LIBERA UNIVERSITA' INTERNAZIONALE DEGLI STUDI SOCIALI

Department of Economics and Finance, Chair of Financial and Credit Derivatives

BETTING EXCHANGES: A MARKET MAKER PROCESS

(Abstract)

Supervisor:

Professor Federico Nucera

Candidate

Dario Occhipinti 656531

Co-supervisor:

Professor Raffaele Oriani

Table of contents

Since ancient times, people have wagered money on every kind of events or games whether as leisure activities or as a way to enrich themselves.

During centuries, new forms of business have developed in order to satisfy the never-ending demand of gambling and betting.
In 1790, Harry Ogden opened a business that has been recognised as the first bookmaker in the United Kindom. Until 2000, betting with a bookmaker was the typical way to wager on an event.
Since 2000, however, a new player has arisen in the industry: betting exchange.

In this thesis we will analyse the implementation of a Market Maker (MM) model, which trades on the betting exchange market Betdaq and allows a bookmaker to reduce the negative liability and the level of active risks. The creation of a MM model is needed considering the relative dimension of betting exchange and bookmakers.

Considering the vastness of the events, this thesis will focus only on the horse races sector. This sector, even though it is decreasing, continues to be the most important category of betting in the United Kingdom. From October 2013 to September 2014, bettors have wagered around 5.42£ billions on horse races, 1.36£ billions on football matches and around 1.31£ billions on dogs races[1].

I have used real data provided by Ladbrokes plc. The databases contain:

- 1-year sample of orders on the betting exchange Betdaq;
- 1-year sample of all the bets wagered by Ladbrokes's customers on horse races.

In **Chapter 1**, all the notions needed to fully understand the betting industry are explained.
A bet is a contract on some future cash flow based on the outcome of a given event (e.g. horse race, football match and so on). Therefore, the cash flow is determined by the outcome of the underling itself and by the price of the contract, i.e. the odd. The fixed-odds betting is the most common way to wager on sports events and the main characteristic is that the bettor knows the odds at the time he places a wager.
The two main typologies of business of this industry are: Bookmaker and Betting Exchanges.
The first, widely acknowledged, bookmaker in the United Kingdom is Harry Ogden[2], who sets-up a business in the 1790s.
Bookmakers provide public odds[3] at which they will accept any amount bet. Considering these characteristics, the bookmaker market is a quote-driven market in which bettors can only decide to bet at the price decided by the bookmaker. The transparency on this kind of markets is very low, as only the last odds are public. There are no information about the price's formation rule or about the historical price changes.
The odds quoted by the bookmarkers contain already the so-called overround, which is their compensation for the services that they are providing and for bearing the risk of unfavourable outcomes.

---

[1] U.K. Gambling Commission- Industry Statistics, June 2015

[2] Munting R. "An Economic and Social History of Gambling in Britain and the USA"

[3] The set of odds provided for one event is called Sportbook or simply book.

Betting Exchanges, instead, *"operate as order-driven markets, where buyers and sellers trade directly with each other in a continuous double auction without the intermediation of market makers"*(Flepp, Nüesch, Franck, 2013[4]).
Betting exchange, thus, allow bettors not only to bet on a positive outcome (e.g. the victory of a team or of a horse), but also to act as a bookmaker and to sell the bet, thus it allows to bet on a negative result (e.g. the loss of a team, horse and so on). As in financial markets, bettors can place a limit order and wait that someone hits their order or they can submit a market order and match an already existing bet.
In the thesis are shown some examples, which help to understand the exchange dynamics.

These two types of business differ for many elements. The most important one is the different level of risk faced.
The owner of the exchange does not take any trading position, thus, he is not bearing any inventory risk. His revenues derive from the commission charge on the net profit, thus it is a certain cash flow. Whereas, the bookmaker takes the opposite side of every transaction, therefore, he needs to manage an active risk. His profit is not certain. He can lose a considerable amount of money in case of over-exposition on the final result.

Chapter 2, "Market Analysis and descriptive statistics", contains liquidity data and descriptive statistics of the betting exchange market Betdaq.
The dataset, provided by Ladbrokes plc, contains the data of 12571 horse races between 01/01/2013 and 31/12/2013.
The amount placed and matched has been converted in dollars using the exchange rate provided within the database.

The analyses made are:

1 Horse race classification:

It is important to underline that horse races have various importance and consideration across the gambling community. Some races, as the Royal Ascot, have a predominant role, while others are just considered "local" events. The amount of money wagered and matched is directly proportioned with the importance of the race.
In order to classify the data, after computing the average total matched amount of all races equal to 131,099.38$, the races have been grouped based on their total amount matched[5]:

1 Class 5 Races with lower than average betting volume (4164[6])
2 Class 4 Race with an average betting volume (5118);
3 Class 3 Races with a moderate betting volume (3013);
4 Class 2 Races with higher than average betting volume (203);
5 Class 1 Races with very high betting volume (73).

[4] Flepp R., Nüesch S., Franck E. "The liquidity Advantage of Quote-driven Markets: Evidence from the Betting Industry". University of Zurich, 2013
[5] We are following the idea of Michael A. Smith & David Paton & Leighton Vaughan Williams, 2006. "Market Efficiency in Person-to-Person Betting," Economica, London School of Economics and Political Science, vol. 73(292), pages 673-689, November.
[6] Number of races in that class.

2 Liquidity study

The betting exchange is an order-driven market, thus liquidity is the engine of the exchange. Moreover, the “stock” traded in this kind of market (i.e. the bet on a horse) has a maturity, which is the end of the races, and it strongly influences the way customers trade on the exchange.
In order to study the liquidity, I have computed the ratio between the total amount matched[7] until time *t* and the total amount matched in the whole race (see formula 1). Doing so for a standardized timespan (i.e. every one millisecond), we can study the evolution of the total matched amount during the race’s life and we can make a comparison among different races.

$$Liquidity\ ratio = {}^{AM_t}/_{TL} \qquad (1)$$

Where:
AM_t is the total amount matched until time *t;*
TL is the total amount matched in the race.

The analysis has been done on 73 races of class1 and 203 races of class 2. The results show that the betting exchange is illiquid until 50 minutes before the start of the race (e.g. the % amount is equal or less than 10% for both the classes), after that it increases in an exponential way reaching around 95% at the beginning of the race. The results are nearly the same for both the classes.

3 The average bet’s size;

This study is important to define a feasible bet’s size that will be used for the MM model.
One characteristic of the bet’s size is that it is different for each kind of horses (i.e. the size on the favourite horse is bigger than the one of the last favourite).
Therefore, in order to take into account this characteristic, I am assuming that it is possible to classify and aggregate each horse of every race by their average matched price of the pre-live period.
In order to clarify, we can consider the favourite horse case. The assumptions based on real observations are: it has the lowest average price in the pre-live period and each race has a favourite runner. Thus, it is possible to aggregate the favourite horses of each race that we want to analyse.
Thanks to this classification it was possible to study every kind of representative horse and obtain a standardized bet’s size for each of them.

4 Volatility study.

An important element for every financial model is the price volatility. Effectively, a higher volatility will require a thicker spread to protect the MM from the uncertainty linked with the price. The procedure to compute the standard deviation is explained in the **appendix 1** of the thesis.
The results show that the S.D. is generally low. It tends to increase as we reach the start of the race. This can be explained by the fact that, until 25 minutes of the race’s start, there are no trades on the exchange, hence, the price remains stable.

[7] The amount matched is the total amount wagered by costumer (i.e. the total amount backed which has been accepted in laid).

It was discovered that, in some cases (mainly on the last favourite horse), the S.D. can be extremely high. This happens when the back-lay spread on the market is wide and the price matched bounces from the lower bound to the higher bound.

In **chapter 3**, "The Market-Maker model", it is presented the MM model developed.
The first element to define is the model's strategy. The best way to hedge the liability[8], while acting as market maker, is to back and lay on all the horses in an asymmetrical way. This means that we want to back more than lay[9] on the horse with the higher negative liability and we want to lay more than back on the other horses.
The formula (3) shows how to compute the liability on one horse considering the model' strategy.

$$TL_k = \sum_{i=1}^{N} X_i - \sum_{i=1}^{N} Y_i + \mathbf{1}_{\boldsymbol{k}}\left[\sum_{l=1}^{L}\left(Y_{k,l} * BP_{k,l}\right) - \sum_{j=1}^{J}\left(X_{k,j} * LP_{k,j}\right)\right] \qquad (3)$$

Where:

TL_k is the total liability on the k-th horse;
N is the number of horses;
X_i is the total amount wagered, by the bookmaker's customers and by the bookmaker itself with the lay orders on the exchange, on the i-th horse;
$\sum_{i=1}^{N} X_i$ is the amount wagered on all the horses;
$\sum_{i=1}^{N} Y_i$ is the amount the bookmaker wagered on the exchange on all the horses,
$\mathbf{1}_\mathbf{k}$ is the indicator function. It is equal to 1 if the k-th horse wins the race, 0 otherwise;
L is the total number of back-order made on the k-th horse;
$Y_{k,l}$ is the amount betted on the l-th order on the k-th horse;
$BP_{k,l}$ is the back price available on the exchange for the k-th horse and the l-th order;
$\sum_{l=1}^{L}(Y_{k,l} * BP_{k,l})$ is the amount the bookmaker will receive if k-th horse wins the race;
J is the total number of lay-order made on the k-th horse;
$X_{k,j}$ is the amount betted on the j-th order on the k-th horse;
$LP_{k,j}$ is the lay price (made on the exchange or by the bookmaker[10]);
$\sum_{j=1}^{J}(X_{k,j} * LP_{k,j})$ is what the bookmarker has to pay in case of victory of the k-th horse.

With this strategy, the model is reducing the negative liability in two ways[11]:

1. Backing on the horse with the negative liability. Thus, if that horse will win the race, the bookmaker will receive money from the exchange for winning the bets. In the formula (2), the $\sum_{l=1}^{L}(Y_{k,l} * BP_{k,l})$ for the k-th horse and $\sum_{i=1}^{N} Y_i$ for all horses are increasing. For the horse with a negative liability, the positive effect of the first term is higher than the negative

[8] The liability is the amount that the bookmaker could win or lose when that horse wins the race. When it is positive is a profit, whereas, when it is negative it is a loss.
[9] *Backing* a horse is betting on the victory of that horse. Whereas, *laying* is betting on the outcome that the horse will lose the race.
[10] As we known, (from chapter 1) the exchange's odds are different from the bookmaker's one. Thus, $LP_{k,j}$ will be the price offer by the bookmaker when a customer bet on the sportbook. Whereas, it will be the lay price when the bookmaker place a lay order on the exchange.
[11]We are considering just the final effect. The model will back and lay on all horses, however, thanks to the asymmetrical betting, the net effect will be just back orders for the horses with negative liability and lay orders for the horses with a positive one.

effect of the second term. For all the other horses, we have just the negative effect of the second term which is small;

2. Laying on all the other horses. In this way, the model is increasing the cash inflow from the other horses[12]. In the formula (2) the $\sum_{i=1}^{N} X_i$ term for all the horses and $\sum_{j=1}^{J}(X_{k,j} * LP_{k,j})$ for the k-th horse are increasing. Therefore, for the horse with a high negative liability there is a positive effect, which helps to complete the hedge. Whereas, for all the other horses the negative effect will be greater and, as result, the positive liability on those horses will decrease[13].

Furthermore, the strategy followed by the model must be dynamic. If the liability on one horse, during the "lifecycle" of the race, changes it sign (i.e. it moves from negative to positive), the model should change dynamically the "direction" of its bets (e.g. when the liability is negative the model will back more than lay, however, when it will become positive the MM will start to lay more than back).
In addition, we should keep in mind that the overround times the amount wagered, which can called "limit", by Ladbrokes's customers is the amount that, in an ideal situation, we will earn regardless the result of the race.

Therefore, for the above cited characteristics, the model will change its behaviour based on the position and the limit, which can be derived from multiplying the overround with the wagered amount. In table 16, we can see the MM behaviour[14].
It is important to underline that on the favourite horse we want to back more than lay regardless the liability sign[15].

Table 1 Model behaviour

Scenario	MM behaviour
$Liability_k < 0$	Backing more than laying
$0 < Liability_k < limit$	Do not trade on the horse
$Liability_k > limit$	Laying more than backing

While the first and third scenarios are easy to understand, instead it is better to explain why the mode will not trade in the second scenario. When the liability is positive and higher than the limit, the model will lay more than back. Doing so the liability will be reduced over time and, without any non-trading area, it can easily changes sign and become a negative value. If the model stops to trade when the liability is lower than the limit, we can avoid to transform a positive liability to a negative one.
On the other hand, if we start with a negative liability when it will become positive, without a non-trading area, the model will start to lay more than back and the result will be a negative value again. Implementing a non-trading area is important to make the model more efficient.

[12] As we said, when you made a lay order on the exchange you will receive the amount wager, which is the bookmaker's turnover.
[13] That is a logical result considering that we are implementing a hedging strategy.
[14] $Liability_k$ is the value of the liability of the k-th horse.
[15] This is because the final liability on the favourite horse will be, in most of the cases, negative and the liquidity on the exchange is not enough to hedge the liability completely.

Now that the model's strategy and behaviour has been explained, we need to define the other key elements to implement "materially" the MM model:

1. The expected mid-price

   In this thesis, the estimation of the mid-price has been computed as in a simple weighted average (formula 4).

$$Ep = Best\ BP * (1 - Pup) + Best\ LP * Pup \tag{4}$$

   Where:

   Ep is the Expected price,
   $Best\ BP$ is the best back price available on the exchange,
   $Best\ LP$ is the best lay price available on the exchange,
   Pup is the probability of an upward price movement ,
   $(1 - Pup)$ is the probability of a downward price movement.

   To compute the Pup, following the methodology of Avelanda et al, (2011)[16], it was exploed the information implied in the level 1 quotes of the Order Book (OB).

   The main idea of the work of Avelanda et al,(2011)[16] is that "*the dynamics leading to a price change may thus be viewed as a race to the bottom*" In effects, if the back market orders deplete the best lay size before the lay market orders deplete the best back size, we will have a price's increase. Otherwise, if the lay market orders deplete the best back size before the back market orders deplete the best lay, the price will decrease.

   Thus, the probability of an up-movement was computed using the formula (5):

$$Pup = \frac{Best\ Back\ Volume}{(Best\ Back\ Volume + Best\ Lay\ Volume)} \tag{5}$$

   In order to test the formula (4) and (5), it was made a Matlab code to obtain the order book and to apply the formulas in the horse races world. The result, shown in table 17 and 18 of the thesis, are extremely positive. The general prediction error is around 0.5 ticks and the standard deviation is low.

2. The back-lay spread

   The spread should be a function of the following parameters:

   - Time and liquidity. Less liquid markets require a bigger spread. Considering that liquidity is also a function of time in the betting exchange, the spread must be itself a function of time;
   - Volatility. A more volatile market means that the risk of acting as market maker is bigger. Thus, a bigger spread is needed;
   - Position. When the exposure on one horse increases considerably, it is possible to increase or decrease the spread to handle it in the best way possible.

[16]Avellaneda, M., Reed, J. & Stoikov, S., 2011. Forecasting Prices from Level-I Quotes in the Presence of Hidden Liquidity. Algorithmic Finance, 29 June, 1(1)

Taking the first-two elements in consideration, I have computed the spread in the following way:

$$S_{T+1} = max(1, round(e^{-X} * \frac{As_t}{2} * (1 + \sigma)) \quad (6)$$

Where:

S_{T+1} Is the spread at time t+1 (expressed in number of tick);
As_t Is the actual spread;
σ Is the historical price's standard deviation;
X Is the historical mean percentage value of amount matched during time (is the value of the liquidity ratio defined in chapter 2.2);

In order to handle the "Inventory risk", instead, a bigger spread will be applied to the orders, which will worsen our position[17].
Therefore, for the horses in which we want to back more than lay, we will compute the back and lay price[18] as:

$$Back\ price = Ep + S * v \quad (7)$$
$$Lay\ price = Ep - (S + 1) * v \quad (8)$$

Whereas, for the horses in which we want to lay more than back, it will be:

$$Back\ price = Ep + (S + 1) * v \quad (9)$$
$$Lay\ price = Ep - S * v \quad (10)$$

Where:
Ep Is the Expected price;
S Is the spread;
v Is the value of the tick.

3. The orders size

   In order to decide a size for the model's limit orders, the results obtain in the AVG bet's size were used. In order to implement the strategy of asymmetrical betting the size obtain from the studied was multiplied by $1 + \beta$. For the horses with a high negative liability, β is equal to 2 for the back orders and equal to 1 for the lay orders. For the horses with a positive liability, instead, it will bell exactly the opposite.

In **chapter 4,** the results of the application and testing of the model are presented.
The first test was made in order to see, through graphs, if the formula (6) can produce feasible back and lay prices. The overall results are extremely positive, in effect, it is shown (figure 9-10) that on horses with a negative liability the formulas (7) and (8) allows the MM to back more than lay. Whereas, on horses with positive liability the formulas (9) and (10) allows to lay more than back (see figure 9 and 10)

---

[17] E.g. if the liability on one horse is negative, we want to reduce the possibility that a lay order will be hit. Thus, we will subtract a higher spread to the Ep.
[18] Note: Here we are seeing the back and lay price by the market maker point of view. In the market the Lay price would be the back price and the back price the lay price. In financial markets the MM buys at bid and sells at ask price.

The second test has been done in order to see the effect of the MM model on the Ladbrokes's liability and profitability. In the thesis are shown the results on two different races with graphs and tables that presents the evolution of the matched price, the liability, all the trades made on the exchanges as well as the final effects on the Ladbrokes's profitability for each horse. In this abstract we will see only the table 19, which sum-up the effects of the MM model.

In table 19, the winning probability column shows the implied probability derived by the bookmaker's price. The true probability column, instead, shows the winning probability without the overround[19]. The Liability pre MM is the liability faced the bookmaker without the model. Whereas the Liability post MM is the final liability that we will obtain thanks to the model and it was computed following the formula (3). The value obtain in these columns are the profit/loss of the bookmaker in case that horse win the race (e.g. if horse 1, in race 1, win the race the bookmaker will lose 20,334$ pre MM and 5,431$ post MM). The MM effect column shows the change in the liability. The sportbook turnover column shows the amount betted by Ladbrokes' customers, while the exchange turnover column shows the turnover generated on the exchange. It is important because, considering that Betdaq was bought by Ladbrokes, that amount times 1.2%[20] is an extra profit that the model generated[21]. The amount betted column shows the amount of £ that the model need to perform, when negative, or generate, when positive, during the 25 minutes before the start of the race[22].
The last two columns show the profit's expected value (EV) with and without the MM model. In order to compute so, I have multiplied, for each horse, both sportbook and hedged liability by the true probability of winning the race and I have summed-up the single results. In order to make a complete comparison, I have computed also the relative standard deviation, the range of EV[23] and the actual bookmaker's profit/loss.

[19] We can obtain the true probability from the winning probability with a simple proportion. See chapter 1.3.
[20] The owner of the exchange charges a commission fee on the net profit. Betdaq charges a standard rate of 5%, but it can be reduced as low as 2%, depending on the total amount wagered on the exchange during time. The value of 1.2% it consider also tax on the profit.
[21] This, of course, is valid just in the case that the bookmaker is also the owner of the exchange.
[22] This is what we defined as cash in-flow and outflow in chapter 3.1 .
[23] For the range of EV it was taken the minimum and the maximum EV pre and post MM.

