Long-range correlations in an online betting exchange for a football tournament

Stephen J Hardiman¹, Peter Richmond and Stefan Hutzler
School of Physics, Trinity College Dublin, Dublin, Ireland
E-mail: hardimas@tcd.ie

New Journal of Physics 12 (2010) 105001 (14pp)
Received 18 December 2009
Published 7 October 2010
Online at http://www.njp.org/
doi:10.1088/1367-2630/12/10/105001

Abstract. We analyze the changes in the market odds of football matches in an online betting exchange, Betfair.com. We identify the statistical differences between the returns that occur when the game play is under way, which we argue are driven by match events, and the returns that occur during half-time, which we ascribe to a trader-driven noise. Furthermore, using detrended fluctuation analysis we identify anti-persistence (Hurst exponent $H < 0.5$) in odds returns and long memory ($H > 0.5$) in the volatilities, which we attribute to the trader-driven noise. The time series of trading volume are found to be short-memory processes.

Contents

1. Introduction 2
2. Long-range correlations in financial markets 3
3. Data analyzed 4
4. In-play versus half-time returns 4
5. Testing for long-range correlations in volatilities and trading volume 8
   5.1. Detrended fluctuation analysis (DFA) 8
   5.2. Lo’s modified rescaled range test 11
6. Summary and conclusions 13
Acknowledgments 13
References 13

¹ Author to whom any correspondence should be addressed.
1. Introduction

Traditionally, if one wanted to make a bet on a certain outcome (such as the outcome of a football game), the bet would be placed with a bookmaker or with a parimutuel betting market. The bookmaker is an independent agent who decides the odds that he/she feels are appropriate for the expected probability of the outcome and the customer can place bets against those odds. Bookmakers can expect to make profits in the long run by providing odds to their customers which slightly overestimate the true probability of a specific outcome. A parimutuel betting market is a system in which all bets on the outcome of a particular event are placed together in a pool, and the payoff odds are calculated by sharing the pool among all winning bets.

A betting exchange, like Betfair, is different from both of these systems and operates in a manner more akin to a stock exchange. Here, there is no single agent acting as a bookmaker, but agents can play both the roles of the customer and the bookmaker by choosing to either place a bet on an outcome (‘back’) or offer a bet (‘lay’) that can be backed by another agent. An agent who bets $1 on an outcome (a ‘‘back’’ bet) at 3 : 1 odds can expect to get $4 back if they win ($3 profit), but lose their $1 stake in the event of a loss. The converse of backing is to lay a bet. In this case, the agent plays the role of the bookmaker by taking up the bet of another agent. In the same example, the layer must pay out $4 dollars to the successful bettor in the event of a winning outcome, but retains the $1 stake in the event of a loss. Laying is thus equivalent to betting against an outcome.

The match-making is managed with a double auction order book as with a regular financial market, except that the bid and ask columns are replaced by back and lay. It is important to note, however, that unlike a stock market, a betting market has a definite conclusion. The odds will inevitably move towards infinity or zero as the outcome becomes a certainty.

The data used in our study are extracted from the online betting website, Betfair.com, based in London, United Kingdom. Since Betfair was launched in June 2000, it has become the largest online betting company in the UK and the largest betting exchange in the world. The website claims over two million clients from all over the world.

Betfair acts as a kind of prediction market [1, 2]. Prediction markets are speculative markets where assets traded have a value that is tied to a particular event (e.g. will the next US president be a Republican?). Previous studies of prediction markets have shown that the market price of a contract can provide a good reflection of the true odds of that event taking place, or at least the mean market belief [1], [3]–[7]. One of the oldest and most famous prediction markets, the Iowa Political Stock Market, has been shown to beat opinion polls at forecasting the outcomes of presidential elections [7]; however, differing viewpoints [8] do exist in the literature. Furthermore, there is evidence to suggest that parimutuel betting markets have the capacity to aggregate widely dispersed information held by the agents [9], and this may also be the case with the Betfair betting exchange.

Some Betfair traders may be speculative traders, hoping to back high and lay low. While others may simply bet on their favorite team, regardless of the quoted odds. Odds fluctuations may or may not be a reflection of changing probabilities of the event taking place.

