Physica A 392 (2013) 6169–6188

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Using conditional probability to identify trends in intra-day high-frequency equity pricing

Michael Rechenthin *, W. Nick Street

The University of Iowa, Tippie College of Business, Iowa City, IA 52242-1000, United States

ARTICLE INFO

Article history:
Received 18 November 2011
Received in revised form 25 March 2013
Available online 21 August 2013

Keywords:
Conditional probability
Stock prediction
Intra-day trading
High-frequency trading

ABSTRACT

By examining the conditional probabilities of price movements in a popular US stock over different high-frequency intra-day timespans, varying levels of trend predictability are identified. This study demonstrates the existence of predictable short-term trends in the market; understanding the probability of price movement can be useful to high-frequency traders. Price movement was examined in trade-by-trade (tick) data along with temporal timespans between 1 s to 30 min for 52 one-week periods for one highly-traded stock. We hypothesize that much of the initial predictability of trade-by-trade (tick) data is due to traditional market dynamics, or the bouncing of the price between the stock's bid and ask. Only after timespans of between 5 to 10 s does this cease to explain the predictability; after this timespan, two consecutive movements in the same direction occur with higher probability than that of movements in the opposite direction. This pattern holds up to a one-minute interval, after which the strength of the pattern weakens.

1. Introduction

In this paper we examine the predictability of market directional movements using high-frequency data in a popularly-traded stock by examining the prior movements. If the upward and downward price movements are systematic, rather than random in nature, this knowledge should help traders better predict the direction of the stock and make more informed trading decisions.

Mainstream finance theory has traditionally held the view that financial prices are efficient and follow a random walk. The definition of a random walk is a process where the changes from one time period to the next are independent of one another, and are identically distributed. One of the earliest studies of market efficiency was done by Fama in 1965 to describe how equity prices at any point in time best represent the actual intrinsic value, with the prices updating instantaneously to information [1]. Efficiency is associated with a trendless and unpredictable financial market.

According to the theory of random walks and market efficiency, the future direction of a stock is no more predictable than the path of a series of cumulative random numbers [1]. Statistically it can be said that each successive price is independent of the past; each series of price changes has no memory. If testing for market independence, the probability of market directional-movement at time t is compared against time $t-1$. The same should hold as more prior information is added since, according to market efficiency, the past cannot be used to predict the future.

The theory of random walks and efficiency of market prices was expanded by Fama in Ref. [1] to the Efficient Market Hypothesis (EMH) in the 1960's. The theory states that the current market's price is the correct one, and any past information is already reflected in the price. According to the EMH, although no market participant is *all knowing*, collectively they know as much as can be known; for as a group, they *are* the market. These individuals are constantly updating their beliefs about

* Corresponding author. Tel.: +1 773 209 5449.
E-mail address: mrechenthin@gmail.com (M. Rechenthin).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.08.003

the direction of the market, and although they will disagree on the direction of the stock, this will lead, as noted by Fama, to "a discrepancy between the actual price and the intrinsic price, with the competing market participants causing the stock to wander randomly around its intrinsic value [1]". If markets are indeed efficient, then it implies that markets never overreact or underreact during the trading day. Any effort that an average investor dedicates to analyzing and trading securities is wasted, since one cannot hope to consistently beat the market. Any attempt to predict future prices is futile and although high rates of return may be achieved, they are on average, proportional to risk. In addition, high risk may achieve high rates of return, but it also can deliver high rates of loss. For example, when flipping a fair coin, a roughly 50% chance of getting heads would be expected; however, expecting heads ten times consecutively would come with high risk. The concept of an efficient market implies that consistently predicting the market carries a high risk.

For those who believe that markets are predictable, there are two main schools of thought: the "technician" and the "fundamentalist". Fundamental analysts look at the external and economic factors to determine price change. The belief is that since stocks are shares of a corporation, examining the fundamental indicators such as profits, sales, debt levels, and dividends should provide an outlook into the future direction of the price. Technical analysts believe that the past performance of equities can help forecast future price movements. They study historical prices to try and understand the psychology of other market participants (the crowds). The technician attempts to identify regularities in the time-series of price or volume information; the thought is that price patterns move in trends, and that these patterns often repeat themselves [2,3]. There is a skew toward technical analysis over fundamental analysis when considering shorter (intra-day) time horizons. This paper examines the technical approach.

Trading filters, as first used by Alexander [2], attempted to show predictability, and therefore the existence of trends, by using quantitative rules based on prior price history to create profits by buying and selling. If markets are random, zero profits would be expected over a baseline amount; however, if a model can be introduced that shows apparent profitability, then this opens the possibility of markets that occasionally trend. According to Granger and Timmermann [4], the existence of a single successful trading model would be sufficient to demonstrate a violation of the market efficiency hypothesis. A number of empirical studies using daily data, such as Neely et al. [5], Osler and Chang [6], Levich and Thomas [7], and Sweeney [8] found profitability of trading rules in excess of the risks taken. The consensus of these papers is that the market is predictable, by way of trading rule profitability, at least part of the time.

Timmermann [9] however, found forecasting models that use daily and longer interval data to predict stock returns mostly performed poorly. He did find some evidence of short-lived instances of predictability, thus requiring the examination of intra-day trading data. The theory is that if there are more instances of a particular high-probability pattern during a timespan, they will more likely be spotted by other traders and implemented in their trading strategy. This widespread adoption of a particular trading approach drives the asset price either up or down enough to eliminate the pattern. Furthermore, while it is common for professional traders to use intra-day data, this short time horizon is often under-represented in the academic literature.

Ohira et al. [10], Tanaka-Yamawaki [11], Sazuka [12], and Hashimoto et al. [13] examined market data at the lowest intra-day level available, trade-by-trade (sometimes known as tick data) and found extremely high levels of predictability. For example, in Refs. [10,11] the authors report predictability as high as 79.7% and 75.0% respectively. While the movements are clearly predictable and raise doubt as to the efficiency of the currency market, we theorize here that much of the predictability in those two papers can be explained by the noisy continuation[1] of the bid–ask market dynamics.[2] While the *bid–ask bounce* has been discussed in academic literature previously, we believe this is the first study of this size (dataset includes 15 billion in share volume) and level of detail (number of intervals examined) that examines when a stock escapes the confines of the bid and ask spread. To escape the noisy influence of bid–ask market dynamics, some researchers have sampled the market at even intervals such as 5, 10, and 60 min intervals. A paper by Reboredo et al. [16] found profitability over a benchmark for 5, 10, 30, and 60 min intervals of intra-day data using Markov switching, artificial neural networks and support vector machine regression models. Additionally Wang and Yang [17] found intra-day market inefficiency in the energy markets using 30 min intra-day prices.

Our research demonstrates however, in most cases, the market has gone back to efficiency after a one-minute timespan. We empirically examine the conditional probabilities of upward versus downward movements by using intra-day timespans of trade-by-trade (tick) data along with nine temporal timespans of 1, 3, 5, 10, 20, and 30 s and 1, 5, and 30 min for 52 separate one-week periods in 2005 of a popularly traded stock, the Standard and Poor's 500 index (symbol: SPY). By investigating the conditional probabilities, we find that the market escapes the confines of the bid–ask spread after a 5–10 s timespan. An additional contribution of this paper is the observation of trends with seemingly high predictability; trends that have high occurrences of continuing rather than going against the trend, unless the trend is broken.