Figure 1 Spread on horse 1

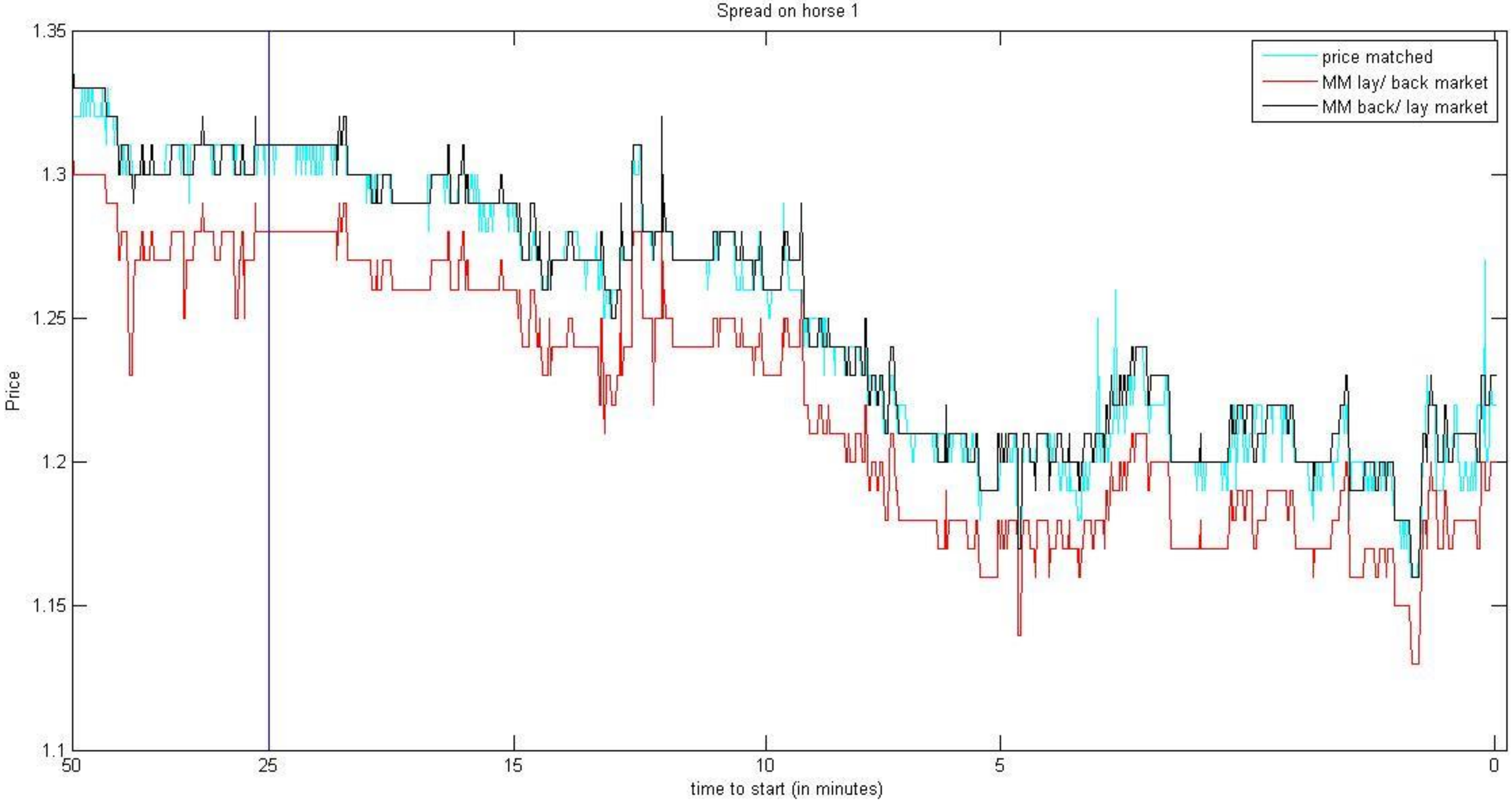


Figure 2 Spread on a normal horse 7

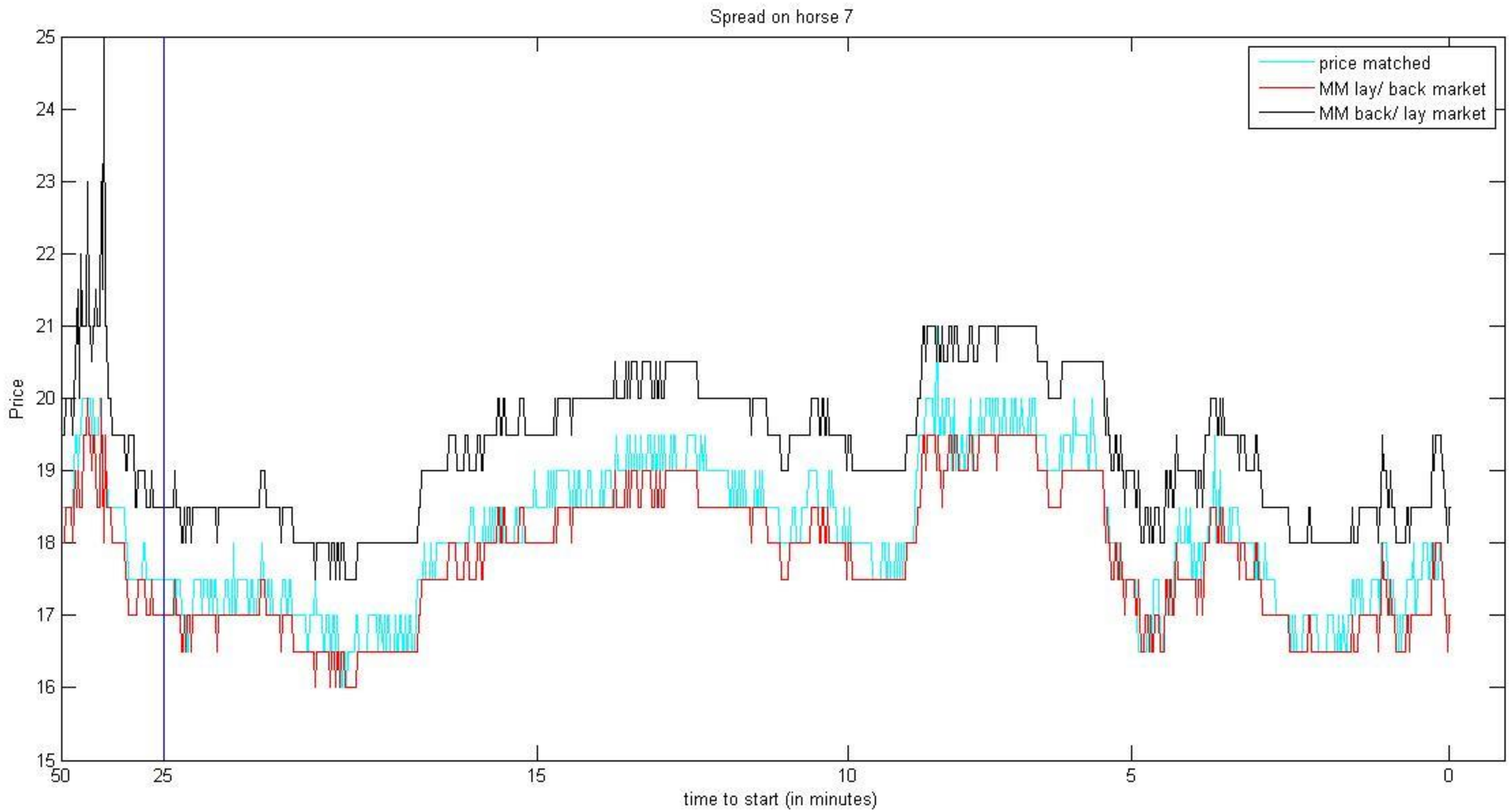


Table 2 Sum-up table on race 1 (Winner horse 1)

Horse	Winning probability	True Probability	Cumulative probability	Liability Pre MM	Liability Post MM	MM Effect	Turnover Sportbook	Turnover Exchange	Amount betted	profit's EV pre MM	profit's EV post MM
1	50,00%	43,84%	43,84%	-$ 27.787,72	-$13.387,64	$14.400,08	$ 84.822,53	$ 4.315,27	-$ 4.010,96	-$12.182,25	-$ 5.869,20
2	16,67%	14,62%	58,46%	$ 24.466,31	$20.451,83	-$ 4.014,48	$ 18.133,83	$ 1.950,60	$ 1.028,22	$ 3.576,09	$ 2.989,32
3	13,33%	11,69%	70,14%	$ 5.271,03	$11.941,74	$ 6.670,71	$ 17.303,31	$ 493,00	-$ 493,00	$ 616,07	$ 1.395,73
4	12,50%	10,96%	81,10%	$ 82.586,32	$60.026,86	-$22.559,46	$ 6.525,96	$ 2.730,30	$ 2.730,30	$ 9.051,55	$ 6.579,01
5	9,09%	7,97%	89,07%	$ 34.263,60	$28.777,32	-$ 5.486,28	$ 8.864,66	$ 870,00	$ 609,38	$ 2.730,87	$ 2.293,61
6	4,76%	4,17%	93,25%	$ 94.511,97	$68.203,96	-$26.308,01	$ 1.813,58	$ 1.093,50	$ 1.075,50	$ 3.944,56	$ 2.846,57
7	4,76%	4,17%	97,42%	$104.117,08	$75.273,28	-$28.843,80	$ 1.087,98	$ 967,72	$ 967,72	$ 4.345,44	$ 3.141,61
8	2,94%	2,58%	100,00%	$124.827,31	$92.191,78	-$32.635,53	$ 541,68	$ 623,51	$ 615,94	$ 3.217,82	$ 2.376,54
total	114,05%	100,00%					£139.093,53	£13.043,90	£ 2.523,09	£15.300,15	£15.753,18
				Profit's EV	Relative STD	Min EV	Max EV	Actual Profit/loss			
			Pre MM	$ 15.300,15	47%	-$12.182,25	$ 9.051,55	-$27.787,72			
			Post MM	$ 15.909,71	26%	-$ 5.869,20	$ 6.579,01	-$13.387,64			

As said in the introduction, the aim of the MM model is to reduce the liability on the horses with a negative one and, therefore, to reduce the level of active risk faced by the bookmaker.

As we can see, the results are positive. The model succeeds in hedging the negative liability on the horse 1 (there is an improvement of 14,402$ which means a relative improvement equal to 51.80%). The level of active risk, measured with the relative STD, is remarkably lower after the MM model, nearly half than the case without the model (from 47% to 26%), while the profit's EV is just a bit higher. Of course, these positive effects comes with a price. The positive liability on the other horses is nearly always lower with the MM maker model. It is a logic and common result considering the general aim of hedging strategies: reducing or nullifying possible losses in case of negative outcome, renouncing a higher profit in case of positive outcome. This idea is well expressed in the min-max profit's EV[24]. We can see that with the model the minimum EV increases from -12.182,25$ to -5869,20$, whereas the maximum value decreases from 9.051,55$ to 6.579,01$. In this case, the favourite horse have won the race and, as result, the bookmaker, with the implementation of the MM model, would have lose only -13.387,64$ instead of -27.787,72$.

[24] The profit's expected value was computed using the true probability of winning the race.

The overall results of the MM model built in this thesis are very positive. In effect, the model can reduce dramatically (e.g. half and even more) the race's level of risk without reducing the profit's expected value.
Furthermore, it has proven that when the winning probability is well balanced among many horses, it is better not to operate on the exchange. This is, mainly, due to the high unpredictability of the race's final result. In theory, in this case, the optimal solution for a bookmaker would be not to accept any bets. However, it will probably result in loosing markets shares, which is not acceptable for any business. The logical strategy to implement will be to reduce as much as possible the number of bets. Thus, it is better not to operate on the exchange.
The model continues to work, when there are many horses but the winning probability is not well balanced among them

References

Amihud Y., Mendelson H "DEALERSHIP MARKET. Market-making with Invetory", Journal of Financial Economics 8 (1980), 31-53

Avellaneda, M., Reed, J. & Stoikov, S., 2011. Forecasting Prices from Level-I Quotes in the Presence of Hidden Liquidity. Algorithmic Finance, 29 June, 1(1)

Betfair Annual Report, 2014

Brown A., Yang F. "Have Betting Exchange Corrupted Horse Racing?", (2014)

Copeland T., Galai D. "Information Effects on the Bid-ask Spread", The Journal of Finance, Vol.XXXVIIII, no.5 Decembr 1983

Croxson, K., & Reade, J. "Exchange vs. dealers: A high-frequency analysis of in-play betting prices". *Working Paper, University of Oxford*. (2011)

Flepp R., Nüesch S., Franck E. "The liquidity Advantage of Quote-driven Markets: Evidence from the Betting Industry". University of Zurich, (2013)

Franck E., Verbeek E., Nüesch S "Inter-market Arbitrage in Sports Betting", NCER Working Paper Series, Sportometrics Pogram, (2009)

Gambling Commission- Industry Statistics June 2015

Garman M. "MARKET MICROSTRUCTURE" Journal of Financial Economics 3 (1976) 257-275

Glosten L., Milgrom P. "BID, ASK AND TRANSACTION PRICES INA SPECIALIST MARKET WITH HETEROGENEOUSLY INFORMED TRADERS", Journal of Financial Economics 14(1985),71-100,

Ho T., Stoll H. "OPTIMAL DEALER PRICING UNDER TRANSACTIONS AND RETURN UNCERTAINTY" Journal of Financial Economics 9 (1981) 47-73

Luckner S., Schröder J., Slamka C., Franke M. Geyer-Schulz A.,Skiera Martin Spann B., Weinhardt C. " Prediction Markets. Fundamentals,Designs, and Applications", Gabler, 2012

Luckner S., Schröder J., Slamka C., Franke M. Geyer-Schulz A.,Skiera Martin Spann B., Weinhardt C. " Prediction Markets. Fundamentals,Designs, and Applications", Gabler, 2012

Michael A. Smith & David Paton & Leighton Vaughan Williams, 2006. "Market Efficiency in Person-to-Person Betting," Economica, London School of Economics and Political Science, vol. 73(292), pages 673-689, November.

Ozgit A. “The Bookie Puzzle: Auction versus Dealer Markets in Online Sports Betting”, 2005

Smith M., Paton D., Williams L.V., “Market Efficiency in Person-to-Person Betting”, Economica (2006) 73, 673-689

Smith, M., Paton, D., & Vaughan Williams, L. (2009). Do bookmakers possess superior skills to bettors in predicting outcomes? *Journal of Economic Behaviour & Organization, 71* , 539-549.

Stoll H, “THE SUPPLY OF DEALER SERVICES IN SECUIRITIES MARKETS” The Journal of Finance VOL. XXXIII, NO 4, September 1978

Sites:

http://www.betdaq.com/GBE.Help/BrokerDirectory/BETDAQ/en/Flash_Help/help_centre.htm#about_us.htm

LUISS Guido Carli
LIBERA UNIVERSITA' INTERNAZIONALE DEGLI STUDI SOCIALI

Department of Economics and Finance, Chair of Financial and Credit Derivatives

BETTING EXCHANGES: A MARKET MAKER PROCESS

Supervisor:

Professor Federico Nucera

Candidate

Dario Occhipinti 656531

Co-supervisor:

Professor Raffaele Oriani

Acknowledges

I would like to thank the Ladbrokes plc, represented by Doctor Francesco Borgosano, and the Centro Arcelli per gli Studi Monetarie e Finanziari (CASMEF), represented by Professor Marco Spallone, for the agreement that allowed me to use real data for my analysis.

I would like, also, to express my deep gratitude to Professor Federico Nucera and Professor Raffaele Oriani, my thesis supervisors, for their patient guidance, enthusiastic encouragement and useful critiques of this research work

Finally, I wish to thank my family for their never-ending support and encouragement throughout my studies.

Table of Contents

List of Figures

List of Tables

Introduction

Since ancient times, people have wagered money on every kind of events or games whether as leisure activities or as a way to enrich themselves.
During centuries, new forms of business have developed in order to satisfy the never-ending demand of gambling and betting.
In 1790, Harry Ogden opened a business that has been recognised as the first bookmaker in the United Kindom.
Bookmakers provide public odds at which they will accept any amount bet. Considering these characteristics, the bookmaker market is a quote-driven market in which bettors can only decide to bet at the price decided by the bookmaker, who provides all the liquidity in the market. Until 2000, it was the typical way to wager on an event.
Since 2000, however, a new player has arisen in the industry: betting exchange.
A betting exchange is an order-driven market, where bettors trade directly with each other in a continuous double auction. The liquidity is provided only by the traders.

These two types of business differ for many elements. The most important one is the different level of risk faced.
The owner of the exchange does not take any trading position, thus, he is not bearing any inventory risk. His revenues derive from the commission charge on the net profit, thus it is a certain cash flow. Whereas, the bookmaker takes the opposite side of every transaction, therefore, he needs to manage an active risk. His profit is not certain. He can lose a considerable amount of money in case of over-exposition on the final result.

In this thesis we will analyse the implementation of a Market Maker (MM) model, which trades on the betting exchange market Betdaq and allows a bookmaker to reduce the negative liability and the level of active risks. The creation of a MM model is needed considering the relative dimension of betting exchange and bookmakers.

Considering the vastness of the events, this thesis will focus only on the horse races sector. This sector, even though it is decreasing, continues to be the most important category of betting in the United Kingdom. From October 2013 to September 2014, bettors have wagered around 5.42£ billions pounds on horse races, 1.36£ billions on football matches and around 1.31£ billions on dogs races[25].

I have used real data provided by Ladbrokes plc. The databases contain:

- 1-year sample of orders on the betting exchange Betdaq;
- 1-year sample of all the bets wagered by Ladbrokes's customers on horse races.

The thesis is organized as follows. In chapter 1, all the notions needed to fully understand the betting industry will be explained. Chapter 2 will be dedicated to the analysis made on the

[25] U.K. Gambling Commission- Industry Statistics, June 2015

betting exchange. In chapter 3 we will see the assumption and the estimation of each parameter of the MM model. Chapter 4 will present the results of the application and testing of the model.

The overall results of the MM model built in this thesis are very positive. In effect, the model can reduce dramatically (e.g. one third and even more) the race's level of risk without reducing the profit's expected value.
Furthermore, it has proven that when the winning probability is well balanced among many horses, it is better not to operate on the exchange. This is, mainly, due to the high unpredictability of the race's final result. In theory, in this case, the optimal solution for a bookmaker would be not to accept any bets. However, it will probably result in loosing markets shares, which is not acceptable for any business. The logical strategy to implement will be to reduce as much as possible the number of bets. Thus, it is better not to operate on the exchange.
The model continues to work, when there are many horses but the winning probability is not well balanced among them.

1 Betting Industry

1.1 Introduction

This chapter is dedicated to the explanation of the fundamentals needed to understand the analysis done and the Market Maker model for Betting Exchanges.
Therefore, it will be explained what a bet is and how it can be displayed.
Moreover, there will be an explanation of the two types of business that operate in the Betting Industry and how their coexistence is possible even though, theoretically, one is superior to the other.

1.2 Fixed-odds gambling

A bet is a contract on some future cash flow based on the outcome of a given event (e.g. horse race, football match and so on). Therefore, the cash flow is determined by the outcome of the underling itself and by the price of the contract, i.e. the odd.

The fixed-odds betting is a form of wagering offered by a bookmaker or on a betting exchange. The main characteristic of this type of betting is that the bettor knows the odds at the time when he places a wager.

There are three main types of odds display[26]: Fractional, also known as Traditional or British, Decimal, known as European, and Moneyline, known as American.
The Fractional odds are manly used in the United Kingdom and Ireland. They are displayed as "3/1" or "7/4" and the numerator express the amount of money you will earn if you bet the amount express by the denominator. Thus, in the first case, for every euro you wager, the profit will be, in case of positive result, 3£, whereas for the second example it will be 7£ for every 4£ betted.
Furthermore, if we consider the odd as "chances against"/"chances in favour" we can obtain the implied probability of winning the bet[27].
In the first case, "3/1", the total chances are 4 (i.e. 3+1), the implied probability of victory is $\frac{1}{4} = 25\%$, whereas the probability of losing the bet is $\frac{3}{4} = 75\%$. In the latter example,"7/4", we have 36.36% probability of winning the bet and 63.64% probability of losing.

The second way to display odds is the Decimal or European odds. They are displayed as "1.70" or "2.50" and they tell the total return the customer will receive if she places a bet at that price. Thus, with a 10£ wager, the cash flow will be 17£ and 25£ respectively.

---

[26] There are also the Hong Kong, Indonesia and Malay odds; however, they play a marginal role in the gambling industry.
[27] When the odds are offered by a bookmaker, the implied probability it is not the true probability of victory. As we will see later, the bookmaker add the so-called overround. As result, the implied probability is higher than the true probability. Whereas, when the odds are offered on the betting exchange we obtain directly the true probability of victory (there is not overround on the exchange).

In this case, it is possible to compute the implied probability of victory dividing 1 by the odd. Therefore, an odd equal to "1.70" has a winning probability of 58.82% (i.e. 1/1.70), whereas in the second case the probability is 40% (i.e. 1/2.50).

The last way is the Moneyline or American one. In this case, the quote can be expressed either positive or negative figure[28] (+150 or -120).
If the figure is positive, e.g. +150, it shows the amount of money you will win on a 100$ bet. Whereas, if we have a negative odd, e.g. -120, it expresses the amount of money you need to wager in order to win 100$.
In order to obtain the implied probability, we divide, in absolute value, the money we wager by the money we will receive[29]. Thus, the probability are 60% and 54.54% respectably[30].