A Betfair market is a complex system with many heterogeneous interacting agents, which in many ways operates in a manner akin to a regular stock exchange. We may therefore expect to see similar behavior in the dynamics of its market observables. To this end, we aimed to

2  http://www.betfair.com/
investigate the presence of long-range temporal correlations in Betfair, a phenomenon that has been observed in financial markets and that has also recently been observed in a political prediction market [10]. What we can learn about the dynamics of a betting market may provide an insight into the dynamics of a financial market and vice versa.

2. Long-range correlations in financial markets

To investigate the independence of changes of the price, $S(t)$, of a financial asset in time, one typically calculates the autocorrelation function, $\gamma(\tau)$, for the log-returns, $R(t)$, which, if assumed weakly stationary with zero mean, is given by

$$\gamma(\tau) = \frac{\langle R(t)R(t+\tau) \rangle - \langle R(t) \rangle^2}{\sigma^2},$$

(1)

where

$$R(t) = \ln(S(t+\tau)) - \ln(S(t))$$

(2)

and the variance, $\sigma^2$, is given by

$$\sigma^2 = \langle R(t)^2 \rangle - \langle R(t) \rangle^2.$$  

(3)

Here, the logarithm of the price is used to account for the changing scale of prices [11]. This autocorrelation function for financial returns is usually characterized by an exponential function with a decay time of the order of minutes [12]. A slowly decaying autocorrelation function would allow traders to predict the direction of future price movements with confidence and to buy and sell within the time window of this correlation time to ensure profits. But the rapidly decaying autocorrelation observed would seem to confirm the efficient market hypothesis, which states that a market is ‘efficient’, meaning that all available information about an asset is instantly processed when it reaches the market and is immediately reflected in the price.

However, despite this apparent efficiency of financial markets, we see strong correlations in the volatility of price movements and this is already clear from a visual inspection of a price return time series. Small price changes tend to be followed by small price changes, and large price changes tend to be followed by large price changes. In fact, the volatilities of financial returns have been reported to be long-memory processes [13].

A time series, $x_i$, with autocorrelation function $\gamma(\tau)$ is said to exhibit long-memory if, in the limit $\tau \to \infty$, the following scaling relation is valid:

$$\gamma(\tau) \sim \tau^{-\alpha}L(\tau).$$

(4)

Here, $0 < \alpha < 1$ and $L(\tau)$ is a slowly varying function at infinity. In such a process, values from the distant past can have a significant effect on the present. Furthermore, the resulting non-integrability of the autocorrelation function implies anomalous diffusion. For normal diffusion, the standard deviation of running sums $\sigma(t) = \sqrt{\langle \sum_{i}^{t} x_i \rangle^2}$ asymptotically increases as $t^{0.5}$, whereas for diffusion processes with long-memory increments, the standard deviation asymptotically increases as $t^H$ with $H$ in the interval $(0.5, 1)$. $H$ is called the Hurst exponent and is related to the exponent $\alpha$ in the autocorrelation function by $H = 1 - \frac{\alpha}{2}$ [13].
The Hurst exponent was originally developed in hydrology for the practical purpose of determining optimal dam size for the Nile river’s volatile rain and drought conditions. Hurst observed the seasonal flooding of the Nile to be a long-memory process. Long-memory processes are typically modeled with fractional Brownian motion [14] or ARFIMA models [15].

Besides in the volatility of financial price change, long memory has also been observed in the time series of trading volumes associated with traded assets [16]. Furthermore, the time series of the order signs of trades on financial markets have been shown to be correlated with long memory [13].

If, however, the standard deviation of running sums in a time series increases as $t^H$ with $0 < H < 0.5$, then such a process is called a mean-reverting or anti-persistent process and possesses anti-correlations at all scales. Such anti-persistent auto-correlations have been observed in the returns of spot prices in commodity markets [17, 18].

3. Data analyzed

The dataset used in our analysis comprises snapshots of the Betfair.com order book for the activity of betting markets for the outcomes of 146 matches that took place during the 2008 Champions League football tournament. For each match, there is a market associated with each potential outcome: win, lose or draw. These data were collected using front-end software provided by Fracsoft\(^3\) which interfaces with Betfair.com. The data are resolved to one second accuracy. Also included is a value for the ‘odds’ at which the last bet for each market was matched during that second and the value of bets matched to date for that outcome. It is important to note that the odds with which Betfair users choose to back or lay outcomes can only take on values from a specified set of numbers between 1.01 and 1000 imposed by the Betfair user interface. If a bet is matched at Betfair ‘odds’ of 3.0, the backer will triple his/her money in the event of a payout; such a payout would be more traditionally represented in gambling parlance as ‘2-to-1’.

