes of the line of suspicious matches 1, 5 and 8, particularly in the context of the presumed outsider. Whatever the player tracking is (favourite or outsider), it is observed that the average spreads of suspicious matches 1 and 8 have in common to be above the reference in period 1 and lower in period 2, which is not entirely the case with the suspect match n°5. Concerning the lines, we can see that on this last game, the line is, in all cases, almost parallel to the reference sample line. In contrast, the line of the suspect match n°8 is flatter than others and the line of the suspect match n°1 evolves in a reverse direction in the case of the presumed outsider. Observing these three cases we understand that fraud intensity is different in the

studied games, however, some trends seem to emerge. Results of Mann Whitney U tests corroborate these visual insights. They show that average spreads of suspicious matches, except for games 3, 5 and 10 are significantly different from the average spreads of the reference sample.

Table 14. Mann Whitney test results for each sample and the presumed **favourite**

Average of quoted spreads										
P.value associated with the Mann Whitney test										
Suspicious matches	1	2	3	4	5	6	7	8	9	10
Period 1	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.186
Period 2	<0.0001	<0.0001	<0.0001	0.002	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Average of effective spreads										
P.value associated with the Mann Whitney test										
Suspicious matches	1	2	3	4	5	6	7	8	9	10
Period 1	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.093
Period 2	<0.0001	<0.0001	0.0000	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Table 15. Mann Whitney test results for each sample and the presumed **outsider**

Average of quoted spreads										
P.value associated with the Mann Whitney test										
Suspicious matches	1	2	3	4	5	6	7	8	9	10
Period 1	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Period 2	<0.0001	<0.0001	0.7590	<0.0001	0.1560	<0.0001	0.0008	<0.0001	<0.0001	<0.0001
Average of effective spreads										
P.value associated with the Mann Whitney test										
Suspicious matches	1	2	3	4	5	6	7	8	9	10
Period 1	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Period 2	<0.0001	<0.0001	<0.0001	<0.0001	0.0430	<0.0001	0.1050	<0.0001	<0.0001	<0.0001

In terms of trends, firstly, for the period 1, average spreads listed for most suspicious matches are superior to our reference sample. This phenomenon seems aligned with most empirical results highlighted on the financial markets, namely that the positions of the insiders have the effect of significantly expanding the spread. But the most interesting phenomenon that this study suggests is this impact will be even stronger than expected when the proportion of uninformed bettors is low. Indeed, during period 2, suspicious matches average spreads is significantly lower than the reference sample, which corresponds to the explosion in volumes traded bets. In agreement with the results of Cornell and Sirri (1992) within the financial markets, we can assume that in period 2, the proportion of uninformed bettors clearly

increased. It implies asymmetry information which was offset by higher trading volumes uninformed bettors. As a result, insiders will have the opportunity to make their transactions directly with the uninitiated traders which will reduce the immediate impact of adverse selection costs and therefore reduce the spread. It is possible to refine these observations based on the standard deviation of the spread. Concerning the presumed outsider, the standard deviation of the spreads associated with suspicious matches is in most cases, higher than our reference sample for the period 1 (the results are more mixed for the presumed favourite), suggesting that insiders will tend to place their bets just before the start of the meeting in order not to significantly influence the other punters who have much lower level of information. This idea is consistent with that made by Cain, Law, Peel (2001) in connection with horse racing. In this industry, in order to maximize their profits, according to the structure of the market, insiders will make every means at their disposal to not disclose their information to other bettors.

Conclusion

Study and comparison of the universe of sports betting with the financial markets has highlighted interesting similarities between these two worlds, confirming this point in previous studies. These similarities allow to observe the behavior in one of two universes-in this case that of sports betting- to try, by similarity, to understand what may happen in the other. In this context, it is the study of the spread as part of the demonstration of insider trading that has been emphasized. In this context, several observations have been highlighted.

First, there is a similarity of the spread's behavior on the financial markets and the sports betting market. Under the presumed favourite's case, once the match starts (period2), quoted and effective spreads is elevated because there is no strong consensus on the conduct of the match. Then, it narrows slightly before expanding gradually as the match issue doesn't make doubts. Transposed to financial markets, we can understand the daily *smile effect* of the spread: early in the day, the consensus is uncertain, then, as trades increases, investors converge on the price and, finally, at the end of the session, all important exchanges are treated and the spread widens again because it only stays in the order book the orders of operators who diverge on the stock price. Figure 8 shows the quotes of a stock, Rodriguez, and clearly illustrates this phenomenon.