In this thesis, we will use only the Decimal odds. In table 1, it is possible to see how to convert odds from one way to another.

Table 3 Odds conversion formulas

From	To	Computation (steps)
Fractional	Decimal	1) Solve fraction 2) Add 1
Fractional	Moneyline	1) Solve fraction 2) If the solution is >=1 → solution*100 If solution is <1 → -100/ solution
Decimal	Fractional	1) Deduct 1 2) Convert to fraction
Decimal	Moneyline	1) If Decimal odd >2 → 100*(Decimal odd-1) If Decimal odd <2 → -100*(Decimal odd-1)
Moneyline	Decimal	1) If Moneyline odd >0 → (Moneyline Odd/100)+1 If Moneyline odd <0 → (-100/Moneyline Odd)+1
Moneyline	Fractional	1) If Moneyline odd >0 → Moneyline Odd/100 If Moneyline odd <0 → -100/Moneyline Odd

1.3 Bookmaker markets

The betting industry, which is one sector of the gambling world, was traditionally operated by bookmaker. The first, widely acknowledged, bookmaker in the United Kingdom is Harry Ogden[31], who sets-up a business in the 1790s.

[28] If the implied probability is higher than 50% the figure is positive, otherwise is negative.
[29] E.g. if the quote is +150 it means that when you bet 100$ you will receive 250$ which means a profit of 150$
[30] I.e. 150/250 = 60% and 120/220=54.54%.
[31] Munting R. "An Economic and Social History of Gambling in Britain and the USA".

Bookmakers provide public odds[32] at which they will accept any amount bet. Considering these characteristics, the bookmaker market is a quote-driven market in which bettors can only decide to bet at the price decided by the bookmaker. The transparency on this kind of markets is very low, as only the last odds are public. There are no information about the price's formation rule or about the historical price changes.
The odds quoted by the bookmarkers contain already the so-called overround, which is their compensation for the services that they are providing and for bearing the risk of unfavourable outcomes.
If we want to retrieve the overround from the odds, we just need to compute the implied probability of victory for each selections and sum them up. The percentage over 100% is the overround applied at that race (see Table 2 for an example on the horse races).

Table 4 Overround

	Odds	Implied probability	True probability	Overround
Horse 1	2	50.00%	44.97%	5.03%
Horse 2	4,5	22.22%	19.99%	2.23%
Horse 3	6,5	15.38%	13.84%	1.55%
Horse 4	10	10.00%	8.99%	1.01%
Horse 5	13	7.69%	6.92%	0.77%
Horse 6	17	5.88%	5.29%	0.59%
Total		111.18%	100%	11.18%

In this case, the overall overround is 11.18%. It is possible to see that it is not distributed homogenously on each horse but it is higher on the horses with a higher probability of victory. This is a logical result. Bettors will wager more on the horses with a higher probability of winning the race, thus, the bookmakers need to increase their protection, provided by the overround, on those horses. It is possible to obtain the true probability, which is the probability without the overround, with a simply proportion.
Example of Horse 1 (table 2):

$$111.18\%: 50\% = 100\%: x \ \rightarrow x = 44.97\%$$

The overround has also a second interpretation: it is the margin that the bookmaker will earn[33], regardless the final result of the event, if the customers divide their bets following the true winning probabilities. When it happens, the bookmaker's book is called "Dutch book" [34].
Continuing the example of table 2, if the customers place 44.97% of bets on the horse 1, 19.99% on the horse 2, 13.84% on the horse 3 and so on, the bookmaker will earn the overround times

[32] The set of odds provided for one event is called Sportbook or simply book.
[33] The profit of the bookmaker is equal to the amount customers wagered minus the amount wagered on the winner horse times the odds provided.
[34] The "Dutch book" is a set of odds and bets which guarantees a certain profit, regardless of the outcome of the gamble.

the total amount wagered in any scenario. Thus, if the total amount wagered by bettors is 70,000£, the bookmaker will have a profit equal to 7,037.32 £[35].
In table 3, we can see all the data of the example[36].

Table 5 Dutch Book

	Odds	Implied probability	True probability	Overround	Bookmaker's Profit
Horse 1	2	50.00%	44.97%	5.03%	£ 7,037.32
Horse 2	4.5	22.22%	19.99%	2.23%	£ 7,037.32
Horse 3	6.5	15.38%	13.84%	1.55%	£ 7,037.32
Horse 4	10	10.00%	8.99%	1.01%	£ 7,037.32
Horse 5	13	7.69%	6.92%	0.77%	£ 7,037.32
Horse 6	17	5.88%	5.29%	0.59%	£ 7,037.32
Total		111.18%	100%	11.18%	

1.4 Betting Exchange

Since 2000, new players have arisen: the betting exchanges[37].
"They operate as order-driven markets, where buyers and sellers trade directly with each other in a continuous double auction without the intermediation of market makers"(Flepp, Nüesch, Franck, 2013[38]).
Betting exchange, thus, allow bettors not only to bet on a positive outcome (e.g. the victory of a team or of a horse), but also to act as a bookmaker and to sell the bet, thus it allows to bet on a negative result (e.g. the loss of a team, horse and so on). As in financial markets, bettors can place a limit order and wait that someone hits their order or they can submit a market order and match an already existing bet.
The role of the exchange itself is to provide a platform in which people can trade. The provider of the platform charges a commission fee, usually between 2.5% and 5%, on the net profit.
There is a high degree of transparency as it is possible to know the whole limit order book as well as the matched price history. The market leader is Betfair with around 53£ billion[39] of amount matched[40] in 2014. The second bigger exchange is Betdaq, with a weekly turnover equal to 75£ million[41].

---

35 In theory, the 11.18% of 70,000£ is 7,826£ and not 7,037.32. In this example, to simplify, it is assumed a uniform distribution for the overround, whereas in reality is a little different. In any case, this assumption does not change the general result.
36 Note: in the table the value are shown with only two digits after decimal point, however, to obtain exactly 7,037.32£ for every horse we need to use all the digits.
37 Betting Exchanges are prediction markets. To have more information on the topic read: Luckner S., Schröder J., Slamka C., Franke M. Geyer-Schulz A.,Skiera Martin Spann B., Weinhardt C. " Prediction Markets. Fundamentals, Designs, and Applications", Gabler, 2012
38 Flepp R., Nüesch S., Franck E. "The liquidity Advantage of Quote-driven Markets: Evidence from the Betting Industry". University of Zurich, 2013
39 Betfair Annual Report, 2014.
40 The amount matched is the total amount wagered by costumer (i.e. the total amount backed which has been hit in laid)
41 http://www.betdaq.com/GBE.Help/BrokerDirectory/BETDAQ/en/Flash_Help/help_centre.htm#about_us.htm

In figure 1 we can see the interface of Betdaq.

Figure 3 Betdaq exchange


15:00 Pontefract - Win Market - 6f [-]

Spindrifter EBF Conditions Stakes (Plus10) 18000 GBP Added, 0m 6f 0y

29-06-2015 15:00 Tip Check Price Check Trading History Matched: £7,237

Selections=5		101.50%	Back	Lay	99.19%	
3 Gold Medallion	2.36 £5	2.4 £46	2.42 £24	2.46 £2	2.5 £150	2.54 £11
1 Age Of Empire	2.94 £14	2.96 £38	3 £8	3.05 £70	3.1 £8	3.15 £11
2 Dodgy Bob	7.8 £5	8 £4	8.2 £3	8.4 £8	8.6 £5	8.8 £5
4 Young John	9.8 £38	10 £3	10.5 £14	11 £7	11.5 £3	12 £3
5 Mon Beau Visage	18.5 £1	19 £3	19.5 £3	21 £3	22 £2	23 £2

To clarify the market dynamics, consider the following simplified example in horseracing[42]:

Table 6 Betting exchange example n 1

	Back			Lay		
Horse 1	2.02 200£	2.10 40£	2.12 50£	2.14 20£	2.16 40£	2.24 350£
Horse 2	1.80 180£	1.88 60£	1.9 40£	1.92 60£	1.94 40£	2 240£

Backing a horse is betting on the victory of that horse. Whereas, *laying* is betting on the outcome that the horse will lose the race. If we consider as "standard outcome" the victory of the race, we can make a sort of parallelism between betting exchanges and financial exchanges. In this case, the *back price* is the price at which it is possible to buy a bet on the winning result (e.g. the horse 1 win the race) which is the *ask price* in financial markets. While the *lay price* is the price at which it is possible to sell the bet which is the *bid price* in financial markets[43].The numbers under the odds correspond to the available volume offered by the counterparty, who had placed a limit order. The grey area is the top of the market (i.e. the highest price at which it is possible to buy the bet and the lowest price at which it is possible to sell the bet).

Therefore, in this example, it is possible to bet up to 50£ at 2.12 on the victory of the horse 1. If that horse will win the race, the cash flow will be equal to amount betted times odd (e.g. (50*2.12)=106£).

[42] A similar example on basketball matches can be found in Ozgit A. "The Bookie Puzzle: Auction versus Dealer Markets in Online Sports Betting", 2005

[43] It would seem that it is possible to sell a bet an higher price than the one to buy it, however, we should remember that the lay price is what we have to pay in case that horse win the race. To obtain a profit you need to sell the bet at one price lower than the one bought (i.e. if you lay 5£ at 2.00 and you back 5£ at 2.02 you will obtain (5*2.02)-(5*2.00)=0.10£ as profit in case of victory of that horse, otherwise the final profit will be 0).

Otherwise, if the bettor thinks that the horse 1 will lose the race, he can lay at 2.14. In this case, the bettor will earn the amount betted if the horse will lose, otherwise he will pay the amount betted times odd to the backer (e.g. (20*2.14)=42.8£ , thus he lose (42.8-20)=22.8£).
In our example, with just two horses, the bettor can also decide to back on horse 2 to obtain a less risky payoff on the same outcome of laying on horse 1 (i.e. if he bets 20£ his profit will be (20*1.9-20)=18£ in case the horse 2 win the race, otherwise he will lose the 20£ betted). However, in the real world you need to back on all the other horses (up to 18 in some race) to achieve the same result of laying on horse 1, which is more affordable.

If a bettor places an order with an amount bigger than the one available on the market, the unmatched part will be shown on the opposite side of the market of the same horse. For example, if someone wants to back 80£ at 2.12 on the horse 1, the limit order book will be:

Table 7 Betting exchange example n2

	Back			Lay		
Horse 1	2.00 450£	2.02 200£	2.10 40£	2.12 30£	2.14 20£	2.16 40£
Horse 2	1.80 180£	1.88 60£	1.9 40£	1.92 60£	1.94 40£	2 240£

As we can see, the unmatched volume (i.e. 30£) is shown in the Lay column.
The same happens in case of lay orders. For example, if a bettor wants to lay 90£ ad 1.92 on the horse 2, the limit order book will be:

Table 8 Betting exchange example n3

	Back			Lay		
Horse 1	2.00 450£	2.02 200£	2.10 40£	2.12 30£	2.14 20£	2.16 40£
Horse 2	1.80 180£	1.88 60£	1.92 30£	1.94 40£	2 240£	2.02 400£

A key difference between the betting exchange and the financial markets exists: the minimum tick value changes in relation to the price. In table 7, we can see the prices' range and the value of 1 tick.

Table 9 Tick value

Range in £	Tick value in £
Under 2.00	0.01
2.00-3.00	0.02
3.00-4.00	0.05
4.00-6.00	0.1
6.00-10.00	0.2
10.00-20.00	0.5
20.00-50.00	1.00
50.00-200.00	2.00
200-1000	5.00

1.5 Bookmakers vs Betting Exchanges

Bookmakers and betting exchanges differ for many elements.
The main one is the overround. There is none on the exchange and, as a result, the odds on the exchange are higher than the ones made by the bookmaker. Thus, for a bettor, betting on the exchange is the best solution because it yields a higher return than the one offered by the bookmaker's odds. This difference is mainly due to the different level of risk faced by these two types of business. The owner of the exchange does not take any trading position, thus, he is not bearing any inventory risk. His revenues derive from the commission charge on the net profit[44], thus it is a certain cash flow. Whereas, the bookmaker takes the opposite side of every transaction, thus, he needs to manage an active risk. His profit is not certain. He can lose a considerable amount of money in case of over-exposition on the final result. There is also a different degree of transparency. In the betting exchanges, it is possible to know the whole limit order book as well as the matched price history. Whereas, with the bookmakers only the last public odds are shown. The knowledge of the entire limit order book can be useful, for example, to predict prices or, in general, to gather as much information as possible on the event. Moreover, several studies[45] demonstrate that the betting exchanges' odds have a more predictive power than the bookmakers' ones.

Betting Exchanges are considered as a superior business model to the traditional bookmakers, for all the above cited reasons. Nevertheless, the latter continues to be more successful than the first.

[44] E.g. Cash in-flow minus the amount betted.
[45] Smith, M., Paton, D., & Vaughan Williams, L. (2006). Market efficiency in person-to-person betting. Economica, 73, 673-689.
Smith, M., Paton, D., & Vaughan Williams, L. (2009). Do bookmakers possess superior skills to bettors in predicting outcomes? Journal of Economic Behavior &Organization, 71 , 539-549.

In literature, this has been explained for three main reasons:

1. The first is the learning costs to switch to a betting exchange structure[46]. Bettors are used to wagering at a traditional bookmaker, which is simpler than trading on an exchange. In the first case, the only decision to make is to bet or not at one price decided by the bookmaker. Whereas, in the exchange, the customers face different odds and options (limit or market orders, back or lay orders) and it takes time and effort to get used to new opportunities;
2. Secondly, Ozgit (2005)[47] provides a trade-size reason. It can easily happen that the best odds volume is not enough to deplete an order. In this case, the bettor must decide if executing the remain part at the second best and maybe the third best or placing a limit order and wait to be matched. In the first case, the extra-return provided by the exchange's odds will disappear, while in the latter, the order can remain unmatched. This problem does not exist with the bookmakers because they instantaneously accept any amount of money;
3. The last reason is the different level of liquidity in these two types of market[48]. As said before, the betting exchanges are order-driven market and all the liquidity depends by the bettor's orders, while in the bookmakers' markets the liquidity is guaranteed by the bookmakers. Therefore, when the liquidity on the exchange is low, the extra-returns will disappear and the bookmakers odds become more "profitable". This happens frequently for low-grade events (e.g. local event) which do not attract many bettors.

[46] Croxson, K., & Reade, J. "Exchange vs. dealers: A high-frequency analysis of in-play betting prices". Working Paper, University of Oxford. (2011).

[47] Ozgit A. "The Bookie Puzzle: Auction versus Dealer Markets in Online Sports Betting", 2005

[48] Flepp R., Nüesch S., Franck E. "The liquidity Advantage of Quote-driven Markets: Evidence from the Betting Industry". University of Zurich, (2013).

2 Market Analysis and descriptive statistics

This chapter contains liquidity data and descriptive statistics of the betting exchange market Betdaq.
The dataset, provided by Ladbrokes plc, contains the data of 12571 horse races between 01/01/2013 and 31/12/2013.
The amount placed and matched has been converted in dollars using the exchange rate provided within the database.

The analyses made are:

2 Horse race classification;
3 Liquidity study;
4 The average bet’s size;
5 Volatility study.

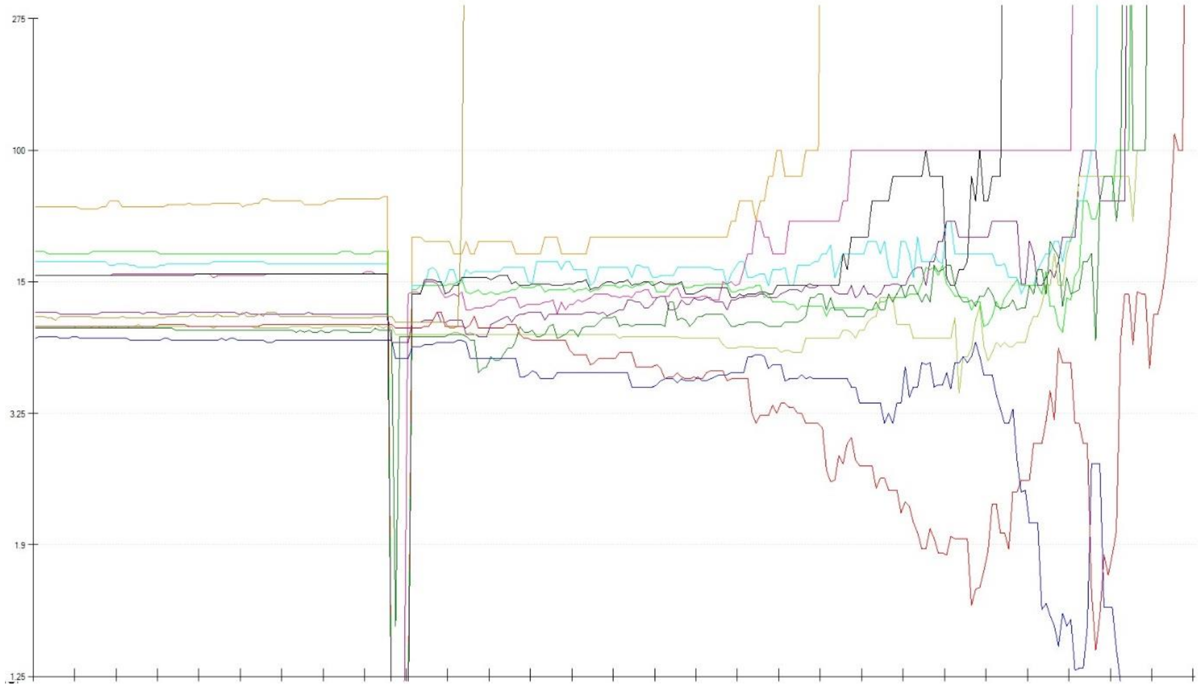
It is possible to divide the “lifecycle” of the horse race in two main parts: *pre-live* and *in-play*. The *pre-live* is the time period which goes from the first instance that is possible to bet until the start of the race. Whereas the *in-play* period goes from the event-off until the end of the race.

The above-mentioned market phases have different characteristics. The first, and the most obvious, is the timespan. The pre-live period can begin even some days before the start of the race, while the in-play period is usually around 3 minutes (it depends on the type of race). The relative weight, measured as the amount of money wagered on each part, is completely different too. Later in this chapter, we will see that usually more than 90% of the matched amount is wagered on the pre-live period. The last but not least important difference is that during the pre-live period the price volatility is low, whereas it is high in the in-play period. This can be explained by the high uncertainty related to possible unpredictable events during the race (e.g. if a horse falls down during the race his price will immediately hike to 1000£, which is the maximum, as his probability of victory will be 0%).

We can see in figure 2 the price dynamics of the in-play period with 11 horses.

Taking into account the above-cited characteristics of the betting exchange market, the analysis will be focused on the pre-lived period only.

Figure 4 Price dynamics of the in-play period (price on y-axes, time on x-axes, 11 horses)



2.1 Horse race classification

It is important to remark that horse races have various importance and consideration across the gambling community. Some races, as the Royal Ascot, have a predominant role, while others are just considered “local” events. The amount of money wagered and matched is directly proportioned with the importance of the race.