We take the reciprocal of the Betfair ‘odds’ to give a value that reflects the probability of that particular outcome. We consider the change in this probability with respect to time to be a stochastic variable analogous to financial price returns and call it the \textit{implied probability return}. Furthermore, by subtracting the current from the previous value of total bets matched on the market, we deduce the 1 s trading volume, which is the total monetary value of bets matched on the market during any given second.

The market remains open for Betfair users to trade bets up until the very end of a match, when the conclusion is known and the market odds naturally move towards either 1.01 or 1000. A typical dataset is composed of about 6500 records representing the state of the betting market at each second during the length of the match, including approximately 15 min of half-time and typically 3 min of stoppage time.

4. In-play versus half-time returns

In figure 1, we show the evolution of the implied probability for the three possible outcomes (win, lose or draw) for an example market in our dataset. As the match moves towards its conclusion, we see the implied probability move towards 1.0 in the case of Manchester United,

http://www.fracsoft.com/
Figure 1. (a) The evolution of the implied probability (reciprocal of match odds) for the three possible outcomes of a Champions League football match. As the match draws toward its conclusion, the implied probabilities for the outcomes tend toward 0 and 1 as the market odds tend toward 1000 and 1.01. We see a large probability change for each outcome at the $t \approx 1400$ s mark corresponding to a goal scored by Manchester United (Man Utd.). We see little variation in the implied probability during half-time. (b) The one-second return (change in the implied probability) as a function of time remaining.

but 0.0 in the case of Roma and the draw. As expected, it was Manchester United who won the match. We see a large change in the probability at roughly $t = 1400$ s remaining corresponding to a goal scored. We also see in figure 1(a) that since no match events are occurring, the implied probability remains approximately constant at half-time. In the analysis that we present from now on, graphs will be the result of averaging over all such markets in our dataset.

In this section, we aim to compare the statistics of Betfair implied probability returns with the model of [10] that describes the evolution of the price of a binary option contract in the Iowa Political Stock Market. A binary option is a contract that pays out $1 if an event occurs and $0 otherwise. Assuming an efficient market, the value of the contract should reflect the probability that the event will occur. The price of a contract in a binary option market is thus likely to behave similarly to the implied probability in a Betfair market.

In [10], the authors’ model changes in what they consider the ‘true’ price of a contract in a binary option market, by the following process:

$$p(t_{i-1}) = \begin{cases} p(t_i) + \frac{1 - p(t_i)}{(t_i/T_a)^\gamma} & \text{with probability } p(t_i), \\ p(t_i) - \frac{p(t_i)}{(t_i/T_a)^\gamma} & \text{with probability } 1 - p(t_i), \end{cases}$$

(5)

where $t_i$ is the time at which the $i$th last trade takes place, $p(t_i)$ is the price of the contract at the $i$th last trade, $T_a$ is the average time between consecutive trades and $\gamma$ is a free parameter that
describes how the magnitude of price changes scales with time remaining. By fitting the model to their empirical market data, they estimate $\gamma = 0.49$.

It can be shown that when $t_i$ is large and price changes are small (such that we can consider both $p(t_i)$ and $t_i$ to be constant), the average magnitude of partial sums, $\Delta p(t_i, n) = p(t_i - n) - p(t_i)$, has the following form:

$$\langle |\Delta p(t_i, n)| \rangle = \left( \frac{T_a}{t_i} \right)^{\gamma} \frac{2n}{\pi} \sqrt{\frac{1}{4} - (p(t_i) - 1/2)^2}. \quad (6)$$

For simplicity, we take $\gamma = 0.5$, which gives

$$\langle |\Delta p(t_i, n)| \rangle = \frac{\alpha}{\sqrt{t_i}} \sqrt{\frac{1}{4} - (p(t_i) - 1/2)^2}, \quad (7)$$

where $\alpha = \sqrt{\frac{2nT_a}{\pi}}$ is a constant.