In Section 2, we explore efficiency and conditional probability within high-frequency stock data and why market dynamics may explain some of the market predictability. Within this section we also explain how sampling methods have been used to eliminate the noise associated with the bid–ask spread. In Section 3, we describe our dataset and demonstrate our

1 "Continuation" is a term used by Ref. [14] and refers to the pattern where the signs of at least two non-zero consecutive changes are in the same direction. See also Section 3.3.

2 While the currency-spot market is different from the equity market, such as the absence of a reported last trade/transaction, market dynamics still apply [15]. The large number of participants and lack of a centralized reporting facility cause the bid and ask to fluctuate in the currency-spot market, similar to the last trade/transaction in the equity market bouncing between the bid and ask.

two experiments. The first experiment is a test of market independence for trade-by-trade along with 9 temporal timespans. The second experiment determines where the market escapes the confines of the bid and ask. Lastly we examine two trends with high levels of predictability and how these trends remain relatively stable over the course of the examined year.

2. Background

2.1. Efficiency and conditional probabilities

To demonstrate statistical efficiency (or inefficiency) within the equity market, conditional probabilities of upward versus downward market movements given prior price movement are examined. In this paper we focus on the binary representation of price movements, which can be written as $\Pr(\triangle p_t = \{\text{up}, \text{down}\} | \triangle p_{t-1} = \{\text{up}, \text{down}\}, \triangle p_{t-2} = \{\text{up}, \text{down}\}, \ldots, \triangle p_{t-m} = \{\text{up}, \text{down}\})$ where p is the price and $\triangle p_t = p_t - p_{t-1}$. Market independence would have us believe that $\Pr(\triangle p_t = \text{up} | \triangle p_{t-1} = \text{down})$ should equal $\Pr(\triangle p_t = \text{up})$. Upward movements are abbreviated as $(+)$ and downward movements as $(-)$; for example, the conditional probability of an upward movement given two previous downward movements is written as $\Pr(+|-, -)$.

To illustrate statistical efficiency, we offer a brief comparison between random and actual data. Fig. 1(a) shows the appearance of a downward trend, which appears simplistically predictable. However, this chart was created by randomly choosing, with equal probability, an upward or downward movement. Fig. 1(b) is actual 1-min intra-day data for the stock SPY over the period January 3 through December 30, 2005. The two charts appear remarkably the same in terms of the existence of trends and potential predictability. But when the 2^m conditional probabilities for memory depth m are computed for both datasets displayed in Fig. 1, the results are very different. For the random data there is, as expected, a roughly 0.50 conditional probability that the market will go up given prior information. However, for the actual intra-day 1-min data, a probability of an upward movement occurs with roughly 0.50 probability, but only 0.422 that the market will go up given a downward movement in price, $\Pr(+|-)$. With entirely independent data, this should not be true. Our intra-day analysis in Section 3 explores the conditional probabilities of all 52 weeks in 2005 separately and describes how these probabilities differ over varying timespans.

The first paper that we are aware of that used probabilities to examine the existence of market trends and thus market inefficiencies in high-frequency data was Niederhoffer et al. in 1966 [14]. The authors found the stock examined had a higher probability to reverse directions from the previous price change than to continue in the same direction. Much of the existence of predictability was explained by traditional market dynamics, or the bouncing of prices between the bid and ask. Niederhoffer et al. calls this "the natural consequence of the mechanics of trading on the stock exchange". An explanation of market dynamics follows.

2.2. Market dynamics

Within financial markets, there is a small region of price that brackets the underlying value of the asset which is called the bid–ask spread. The bid is the highest price an individual is willing to pay, and the ask is the lowest price an individual is willing to sell his or her stock at the moment. The "value" can be thought of as somewhere between the bid and ask (see Fig. 2). The bid and ask fluctuate depending on supply and demand—more demand sends the price up, while more supply sends the price down.

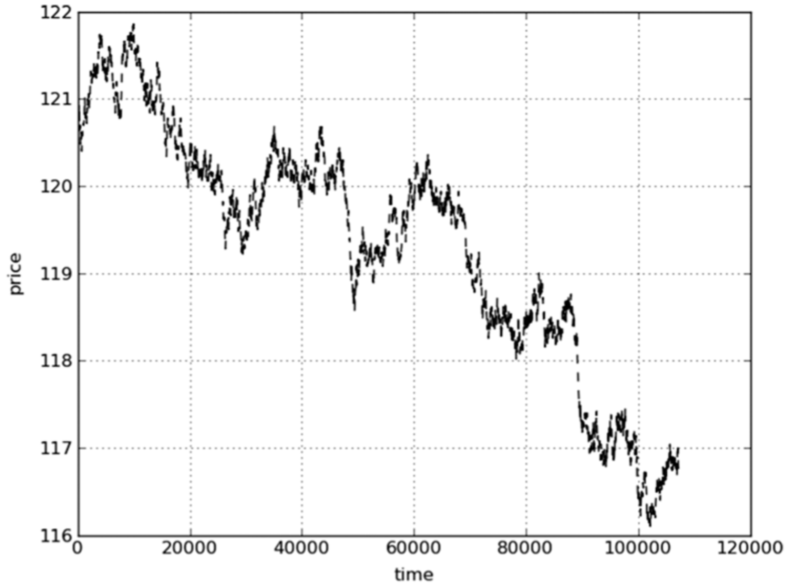
For example, let us consider that Participant 1 wants to buy 100 shares of stock at a price of $10.01; this is currently the best bid. Another participant, Participant 2, is attempting to sell his or her 100 shares at a price of $10.02; this is currently the best ask. A trade does not take place until Participant 1 either pays Participant 2's ask of $10.02, or Participant 2 lowers his ask to Participant 1's bid of $10.01. Of course in an actual market, there are often hundreds or even thousands of participants at any given time who can participate in transactions. In an efficient market, the bid and ask fluctuate randomly [18].

Furthermore, in widely traded stocks with multiple active participants, there may be thousands of shares available at a bid or ask at any given time. In the short run, these orders act as a barrier to continued price movement in either direction. The larger the number of orders, or participants, at a given price level, the longer the price will stay constrained within a small price bound. Only after the bid or ask is eliminated, will the stock move to another price point [14].