In order to classify the data, after computing the average total matched amount of all races equal to 131,099.38$, the races have been grouped based on their total amount matched[49]:

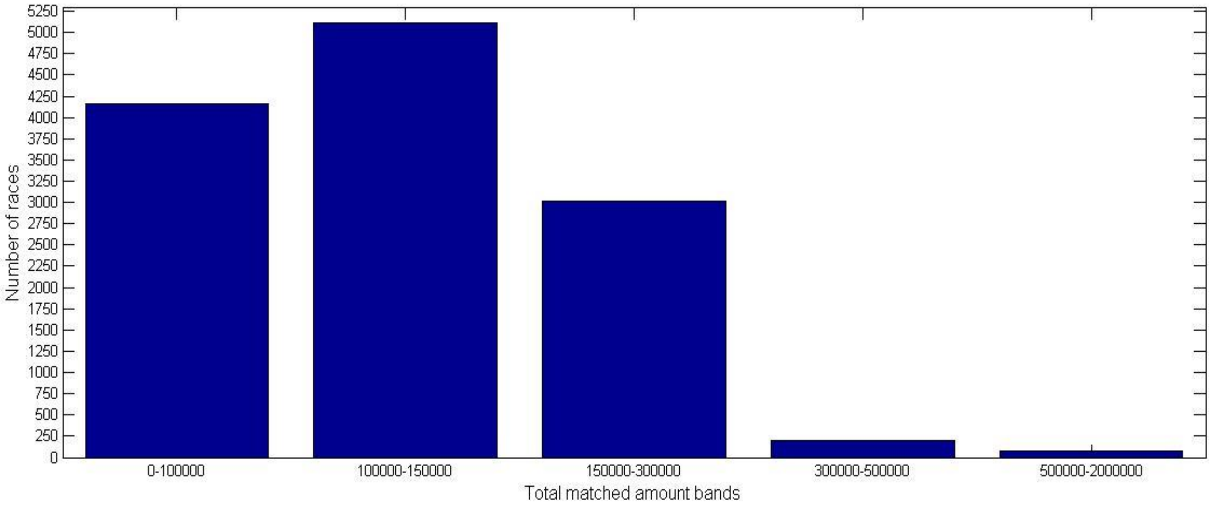
1 Class 5 Races with lower than average betting volume (4164[50])
2 Class 4 Race with an average betting volume (5118);
3 Class 3 Races with a moderate betting volume (3013);
4 Class 2 Races with higher than average betting volume (203);
5 Class 1 Races with very high betting volume (73).

In figure 3 it is possible to see the result of the classification, with the band’s size.

[49] We are following the idea of Michael A. Smith & David Paton & Leighton Vaughan Williams, 2006. "Market Efficiency in Person-to-Person Betting," Economica, London School of Economics and Political Science, vol. 73(292), pages 673-689, November.

[50] Number of races in that class.

Figure 5 Horse race classification



The analysis, which will be discussed in this chapter, has been done for class 1 and 2. We will see that races of different classes behave, nearly, in the same manner.

2.2 Liquidity study

As said in chapter 1, the betting exchange is an order-driven market, thus liquidity is the engine of the exchange. Moreover, the "stock" traded in this kind of market (i.e. the bet on a horse) has a maturity, which is the end of the races, and it strongly influences the way customers trade on the exchange.

In order to study the liquidity, I have computed the ratio between the total amount matched[51] until time t and the total amount matched in the whole race (see formula 1). Doing so for a standardized timespan (i.e. every one millisecond), we can study the evolution of the total matched amount during the race's life and we can make a comparison among different races.

$$Liquidity\ ratio = {AM_t}/{TL} \qquad (1)$$

Where:

AM_t is the total amount matched until time *t;*

TL is the total amount matched in the race.

In Figure 4-7, we can see the evolution of the total matched amount, mean value and percentile 5%-95% of the distributions, for class 1 (73 races) and class 2 (203 races). Table 8 and Table 9 show the value of the ratio at a fixed time.

[51] The amount matched is the total amount wagered by costumer (i.e. the total amount backed which has been accepted in laid).

As we can see, until 50 minutes before the start, the market is, basically, illiquid (e.g. the % amount is equal or less of 10% for both the classes), after that it increases in an exponential way reaching around 95% at the beginning of the race.
This characteristic will strongly influence the market maker model (e.g. how to set the frequency, how to calibrate the spread, and so on), as well as any other strategies, which will be implemented in the future on betting exchanges.

Table 10 Class 1 liquidity evolution

Time-range (in minute)									
	750	500	250	100	50	25	15	10	5
5th percentile	0.12%	0.34%	1.30%	2.35%	4.31%	7.88%	19.68%	27.21%	49.23%
Mean Value	1.26%	1.51%	4.26%	8.16%	10.85%	17.59%	30.90%	44.11%	65.47%
95th percentile	2.55%	2.84%	6.67%	14.21%	18.67%	27.62%	44.62%	56.35%	81.42%

Table 11 Class 2 liquidity evolution

Time-range (in minute)									
	750	500	250	100	50	25	15	10	5
5th percentile	0.00%	0.00%	0.14%	0.85%	1.89%	3.46%	8.37%	15.83%	37.04%
Mean Value	0.55%	0.71%	2.53%	5.50%	7.93%	12.82%	22.34%	33.62%	57.31%
95th percentile	1.25%	1.53%	5.68%	11.71%	14.60%	21.82%	31.68%	43.77%	70.14%

In Figure 8, instead, we can see the evolution of the price matched and the volume matched on each seconds (from the last 50 minutes before the start of the race) of the favourite horse. Both the frequency of orders and the volume are higher in the last 15 minutes than in the previous part.

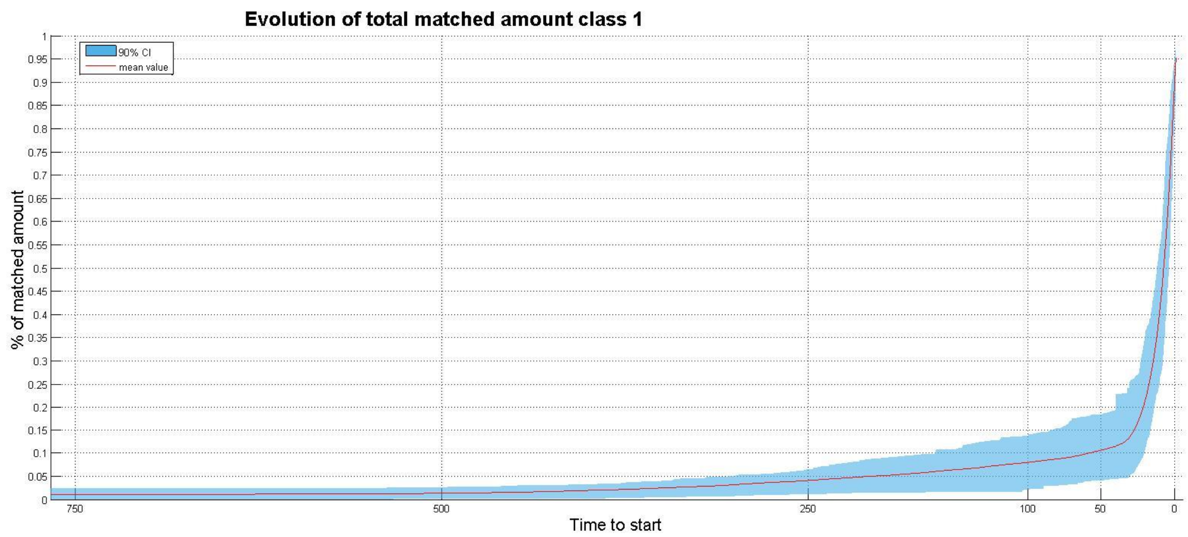


Figure 6

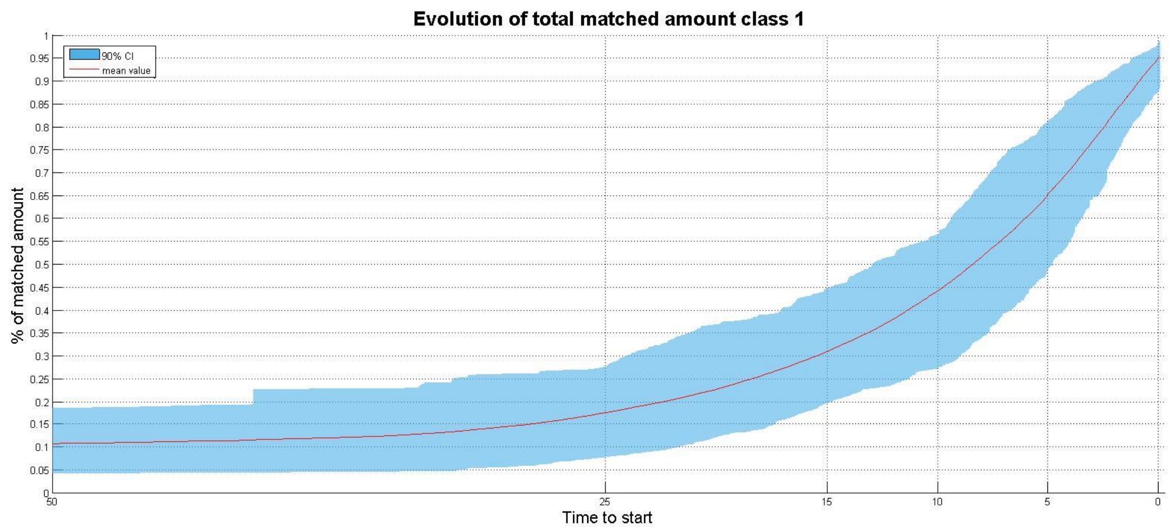


Figure 7

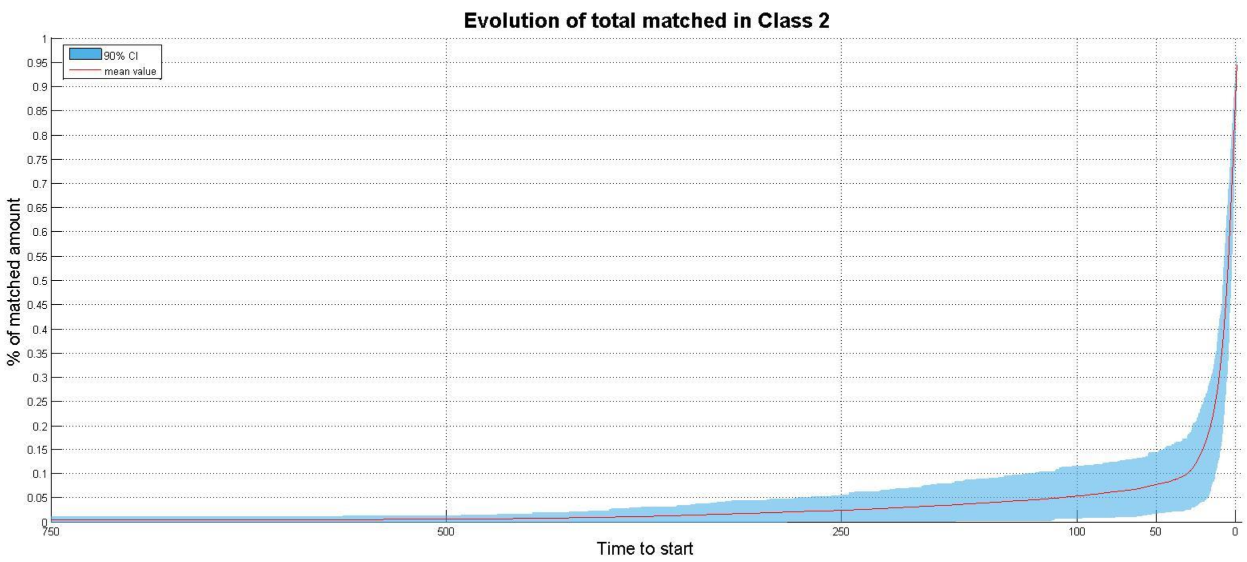


Figure 8

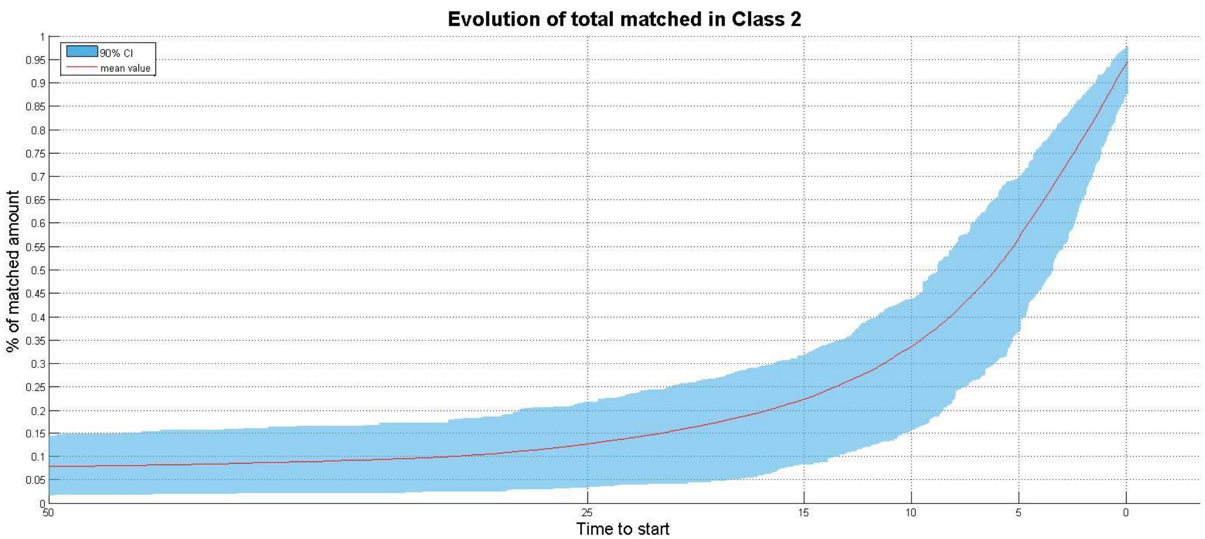
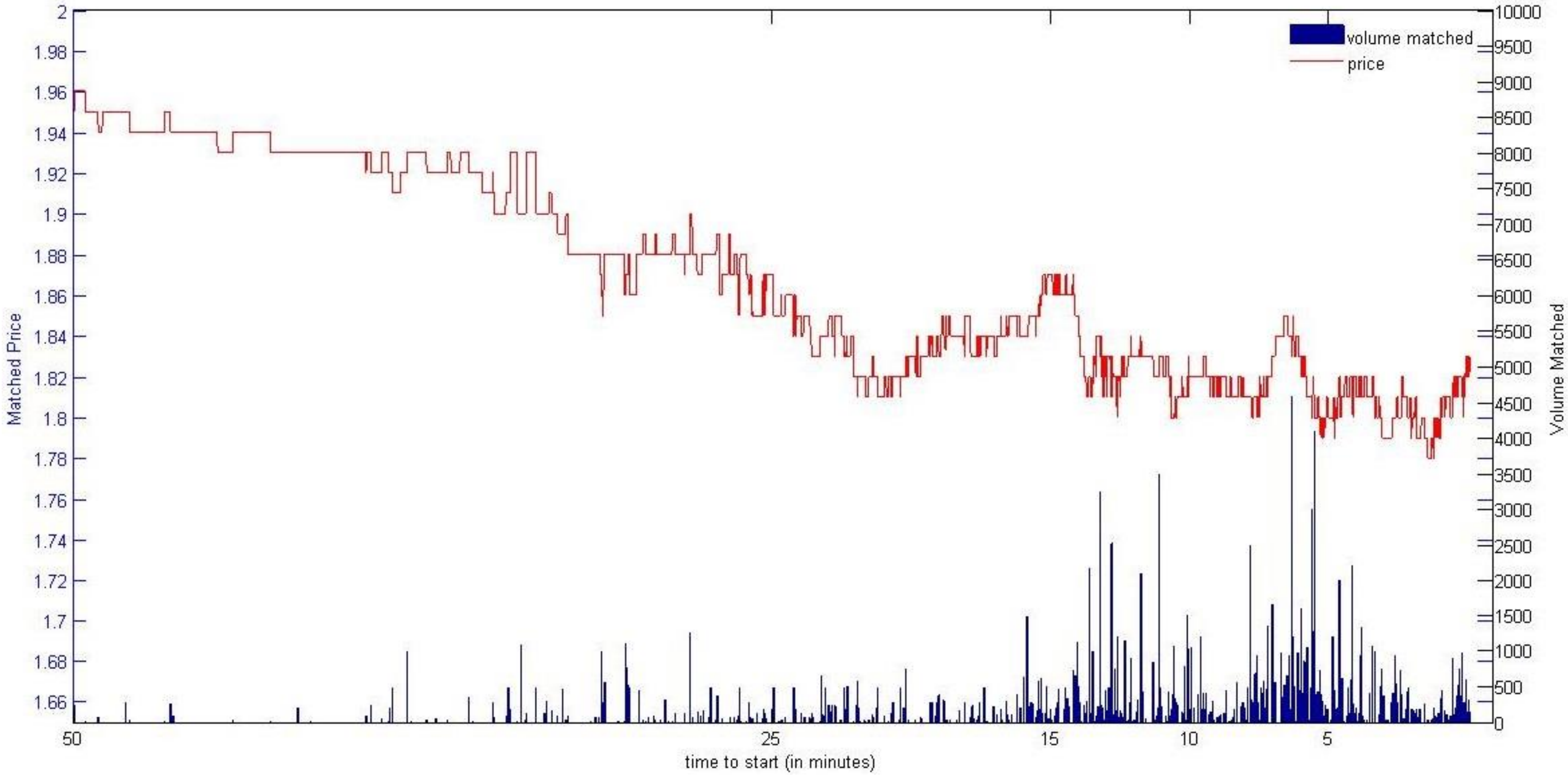


Figure 9

Figure 10 Matched price and volume evolution (left axes matched price, right axes volume matched)



2.3 The Average bet's size

The second analysis that is important to perform is the estimation of the average bet's size in monetary terms. Having an idea of this value will help to define a feasible bet's size that will be used for the MM model.

It is important to consider that the bet's size is different for each kind of horse (i.e. the size on the favourite horse is bigger than the one of the last favourite).
Therefore, in order to take into account this characteristic, I am assuming that it is possible to classify and aggregate each horse of every race by their average matched price of the pre-live period.
In order to clarify, we can consider the favourite horse case. The assumptions based on real observations are: it has the lowest average price in the pre-live period and each race has a favourite runner. Thus, it is possible to aggregate the favourite horses of each race that we want to analyse.
Once obtained each group of representative horses[52], I have computed the average bet's size as the total matched amount divided by the numbers of orders. Then, I have sorted the result (see table 10 for an example on horse 1) and, finally, I was able to compute the mean, min, max value and AVG size's S.D. for each percentile required for the study. I have analysed 73 races of class 1 and 126 of class 2.

Table 12 Horse 1, AVG bet sorted

Race	total amount matched	number of bets	AVG size
1	$ 146,094.65	4,836	30.27
2	$ 88,002.52	2,628	33.52
3	$ 113,820.48	3,278	34.81
4	$ 122,662.31	3,490	35.20
5	$ 155,038.89	3,878	40.00
6	$ 136,600.62	3,372	40.52
7	$ 182,808.70	4,310	42.57
…	…	…	…

In table 11 we can see the result of class 1, horse 1.
In order to clarify, "Bottom 5%" refers on the first 5% of the sorted races (i.e. Table 10), "Bottom 25%" refer to the first 25% of the sorted races and so on. Thus, for class 1 horse 1, the mean value, min value and max value of "Bottom 5%" refer to the first 4 races of our population (73 races, 5%*71=4 races).
Table 11 shows that the value goes from a minimum of 30.27$ to a maximum of 339.34$. The latter case happens when nearly the total amount wagered is concentrated on the favourite horse (e.g. around 75% of the total matched amount). The same behaviour is observed for class 2 races (see table 12).

[52] Just the first eight horses will be considered because to total amount betted on these horses is around 97% for both class 1 and 2.

Considering the above cited results, the mean value of the first half sample can be selected as the standard bet's size. This will allow the final model to place feasible orders on the exchange. In Table 13 and 14 we can see the estimated value for the other horses.

Table 13 Class 1 horse1 average bet size

Percentiles	Mean value	Min value	Max Value	S.D.
Bottom 5%	33.45	30.27	35.20	2.24
Bottom 25%	43.05	30.27	51.79	6.25
Half	51.00	30.27	70.88	9.97
Bottom 75%	61.49	30.27	99.14	18.08
Bottom 95%	77.62	30.27	229.54	40.15
Whole Population	89.74	30.27	339.34	63.75

Table 14 Class 2 horse1 average bet size

Percentiles	Mean value	Min value	Max value	S.D.
Bottom 5%	27.45	22.49	31.95	3.94
Bottom 25%	40.79	22.49	53.18	8.81
Half	50.77	22.49	67.61	12.30
Bottom 75%	61.44	22.49	96.26	18.58
Bottom 95%	74.53	22.49	174.49	32.20
Whole Population	82.41	22.49	341.44	48.50

Table 15 Class 1 horse from 2 to 8 average bet size

Horse	CLASS 1 Mean value	Min value (population)	Max value (population)	S.D. (population)
Horse 2	29.75	11.26	81.47	14.16
Horse 3	21.46	3.97	63.06	9.89
Horse 4	16.66	5.50	53.66	9.33
Horse 5	13.84	3.73	37.60	6.14
Horse 6	11.59	2.41	39.94	6.14
Horse 7	8.89	1.71	27.09	5.97
Horse 8	8.67	2.87	50.12	7.22

Table 16 Class 2 horse from 2 to 8 average bet size

Horse	CLASS 2 Mean value	Min value (population)	Max value (population)	S.D. (population)
Horse 2	37.20	11.26	81.47	14.16
Horse 3	15.97	5.85	46.05	8.83
Horse 4	13.70	1.68	41.69	7.68
Horse 5	8.23	1.70	33.73	6.33
Horse 6	6.18	0.83	27.23	6.33
Horse 7	5.39	1.19	32.50	5.60
Horse 8	4.02	1.42	27.39	5.18

2.4 Price standard deviation analysis

An important element for every financial model is the price volatility. For MM models, the price's standard deviation is essential for the computation of the spread.[53] Effectively, a higher volatility will require a thicker spread to protect the MM from the uncertainty linked with the price.
The procedure followed to compute the standard deviation is explained in the appendix 1. In this chapter we will focus our attention to the results.