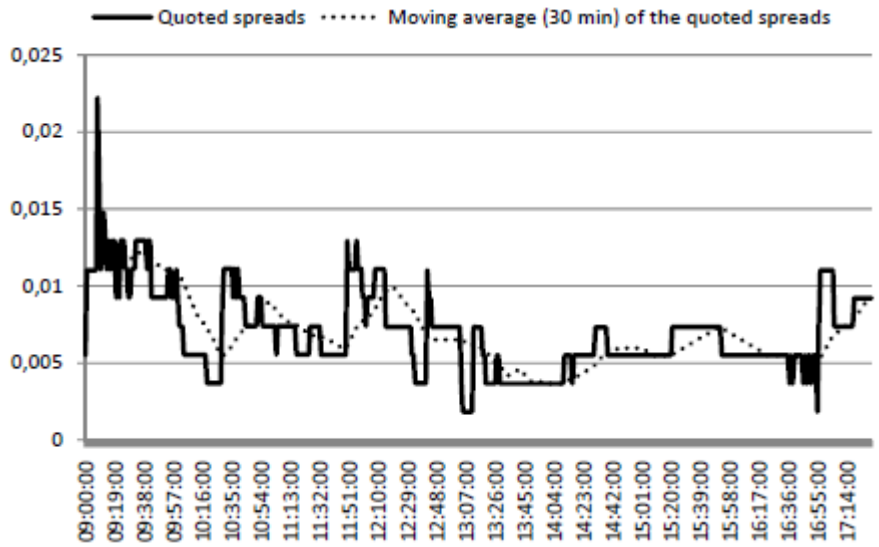

Figure 8. Evolution of the quoted spread (format 1 minute) and its moving average (format 30 minutes) of a stock (Rodriguez) randomly selected during a trading day[14].

In the world of sports betting, the *end of match* effect imply that spread (quoted or effective) increases in an important way towards the end of the match in the favourite's case. Nevertheless, it should be noted that in the presumed outsider's case, spread behaves in an inverse way and is reduced. Secondly, and this is the answer to the original question of this article, the spread's behavior in the world of sports betting in case of fraud is quite similar to observations made in the financial markets in case of insider trading. The use of the spread as an indicator of the existence of insider trading is therefore reinforced.

Deformation of the range appears to be a specific reaction to the occurrence of an event, in our case of insider trading. In the sports betting market, it helps to attract the attention of bettors, but does not directly alter their wealth. In financial markets, it is important that this tool will be more systematically used by market authorities in the search for signs of insider trading.

If we use the spread like a tool for reporting suspicious matches in the analyzed sample analyzed, we will obtain the following results with the effective spreads:

F = Fraud ; NF = No Fraud ; PF = Possible Fraud

	Effective spreads		
Match title	Favourite	Outsider	Verdict
Suspect1	F	F	F
Suspect2	F	F	F
Suspect 3	F	F	F
Suspect4	F	NF	PF
Suspect5	NF	NF	NF
Suspect6	F	NF	PF
Suspect7	NF	NF	NF
Suspect8	F	F	F
Suspect9	NF	NF	NF
Suspect10	NF	NF	NF

If one assumes that a match which have to be studied in details is a game where the spread is greater than the reference sample in period 1 and then smaller in period 2 in the case of the favourite and the outsider, the matches 1, 2, 3 and 8 should be considered highly suspect when a suspicion thinner could be worn on the matches 4 and 6, and finally; matches 5, 7, 9 and 10 could pass our test.

Appendix 1

Matches	Players	ATP J1	ATP J2	Date	Tournament	Round	Score
Suspect1	Arguello - Davydenko	87	4	02/08/2007	Sopot	2nd Round	2-6 6-3 2-1
Suspect2	Prysiesny - Volandri	207	28	30/07/2007	Sopot	1st Round	6-4 2-6 7-5
Suspect3	Pashanski - Starace	120	32	30/07/2007	Sopot	1st Round	6-7 6-2 3-1
Suspect4	Tipsarevic - Murray	68	18	11/10/2007	Moscou	2nd Round	6-4 7-6
Suspect5	Berrer - Youzhny	72	17	11/10/2007	Moscou	2nd Round	7-5 6-4
Suspect6	Benneteau - Simon	68	32	25/10/2007	Lyon	2nd Round	7-5 6-4
Suspect7	Mahut - Haas	71	11	27/09/2007	Bangkok	2nd Round	6-4 6-4
Suspect8	Monfils Davydenko	57	5	27/09/2007	Gstaad	1st Round	3-6 6-4 7-5
Suspect9	Rochus - Davydenko	34	4	28/02/2007	Dubai	2nd Round	4-6 6-4 6-2
Suspect10	Soderling - Davydenko	26	3	14/02/2007	Marseille	1st Round	3-6 6-4 6-1

Comments:

All tournaments involving players are minor tournaments, defeated by the favourites has virtually no effect on the ATP rankings.

On our ten suspicious matches detained, the name Davydenko appears four times.