In the following analysis, we make a distinction between the implied probability changes that are observed when the game play of the football match is under way, which we call in-play returns, and returns that occur at half-time. When we factor out the time-dependent component, by multiplying the empirical in-play 100 s Betfair returns by the square root of time remaining in the match, the average result is well modeled by the probability-dependent component of equation (7). This is shown in figure 2(a). Furthermore, when we factor out the probability-dependent component, by dividing the returns by $\sqrt{\frac{1}{4} - (p(t_i) - 1/2)^2}$, the result is well approximated by the inverse square root law of equation (7), as seen in figure 3. In-play

\[ \textbf{Figure 2.} \] (a) $\langle |\Delta p(t_i)\sqrt{t_i}| \rangle$ for in-play (circles) and half-time (triangles) implied probability returns as calculated for 100 and 1 s intervals. The average is taken over all non-overlapping returns in each market of the dataset. The broken line is the best fit of the in-play points to the functional form $f(p) \propto \sqrt{1/4 - (p_{imp} - 1/2)^2}$, which is the expected result for the binary option model of [10]. The model is a poor fit to the average half-time returns, which are significantly smaller. (b) When the in-play implied probability is sampled at 1 s frequency, the returns are influenced by a trader-driven noise, which increases with implied probability.
implied probability returns for Champions League football returns thus exhibit the ‘conditional diverging volatility’ reported in [10].

However, we see a very different result for $\langle |\Delta p_{\text{imp}}(t)\sqrt{1/4 - (p(t) - 1/2)^2}| \rangle$ in the case the 100 s half-time returns (figure 2(a)) and the model is also a poor fit in the case of the 1 s in-play returns. In particular, the average magnitude of 1 s in-play returns is seen to systematically increase with implied probability. This is shown in figure 2(b). We propose that whereas for low frequencies the model may capture the dynamics of the changing probability of the outcome, at very high frequencies in-play returns are influenced by a trader-driven noise.

The authors of [10] do include noise in their full model that amounts to multiplying the price of equation (5) by an error term $\exp \epsilon(t_i)$, where $\epsilon(t_i)$ is a Gaussian random variable with a constant variance fit to experiment. Our results suggest that such a noise term for Betfair market returns should have a variance that is a function of the implied probability. This assertion is supported by the results of figure 4 in which we see a strong dependence of trading volume on implied probability (figure 4(a)) and a square root law relating half-time trading volume to volatility (figure 4(b)).

At half-time, when no match events are occurring, we expect to see little change in the ‘true’ probability of the outcome, so both 1 and 100 s half-time returns can be attributed entirely to trader-driven noise. We therefore separate the analysis in the next section into both in-play and half-time. We expect in-play market observables to be affected by systematic trends related to the changing probability of the outcome of the match, but the half-time market observables will result from more stationary trader-driven noise. This is in agreement with the results of [22], in which we find very different probability distributions for the returns that occur at half-time and the returns that occur when the game play of the match is under way. In particular, the in-play returns are seen to exhibit fatter tails, indicating that large probability changes are much more likely to occur when the match is under way, due to significant
match events such as goals scored creating dramatic changes in the implied probabilities, as seen in figure 1.

5. Testing for long-range correlations in volatilities and trading volume

5.1. Detrended fluctuation analysis (DFA)

Due to their non-stationary nature, to investigate long-range correlations in Betfair market observables, following [10], we employ DFA. DFA is a well-established method for determining the scaling behavior of noisy data in the presence of trends without knowing their origin and shape and it works as follows [23].

Given a time series \( x_i \), we first integrate it to form \( X(t) = \sum_{i=1}^{t} x_i \). The integrated time series, \( X(t) \), is then divided into blocks of equal length \( m \) and a linear ordinary least squares regression is performed to the points in each block to produce the function \( X_m(t) \). We then detrend the integrated time series \( X(t) \), by subtracting the local trend, \( X_m(t) \), in each block and calculate the root mean square fluctuation,

\[
F(m) = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (X(t) - X_m(t))^2}. \tag{8}
\]

This procedure is repeated over a variety of box sizes, \( m \), to construct the function \( F(m) \). The Hurst exponent, \( H \), of the time series is then obtained by fitting the function \( F(m) \) to the form \( F(m) \propto m^H \). In the analysis that follows, \( F(m) \) is the result of averaging this procedure over all 438 markets in the study.
Figure 5. The results of DFA on the magnitude of implied probability returns during half-time (circles) and in-play (triangles). We see power-law scaling with a Hurst exponent of \( H \approx 0.61 \). For the in-play returns the scaling breaks down as the bin size approaches the length of the match. The dashed line indicates the slope corresponding to \( H = 0.5 \).