In table 15 we can see the result of the estimations for Class 1.
As in the AVG bet's size study, in table 15 "bottom 5 %" represents the first 5% of sorted sample.

As we can see, the standard deviation is generally really low. However, the S.D. can be extremely high in some cases. This happens when the back-lay spread on the market is wide and the price matched bounces from the lower bound to the higher bound. The results for class 2 are similar (not reported in this thesis). The only difference is that the standard deviation for the least favourite horses is a bit bigger.
It is important to notice the S.D. increases as we reach the start of the race. This result is common for every type of representative horses. This is meanly due to the illiquidity of the market until 25 minutes before the start[54]. In effect, the price tends to remain stable because there are no trades on the exchange. As we reach the start of the race, the number of trades increases exponentially and the volatility increases too[55].

[53] Bid-ask spread in financial markets, back-lay spread in exchange markets.
[54] As we have seen in the liquidity study.
[55] We can see this characteristic clearly in figure 8.

Table 17 Class 1 price standard deviation in ticks, timespan: 1 second

Percentiles	Horse 1				Horse 2			
Time range (minutes)	50-0	25-0	10-0	5-0	50-0	25-0	10-0	5-0
Bottom 5%	0.24	0.34	0.45	0.45	0.21	0.28	0.36	0.38
Half	0.35	0.46	0.57	0.63	0.29	0.39	0.49	0.55
Bottom 95%	0.73	0.89	1.16	1.39	0.36	0.48	0.62	0.77
Max value	0.87	1.07	1.31	1.48	0.39	0.53	0.70	0.83
	Horse 3				Horse 4			
Time range (minutes)	50-0	25-0	10-0	5-0	50-0	25-0	10-0	5-0
Bottom 5%	0.20	0.27	0.33	0.40	0.18	0.25	0.32	0.36
Half	0.26	0.36	0.46	0.53	0.26	0.35	0.47	0.52
Bottom 95%	0.33	0.46	0.58	0.66	0.34	0.46	0.62	0.71
Max value	1.17	1.66	2.60	3.66	0.38	0.51	0.68	0.86
	Horse 5				Horse 6			
Time range (minutes)	50-0	25-0	10-0	5-0	50-0	25-0	10-0	5-0
Bottom 5%	0.19	0.24	0.32	0.29	0.16	0.22	0.29	0.33
Half	0.24	0.33	0.43	0.50	0.24	0.33	0.45	0.51
Bottom 95%	0.38	0.48	0.61	0.73	1.30	1.74	2.40	2.16
Max value	0.54	0.73	0.79	0.96	10.05	12.89	13.99	15.64
	Horse 7				Horse 8			
Time range (minutes)	50-0	25-0	10-0	5-0	50-0	25-0	10-0	5-0
Bottom 5%	0.16	0.20	0.29	0.33	0.16	0.22	0.27	0.30
Half	0.25	0.34	0.45	0.52	0.26	0.34	0.44	0.54
Bottom 95%	4.48	5.88	6.91	8.49	1.30	1.64	2.24	3.08
Max value	12.44	17.15	22.25	21.48	30.32	37.62	56.05	41.34

2.5 Conclusion

After the computation of these descriptive statistics, we have a clearer picture of Betdaq. Firstly, we saw how it is possible to classify the races by their matched amount. Secondly, we studied the liquidity evolution through time and understood that the market is illiquid until 25 minutes before the start of the race and that the liquidity increases exponentially in the last part. Moreover, we discovered that the price's S.D. is low. These two characteristics will influence how the Market Maker model will behave. Finally, we studied the average order's size with the aim of allowing our model to place bet with a feasible size coherent with the market conditions.

3 The Market-Maker model

In this chapter, the market maker model, developed with the special aim of hedging an already existing liability for the bookmaker, will be presented.

Before doing so, it is important to remark and specify some ideas.
When a bettor places a lay order he is acting as a bookmaker. He will receive the amount betted, which is the bookmaker's turnover, if that horse loses the race. Otherwise, he will pay the amount wagered times the odds to the counterparty. In parallel, when the bookmaker accepts a bet it is like he is betting on the opposite outcome. For instance, when a customer bets on the victory of one horse, the bookmaker is betting, implicitly, on the defeat of that horse. Thus, following the betting exchange vocabulary, he his laying on that horse. This parallelism is extremely important for the market maker strategy.
Thus, to clarify, the amount wagered by the bookmaker itself on a lay order on the exchange, or wagered by the bookmaker's customers on the sportbook will be considered as a cash inflow for the bookmaker. Whereas, a back order made by the bookmaker on the exchange will be considered as a cash outflow.

3.1 Liability

Before explaining the main idea behind the market-maker model, it is better to formally define what "liability" on one horse means for the bookmaker.

The liability is the amount that the bookmaker could win or lose when that horse wins the race:

$$TL_k = \sum_{i=1}^{N} X_i - \mathbf{1}_{\boldsymbol{k}} \sum_{j=1}^{J} (X_{k,j} * P_{k,j}) \qquad (2)$$

Where:

TL_k is the total liability on the k-th horse;
N is the number of horses;
X_i is the total amount wagered, by the bookmaker's customers, on the i-th horse;
$\sum_{i=1}^{N} X_i$ is the amount wagered on all the horses;
J is the total number of order made on the k-th horse;
$X_{k,j}$ is the j-th amount betted on the k-th horse;
$P_{k,j}$ is the price made by the Bookmaker for the j-th orders[56] on his sportbook;
$\mathbf{1_k}$ is the indicator function. It is equal to 1 if the k-th horse wins the race, 0 otherwise;
$\sum_{j=1}^{J} (X_{k,j} * P_{k,j})$ is what the bookmarker has to pay in case of victory of the k-th horse.

When TL_k is negative, it means that in case of victory of that horse, the bookmaker is going to pay an amount bigger than his revenues of all the horses. Whereas, in case of positive value, it means that the revenues are bigger than the costs, thus he will make a profit.

[56] The bookmakers tend not to change their prices too often because it will be seen as not "correct" for the customers who already have betted.

3.2 The model

As we have seen in the previous chapter, the percentage amount matched in the pre-live period is around 93% for both class 1 and 2. Moreover, the prices tend to remain relative stable during the pre-live, while they are extremely volatile and chaotic during the in-play period.
Taking these two main factors in consideration, the model developed in this thesis will work only in the pre-lived period. More specifically it will start to act as a market maker during the last 25 minutes before the start of the race only.

The best way to hedge the liability, while acting as market maker, is to back and lay on all the horses in an asymmetrical way. This means that we want to back more than lay[57] on the horse with the higher negative liability and we want to lay more than back on the other horses.
In this way, the formula (2) becomes:

$$TL_k = \sum_{i=1}^{N} X_i - \sum_{i=1}^{N} Y_i + \mathbf{1}_{\boldsymbol{k}}\left[\sum_{l=1}^{L}\left(Y_{k,l} * BP_{k,l}\right) - \sum_{j=1}^{J}\left(X_{k,j} * LP_{k,j}\right)\right] \qquad (3)$$

Where:

TL_k is the total liability on the k-th horse;
N is the number of horses;
X_i is the total amount wagered, by the bookmaker's customers and by the bookmaker itself with the lay orders on the exchange, on the i-th horse;
$\sum_{i=1}^{N} X_i$ is the amount wagered on all the horses;
$\sum_{i=1}^{N} Y_i$ is the amount the bookmaker wagered on the exchange on all the horses,
$\mathbf{1}_{\mathbf{k}}$ is the indicator function. It is equal to 1 if the k-th horse wins the race, 0 otherwise;
L is the total number of back-order made on the k-th horse;
$Y_{k,l}$ is the amount betted on the l-th order on the k-th horse;
$BP_{k,l}$ is the back price available on the exchange for the k-th horse and the l-th order;
$\sum_{l=1}^{L}(Y_{k,l} * BP_{k,l})$ is the amount the bookmaker will receive if k-th horse wins the race;
J is the total number of lay-order made on the k-th horse;
$X_{k,j}$ is the amount betted on the j-th order on the k-th horse;
$LP_{k,j}$ is the lay price (made on the exchange or by the bookmaker[58]);
$\sum_{j=1}^{J}(X_{k,j} * LP_{k,j})$ is what the bookmarker has to pay in case of victory of the k-th horse.

---

57 Recall: *Backing* a horse is betting on the victory of that horse. Whereas, *laying* is betting on the outcome that the horse will lose the race.

58 As we known, (from chapter 1) the exchange's odds are different from the bookmaker's one. Thus, $LP_{k,j}$ will be the price offer by the bookmaker when a customer bet on the sportbook. Whereas, it will be the lay price when the bookmaker place a lay order on the exchange.

With this strategy, the model is reducing the negative liability in two ways[59]:

3. Backing on the horse with the negative liability. Thus, if that horse will win the race, the bookmaker will receive money from the exchange for winning the bets. In the formula (2), the $\sum_{l=1}^{L}(Y_{k,l} * BP_{k,l})$ for the k-th horse and $\sum_{i=1}^{N} Y_i$ for all horses are increasing. For the horse with a negative liability, the positive effect of the first term is higher than the negative effect of the second term. For all the other horses, we have just the negative effect of the second term which is small;
4. Laying on all the other horses. In this way, the model is increasing the cash inflow from the other horses[60]. In the formula (2) the $\sum_{i=1}^{N} X_i$ term for all the horses and $\sum_{j=1}^{J}(X_{k,j} * LP_{k,j})$ for the k-th horse are increasing. Therefore, for the horse with a high negative liability there is a positive effect, which helps to complete the hedge. Whereas, for all the other horses the negative effect will be greater and, as result, the positive liability on those horses will decrease[61].

Furthermore, the strategy followed by the model must be dynamic. If the liability on one horse, during the "lifecycle" of the race, changes it sign (i.e. it moves from negative to positive), the model should change dynamically the "direction" of its bets (e.g. when the liability is negative the model will back more than lay, however, when it will become positive the MM will start to lay more than back).
In addition, we should keep in mind that the overround times the amount wagered, which can called "limit", by Ladbrokes's customers is the amount that, in an ideal situation, we will earn regardless the result of the race.

Therefore, for the above cited characteristics, the model will change its behaviour based on the position and the limit, which can be derived from multiplying the overround with the wagered amount. In table 16, we can see the MM behaviour[62].
It is important to underline that on the favourite horse we want to back more than lay regardless the liability sign[63].

Table 18 Model behaviour

Scenario	MM behaviour
$Liability_k < 0$	Backing more than laying
$0 < Liability_k < limit$	Do not trade on the horse
$Liability_k > limit$	Laying more than backing

While the first and third scenarios are easy to understand, instead it is better to explain why the mode will not trade in the second scenario. When the liability is positive and higher than the limit, the model

[59] We are considering just the final effect. The model will back and lay on all horses, however, thanks to the asymmetrical betting, the net effect will be just back orders for the horses with negative liability and lay orders for the horses with a positive one.
[60] As we said, when you made a lay order on the exchange you will receive the amount wager, which is the bookmaker's turnover.
[61] That is a logical result considering that we are implementing a hedging strategy.
[62] $Liability_k$ is the value of the liability of the k-th horse.
[63] This is because the final liability on the favourite horse will be, in most of the cases, negative and the liquidity on the exchange is not enough to hedge the liability completely.

will lay more than back. Doing so the liability will be reduced over time and, without any non-trading area, it can easily changes sign and become a negative value. If the model stops to trade when the liability is lower than the limit, we can avoid to transform a positive liability to a negative one.
On the other hand, if we start with a negative liability when it will become positive, without a non-trading area, the model will start to lay more than back and the result will be a negative value again. Implementing a non-trading area is important to make the model more efficient.
In order to clarify the behaviour of the model, we can make a numerical example.
Considering that the customers wager 80,000$, the overround is 15% and the limit is 12,000$[64]. When the liability on the horse is negative, the model will back more than lay. When it is bigger than 0$, but lower than our limit (0<liability<12,000$), the model will not trade. When it is bigger than our limit, the model will lay more than back (liability >12,000$).
Implementing these conditions will allow the model to dynamically handle the changes in liability sign.

Now that the model's strategy and behaviour has been explained, we need to define the other key elements to implement "materially" the MM model:

4. The expected mid-price;
5. The back-lay spread;
6. The orders size;

The first step will be to compute the expected mid-price, after that the model will apply a certain number of tick (i.e. the back-lay spread) in order to obtain the back and lay prices. Once the prices are computed, the MM model will place limits order at that prices with the determined size.
In the next paragraphs, we will discuss in detail the estimation of these parameters.

3.2.1 Estimation of the mid-price

First, we need to estimate the expected mid-price. This is the foundation of the model.
In each time-period the price is allowed to increase, decrease or to remain unchanged. If we are able to compute the probability of a price's increase, we can estimate the mid-price as in a simple weighted average:

$$Ep = Best\ BP * (1 - Pup) + Best\ LP * Pup \quad (4)$$

Where:

Ep is the Expected price,
$Best\ BP$ is the best back price available on the exchange,
$Best\ LP$ is the best lay price available on the exchange,
Pup is the probability of an upward price movement ,
$(1 - Pup)$ is the probability of a downward price movement.

[64] The limit is equal to the total amount wagered by Ladbrokes' costumers times the overround.

To compute the Pup, following the methodology of Avelanda et al, (2011)[65], I am exploiting the information implied in Order Book (OB)[66].
More specifically, the level 1 quotes are being considered:

- The best back price and its volume,
- The best lay price and its volume.

During the trading period, we can have two main events:

- The best back size is depleted and the price decreases (i.e. the second best back becomes the new top of the market and, of course, the price is lower than the previous one);
- The best lay size is depleted and the price increases (i.e. the second best lay becomes the new top of the market and the price is higher than the previous one).

Thus "*the dynamics leading to a price change may thus be viewed as a race to the bottom*" (Avellanda et al[41]) In effects, if the back market orders deplete the best lay size before the lay market orders deplete the best back size, we will have a price's increase. Otherwise, if the lay market orders deplete the best back size before the back market orders deplete the best lay, the price will decrease.
Avellanda et al use the following equation to compute the probability of a price's increase based on this "race to the bottom":

$$Pup = \frac{Best\ Back\ Volume}{(Best\ Back\ Volume + Best\ Lay\ Volume)} \qquad (5)$$

In order to see if the formulas (4) and (5) are feasible for our Betting Exchange, I have done some back testing.
I have built a code in Matlab in order to obtain the Order Book and, consequently, the top of the market from the Betdaq Database[67].
After obtained the top of the market, the formulas (4) and (5) have been applied in order to compute the expected price. The final step was to compare the EP with the matched price one-step ahead and computed the forecast error. In this way, it is possible to see the robustness of the estimation.

The test has been done on 120 horses of 16 races of class 1 event.
In order to have a clearer picture, the horses have been classified and aggregated by their average price in the pre-live period (e.g. Horse 1 will be the favourite horse and so on)[68]. Thus, the analysis has been done for each representative horse. Furthermore, the results have been standardized by a tick, in order to make them comparable, considering that their values change in relation on the price.

In table 17 and table 18 we can see the results or our estimation.

---

[65]Avellaneda, M., Reed, J. & Stoikov, S., 2011. Forecasting Prices from Level-I Quotes in the Presence of Hidden Liquidity. Algorithmic Finance, 29 June, 1(1)
[66] The Order book is a list, organized by price level, of available buy and sell orders for a specific financial instrument.
[67] The code will not be explained in detail due to its complexity. However, in order to have an idea I have recreated the order book from the first instance until the beginning of the race. To do so have used as input all the orders, in chronologic order, made on the exchange for every horse and the top of the market, after each order, has been saved.
[68] As in the study of the average bet's size.

The results are extremely positive: as shown[69], the mean estimation error is around 0.5 ticks and the standard deviation is low. Thus, estimating the Expected price and the Pup with formulas (3) (4) is a reliable methodology.
We can see that the prediction error of the horses number 7 and 8 (the second to last and last favourite) can become quite big (i.e. max error for horse 7 is 97 ticks). This happens when the actual spread on the exchange is big (e.g. 30-50 ticks) and the matched price bounces from the lower bound to the higher bound.

Table 19 Class 1 EP's prediction error (value in ticks, timespan: last 50 minutes)

Percentile[70]	Horse 1			Horse 2		
	Max Prediction Error	Average Prediction Error	Standard deviation	Max Prediction Error	Average Prediction Error	Standard deviation
5th percentile	1	0.370	0.465	2	0.328	0.459
50th percentile	3	0.491	0.552	2	0.467	0.505
95th percentile	5	0.728	0.723	3	0.545	0.578
	Horse 3			Horse 4		
	Max Prediction Error	Average Prediction Error	Standard deviation	Max Prediction Error	Average Prediction Error	Standard deviation
5th percentile	1	0.232	0.430	2	0.337	0.499
50th percentile	2	0.458	0.526	2	0.527	0.544
95th percentile	4	0.531	0.581	3	0.542	0.625
	Horse 5			Horse 6		
	Max Prediction Error	Average Prediction Error	Standard deviation	Max Prediction Error	Average Prediction Error	Standard deviation
5th percentile	1	0.428	0.478	2	0.451	0.514
50th percentile	2	0.526	0.533	2	0.475	0.534
95th percentile	5	0.562	0.602	3	0.565	0.559
	Horse 7			Horse 8		
	Max Prediction Error	Average Prediction Error	Standard deviation	Max Prediction Error	Average Prediction Error	Standard deviation
5th percentile	2	0.451	0.548	2	0.455	0.532
50th percentile	3	0.607	0.669	3	0.670	0.660
95th percentile	235	18.401	33.644	390	57.868	63.120

[69] The timespan considered were: 50-0, 25-0, 10-0 and 5-0 minutes. In the thesis are presented the first and the last range. The results are nearly the same.
[70]Each group of representative horses contains the data of 16 different races. Therefore, after computing the metrics (i.e. max prediction error, average prediction error and standard deviation) for each races and sorted it, I have taken the values for the desired percentiles as in the study of the average bet's size.

Table 20 Class 1 EP's prediction error (value in ticks, timespan: last 5 minutes)

Percentile[46]	Horse 1			Horse 2		
	Max Prediction Error	Average Prediction Error	Standard deviation	Max Prediction Error	Average Prediction Error	Standard deviation
5th percentile	1	0.352	0.447	1	0.202	0.313
50th percentile	3	0.522	0.555	2	0.534	0.515
95th percentile	4	0.723	0.698	3	0.648	0.638
	Horse 3			Horse 4		
	Max Prediction Error	Average Prediction Error	Standard deviation	Max Prediction Error	Average Prediction Error	Standard deviation
5th percentile	1	0.322	0.474	1	0.329	0.492
50th percentile	2	0.532	0.555	2	0.575	0.579
95th percentile	3	0.660	0.780	3	0.736	0.688
	Horse 5			Horse 6		
	Max Prediction Error	Average Prediction Error	Standard deviation	Max Prediction Error	Average Prediction Error	Standard deviation
5th percentile	1	0.464	0.467	2	0.468	0.515
50th percentile	2	0.511	0.525	2	0.576	0.584
95th percentile	3	0.766	0.785	3	0.658	0.711
	Horse 7			Horse 8		
	Max Prediction Error	Average Prediction Error	Standard deviation	Max Prediction Error	Average Prediction Error	Standard deviation
5th percentile	1	0.479	0.502	1	0.256	0.439
Mean/50th	3	0.746	0.692	3	0.731	0.644
95th percentile	97	13.360	20.623	25	5.679	6.524

3.2.2 The spread

The spread should compensate the Market Maker for his service of providing liquidity to the market. Moreover, it should defend him against several risks such as:

- Asymmetric information. This risk might arise when the market maker is trading with someone who has superior information about the race or a superior ability to estimate true probability of victory of the horses.
- Inventory risk. It may arise when the majority of customers want to bet on one horse and the exposure to that horse becomes too big and too risky;

There are many approaches to compute the spread. One is a Naïve approach, which will just require choosing a fixed number of ticks (i.e. 1) as a spread considering that, in theory, the odds offered in

the exchange are higher than the bookmaker's ones thus, operating in the exchange should be relatively safe for the bookmaker.
In this work, we will compute the spread in a more robust way, based on the analysis made on the exchange.