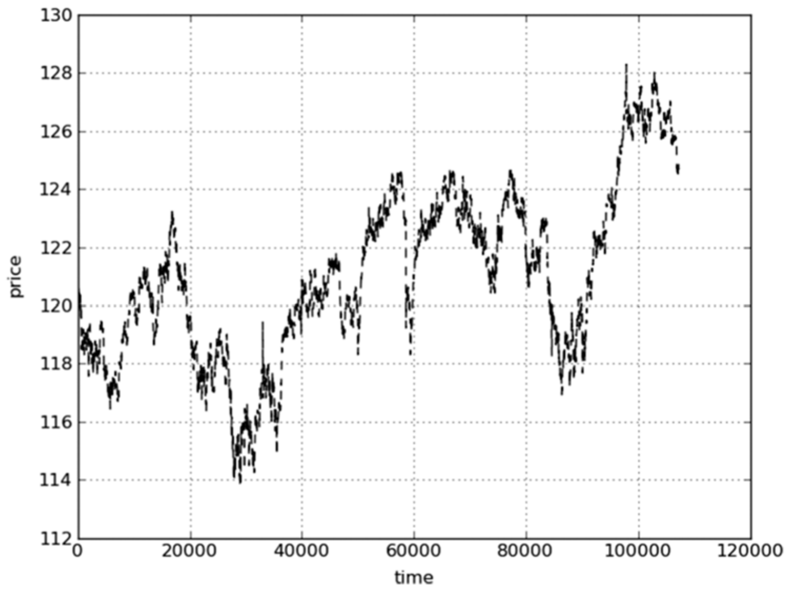
2.3. Irregular intervals

With trade-by-trade data, transactions occur at irregular intervals. A common scheme is to sample or aggregate transactions at regular intervals, such as by minute or by hour. If a chosen interval is too narrow, there may be a lack of transactions or the continuing problem of the noise associated with the price fluctuating between the bid and ask. However, as we find later in this paper, if a chosen interval is too large, the predictability associated with the underlying asset can vanish. Additionally, market structure may be eliminated with a large interval. Oomen in Ref. [19] discusses additional sampling methods.

An alternative to fixed interval sampling/aggregation is the popular autoregressive conditional duration (ACD) model described by Engle and Russel in Ref. [20]. This model compensates for the varying level of transactions through the day and the related problem of transaction sparseness. The duration in the model is defined as the time intervals between two consecutive transactions. A large duration would indicate sparseness of transactions and therefore a lack of trade activity or new information. A short duration would indicate the opposite.



(a) Random data. $\Pr(+) = 0.499$ and $\Pr(+|-) = 0.498$.



(b) Minute data from January 3–December 30, 2005. $\Pr(+) = 0.496$ but $\Pr(+|-) = 0.422$.

Fig. 1. Initial comparison of random data versus actual intra-day prices.

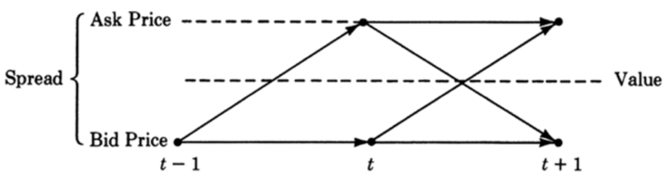


Fig. 2. Bid–ask spread schematic [18].

For our particular problem, sparseness of transactions was not an issue because the stock chosen for our experiments has some of the highest volumes in the world (discussed further in Section 3.1). Additionally, the authors of this paper interviewed two full-time Chicago equity traders and found that regular interval timespans are used frequently in the trading decision-making process among professionals.

3. Stock intra-day analysis

The purpose of this section is to test for market efficiency, and therefore the existence of trends, in a popularly traded stock using conditional probabilities, similar to the methods previously used by Refs. [10,11]. How does the predictability change when examining different timespans? When is predictability representing actual change, rather than movements between the bid and ask? First we present the dataset used, along with preprocessing steps, and then follow by describing the experiments.

3.1. Dataset and preprocessing steps

The stock that was used to examine conditional probabilities of upward versus downward price movements is one of the most widely traded stocks in the world, the Standard and Poor's 500 Index (symbol: SPY). It is an electronically traded fund (ETF) that holds all 500 Standard and Poor's stocks and is considered representative of the overall US market. The sheer number of transactions makes this an interesting stock to observe, and makes analysis easier given the need to examine longer-length series. The problem with spareness, or the lack of transactions when sampling at narrow time intervals, is minimized since volumes per day for SPY in 2005 averaged $63, 186, 191 \pm 19, 474, 197$; the average number of transactions per day was $91, 981 \pm 23, 332$. This large volume of transactions leads to a more efficient and unpredictable stock as a greater number of participants are driving the stock to an equilibrium; findings of predictability would be especially noteworthy.

Trade-by-trade data was retrieved from Wharton Research Data Services for the period January 3, 2005–December 31, 2005. As noted in Refs. [21,22], high-frequency trading data, such as the type used in this paper, requires special consideration. All late-trades, trades reported out-of-sequence, or trades with special settlement-conditions are excluded since their prices are not comparable to adjacent trades. The data was then reduced to temporal timespans of 1, 3, 5, 10, 20 and 30 s and 1, 5 and 30 min data using a volume-weighted average price approach (VWAP). This is calculated using the following formula:

$$P_{\text{VWAP}} = \frac{\sum_j P_j V_j}{\sum_j V_j}$$

where j are the individual trades that take place over the period of time and P_j is the price and V_j is the volume of trade j. Using a volume-weighted average price allows for a more realistic analysis of price movements, rather than sampling the last reported execution during a specific timespan. In addition, half-trading days such as the day after Thanksgiving and before Independence day were eliminated.

Trades were next encoded as either upward $(+)$ or downward $(-)$ as compared against the previous transaction. The data was split into one-week periods covering all 52 weeks in 2005 which allowed for enough prior instances of memory depth 5 (our longest depth used in this paper). As explained in Ref. [4], the existence of predictability in markets will eventually lead to their decline once those anomalies become "public knowledge". Traders who use forecasting models will bid up prices of stocks that are expected to rise, and sell off stocks that are expected to drop, thus eliminating their predictability. To prevent this elimination of possible predictability – the reversion back to randomness – we used 52 one-week timespans. From here, 2^m conditional probabilities for depth $m = 5$ are computed. The computation of separate weekly conditional probabilities allows us to analyze how the predictability changes over a weekly period.

Since the stock's previous day's closing price and the current day's opening price are often different, conditional probabilities for individual days are computed separately. This discontinuity is a distinct disadvantage of equity data over currency data, such as was used in Refs. [10,11]. The worldwide nature of currency trading allows for the market to be continually open somewhere on Earth, except for weekends.

The conditional probabilities of directional movements for the timespans can be seen in the Appendix in Tables A.4–A.12. The 30-minute timespan probabilities were not included in this paper because of the lack of data for all of the events, and as this paper demonstrates in the next section, predictability cannot be assumed for the 30-min timespan.

3.2. Experiment 1: Test of market independence

As explained in Section 2.1, to test for market efficiency, the probability of a given market directional-movement at time t is compared to the directional-movement at time $t - 1$. Under the assumption of efficiency, more prior information should not increase predictability. A violation of independence allows the possibility of predictable trends based on prior information.