5.1.1. Volatility. In figure 5, we show the results of DFA on the magnitude of implied probability returns as drawn from half-time and from times in the match when the game play is under way. In both cases, we see scaling behavior characterized by a straight line on a log–log plot and estimate a Hurst exponent of approximately \( H \approx 0.61 \) from a least squares fit to the graph. This value differs somewhat from the value \( H \approx 0.66 \) and 0.71 for the volatility in the binary option prediction market of [10].

Since, in the case of the in-play returns, we expect the volatility to systematically increase during the match, the scaling breaks down for large box length \( m \) as we approach the length scale of the football match. We conclude that Betfair volatility is a long-memory process, and since the same behavior is also observed during half-time, when little news is hitting the market, the scaling results from the trader-driven noise.

5.1.2. Trading volume. In figure 6, we show the results of DFA on the 1 s trading volumes as drawn from half-time and in-play. In both cases, we see power-law scaling with an exponent \( H \approx 0.5 \) corresponding to short-term memory. Again, in the case of the in-play trading volumes, the scaling breaks down for large box length. This is because of the strong dependence of trading volume on the implied probability, which tends to zero or one at the end of every match. We conclude that unlike what has been observed in the case of financial markets [16], Betfair trading volumes are not a long-memory process.

5.1.3. Implied probability return. Finally, we use DFA to investigate the Hurst exponent of the implied probability returns themselves. The results of the analysis for half-time and in-play are shown in figure 7.

For the half-time returns, we observe power-law scaling with an exponent \( H = 0.21 \). This suggests that the implied probability at half-time is a mean-reverting (anti-persistent) process.
Figure 6. The results of DFA on trading volumes for Champions League Betfair markets during half-time (circles) and in-play (triangles). We see power-law scaling with a Hurst exponent of $H \approx 0.5$ corresponding to short-term memory. For the in-play returns the scaling breaks down as the bin size approaches the length of the match.

Figure 7. The results of DFA on the implied probability return during half-time (circles) and in-play (triangles). For half-time, we see power-law scaling with a Hurst exponent of $H \approx 0.21$ corresponding to anti-persistence. For the in-play returns, $F(m)$ has a very different shape. This is because during game play the implied probability follows the true probability, which fluctuates with match events; furthermore, the implied probability, always tends to zero or one at the end of each match.

Such mean-reverting behavior is to be expected, since if the ‘true’ probability of the outcome remains constant at half-time, then when market odds depart from what is believed to be the true probability, the actions of traders seeking arbitrage will serve to restore the ‘true’ odds.
A surprising result is the self-affinity of the signal; we observe a power-law scaling with $H = 0.21$, spanning at least one order of magnitude.

For $t < 25$ s, $F(m)$ scales with an exponent $H = 0.5$, indicating that over very short timescales the implied probability behaves more randomly. This may be due to information delays in the Betfair system or represent the reaction time with which traders can respond rationally to changes in the Betfair odds.

The plot of $F(m)$ for the in-play returns exhibits mixed behavior. Scaling with $H < 0.5$ is still detectable, but over longer timescales the implied probability return follows the true probability of events, which behaves more like a random walk ($H \approx 0.5$). For very long timescales, the DFA is influenced by the systematic trend in the odds to move towards zero or one at the end of every match.

Such anti-correlations are atypical for financial markets but have been observed in electricity spot prices [17, 18]. However, it should be noted that the implied probability is the reciprocal of the odds at which the last bet was matched. A change in this value does not necessarily mean a change in the market’s best back or lay price. Therefore, the anti-correlation present in the half-time signal cannot necessarily be exploited to beat the market.