The spread should be a function of the following parameters:

- Time and liquidity. Less liquid markets require a bigger spread. Considering that liquidity is also a function of time in the betting exchange[71], the spread must be itself a function of time;
- Volatility. A more volatile market means that the risk of acting as market maker is bigger. Thus, a bigger spread is needed;
- Position. When the exposure on one horse increases considerably, it is possible to increase or decrease the spread to handle it in the best way possible.

Taking the first-two elements in consideration, I have computed the spread in the following way:

$$S_{T+1} = max(1, round(e^{-X} * \frac{As_t}{2} * (1 + \sigma)) \qquad (6)$$

Where:

S_{T+1} Is the spread at time t+1 (expressed in number of tick);
As_t Is the actual spread;
σ Is the historical price's standard deviation;
X Is the historical mean percentage value of amount matched during time (is the value of the liquidity ratio defined in chapter 2.2);

The actual spread is the starting value in (6). That value will be adjusted considering the time evolution of liquidity and the volatility of the horse.

The value of X is linked to the historical percentage amount matched. As we have seen[72], the matched amount is extremely low until 25 minutes before the start of the race and, after that, it increases in an exponentially way. The trend is nearly the same for class 1 and class 2. Therefore, the most logic way to link the spread with the liquidity is to take the inverse of the exponential function, e^-X. In this way, we will have a wider spread in the first stages of the race, when the market is illiquid (i.e. the value of e^-X will be bigger). Whereas, when the market will be more liquid, 25-20 minutes before the start of the race, the spread will be thinner.

Thus, the value of e^{-X} will be set as:

$$e^{-X} = \begin{cases} 0.65, & 10 \leq t < 25 \\ 0.52, & 5 \leq t < 10 \\ 0.3, & 5 < t \end{cases}$$

[71] Chapter 2.2
[72] Chapter 2.3

It will be important to adjust these values with fine-tuning operations to be sure that they fit the real market conditions.

In Chapter 2, price standard deviation has been computed and we have seen that the price is stable. However, when the actual spread in the market was wide, the prices were highly volatile. This situation happens, above all, for the last favourite horse.

Therefore, in order to select the right value for σ, we need to separate the two different cases: when the actual spread is less than a threshold of Z ticks and when it is higher.

In the first case, we can use the historical value, and precisely the 50^{th} percentile's one, as value of σ. Whereas, in the latter case we have two possibilities: we can use the 95^{th} percentile's value or we can decide not to trade on that horse. The main reason to do not trade on horses with an extreme price standard deviation is to avoid potential losses due to the spread mispricing. Moreover, the amount betted on those horses is usually very low.
The value of Z, based on the different calibrations, can be set equal to 20 tick.
It will be extremely important to see if this value fit the market and, if not, to calibrate a better value.

In order to handle the "Inventory risk", instead, a bigger spread will be applied to the orders, which will worsen our position[73].
Therefore, for the horses in which we want to back more than lay, we will compute the back and lay price[74] as:

$$Back\ price = Ep + S * v \qquad (7)$$
$$Lay\ price = Ep - (S + 1) * v \qquad (8)$$

Whereas, for the horses in which we want to lay more than back, it will be:

$$Back\ price = Ep + (S + 1) * v \qquad (9)$$
$$Lay\ price = Ep - S * v \qquad (10)$$

Where:
Ep Is the Expected price;
S Is the spread;
v Is the value of the tick.

3.2.3 The order size

The third element we need to define is the order size.
Before doing so, we need to recall that bettors do not wager in a homogenous way on all the horses: they prefer to bet more on the first three favourite horses[75] than on all the others.
This is because the first three horses have the highest implied probability of victory.

[73] E.g. if the liability on one horse is negative, we want to reduce the possibility that a lay order will be hit. Thus, we will subtract a higher spread to the Ep.
[74] Note: Here we are seeing the back and lay price by the market maker point of view. In the market the Lay price would be the back price and the back price the lay price. In financial markets the MM buys at bid and sells at ask price.
[75] E.g. for the class 1 the average amount betted on the first three horses is around 74%, while for class 2 it is around 79%.

Thus, considering the strategy that should be implemented, the model too need to bet more (both backing and laying) on the first three horses than all the others.

As we have seen in the analysis of the average bet size[76], we can take the mean value of the 50th percentile as bet size.
Furthermore, the size decreases as we move from the favourite horse (horse 1) to the last favourite one (horse 8), so this is coherent with the customers' behaviour.

The last step is to multiply the bet size by $1 + \beta$ to achieve the aim of hedging our liability.
For the horse with a high negative liability, the value of β will be:

$$\beta_k = \begin{cases} 1, & for\ the\ lay\ orders \\ 2, & for\ the\ back\ orders \end{cases}$$

Whereas, for the other horses it will be:

$$\beta_k = \begin{cases} 1, & for\ the\ back\ orders \\ 2, & for\ the\ lay\ orders \end{cases}$$

It will be important to calibrate $the\ \beta$ value to be sure that it fits the real market conditions.

---

[76] Chapter 2.3.

4 Results

In order to test the MM model two different tests have been performed:

1. The first one is focused on the spread estimation. The spread on some different kinds of horse has been computed to see if the equation (6) can produce feasible value;
2. In the second test the MM model was applied to some races to see the results of the model.

In the first test, the formula (6) has been applied to some horses in order to see, through graphs, the back and lay price path created by the model. Moreover, the case of volatile horses has been analysed in order to understand if it is better to use, as volatility, the 95th percentile value of our study, or not to trade.

In figure 9, we can see the effect on one favourite horse (horse 1): in sky blue we have the matched price, in black the market maker back (which is the lay price in the exchange[77]) and in red the market maker lay (which is the back price in the exchange). As we can see, the results are extremely positive. Quite always, the liability on the favourite horse is negative (people prefer to bet on the horse with the higher probability of victory) and we can see that the market model is able to back more than lay. The blue vertical line underlines when our market making model will start to bet on the exchange.

In figure 10, we can see the effect on one "normal" horse 7. It is called normal because the spread was always less than 20 tick, thus the price standard deviation was set as the mean values as in table 14. The line are exactly like before. We can see that the results continue to be positive. On this kind of horse, we usually want to lay more than back and we can see that the matched price is nearer to the lay price than the back price.
In the figure 11, we can see what happens when we have a price with a high standard deviation.
Here, instead of taking the mean value (table 14), we took the 95th percentile value (the spread was bigger than 20 many times). As we can see, the accuracy is not perfect. Moreover, the price path is extremely volatile and this could lead in losses.

Now that we have a clear picture of the results of our spread estimation, it is possible to decide how to manage highly volatile horses. This problem subsists just for the last favourite horses. The amount betted on those horses is usually very low. Therefore, in order to avoid potential losses due to the spread mispricing, the best solution is to not trade in this situation.

[77] The model places limit orders. As we saw in chapter 1 a limit order is shown on the opposite sides of the markets.

Figure 11 Spread on horse 1

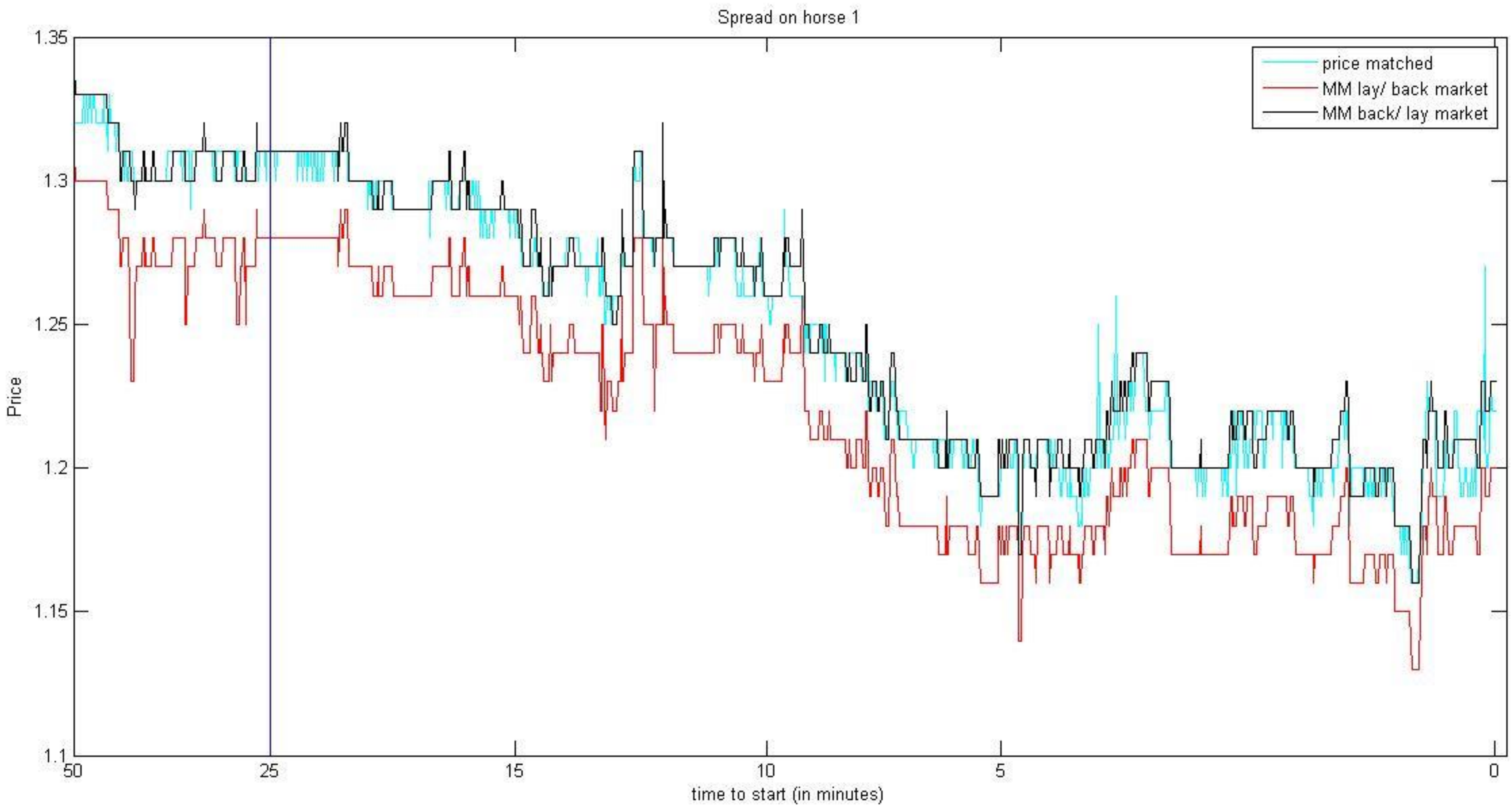


Figure 12 Spread on a normal horse 7

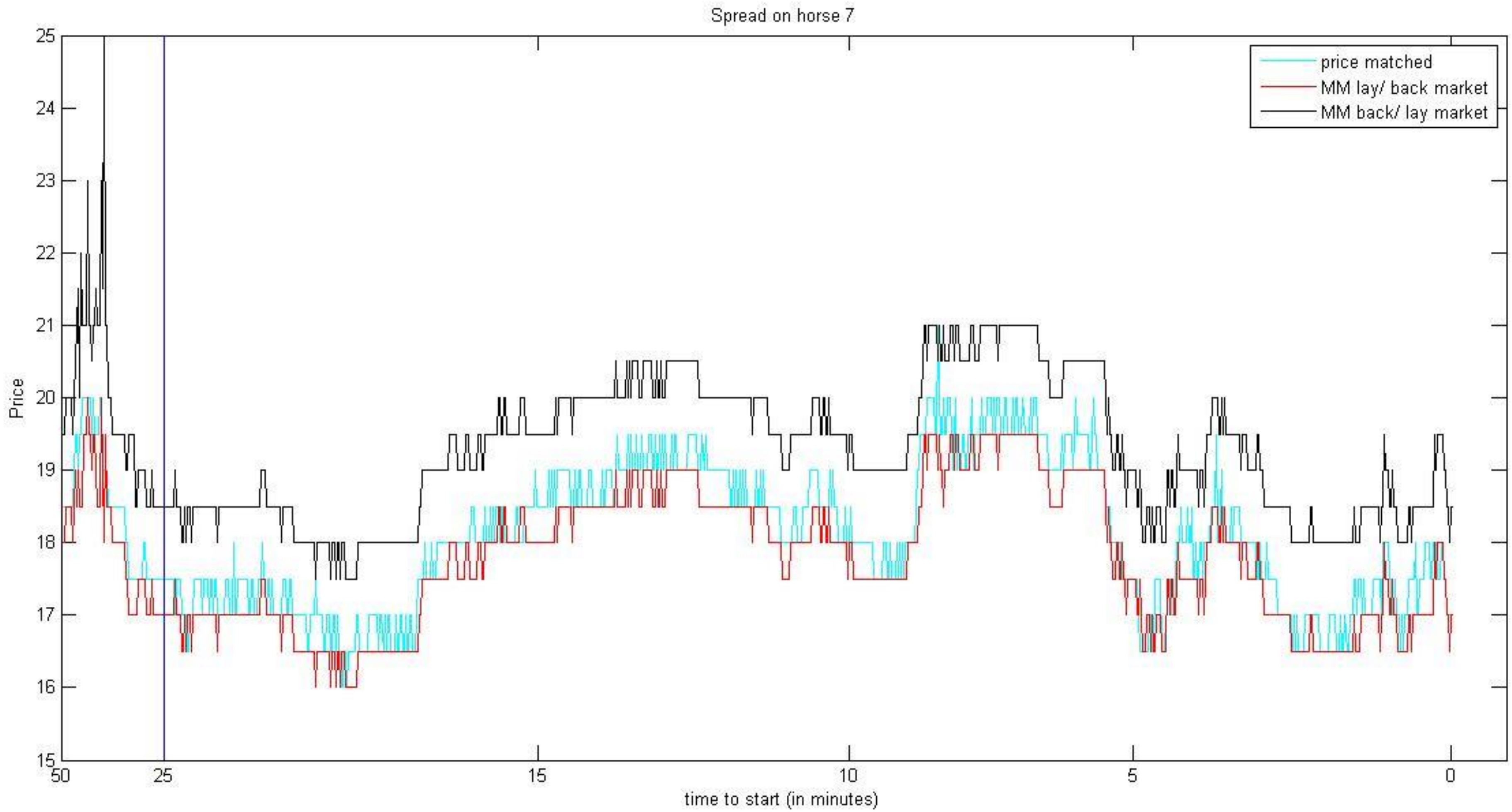
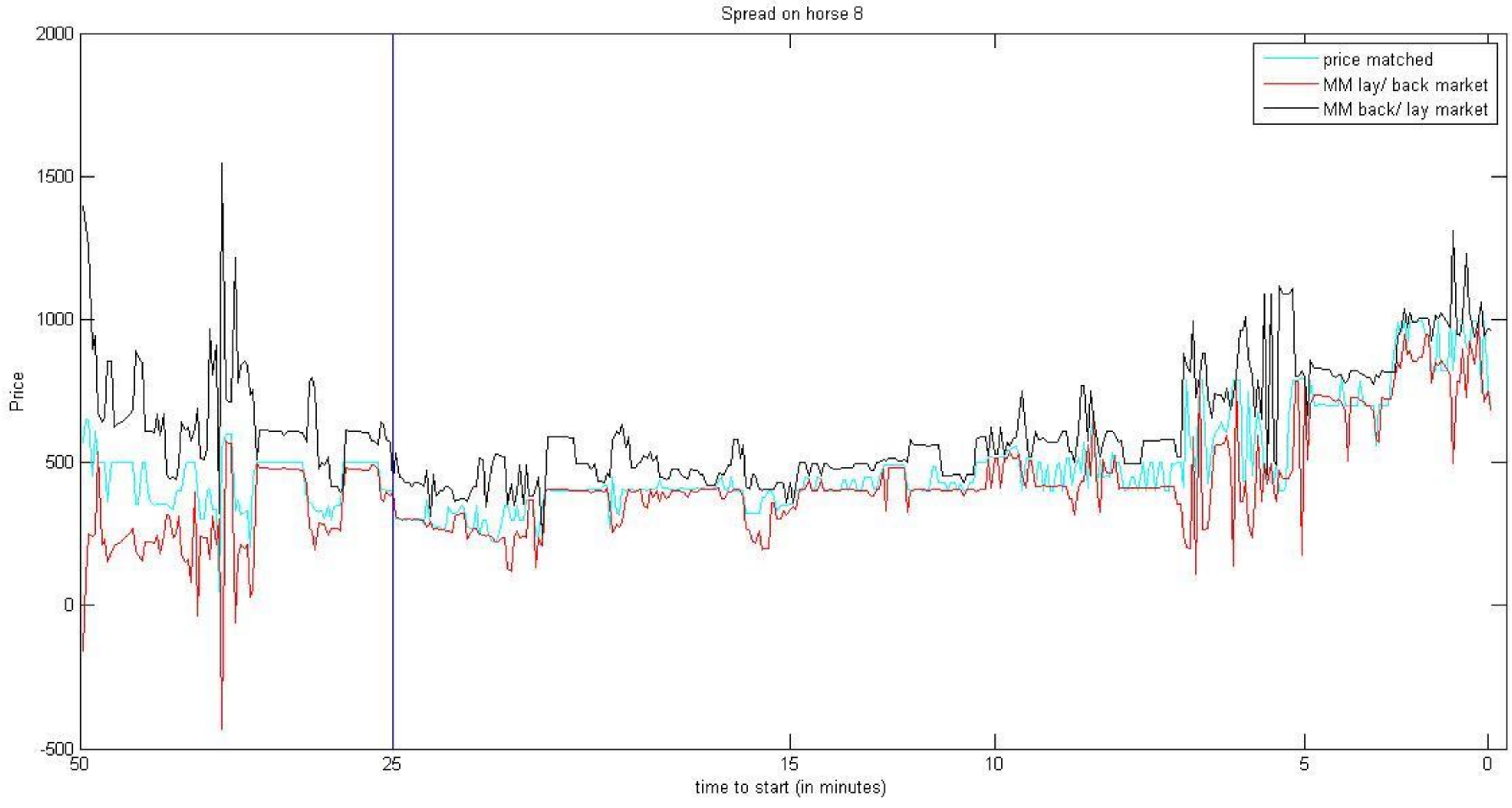


Figure 13 Spread on horse 8



The second test, as said before, has been done in order to see the effect of the MM model on the Ladbrokes's liability and profitability.
To perform this test, another database provided by Ladbrokes plc has been used. This database contains all the wager places by Ladbrokes's customers.
I have built an algorithm[78] in Matlab, which simulates the betting exchange dynamics in order to see the effect of our model.
The procedure followed is the ensuing:

1. First of all, I have taken all the bets, divided by the horses, made from Ladbrokes's customers on the sportbook. In this way, it has been possible to compute the evolution of liability for each horse during time. It has been done for the last 25 minutes with a 1 second step;
2. The second element needed has been the evolution of the top of the market and the matched price. The timespan has been standardized to match with the liability's one.
3. The final step has been the implement of the MM model, linking the sportbook liability with the one generate by the model.