Conditional probabilities of upward versus downward price movement, given prior price movement for each timespan, were computed separately for each of the 52 weeks in 2005 (see Section 3.1). In addition, we calculated binomial 95% confidence intervals to determine the number of weeks that were statistically-significantly outside of the bounds of error. If prices followed a random walk, the probability of an upward movement given prior information would be expected to equal the probability of an upward movement for that particular timespan, while taking into consideration the 95% confidence intervals. The week's probability was determined to be statistically significant if the 95% confidence interval's lower bound is

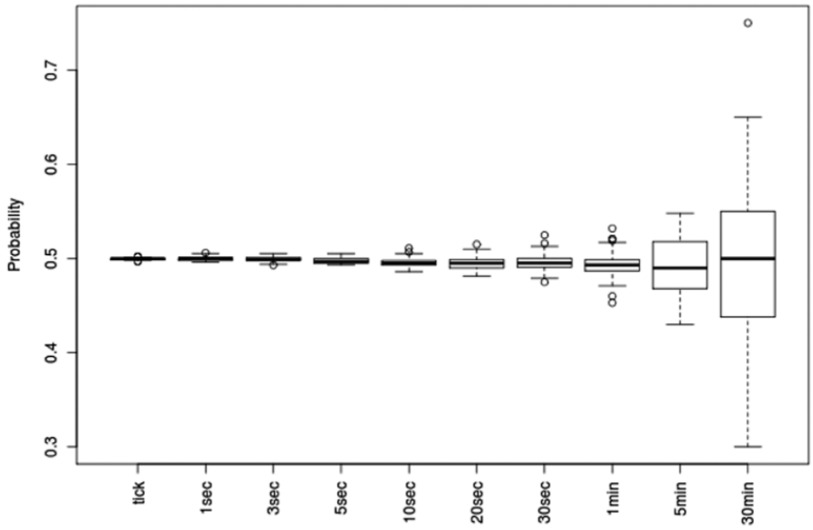


Fig. 3. Boxplot of Pr(+) for different timespans aggregated over 52 separate weeks.

above, or upper bound is below, the probability of the same directional movement. For example, the probability of an upward movement in price, given prior information $(\Pr(+|\{+,-\}\dots\{+,-\}))$, is significantly greater or less than the probability of an upward movement in price $(\Pr(+))$.

In Fig. 3, the probability of an upward movement is plotted for each of the timespans, which is roughly 50% probable that the market will go up. The variance of probabilities increases as the time between spans increases due to the decrease in data points. Fig. 4(a) displays the probability of an upward movement given a downward movement, $\Pr(+|-)$. The trade-by-trade (tick) data shows the highest predictability given prior information with a 79.0% conditional probability of an upward movement given a downward movement. At the 30 min timespan, there is only a 46.2% probability that the market will move up given a previous downward movement. This can be subtracted from 1 to get the probability of a downward given a previous downward movement, $1 - \Pr(+|-) = \Pr(-|-)$. The probabilities over the 52 weeks for the 30 min timespan range from a high of 66.7% to a low of 22.2%.[3]

The number of weeks out of 52 that are statistically significant can be seen in Fig. 4(b). From this chart it can be seen that all 52 weeks of the tick data were statistically significantly above $\Pr(+)$, while with 5 s data a total of 40 weeks were statistically-significant, with 37 weeks significantly above $\Pr(+)$ and 3 weeks below $\Pr(+)$. For the 30 min timespan, only 9 weeks out of the 52 weeks are statistically significantly below that timespan's $\Pr(+)$.

In addition, we use an independent samples t-test with a 95% confidence interval to test the hypothesis below for each of the timespans over the 52 week period:

$H_0 : \Pr(+) = \Pr(+|-)$; accept independence.
$H_1 : \Pr(+) \neq \Pr(+|-)$; reject independence.

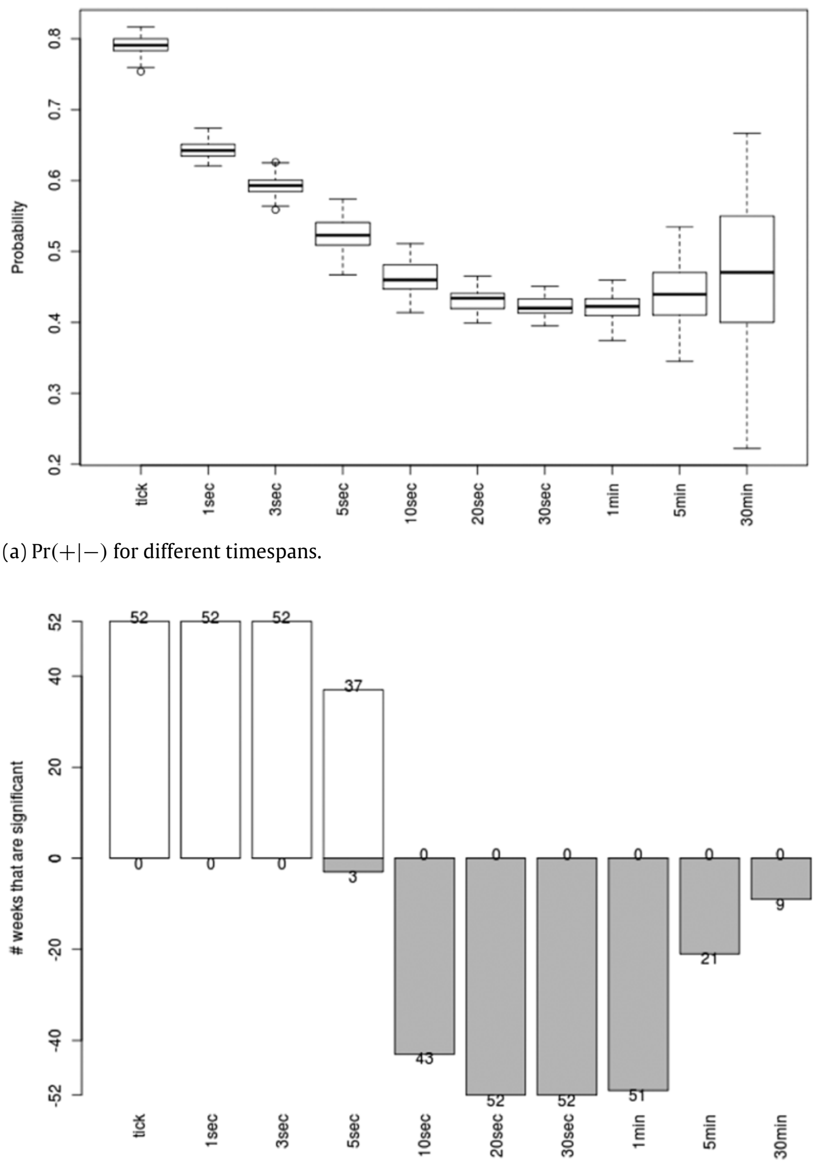
In testing the hypothesis for independence, the null hypothesis is rejected for all but the 30 min timespan (see Table 1). Therefore, in this statistical hypothesis testing, we can reject the independence assumption for the trade-by-trade, 1, 3, 5, 10, 20, 30 s, 1 and 5 min timespans; the independence still holds for the 30 min data. We conclude that the market is inefficient, and prior information impacts price movement, until approximately the 30 min period at which time the market begins to become more efficient.

As previously explained, much of the predictability of, at least, the trade-by-trade data can be explained by the bouncing of the price between the bid and ask. In our examined stock, the average trade size is roughly 700 shares and the median trade is 100 shares; with bid and ask sizes of 5000+ shares per side common, the transaction prices will fluctuate up and down between the bid and the ask until all shares are depleted. The question remains as to when the high levels of predictability cease to be explained by the market dynamics. The next section explores this question further.