5.2. Lo’s modified rescaled range test

Since we expect the implied probability to remain constant during half-time, we expect implied probability returns and trading volumes occurring during half-time to be stationary. We can therefore supplement the DFA with Lo’s modified rescaled range test for long memory in the case of the half-time market data. Lo’s test [24] is a modification of the classical rescaled range test of Mandelbrot [25], which compares the maximum and minimum values of running sums of deviations from the sample mean renormalized by the sample standard deviation. The classical rescaled range test has been shown to be too weak, and due to the influence of short-range correlation it can indicate long memory when it does not exist [24].

Given a sample time series $x_1, x_2, \ldots, x_n$ with sample mean $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$, Lo’s modified $R/S$ statistic is

$$Q_n(q) = \frac{1}{\hat{\sigma}_n(q)} \left[ \frac{\max_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{x}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{x}_n)}{\sum_{j=1}^{q} w_j(q) \hat{\gamma}_j} \right],$$

where

$$\hat{\sigma}_n^2(q) \equiv \hat{\sigma}_x^2 + 2 \sum_{j=1}^{q} w_j(q) \hat{\gamma}_j, \quad w_j(q) \equiv 1 - \frac{j}{q + 1},$$

where $\hat{\sigma}_x^2$ is the sample variance, $\hat{\gamma}_j$ is the sample autocovariance and $q < n$. $Q_n(q)$ differs from the classical $R/S$ statistic of Mandelbrot only in the denominator. Following Lo, we use $q = [k_n]$, where

$$k_n \equiv \left( \frac{3n}{2} \right)^{1/3} \left( \frac{2\hat{p}}{1 - \hat{p}^2} \right)^{2/3} \left( \frac{2n}{1 - \hat{p}^2} \right)^{1/3},$$

with $[k_n]$ being the greatest integer less than or equal to $k_n$ and $\hat{p}$ is the sample first-order autocorrelation coefficient. Provided certain criteria are met [24], $V_n \equiv Q_n/\sqrt{n}$ tends asymptotically to a random variable distributed according to the probability distribution for
Figure 8. Frequency distribution for the value of $V_n = Q_n/n$ of equation (9) for the 438 data time series of trading volume returns (left) and absolute implied probability (right) extracted from the 146 Champions League football matches. The line represents the analytical solution for the distribution of $V_n$ in the case of short-term memory. Also shown is the 95% confidence range for rejecting the null hypothesis of short-range correlations. For trading volumes, the distribution of points compares well with the expectation of a short-memory model. However, in the case of the absolute probability returns, 14% of points lie outside the 95% confidence interval, suggesting that Betfair volatility is a long-memory process.

the range of a Brownian bridge. When $V_n$ is outside the interval $[0.809, 1.862]$, we can reject the null hypothesis of short-range dependence with 95% confidence.

In figure 8, we show the distribution of $V_n$ as calculated for the half-time time series of absolute implied probability returns and trading volumes for all 438 markets studied. In the case of the trading volumes, we reject the null hypothesis of short-range dependence with 95% confidence in 7% of the datasets and we find that data fits well to the analytical expectation for the probability distribution for the range of a Brownian bridge.

$$ P(V_n) = 8 \sum_{n} V_n k^2(4k^2 V_n^2 - 3) e^{-2k^2 V_n^2}. $$ (12)

However, we find larger values of $V_n$ in the case of the absolute returns datasets and their distribution differs much from the analytical solution for the Brownian bridge. In particular, we find that 14% of the volatility results fall outside the 95% confidence interval for short-term memory. 14% may seem small, but it is important to note that Lo’s modified rescaled range test is a very strong test and is heavily biased to accept the null hypothesis of short-range dependence [26].

A Brownian bridge is a continuous-time stochastic process $B(t)$ whose probability distribution is the conditional probability distribution of a Wiener process $W(t)$ (a mathematical model of Brownian motion) given the condition that $B(0) = B(1) = 0$. 

New Journal of Physics 12 (2010) 105001 (http://www.njp.org/)
These results appear to support the hypothesis of long-range dependence in implied probability volatility but short-range dependence in trading volume, as indicated by the DFA analysis in the previous section.

6. Summary and conclusions

We have investigated the temporal correlations in football market data extracted from an online betting exchange, Betfair.com. Unlike financial markets, in which traders buy and sell financial assets, on Betfair it is bets that are instead ‘backed’ or ‘laid’. Despite this significant difference between Betfair and financial markets, we find similar long-range correlations in the market returns.