Suspect 1. Davydenko – Arguello

Suspicion = Match-fixing in favor of the outsider

Resume: The bookmakers suspect that the result is settled in favor of Arguello with the complicity of both players. Davydenko, (ATP n°4), was heavily favored several hours before the meeting with odds between 1,18 and 1,2 which correspond to an estimated probability of winning greater than 80%, therefore, Arguello had the status of Outsider . However just before the match begins, the odd of Arguello fell from 5,75 to 1,51 while that Davydenko mounted to 3. Curious variations were recorded during the match, for example, the odd of Davydenko has remained almost unchanged while the latter had won the first set.

Finality: Arguello won the match following the abandonment of Davydenko at the third set.

Suspect 2. Przysiezny - Volandri

Suspicion = Match-fixing in favor of the outsider

Resume: At the opening of bets, the odd of Volandri (ATP n°29) was 1,10 against Przysisiezny (ATP n°237). His odd was then violently deflected to attach to 1,80 prior to the meeting. Volandri was febrile at critical moments (break points against him), with 5 double faults that have played in favor of his opponent.

Finality: Volandri lost in 3 sets.

Suspect 3. Pashanski - Starace

Suspicion = Match-fixing in favor of the outsider

Resume: After winning the opening set, Starace has shown surprising signs of weakness, does not put up resistance on the service of Pashanski, his winning return percentage fell back to 21%, which is particularly low.

Finality: Pashanski won the match following the abandonment of Starace in the 3rd set.

Suspect 4. Tipsarevic - Murray

Suspicion = Match-fixing in favor of the outsider

Resume: Murray gave the impression to master the game offering himself several break points, but he did not put enough belief and finally converted one out of ten. Tipsarevic certainly made a good game but observers have judged the performance of Murray quite disappointing.

Finality: Tipsarevic won in two sets

Suspect 8. Davydenko – Monfils

Suspicion = Match-fixing in favor of the outsider

Resume: After winning the first set easily, observers have found that Davydenko had moments of weakness in crucial moments of the match. We should add that Davydenko was seeded n°1 in the tournament, and n°5 in the world while his opponent, Monfils was not even seeded world n° 57, and had no fitness level convincing, it has also tilted the next game.

Finality: Monfils won in 3 sets.

Suspect 9. Rochus – Davydenko

Suspicion = Match-fixing in favor of the outsider

Resume: Davydenko showed a relatively low quality of service with 58% of first service managed against 68% for his opponent, as he made 8 double faults while his opponent did not make any.

Finality: Rochus won in 3 sets.

Appendix 2

Matches	Players	ATP J1	ATP J2	Date	Tournament	Round	Score
Match1	Mathieu - Gasquet	33	26	10/02/2009	Rotterdam	1st Round	3-6 7-6 6-3
Match2	Clement - Schuettler	73	31	24/02/2009	Dubai	1st Round	1-6 6-3 6-1
Match3	Schwank - Andreev	154	33	12/02/2008	Costa do Sauipe	1st Round	6-3 6-2
Match4	Johansson - Hernandez	73	51	10/07/2007	Cattella	1st Round	6-4 6-4
Match5	Andreev - Stepanek	24	20	29/07/2008	Cincinnati	1st Round	7-6 6-7 7-6
Match6	Berrer - Vanek	96	66	17/04/2008	Estoril	2nd Round	6-3 3-6 7-5
Match7	Lapentti - Santoro	113	37	09/10/2007	Moscou	1st Round	3-6 6-3 6-2
Match8	Serra - Vliegen	111	66	10/10/2007	Moscou	2nd Round	6-7 6-4 6-2
Match9	Henman - Verdasco	41	26	16/10/2006	Madrid	1st Round	7-5 6-3
Match10	Lopez - Ferrer	40	5	15/10/2008	Madrid	2nd Round	6-4 7-6
Match11	Koubek - Acasuso	67	49	17/07/2007	Stuttgart	1st Round	6-2 7-5
Match12	Querrez - Moya	50	14	21/04/2008	Monte Carlo	1st Round	6-3 1-6 6-3
Match13	Hidalgo - Simon	137	33	22/04/2008	Monte Carlo	1st Round	6-1 6-1
Match14	Vliegen - Safin	52	26	17/04/2007	Monte Carlo	2nd Round	0-6 7-6 6-4
Match15	Youzhny - Berdych	62	24	16/02/2009	Marseille	1st Round	4-6 6-4 7-5
Match16	Gabashvili - Berrer	133	59	12/02/2008	Marseille	1st Round	7-6 6-4
Match17	Clement - Safin	78	26	17/02/2009	Marseille	1st Round	4-6 6-3 6-2
Match18	Soderling - Nieminem	52	26	11/02/2008	Marseille	1st Round	6-3 5-7 7-5
Match19	Zabaleta - Almagro	85	31	30/07/2007	Sopot	1st Round	6-3 7-5
Match20	Gabashvili – Mayer	125	77	26/02/2008	Zagreb	1st Round	7-6 6-0
Match21	Zverev – Seppi	79	33	03/02/2009	Zagreb	1st Round	7-6 3-6 6-2
Match22	Bolleli – Cilic	66	45	27/02/2008	Zagreb	2nd Round	6-4 6-3
Match23	Rochus – Kunitsyn	76	43	09/02/2009	San Jose	1st Round	6-2 7-6
Match24	Becker – Moya	79	15	25/09/2007	Bangkok	1st Round	7-6 6-4
Match25	Meffert – Kunitsyn	228	81	27/09/2007	Bangkok	2nd Round	3-6 7-6 6-0