3.3. Experiment 2: Escaping the confines of the bid and ask

Much of the predictability of trade-by-trade (tick) intra-day data can be explained by market dynamics; the price fluctuating between the bid and ask. As explained in Section 2.2, a *bid* is the best price at which an individual is willing to buy, while an *ask* is the best price at which one is willing to sell. When a stock has a large number of participants placing orders at the same price, but the average transaction size executing against the bid or ask is smaller, it takes time before the stock's bid or ask is eliminated and the stock is allowed to move to the next price point.

[3] For a summary of the conditional probabilities over each of the different timespans, excluding the 30 min timespan, please see the Appendix.



(a) Pr(+|−) for different timespans.

(b) The number of weeks that are statistically significant for the corresponding timespan where the 95% confidence interval's lower bound and upper bound are above or below the appropriate Pr(+) respectively.

Fig. 4. Boxplot of Pr(+|−) for different timespans, along with the number of significant weeks.

Table 1
Results from t-test for the different timespans, assuming unequal variances.

Timespan	t value	p value
Tick	−152.67	<0.0001
1 s	−84.13	<0.0001
3 s	−42.63	<0.0001
5 s	−7.66	<0.0001
10 s	10.32	<0.0001
20 s	28.18	<0.0001
30 s	31.89	<0.0001
1 min	21.30	<0.0001
5 min	7.38	<0.0001
30 min	1.68	0.0968

Using the same terminology as Ref. [14], when the signs of two non-zero consecutive changes are unlike each other, this pattern will be called a *reversal*, and when they are in the same direction, the pattern will be called a *continuation*.

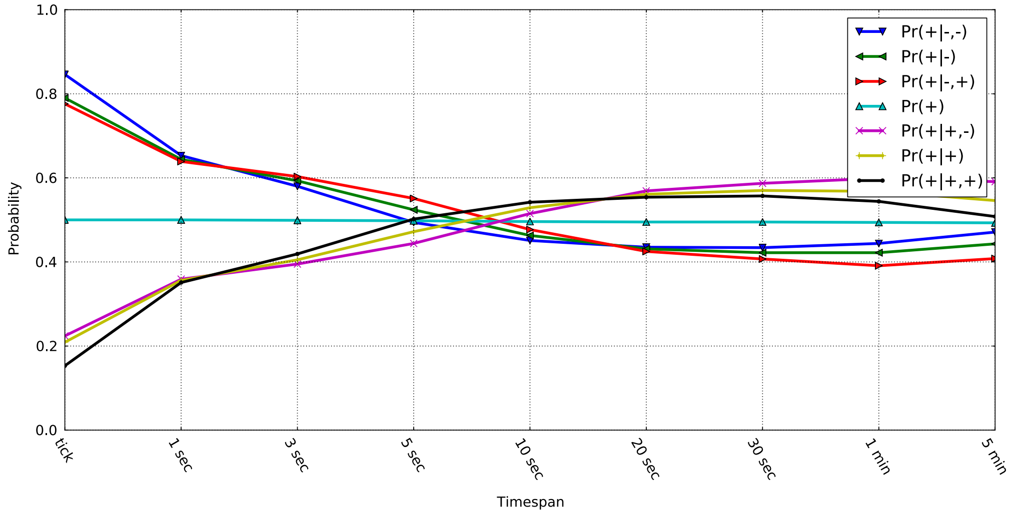


Fig. 5. Mean conditional probabilities of depth 2 for different timespans. Until 5–10 s, predictability is higher for *reversals* of trends, after which *continuation* of trend is higher.

Table 2
Comparing the conditional probabilities of directional movements for reversals versus continuations. The numbers within the brackets are the standard deviations.

	Events	Probability		
		Tick	5 s	20 s
Reversals	$\Pr(+\|-)$	0.79 [0.01]	0.52 [0.02]	0.43 [0.02]
	$\Pr(+\|-,-)$	0.85 [0.02]	0.49 [0.03]	0.44 [0.01]
	$\Pr(+\|-,-,-)$	0.85 [0.03]	0.48 [0.03]	0.44 [0.02]
	$\Pr(+\|-,-,-,-)$	0.84 [0.04]	0.47 [0.03]	0.44 [0.02]
	$\Pr(+\|-,-,-,-,-)$	0.84 [0.07]	0.46 [0.03]	0.44 [0.03]
Continuations	$\Pr(+\|+)$	0.21 [0.01]	0.47 [0.02]	0.56 [0.01]
	$\Pr(+\|+,+)$	0.15 [0.02]	0.50 [0.03]	0.55 [0.01]
	$\Pr(+\|+,+,+)$	0.15 [0.03]	0.52 [0.02]	0.56 [0.02]
	$\Pr(+\|+,+,+,+)$	0.17 [0.03]	0.53 [0.03]	0.56 [0.02]
	$\Pr(+\|+,+,+,+,+)$	0.18 [0.08]	0.53 [0.03]	0.55 [0.03]

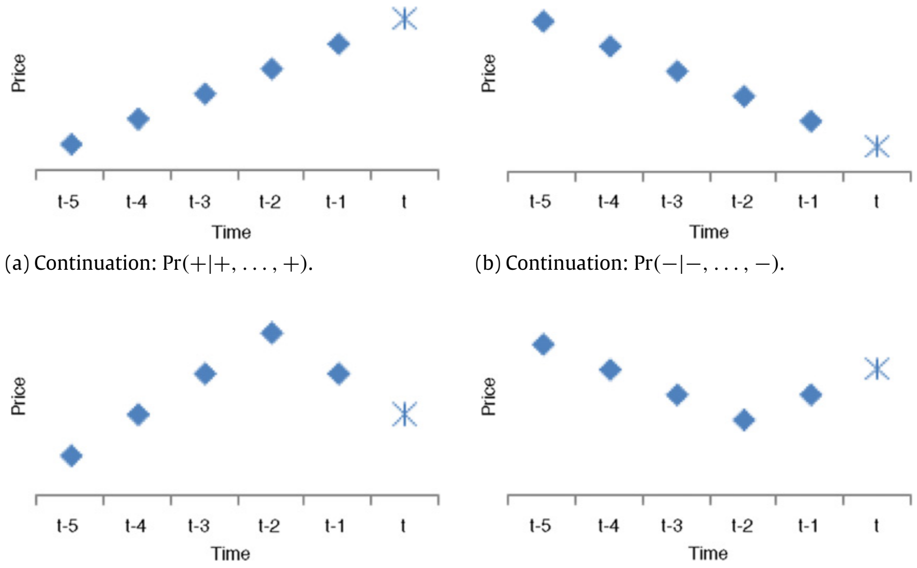
Re-examining Fig. 4(a) and (b), it can be observed that trade-by-trade data, up to a temporal timespan of 5 s, has a higher probability of reversal, rather than a trend continuation. After 10 s, the market has a higher probability of continuation than reversal. This can also be observed in Fig. 5, where the probabilities for different conditional probabilities up to depth 2 are plotted to show how the probabilities change with the increase in time between data points. Thirty-minute intervals were not included because of the lack of independence. As seen from the chart, market reversals ($\Pr(+|-)$, $\Pr(+|-,-)$, $\Pr(+|-,+)$) occur with a greater likelihood than continuations ($\Pr(+|+)$, $\Pr(+|+,+)$, $\Pr(+|+,-)$) until 5–10 s. After this period, continuations occur with greater probability than reversals. Additionally, while the variance increases as the interval between timespans becomes larger, the number of statistically significant weeks remain high and stable over the 52 weeks until a one to five minute timespan.