We show that when sampled at low frequencies, the magnitude of changes in the implied probability can be qualitatively described by the model of [10]. However, when the implied probability is sampled at high frequencies, we see the influence of a trader-driven noise. Furthermore, the magnitude of the trader noise is found to be a function of the implied probability.

Having separated the high-frequency trader noise that dominates the half-time activity from the mixed market behavior of the in-play returns, we perform standard tests for long-range correlations on the in-play and half-time returns and trading volumes. We find that the magnitude of the returns exhibits long-range correlations with a Hurst exponent $H = 0.61$. We find, however, that unlike what has been observed in the case of financial markets, the Betfair trading volumes are a short-memory process with a Hurst exponent $H \approx 0.5$. Finally, detrended fluctuation analysis of the implied probability return reveals that the trader-driven noise gives rise to a self-affine mean-reversion of the implied probability at half-time.

Acknowledgments

SJH is supported by the Irish Research Council for Science, Engineering and Technology funded by the National Development Plan, and PR acknowledges support from the Trinity College Dublin Visiting Professorships and Fellowships Benefaction Fund 2008/2009. Further support was obtained from the European Science Foundation Cost Actions P10 (‘Physics of Risk’) and MP0801 (‘Physics of Competition and Conflicts’). We thank Betfair and Fracsoft for providing us with their data and Steven Tobin for some of the data rendering.

References

[1] Forsythe R, Nelson F, Neumann G R and Wright J 1992 Am. Econ. Rev. 82 1142–61
[2] Wolfers J and Zitzewitz E 2004 J. Econ. Perspect. 18 107
[3] Wolfers J and Zitzewitz E 2006 Discussion Papers 2092 Institute for the Study of Labor
[4] Gjerstad S 2004 Microeconomics (0411002)
[5] Berg J E and Rietz T A 2006 Information Markets: A New Way of Making Decisions (Washington, DC: AEI Press)
[6] Berg J E, Forsythe R, Nelson F D and Rietz T A 2008 The Handbook of Experimental Economic Results vol 1 (Amsterdam: Elsevier)
[7] Berg J E and Rietz T A 2008 Int. J. Forecast. 24 283–98
[8] Erikson R S and Wlezien C 2008 Public Opin. Q. 72 190–215

New Journal of Physics 12 (2010) 105001 (http://www.njp.org/)
[9] Plott C R, Wit J and Yang W C 2003 *Econ. Theory* **22** 311–51
[10] Majumder S R, Diermeier D, Rietz T A and Amaral L A N 2009 *Proc. Natl Acad. Sci. USA* **106** 679–84
[11] Mantegna R N and Stanley H E 2000 *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge: Cambridge University Press)
[12] Gopikrishnan P, Plerou V, Amaral L A N, Meyer M and Stanley H E 1999 *Phys. Rev. E* **60** 1390–1400
[13] Lillo F and Farmer J D 2004 *Stud. Nonlinear Dyn. Econ.* **8** 1
[14] Mandelbrot B B and Van Ness J W 1968 SIAM Rev. **10** 422–37
[15] Granger C W J and Joyeux R 1980 *J. Time Series Anal.* **1** 15–29
[16] Lobato I N and Velasco C 2000 *J. Bus. Econ. Stat.* **18** 410
[17] Weran R and Przybylowicz B 2000 *Physica A* **283** 462–8
[18] Norouzzadeh P, Dullaert W and Rahmani B 2007 *Physica A* **380** 333–42
[19] Lillo F, Farmer J D and Mantegna R N 2003 *Nature* **421** 129
[20] Gabaix X, Gopikrishnan P, Plerou V and Stanley H E 2003 *Nature* **423** 267–70
[21] Farmer J D and Lillo F 2004 *Quant. Finance* **4** C7–11
[22] Hardiman S, Tobin S T, Richmond P and Hutzler S 2010 *Dyn. Socioeco. Syst.* at press
[23] Peng C K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 *Phys. Rev. E* **49** 1685
[24] Lo A W 1991 *Econometrica* **59** 1279
[25] Mandelbrot B B and Wallis J R 1969 *Water Resour. Res.* **5** 967
[26] Teverovsky V, Taqqu M S and Willinger W 1999 *J. Stat. Plan. Inference* **80** 211–27