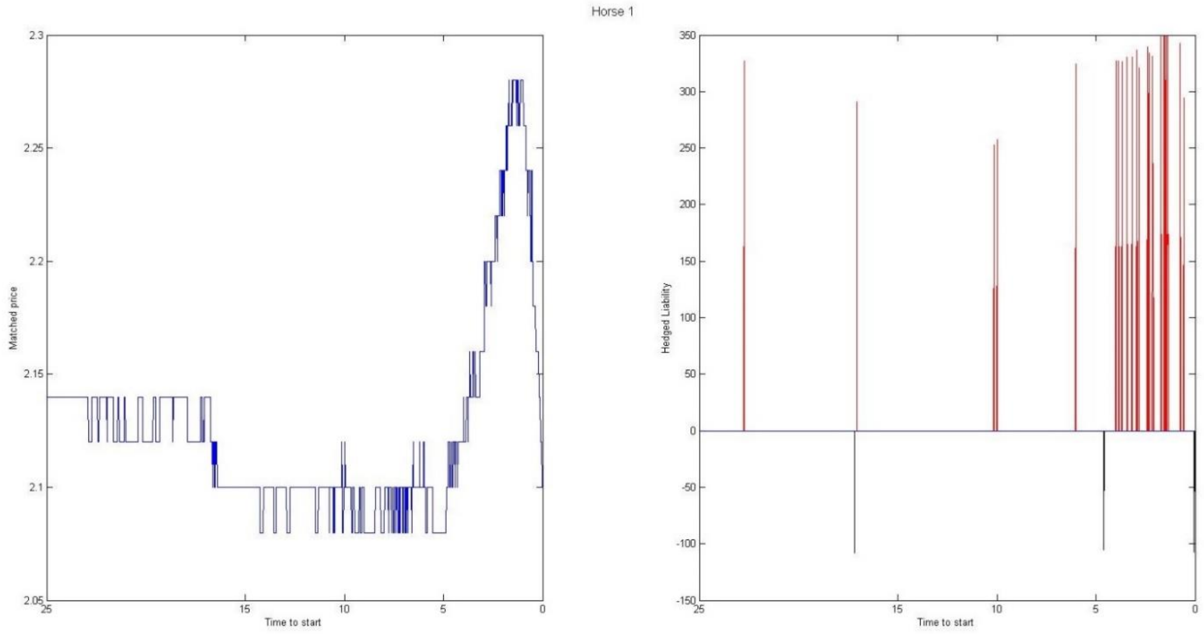
Before explaining the results, it is fundamental to underline one hypothesis that was needed in order to test the model: the price matched is assumed to be the same as the one observed. Our strategy is not affecting the price formation.

Considering that the role of a Market Maker is to provide liquidity to the market, his presence will probably affect the price formation. However, this fact will not compromise the result. The most obvious consequence will be an increase of the number of matched orders and it should result in an improvement of the overall performance.

In Table 19-20 and Figure 13-16 we can see the result on 2 different races.
Figure 13 is divided in 6 sub-groups of graphs. Each sub group represents one horse (e.g. the horse 1 is the favourite horse, the horse 2 is the second favourite and so on). In each subgroup the left graph shows the evolution of the matched price, while the right graph shows the liability generated on the exchange during time while/and the right graph shows the evolution of the liability.
In figure 12 we can see one sub-group of graphs, which represents the horse 1 of the first race.

[78] In this thesis, just a simplified explanation will be presented.

Figure 14 sub-groups for horse 1 of races 1



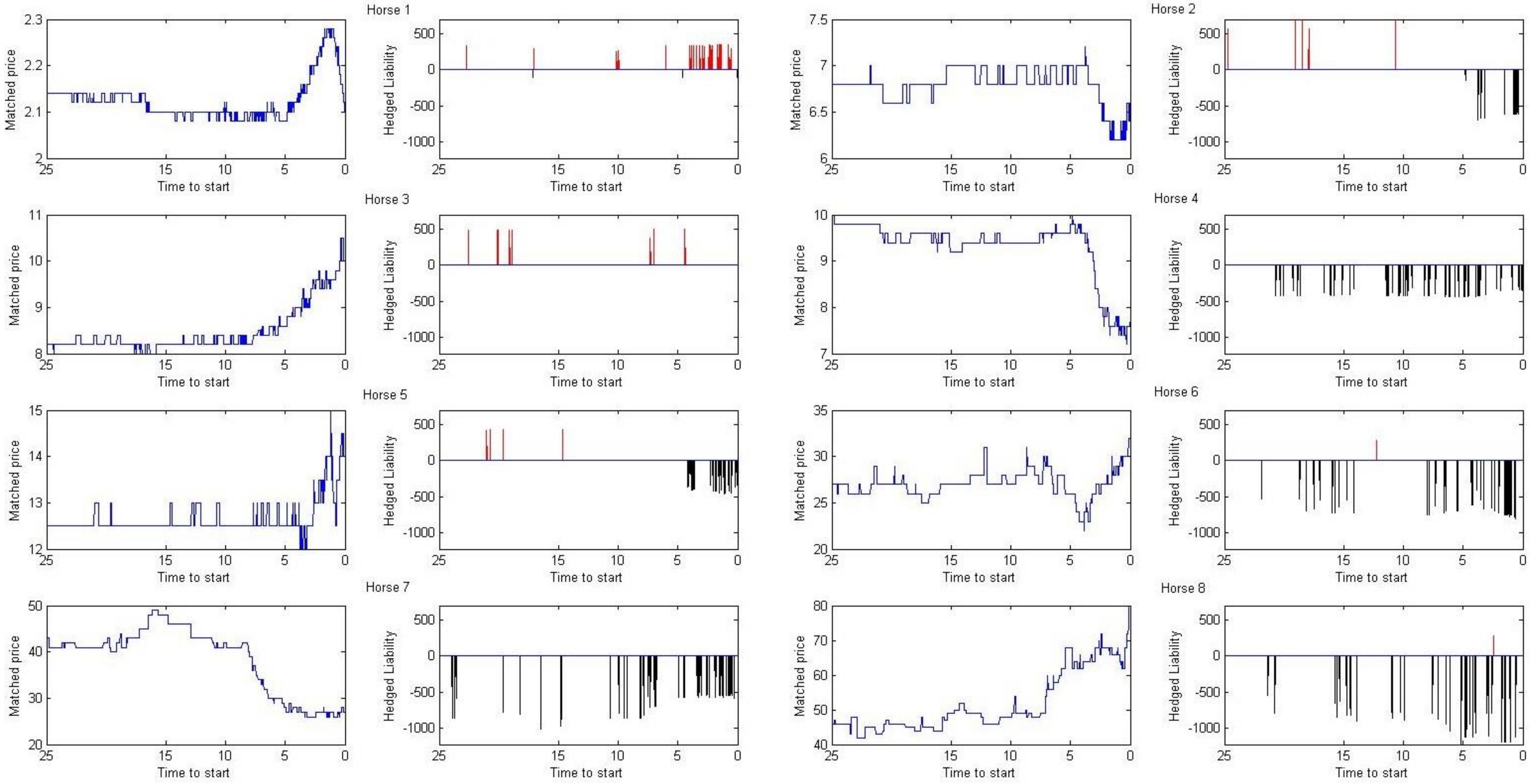
In the right graph, each bar represents the amount backed or laid, by the model on the exchange, with one order. A red bar means that we backed on that horse, whereas a black bar means that we laid on that horse.

The tables help us to understand the effect of the MM model. The winning probability column shows the implied probability derived by the bookmaker's price. The true probability column, instead, shows the winning probability without the overround[79]. The Liability pre MM is the liability faced the bookmaker without the model. Whereas the Liability post MM is the final liability that we will obtain thanks to the model and it was computed following the formula (3). The value obtain in these columns are the profit/loss of the bookmaker in case that horse win the race (e.g. if horse 1, in race 1, win the race the bookmaker will lose 20,334$ pre MM and 5,431$ post MM). The MM effect column shows the change in the liability. The sportbook turnover column shows the amount betted by Ladbrokes' customers, while the exchange turnover column shows the turnover generated on the exchange. It is important because, considering that Betdaq was bought by Ladbrokes, that amount times 1.2%[80] is an extra profit that the model generated[81]. The amount betted column shows the amount of £ that the model need to perform, when negative, or generate, when positive, during the 25 minutes before the start of the race[82].

The last two columns show the profit's expected value (EV) with and without the MM model. In order to compute so, I have multiplied, for each horse, both sportbook and hedged liability by the true probability of winning the race and I have summed-up the single results. In order to make a complete comparison, I have computed also the relative standard deviation, the range of EV[83] and the actual bookmaker's profit/loss.

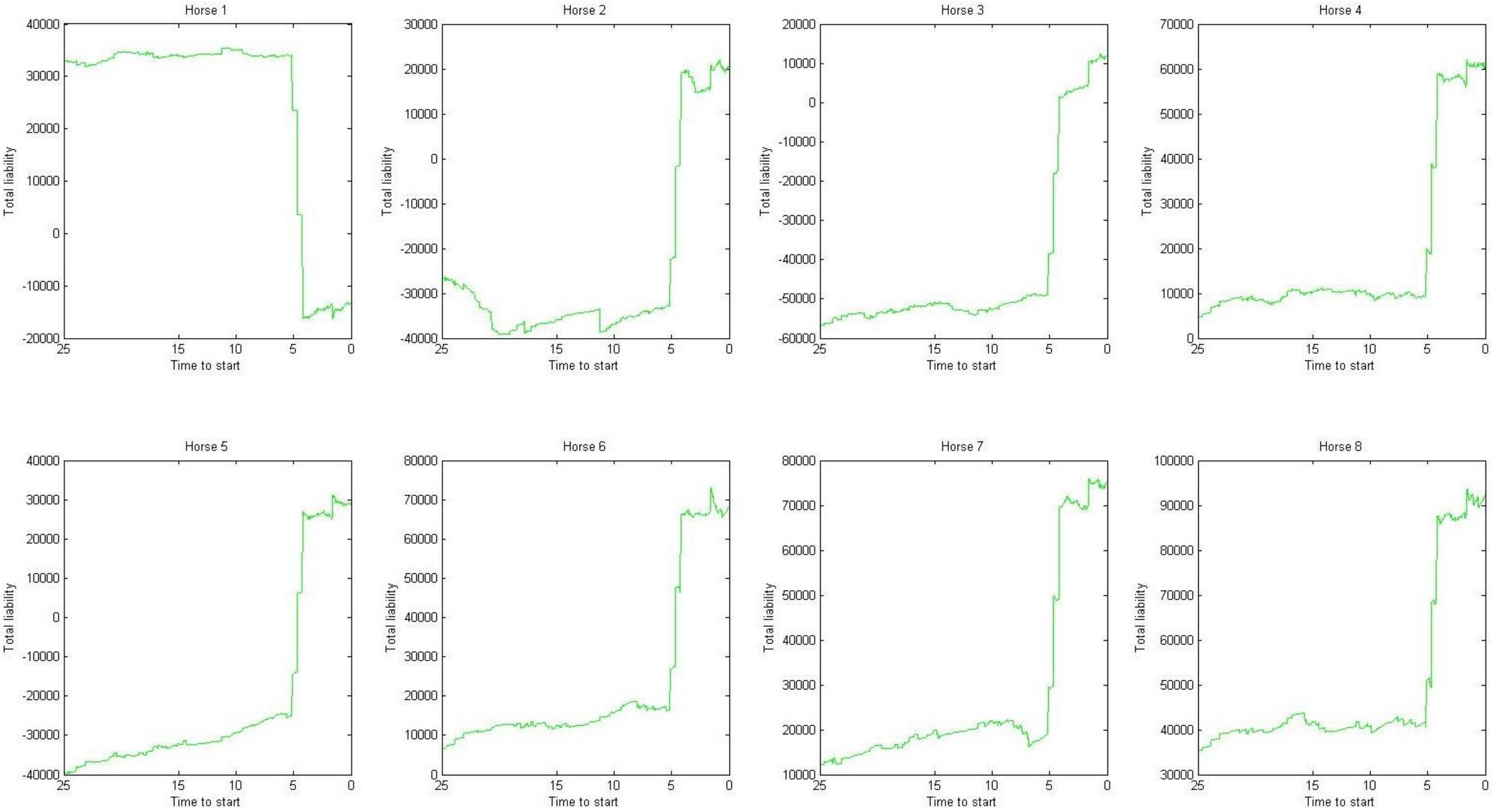
[79] We can obtain the true probability from the winning probability with a simple proportion. See chapter 1.3.
[80] The owner of the exchange charges a commission fee on the net profit. Betdaq charges a standard rate of 5%, but it can be reduced as low as 2%, depending on the total amount wagered on the exchange during time. The value of 1.2% it consider also tax on the profit.
[81] This, of course, is valid just in the case that the bookmaker is also the owner of the exchange.
[82] This is what we defined as cash in-flow and outflow in chapter 3.1 .
[83] For the range of EV it was taken the minimum and the maximum EV pre and post MM.

Figure 15 MM effect on race 1 (Winner: horse 1)



<u>Legend</u>: Each sub-groups represent one horse. Left graph: The blue line show the price evolution during the last 25 minutes before the race start. Right graph: Each bar represent the amount backed or laid (i.e. amount wager times price taken). Red bar is a back order, black bar a lay order.

Figure 16 Liability Evolution on race 1 (Winner: horse 1)



Legend: the green line shows the evolution of the liability, considering both the order that the bookmaker received on the sportbook and the orders made on the exchange by the model, on that horse. When the value is negative in means that, in case of victory of that horse, the bookmaker will lose money, otherwise he will have a profit.

In figure 14, we can see that there is a huge drop in the horse 1 and, consequently, a jump on the liability of the other horses around 5 minutes before the race start. This happened because the bettors wagered, on horse 1, a high amount of money at around 5 minutes the race start so as result we have that pronounced trend.

Table 21 Sum-up table on race 1 (Winner horse 1)

Horse	Winning probability	True Probability	Cumulative probability	Liability Pre MM	Liability Post MM	MM Effect	Turnover Sportbook	Turnover Exchange	Amount betted	profit's EV pre MM	profit's EV post MM
1	50,00%	43,84%	43,84%	-$ 27.787,72	-$13.387,64	$14.400,08	$ 84.822,53	$ 4.315,27	-$ 4.010,96	-$12.182,25	-$ 5.869,20
2	16,67%	14,62%	58,46%	$ 24.466,31	$20.451,83	-$ 4.014,48	$ 18.133,83	$ 1.950,60	$ 1.028,22	$ 3.576,09	$ 2.989,32
3	13,33%	11,69%	70,14%	$ 5.271,03	$11.941,74	$ 6.670,71	$ 17.303,31	$ 493,00	-$ 493,00	$ 616,07	$ 1.395,73
4	12,50%	10,96%	81,10%	$ 82.586,32	$60.026,86	-$22.559,46	$ 6.525,96	$ 2.730,30	$ 2.730,30	$ 9.051,55	$ 6.579,01
5	9,09%	7,97%	89,07%	$ 34.263,60	$28.777,32	-$ 5.486,28	$ 8.864,66	$ 870,00	$ 609,38	$ 2.730,87	$ 2.293,61
6	4,76%	4,17%	93,25%	$ 94.511,97	$68.203,96	-$26.308,01	$ 1.813,58	$ 1.093,50	$ 1.075,50	$ 3.944,56	$ 2.846,57
7	4,76%	4,17%	97,42%	$104.117,08	$75.273,28	-$28.843,80	$ 1.087,98	$ 967,72	$ 967,72	$ 4.345,44	$ 3.141,61
8	2,94%	2,58%	100,00%	$124.827,31	$92.191,78	-$32.635,53	$ 541,68	$ 623,51	$ 615,94	$ 3.217,82	$ 2.376,54
total	114,05%	100,00%					£139.093,53	£13.043,90	£ 2.523,09	£15.300,15	£15.753,18
				Profit's EV	Relative STD	Min EV	Max EV	Actual Profit/loss			
			Pre MM	$ 15.300,15	47%	-$12.182,25	$ 9.051,55	-$27.787,72			
			Post MM	$ 15.909,71	26%	-$ 5.869,20	$ 6.579,01	-$13.387,64			

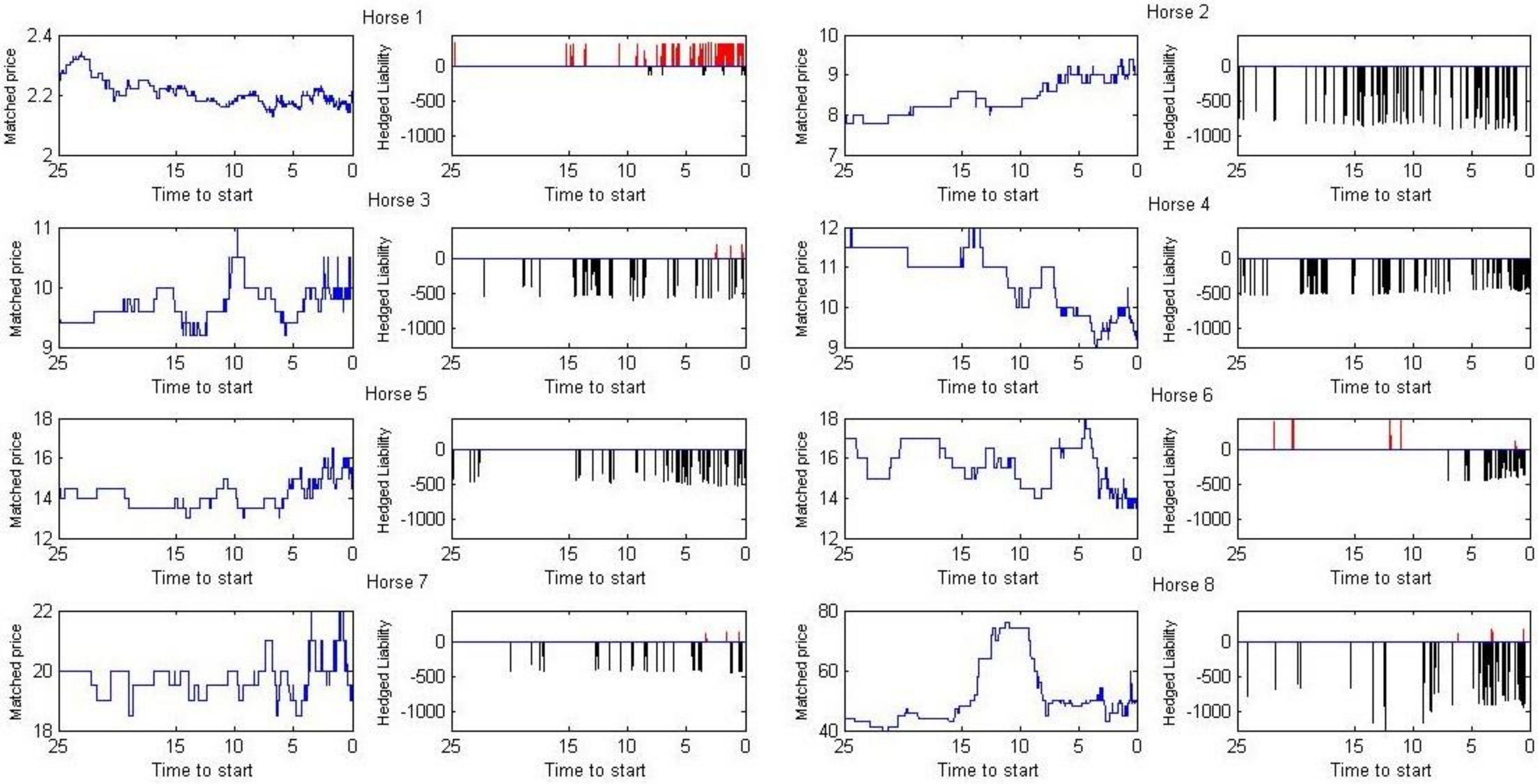
As said in the introduction, the aim of the MM model is to reduce the liability on the horses with a negative one and, therefore, to reduce the level of active risk faced by the bookmaker.

As we can see, the results are positive. The model succeeds in hedging the negative liability on the horse 1 (there is an improvement of 14,402$ which means a relative improvement equal to 51.80%). The level of active risk, measured with the relative STD, is remarkably lower after the MM model, nearly half than the case without the model (from 47% to 26%), while the profit's EV is just a bit higher. Of course, these positive effects comes with a price. The positive liability on the other horses is nearly always lower with the MM maker model. It is a logic and common result considering the general aim of hedging strategies: reducing or nullifying possible losses in case of negative outcome, renouncing a higher profit in case of positive outcome. This idea is well expressed in the min-max profit's EV[84]. We can see that with the model the minimum EV increases from -12.182,25$ to -5869,20$, whereas the maximum value decreases from 9.051,55$ to 6.579,01$. In this case, the favourite horse has won the race and, as result, the bookmaker, with the implementation of the MM model, would has lose only -13.387,64$ instead of -27.787,72$.

The model was tested also in case of a high number of horses (up to 18 horses) per race. The results are mixed.
When the winning probability is allocated in a balance way among the horses, (i.e. true cumulative winning probability on the first 8 horses equal, or lower, to 60%) it is better to not trade on the exchange. This is, mainly, due to the high unpredictability of the race's final result. In theory, in this case, the optimal solution for a bookmaker would be not to accept any bets. However, it will probably results in loosing markets shares, which is not acceptable for any business. The logical strategy to implement will be to reduce as much as possible the number of bets. Thus, it is better not to operate on the exchange.
However, when the probability is distributed in a normal way (e.g. we have a clear favourite runner, or in general the true cumulative winning probability on the first eight horses is higher than 90%), the model continues to produce positive results. We can see an example in figure 14 e table YY.