Table 2 displays the probability of continuations and reversals for trade-by-trade, along with a 5 s and 20 s timespan. Reversals occur with higher probabilities for trade-by-trade data. For example, the $\Pr(+|-)$ occurs with probability 0.79 and $\Pr(+|+)$ occurs with probability of 0.21 which infers a reversal of $1 - \Pr(+|-) = \Pr(-|+) = 0.79$. By the 20 s timespan, the reversal $\Pr(+|-)$ occurs with probability 0.43, which infers a continuation of two downward movements of $1 - \Pr(+|-) = \Pr(-|-) = 0.57$.

We theorize a 5–10 s timespan is the average length of time that the market price breaks the confines of the bid and ask and can freely move outside of these bounds. This observed reversal of directional movements before 5 s reflect the price being trapped between the bid and ask. While these movements are clearly predictable, they do not represent actual market changes, merely bounces of price between the bid and ask.

3.4. Trends with high apparent predictability

An interesting pattern is observed in the data after a 10 s timespan, which requires further analysis. Among some of the highest probabilities observed with the strongest significance over the 52 weeks, is what we call the trend *reversion-to-mean* (Fig. 6(c) and (d)). This can be described as the market trending in one direction, followed by an abrupt change in the opposite direction. The probability is higher that the next market directional movement will move in the same direction as the last movement.



(a) Continuation: Pr(+|+, . . . , +).

(b) Continuation: Pr(−|−, . . . , −).

(c) Reversion-to-mean: Pr(−|−, +, . . . , +).

(d) Reversion-to-mean: Pr(+|+, −, . . . , −).

Fig. 6. Examples of high-probability events: the trend continuation and the trend reversion-to-mean.

Table 3
Comparing the probabilities and the level of significance for reversion-to-mean and trend continuations for a 30 s timespan.

	Events	Prob.	# weeks that are stat. sig. from Pr(+)
Reversion-to-mean	Pr(+|+, −, −)	0.59 [0.02]	49
	Pr(+|+, −, −, −)	0.59 [0.03]	42
	Pr(+|+, −, −, −, −)	0.60 [0.04]	34
	Pr(−|−, +, +)	0.60 [0.03]	50
	Pr(−|−, +, +, +)	0.61 [0.03]	49
	Pr(−|−, +, +, +, +)	0.61 [0.04]	41
Continuation	Pr(+|+, +, +)	0.56 [0.02]	42
	Pr(+|+, +, +, +)	0.55 [0.03]	26
	Pr(+|+, +, +, +, +)	0.55 [0.05]	15
	Pr(−|−, −, −)	0.56 [0.02]	45
	Pr(−|−, −, −, −)	0.56 [0.03]	30
	Pr(−|−, −, −, −, −)	0.56 [0.03]	19

By comparing the trend reversion-to-mean to the trend continuation (Fig. 6(a) and (b)), it can be observed that the probability of trend reversion-to-mean occurs with a greater number of *statistically significant number of weeks* (see Table 3). For example, in a 30 s temporal timespan, Pr(+|+, −, −, −) occurs with a probability of 59.3% and is statistically significant as compared against the probability of an uptick (Pr(+)), 42 out of 52 weeks. We compare this with the probability of Pr(+|+, +, +, +), which occurs with a probability of 55.3%, but is only statistically significant 26 out of 52 weeks. Furthermore, the reversion-to-mean probabilities are larger and occur with a greater number of statistical weeks than continuations of the same depth. This pattern occurs in the 20 s, 30 s, 1 min, and 5 min timespans.

Fig. 7 shows trend continuation events using 30 s interval data (with added 95% confidence intervals) by month. Additionally, all twelve months are statistically significant from a probability of an upward movement during the same period. This further demonstrates the stability of events using high-frequency data.

4. Conclusion

In this paper, market inefficiency was examined empirically by analyzing trade-by-trade data at nine timespans. While statistically significant levels of predictability were found, we question if the high levels were useful or instead simply due to traditional market dynamics of prices fluctuating between the stock's bid and ask. By examining the stock's probability of upward movements (Pr(+)) versus upward given downward movements (Pr(+|−)), it was found that prior to a 5–10 s timespan the probabilities of reversal movements occurred with higher probability than continuation of price movements. After 5–10 s timespans, continuation of price movements became more probable than reversals. We theorize this to be the point at which the stock was escaping the confines of the bid and ask.

The probabilities of market reversions-to-mean were statistically higher than probabilities of continuations of the same depth; this occurred in 20 s, 30 s, 1 min, and 5 min temporal timespans. We also observed higher numbers of statistically

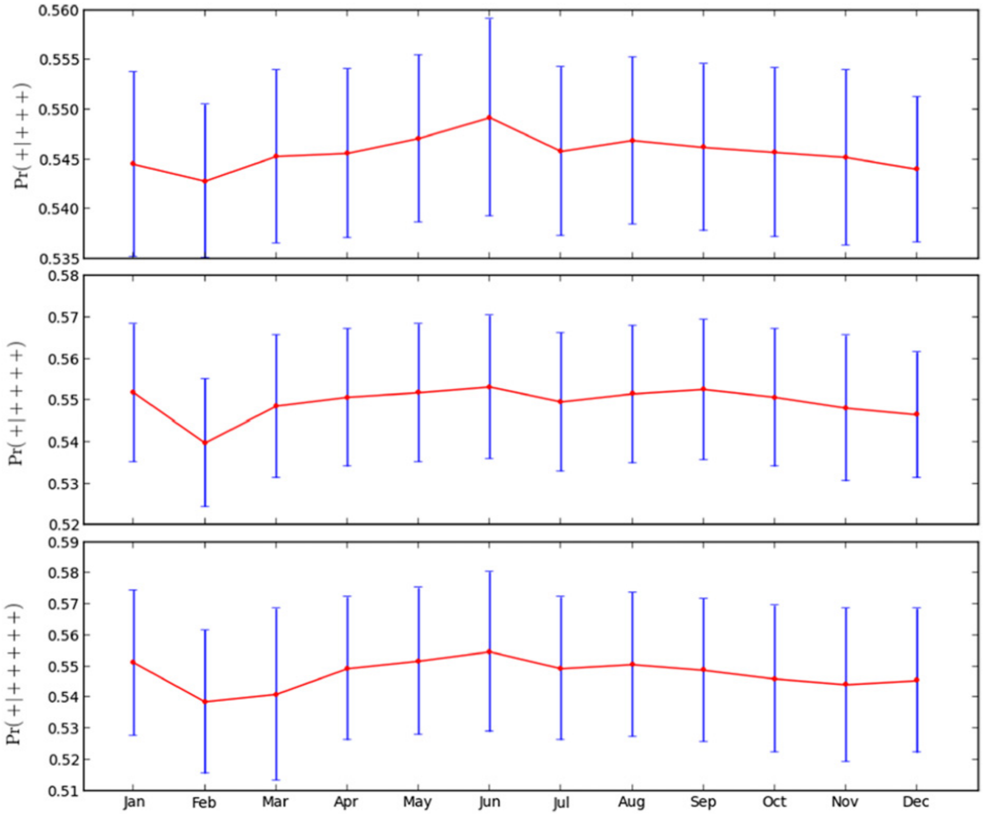


Fig. 7. Examining monthly stability of events using 30 s interval data.

significant weeks of market reversions-to-mean as compared to the number of statistically significant weeks of market continuations for the same depth. This suggests the market being pulled back to the equilibrium price. The information presented in this paper would be useful to traders when deciding to exit or hold a position. If the probability is higher for the market to reverse directions than continue, a trader may decide to close the position, whereas with a higher probability for the market to continue in the same direction, the trader may decide to hold the position longer.