Notes

1. Data are available on the blog midas.org at the following address: http://www.midasoracle.org/2007/08/23/does-betting-exchange-betfair-handle-more-daily-trades-than-the-new-york-stock-exchange.
2. This situation is more comfortable for the researcher because orders can be hidden in financial markets.
3. Betfair applies a loyalty program which allows his best customers to benefit from reduced transaction costs, the minimum being fixed to 2%.
4. i can take two values: 1 if the player is the presumed favorite, 2 if it is the presumed outsider.
5. n can take two values: 1 for the period 1 and 2 for period 2.
6. z can take 25 values depending on the chosen game.
7. The format of our data does not allow us to get the price at the time of the transaction but the last traded price.
8. See Appendix 1.
9. The initial list of 21 suspicious matches retained by the ATP was published on the Italian website Pianeta Tennis before being resumed on numerous forums.
10. By outsider presumed, it refers to a player whose odd is greater than 2 during the first listing registration (2 hours before the game), that is to say a probability of winning less than 50%.
11. See Annex 2
12. The number of spreads selected for the match between Davydenko and Monfils is much larger than the others because it took longer (3 sets) and liquidity has been provide
13. Source: Fracsoft
14. Source: Bloomberg

References

Avery, C. and Chevalier, J. (1999), Identifying Investor Sentiment from Price Paths: The Case of Football Betting , *Journal of Business, 72,* 493-521.

Barraud C. (2012), Le marché des paris sportifs : Un cadre d'Observations Simplifié proche des Marchés, http://chris.barraud.free.fr/, Working Paper, SDFI-LEDA.

Cain, M., Law, D. and Peel, D.A. (2001), The incidence of Insider Trading in Betting Markets and the Gabriel and Marsden Anomaly, *Manchester School, 62,* 197-207.

Chung, K. and Charoenwong, C. (1998), Insider Trading and Bid-Ask Spread, *Financial Review, 33,* 1-20.

Cornell, B. and Sirri, E. (1992), The Reaction of Investors and Stock Prices to Insider Trading, *Journal of Finance, 47,* 1031-1060.

Fishe, R.P. and Robe M.A. (2004), The Impact of Illegal Insider Trading in Dealer and Specialist Markets: Evidence from a Natural Experiment, *Journal of Financial Economics, 71,* 461-488.

Forrest, D. and Simmons, R. (2002), Outcome Uncertainty and Attendance Demand: The Case of English Soccer, *Journal of the Royal Statistical Society (Series D), The Statistician*, *51*, 229-241.

George T.J., Kaul G. and Nimalendran M. (1991), Estimation of the Bid-Ask Spread and its Components: A New Approach , *The Review of Financial Studies*, *4*, 623-656.

Guivarc'h, A. (1996), Conséquences des opérations d'initiés sur la composante d'asymétrie d'information de la fourchette, CEREG, Cahier de recherche n°9605.

Guivarc'h, A. (1997), Contribution des initiés à l'efficience informationnelle, CEREG, Cahier de recherche n°9712.

Huang, R.D. and Stoll, H.R. (1996), Dealer Versus Auction Markets: A paired Combetson of Execution Costs on NASDAQ and the NYSE , *Journal of Financial Economics*, *41*, 313-357.

Jarque C.M. and Bera A.K. (1987), A Test for Normality of Observations and Regression Residuals, *International Statistical Review*, *55*, 163-172.

Kyle, A. (1985), Continuous Auctions and Insider Trading, *Econometrica*, *53*, 1315-1336.

Mann, H. B. and Whitney, D. R. (1947), On a Test of Whether one of two Random Variables is Stochastically Larger than the Other , *Annals of Mathematical Statistics*, *18*, 50-60.

Meulbroeck, L.K. (1992), An Empirical Analysis of Illegal Insider Trading, *Journal of Finance*, *47*. 1661-1699.