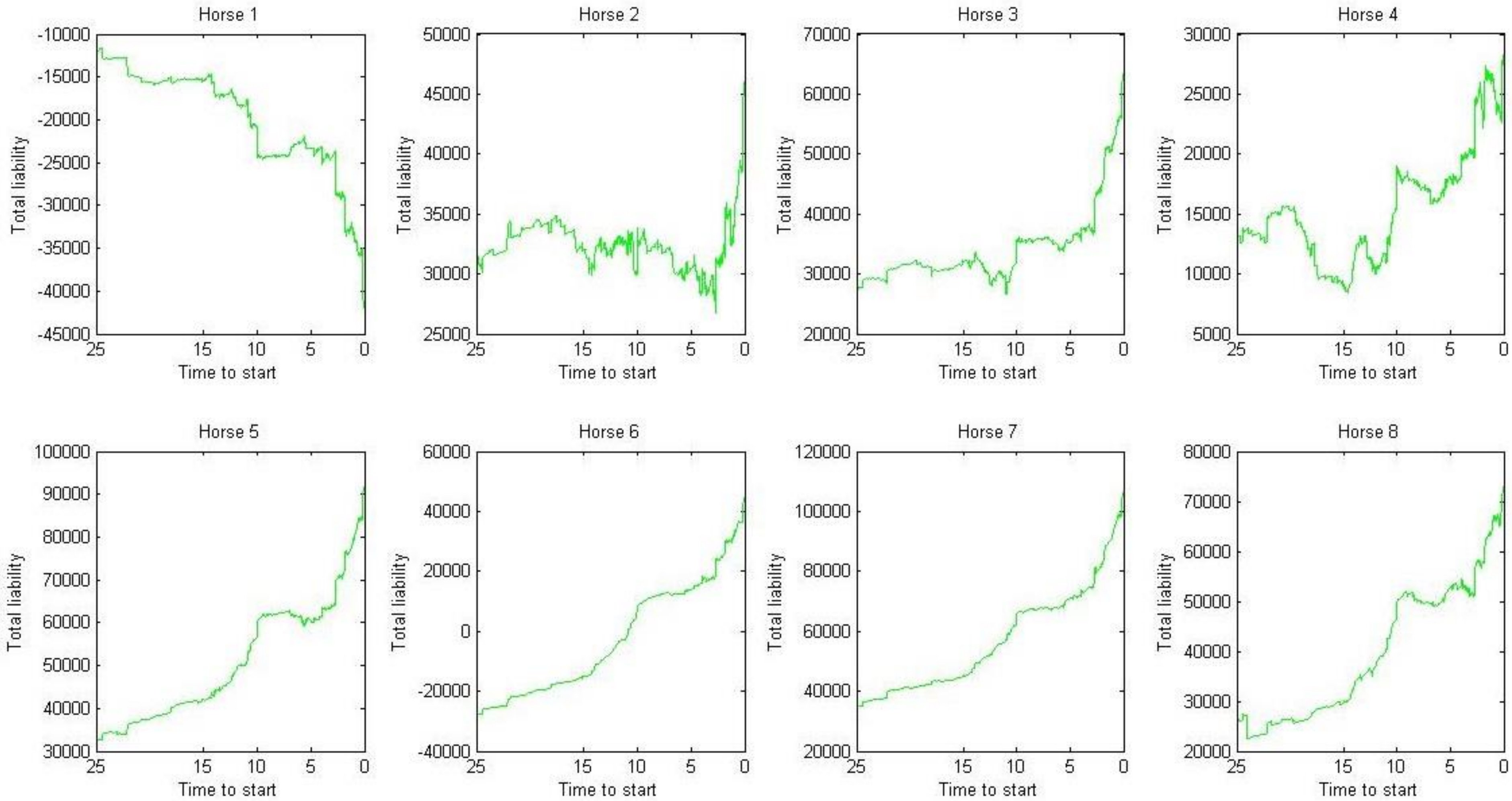
[84] The profit's expected value was computed using the true probability of winning the race.

Figure 17 MM effect on race on race 2 (Winner: horse 6)



<u>Legend</u>*: Each sub-groups represent one horse. Left graph: The blue line show the price evolution during the last 25 minutes before the race start. Right graph: Each bar represent the amount backed or laid (i.e. amount wager times price taken). Red bar is a back order, black bar a lay order.*

Figure 18 Liability Evolution on race 2 (Winner: horse 6)



Legend: the green line shows the evolution of the liability, considering both the order that the bookmaker received on the sportbook and the orders made on the exchange by the model, on that horse. When the value is negative in means that, in case of victory of that horse, the bookmaker will lose money, otherwise he will have a profit.

Table 22 Sum-up table on race 2 (Winner horse 6)[85]

Horse	Winning probability	True Probability	Cumulative probability	Liability Pre MM	Liability Post MM	MM Effect (Delta)	Turnover Sportbook	Turnover Exchange	Amount betted	profit's EV pre MM	profit's EV post MM
1	50,00%	43,22%	43,22%	-$ 71.439,51	-$ 42.047,88	$29.391,64	$103.364,82	$ 10.844,73	-$9.688,38	-$30.872,74	-$18.171,08
2	12,50%	10,80%	54,02%	$ 92.644,99	$ 46.257,67	-$46.387,32	$ 6.093,24	$ 6.378,43	$6.378,43	$10.009,18	$ 4.997,59
3	11,11%	9,60%	63,62%	$ 77.370,74	$ 63.416,54	-$13.954,20	$ 6.503,81	$ 2.465,60	$2.315,57	$ 7.429,46	$ 6.089,52
4	12,50%	10,80%	74,43%	$ 62.755,64	$ 27.703,42	-$35.052,22	$ 7.816,37	$ 4.182,85	$4.182,85	$ 6.780,00	$ 2.993,02
5	7,69%	6,65%	81,07%	$104.767,39	$ 91.437,90	-$13.329,49	$ 2.788,70	$ 1.516,11	$1.516,11	$ 6.963,36	$ 6.077,42
6	7,69%	6,65%	87,72%	$ 45.795,51	$ 44.406,68	-$ 1.388,84	$ 5.804,31	$ 947,92	$ 659,92	$ 3.043,80	$ 2.951,49
7	5,88%	5,08%	92,80%	$108.161,92	$106.375,78	-$ 1.786,14	$ 1.630,08	$ 552,27	$ 511,95	$ 5.496,91	$ 5.406,13
8	2,94%	2,54%	95,34%	$ 99.477,74	$ 73.100,51	-$26.377,23	$ 936,25	$ 715,55	$ 690,82	$ 2.527,78	$ 1.857,52
9-11	5,39%	4,66%	100,00%	$112.066,29			$ 1.664,39				
total	115,70%	100,00%					$136.601,97	$ 27.603,45	$6.567,26	$11.377,76	$12.201,62
				Profit's EV horses 1-8	Relative STD	Min EV	Max EV	Actual profit/loss			
			Pre MM	$ 11.377,76	137%	-$30.872,74	$ 10.009,18	$45.795,51			
			Post MM	$ 12.532,86	75%	-$18.171,08	$ 6.089,52	$44.406,68			

As we can see, the model succeeds in reducing remarkably the negative liability on the horse 1 (nearly 30,000$ improvement) with a decrease of the positive liability on the other horses. The level of active risks is nearly half than the case without the model (from 137% to 75%) and, in this race, the profit's EV is even higher with the MM model than without it. The lower level of risk and, in general, the lower volatility of expected profit is well expressed in the min/max EV range. We can see that the minimum profit's EV is -30.872,74$ pre MM, while is -18.171,08$ post MM. Whereas, the maximum profit's EV is 10.009,18$ pre mm and 6.089,52$ post MM. Finally, we can see that the actual profit/loss have slightly changed (i.e. a decrease equal to 1.388,84$).

The test done on other races, not reported in this thesis, shows nearly always the same result: a lower level of active risk with the same or slightly lower profit's EV.

[85] The sportbook liability for horses 9-11 is an average value, while the winning probability, the true probability and the turnover for horse 9-11 are a cumulative values.

Conclusions

In this thesis, we have analysed the implementation of a Market Maker model, which allows a bookmaker to reduce the negative liability and the risk of horse races. The analysis has been done using the databases provided by Ladbrokes plc.

The final results are extremely positive. In chapter 4, we have seen that, thanks to the implementation of the model, it is effectively possible to reduce the level of active risk faced by the bookmaker having nearly the same expected value of profit.

Moreover, we have seen that when the winning probability is well balanced among many horses it is better to not use the model. This is, mainly, due to the high unpredictability of the race's final result. In theory, in this case, the optimal solution for a bookmaker would be not to accept any bets. However, it will probably results in loosing markets shares, which is not acceptable for any business. The logical strategy to implement will be to reduce as much as possible the number of bets. Thus, it is better not to operate on the exchange.
The model continues to work, when there are many horses but the winning probability is not well balanced among them.

APPENDIX 1: Standard deviation's computation

In this appendix, the methodology followed to compute the S.D. will be explained.

The dataset used to do the analysis consists of 565 horses of 71 classes 1 races and 926 horses of 203 classes 2 races. I have taken the horse identity, the matched price and the time remaining[86] for every matched order. In this way, I have obtained, from the databases provided by Ladbrokes, two matrixes, used in Matlab, with three columns (i.e. horse identity, price, time remaining) and nearly 1.9 million rows for class1 and 2 million for class2.

The methodology followed to compute the S.D. is the following:

1. First of all, as in the average bet's size, I have ranked and classified every horse of each race by their average pre-lived matched price (see table 21);

Table 23 Ranking matrix

Race Identification	Horse Identification	Average Price	Rank
3750293	20539092	3.69	1
3750293	20539089	4.92	2
3750293	20539088	4.93	3
3750293	20539091	6.01	4
3750293	20539090	6.26	5
3750296	20539106	1.49	1
3750296	20539107	3.72	2
3750296	20539108	17.98	3
3750296	20539109	222.33	4
3750296	20539110	473.95	5
3750299	20539121	3.72	1
3750299	20539127	4.34	2
3750299	20539122	5.32	3
3750299	20539123	7.10	4
3750299	20539124	12.28	5
3750299	20539125	20.00	6
3750299	20539126	34.57	7
…	…	…	…

Thus, in each race the horse 1 will be the favourite horse, horse 2 the second favourite and so on[87]. In this way, it is possible to aggregate the horses of difference race in eight different group, one for each representative horse (e.g. Horse 1 will represent, and therefore contain all the data of, every favourite horse of every race)

[86] The time remaining is computed as the race start minus the time the orders was matched and it was transformed in seconds.

[87] We will consider just the first eight horses because to total amount betted on these horses is around 97% for both class 1 and 2.

2. In the second step, I have divided the input matrix (the matrix with the horse identity, price, and time remaining for every matched orders) in eight matrixes: one for each representative horse[88].
   This is a fundamental step because it allows us to study the standard deviation for each group. Once obtained one matrix for each representative horse, I have worked separately on each of them:

   a. For each matrix, I have re-arranged the data in order to have, every three columns, the horse id, the time remaining and the price for every race[89] (see table 22);

Table 24 Data re-arranged for every race

Race 1			Race 2			Race 3			…
Horse Id	Time[90]	Price	Horse Id	Time	Price	Horse Id	Time	Price	…
20868320	77,370	2.19	20971400	59,959	1.4	20986562	71,275	1.5	…
20868320	77,370	2.2	20971400	59,944	1.36	20986562	67,591	1.5	…
20868320	77,366	2.2	20971400	54,280	1.4	20986562	60,675	1.46	…
20868320	77,266	2.18	20971400	53,478	1.4	20986562	26,302	1.43	…
20868320	76,285	2.16	20971400	48,314	1.4	20986562	20,520	1.42	…
20868320	76,148	2.16	20971400	21,806	1.38	20986562	19,461	1.48	…
20868320	74,303	2.14	20971400	20,898	1.36	20986562	19,128	1.43	…
20868320	74,243	2.14	20971400	20,894	1.36	20986562	19,127	1.42	…
20868320	74,063	2.14	20971400	20,892	1.36	20986562	19,127	1.42	…
20868320	74,003	2.14	20971400	20,783	1.35	20986562	19,123	1.42	…
20868320	71,889	2.11	20971400	20,756	1.35	20986562	19,101	1.41	…
20868320	70,597	2.14	20971400	20,039	1.38	20986562	19,096	1.41	…
20868320	70,314	2.14	20971400	18,393	1.39	20986562	18,879	1.41	…
20868320	70,314	2.14	20971400	16,645	1.36	20986562	17,275	1.46	…
20868320	70,176	2.11	20971400	16,134	1.36	20986562	15,273	1.42	…
…	…	…	…	…	…	…	…	…	…

   b. The next step has been done to obtain a unique order timespan. To do so, I have created a new matrix in which the first column is our orders timetable (it starts from the last 3000 seconds, and it decreases with a step equal to 1 second). For the other columns I have taken the prices of the horses at the same time of the timespan[91] (see table 23);

[88] As it has been done for the study of the average bet's size.
[89] Remember: we have the data of the same type of horse (e.g. the favourite one) for all the 71 races of class 1 or 203 races of class 2.
[90] Time is express in second, price in £.
[91] If at time t+1 we do not have any price it means that the matched prices is the same of time t. Thus, we take again the price at time t.

Table 25 Standardize timespan

Time	Race 1's price	Race 2's price	Race 3's price	Race 4's price	Race 5's price	Race 6's price	…
3000	2.34	1.32	1.32	4.2	5.4	4.5	…
2999	2.34	1.32	1.32	4.2	5.4	4.5	…
2998	2.34	1.32	1.32	4.2	5.4	4.5	…
2997	2.34	1.32	1.32	4.2	5.4	4.5	…
2996	2.34	1.32	1.32	4.2	5.4	4.5	…
2995	2.34	1.32	1.32	4.2	5.4	4.5	…
2994	2.34	1.32	1.32	4.2	5.4	4.5	…
2993	2.34	1.32	1.32	4.2	5.4	4.5	…
2992	2.34	1.32	1.32	4.2	5.4	4.5	…
2991	2.34	1.32	1.32	4.2	5.4	4.5	…
2990	2.34	1.32	1.32	4.2	5.4	4.5	…
2989	2.34	1.32	1.32	4.2	5.4	4.5	…
2988	2.33	1.32	1.32	4.2	5.4	4.5	…
2987	2.33	1.32	1.32	4.2	5.4	4.5	…
2986	2.33	1.32	1.32	4.2	5.4	4.5	…
2985	2.33	1.32	1.32	4.2	5.4	4.5	…
2984	2.33	1.32	1.32	4.2	5.4	4.5	…
2983	2.33	1.32	1.32	4.2	5.4	4.5	…
…	…	…	…	…	…	…	…

c. For each horse, I have taken the price differences: $Price_t$ - $Price_{t-1}$. The result has been standardized by tick. The standardization is important in order to have comparable values (see table 24);

Table 26 Price changes standardized by tick

Time	Race 1	Race 2	Race 3	Race 4	Race 5	Race 6	…
…	…	…	…	…	…	…	…
338	-2	0	-2	0	0	0	…
337	0	0	1	-1	0	-1	…
336	0	0	0	0	0	0	…
335	0	0	0	0	1	-1	…
334	0	-1	0	0	0	0	…
333	2	0	0	0	0	1	…
332	0	1	-1	0	-1	0	…
331	0	0	0	0	1	-1	…
330	0	0	1	0	0	0	…
329	0	0	0	0	0	0	…
328	1	0	0	0	0	-1	…
327	0	0	0	0	0	1	…
326	1	0	0	0	0	0	…
325	0	-2	0	0	0	0	…
324	-1	0	0	0	-1	0	…
…	…	…	…	…	…	…	…

Therefore, a value equal to 1 means that the price was one tick higher than the price on a second before. A value equal to -2, instead, means that the price has decreased by 2 tick in comparison of the previously price.

d. The matrix obtained contains the timetable and the price's difference in tick for each horse of the same group (e.g. horse 1). I have computed then the standard deviation of the ticks for each race for every timespan needed for the analysis (e.g. last 50 minutes, last 25 minutes, and so on) and sorted from the smaller to the bigger one's for each timespan (see table 25).

Table 27 Volatility for each races

Time Range (minutes)	50-0	25-0	10-0	5-0
Race 1	0.20	0.27	0.36	0.32
Race 2	0.22	0.28	0.39	0.41
Race 3	0.24	0.31	0.39	0.42
Race 4	0.24	0.34	0.45	0.45
Race 5	0.27	0.36	0.45	0.48
Race 6	0.27	0.36	0.45	0.48
Race 7	0.27	0.37	0.46	0.49
Race 8	0.28	0.37	0.46	0.49
Race 9	0.28	0.37	0.46	0.49
Race 10	0.29	0.38	0.47	0.52
Race 11	0.29	0.38	0.47	0.53
Race 12	0.29	0.38	0.48	0.53
Race 13	0.29	0.39	0.48	0.54
Race 14	0.30	0.39	0.49	0.55
Race 15	0.30	0.40	0.49	0.55
...	..	..	..	..

e. I have computed the 5%-95% percentile and the mean value for each time range;
f. I have repeated these steps for each representative horse in order to obtain the table 14.

References

Amihud Y., Mendelson H "DEALERSHIP MARKET. Market-making with Invetory", Journal of Financial Economics 8 (1980), 31-53

Avellaneda, M., Reed, J. & Stoikov, S., 2011. Forecasting Prices from Level-I Quotes in the Presence of Hidden Liquidity. Algorithmic Finance, 29 June, 1(1)

Betfair Annual Report, 2014

Brown A., Yang F. "Have Betting Exchange Corrupted Horse Racing?", (2014)

Copeland T., Galai D. "Information Effects on the Bid-ask Spread", The Journal of Finance, Vol.XXXVIIII, no.5 Decembr 1983

Croxson, K., & Reade, J. "Exchange vs. dealers: A high-frequency analysis of in-play betting prices". *Working Paper, University of Oxford*. (2011)

Flepp R., Nüesch S., Franck E. "The liquidity Advantage of Quote-driven Markets: Evidence from the Betting Industry". University of Zurich, (2013)

Franck E., Verbeek E., Nüesch S "Inter-market Arbitrage in Sports Betting", NCER Working Paper Series, Sportometrics Pogram, (2009)

Gambling Commission- Industry Statistics June 2015

Garman M. "MARKET MICROSTRUCTURE" Journal of Financial Economics 3 (1976) 257-275

Glosten L., Milgrom P. "BID, ASK AND TRANSACTION PRICES INA SPECIALIST MARKET WITH HETEROGENEOUSLY INFORMED TRADERS", Journal of Financial Economics 14(1985),71-100,

Ho T., Stoll H. "OPTIMAL DEALER PRICING UNDER TRANSACTIONS AND RETURN UNCERTAINTY" Journal of Financial Economics 9 (1981) 47-73

Luckner S., Schröder J., Slamka C., Franke M. Geyer-Schulz A.,Skiera Martin Spann B., Weinhardt C. " Prediction Markets. Fundamentals,Designs, and Applications", Gabler, 2012

Luckner S., Schröder J., Slamka C., Franke M. Geyer-Schulz A.,Skiera Martin Spann B., Weinhardt C. " Prediction Markets. Fundamentals,Designs, and Applications", Gabler, 2012

Michael A. Smith & David Paton & Leighton Vaughan Williams, 2006. "Market Efficiency in Person-to-Person Betting," Economica, London School of Economics and Political Science, vol. 73(292), pages 673-689, November.

Ozgit A. "The Bookie Puzzle: Auction versus Dealer Markets in Online Sports Betting", 2005

Smith M., Paton D., Williams L.V., "Market Efficiency in Person-to-Person Betting", Economica (2006) 73, 673-689

Smith, M., Paton, D., & Vaughan Williams, L. (2009). Do bookmakers possess superior skills to bettors in predicting outcomes? *Journal of Economic Behaviour & Organization, 71* , 539-549.

Stoll H, "THE SUPPLY OF DEALER SERVICES IN SECUIRITIES MARKETS" The Journal of Finance VOL. XXXIII, NO 4, September 1978

Sites:

http://www.betdaq.com/GBE.Help/BrokerDirectory/BETDAQ/en/Flash_Help/help_centre.htm#about_us.htm