While much of traditional mainstream finance has reservations of market inefficiency, or the existence of trends, Malkiel states that there may be a possible explanation of why trends might perpetuate themselves [3]:

> …it has been argued that the crowd instinct of mass psychology makes it so. When investors see the price of a speculative favorite going higher and higher, they want to jump on the bandwagon and join the rise. Indeed, the price rise itself helps fuel the enthusiasm in a self-fulfilling prophecy. Each rise in price just whets the appetite and makes investors expect a further rise.

Given that the chosen stock for this paper (symbol: SPY) is an exchange-traded fund which is comprised of the 500 stocks within the Standard and Poor's 500 index, the results are especially noteworthy. This stock, being one of the most widely traded stocks in the world, would suggest that all inefficiencies were spotted by others and implemented in their trading approach. This widespread adoption of the high-probability events would drive the asset price either up or down enough to eliminate the pattern. However, this is not what was found; stable, high-probability market movements were still found for this popular stock.

Further research would be necessary to determine the timespans at which other stocks become inefficient/efficient. While the 30 min market is the timespan at which the examined stock became efficient, surely this would be different for each stock, for stocks have different levels of trading activity and levels of participation. This continued study would be necessary to understand in order to implement a good trading model that takes advantage of inefficient markets and enable traders to make better trading decisions.

Appendix. Conditional probability tables

Tables A.4–A.12 are the full prior conditional probabilities of market direction movements for trade-by-trade, 1, 3, 5, 10, 20, and 30 s interval timespans, followed by 1 and 5 min timespans. Thirty-minute timespans were not included because of the lack of priors at extended depths. Also included is the number of weeks out of the year that are statistically significant when compared against the probability of an uptick (Pr(+)).

Table A.4
Conditional probabilities of market directional movements for trade-by-trade (tick) data.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.500	0.001	n/a
1	Pr(+|−)	0.790	0.014	52
	Pr(+|+)	0.209	0.014	52
2	Pr(+|−, −)	0.846	0.018	52
	Pr(+|−, +)	0.776	0.014	52
	Pr(+|+, −)	0.224	0.014	52
	Pr(+|+, +)	0.153	0.017	52
3	Pr(+|−, −, −)	0.848	0.025	52
	Pr(+|−, −, +)	0.846	0.017	52
	Pr(+|−, +, −)	0.790	0.016	52
	Pr(+|−, +, +)	0.725	0.009	52
	Pr(+|+, −, −)	0.274	0.009	52
	Pr(+|+, −, +)	0.210	0.016	52
	Pr(+|+, +, −)	0.153	0.017	52
	Pr(+|+, +, +)	0.153	0.025	52
4	Pr(+|−, −, −, −)	0.843	0.035	52
	Pr(+|−, −, −, +)	0.849	0.025	52
	Pr(+|−, −, +, −)	0.845	0.019	52
	Pr(+|−, −, +, +)	0.848	0.016	52
	Pr(+|−, +, −, −)	0.770	0.019	52
	Pr(+|−, +, −, +)	0.795	0.016	52
	Pr(+|−, +, +, −)	0.722	0.010	52
	Pr(+|−, +, +, +)	0.741	0.018	52
	Pr(+|+, −, −, −)	0.256	0.018	52
	Pr(+|+, −, −, +)	0.277	0.010	52
	Pr(+|+, −, +, −)	0.204	0.016	52
	Pr(+|+, −, +, +)	0.230	0.019	52
	Pr(+|+, +, −, −)	0.151	0.015	52
	Pr(+|+, +, −, +)	0.153	0.019	52
	Pr(+|+, +, +, −)	0.151	0.025	52
	Pr(+|+, +, +, +)	0.166	0.034	52
5	Pr(+|−, −, −, −, −)	0.838	0.073	50
	Pr(+|−, −, −, −, +)	0.843	0.040	52
	Pr(+|−, −, −, +, −)	0.845	0.026	52
	Pr(+|−, −, −, +, +)	0.857	0.031	52
	Pr(+|−, −, +, −, −)	0.829	0.024	52
	Pr(+|−, −, +, −, +)	0.849	0.018	52
	Pr(+|−, −, +, +, −)	0.847	0.018	52
	Pr(+|−, −, +, +, +)	0.853	0.023	52
	Pr(+|−, +, −, −, −)	0.754	0.030	52
	Pr(+|−, +, −, −, +)	0.773	0.017	52
	Pr(+|−, +, −, +, −)	0.797	0.017	52
	Pr(+|−, +, −, +, +)	0.787	0.013	52
	Pr(+|−, +, +, −, −)	0.704	0.013	52
	Pr(+|−, +, +, −, +)	0.729	0.011	52
	Pr(+|−, +, +, +, −)	0.739	0.019	52
	Pr(+|−, +, +, +, +)	0.750	0.044	52
	Pr(+|+, −, −, −, −)	0.249	0.043	52
	Pr(+|+, −, −, −, +)	0.257	0.019	52
	Pr(+|+, −, −, +, −)	0.269	0.011	52
	Pr(+|+, −, −, +, +)	0.297	0.014	52
	Pr(+|+, −, +, −, −)	0.213	0.013	52
	Pr(+|+, −, +, −, +)	0.202	0.017	52
	Pr(+|+, −, +, +, −)	0.227	0.017	52
	Pr(+|+, −, +, +, +)	0.245	0.032	52
	Pr(+|+, +, −, −, −)	0.147	0.031	52
	Pr(+|+, +, −, −, +)	0.152	0.017	52
	Pr(+|+, +, −, +, −)	0.149	0.018	52
	Pr(+|+, +, −, +, +)	0.166	0.022	52
	Pr(+|+, +, +, −, −)	0.146	0.035	52
	Pr(+|+, +, +, −, +)	0.153	0.026	52
	Pr(+|+, +, +, +, −)	0.162	0.033	52
	Pr(+|+, +, +, +, +)	0.184	0.077	48

Table A.5
Conditional probabilities of market directional movements for 1 s timespan.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.500	0.002	n/a
1	Pr(+|−)	0.644	0.012	52
	Pr(+|+)	0.356	0.011	52
2	Pr(+|−, −)	0.653	0.016	52
	Pr(+|−, +)	0.639	0.012	52
	Pr(+|+, −)	0.359	0.010	52
	Pr(+|+, +)	0.351	0.016	52
3	Pr(+|−, −, −)	0.634	0.018	52
	Pr(+|−, −, +)	0.663	0.016	52
	Pr(+|−, +, −)	0.640	0.015	52
	Pr(+|−, +, +)	0.636	0.010	52
	Pr(+|+, −, −)	0.360	0.010	52
	Pr(+|+, −, +)	0.358	0.012	52
	Pr(+|+, +, −)	0.338	0.016	52
	Pr(+|+, +, +)	0.374	0.019	52
4	Pr(+|−, −, −, −)	0.621	0.025	52
	Pr(+|−, −, −, +)	0.642	0.019	52
	Pr(+|−, −, +, −)	0.654	0.019	52
	Pr(+|−, −, +, +)	0.679	0.015	52
	Pr(+|−, +, −, −)	0.622	0.017	52
	Pr(+|−, +, −, +)	0.651	0.015	52
	Pr(+|−, +, +, −)	0.630	0.010	52
	Pr(+|−, +, +, +)	0.649	0.016	52
	Pr(+|+, −, −, −)	0.344	0.016	52
	Pr(+|+, −, −, +)	0.368	0.010	52
	Pr(+|+, −, +, −)	0.347	0.013	52
	Pr(+|+, −, +, +)	0.378	0.014	52
	Pr(+|+, +, −, −)	0.320	0.016	52
	Pr(+|+, +, −, +)	0.348	0.018	52
	Pr(+|+, +, +, −)	0.366	0.020	52
	Pr(+|+, +, +, +)	0.388	0.027	52
5	Pr(+|−, −, −, −, −)	0.609	0.039	51
	Pr(+|−, −, −, −, +)	0.629	0.026	52
	Pr(+|−, −, −, +, −)	0.635	0.022	52
	Pr(+|−, −, −, +, +)	0.655	0.027	52
	Pr(+|−, −, +, −, −)	0.640	0.021	52
	Pr(+|−, −, +, −, +)	0.662	0.021	52
	Pr(+|−, −, +, +, −)	0.674	0.016	52
	Pr(+|−, −, +, +, +)	0.690	0.022	52
	Pr(+|−, +, −, −, −)	0.600	0.021	52
	Pr(+|−, +, −, −, +)	0.633	0.019	52
	Pr(+|−, +, −, +, −)	0.647	0.016	52
	Pr(+|−, +, −, +, +)	0.657	0.018	52
	Pr(+|−, +, +, −, −)	0.613	0.016	52
	Pr(+|−, +, +, −, +)	0.640	0.011	52
	Pr(+|−, +, +, +, −)	0.640	0.019	52
	Pr(+|−, +, +, +, +)	0.664	0.022	52
	Pr(+|+, −, −, −, −)	0.335	0.023	52
	Pr(+|+, −, −, −, +)	0.350	0.018	52
	Pr(+|+, −, −, +, −)	0.358	0.013	52
	Pr(+|+, −, −, +, +)	0.385	0.015	52
	Pr(+|+, −, +, −, −)	0.340	0.015	52
	Pr(+|+, −, +, −, +)	0.350	0.014	52
	Pr(+|+, −, +, +, −)	0.364	0.015	52
	Pr(+|+, −, +, +, +)	0.405	0.018	52
	Pr(+|+, +, −, −, −)	0.310	0.026	52
	Pr(+|+, +, −, −, +)	0.325	0.021	52
	Pr(+|+, +, −, +, −)	0.339	0.022	52
	Pr(+|+, +, −, +, +)	0.363	0.018	52
	Pr(+|+, +, +, −, −)	0.353	0.025	52
	Pr(+|+, +, +, −, +)	0.372	0.023	52
	Pr(+|+, +, +, +, −)	0.384	0.030	52
	Pr(+|+, +, +, +, +)	0.395	0.033	50

Table A.6
Conditional probabilities of market directional movements for 3 s timespan.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.499	0.003	n/a
1	Pr(+|−)	0.593	0.016	52
	Pr(+|+)	0.405	0.015	52
2	Pr(+|−, −)	0.580	0.022	52
	Pr(+|−, +)	0.603	0.013	52
	Pr(+|+, −)	0.395	0.014	52
	Pr(+|+, +)	0.419	0.019	52
3	Pr(+|−, −, −)	0.555	0.019	51
	Pr(+|−, −, +)	0.599	0.024	52
	Pr(+|−, +, −)	0.595	0.017	52
	Pr(+|−, +, +)	0.615	0.014	52
	Pr(+|+, −, −)	0.383	0.014	52
	Pr(+|+, −, +)	0.403	0.017	52
	Pr(+|+, +, −)	0.399	0.019	52
	Pr(+|+, +, +)	0.447	0.021	50
4	Pr(+|−, −, −, −)	0.540	0.020	31
	Pr(+|−, −, −, +)	0.567	0.024	50
	Pr(+|−, −, +, −)	0.590	0.025	52
	Pr(+|−, −, +, +)	0.612	0.028	52
	Pr(+|−, +, −, −)	0.578	0.020	52
	Pr(+|−, +, −, +)	0.606	0.017	52
	Pr(+|−, +, +, −)	0.609	0.015	52
	Pr(+|−, +, +, +)	0.623	0.022	52
	Pr(+|+, −, −, −)	0.373	0.021	52
	Pr(+|+, −, −, +)	0.390	0.015	52
	Pr(+|+, −, +, −)	0.391	0.017	52
	Pr(+|+, −, +, +)	0.419	0.020	52
	Pr(+|+, +, −, −)	0.389	0.020	52
	Pr(+|+, +, −, +)	0.406	0.022	52
	Pr(+|+, +, +, −)	0.436	0.023	50
	Pr(+|+, +, +, +)	0.462	0.024	31
5	Pr(+|−, −, −, −, −)	0.533	0.027	12
	Pr(+|−, −, −, −, +)	0.546	0.025	27
	Pr(+|−, −, −, +, −)	0.562	0.025	43
	Pr(+|−, −, −, +, +)	0.575	0.033	43
	Pr(+|−, −, +, −, −)	0.584	0.029	52
	Pr(+|−, −, +, −, +)	0.596	0.027	52
	Pr(+|−, −, +, +, −)	0.613	0.029	52
	Pr(+|−, −, +, +, +)	0.610	0.034	51
	Pr(+|−, +, −, −, −)	0.569	0.021	49
	Pr(+|−, +, −, −, +)	0.583	0.026	52
	Pr(+|−, +, −, +, −)	0.605	0.018	52
	Pr(+|−, +, −, +, +)	0.609	0.021	52
	Pr(+|−, +, +, −, −)	0.601	0.021	52
	Pr(+|−, +, +, −, +)	0.615	0.019	52
	Pr(+|−, +, +, +, −)	0.623	0.024	52
	Pr(+|−, +, +, +, +)	0.623	0.028	52
	Pr(+|+, −, −, −, −)	0.374	0.028	52
	Pr(+|+, −, −, −, +)	0.373	0.023	52
	Pr(+|+, −, −, +, −)	0.387	0.018	52
	Pr(+|+, −, −, +, +)	0.395	0.018	52
	Pr(+|+, −, +, −, −)	0.389	0.020	52
	Pr(+|+, −, +, −, +)	0.393	0.018	52
	Pr(+|+, −, +, +, −)	0.408	0.022	52
	Pr(+|+, −, +, +, +)	0.435	0.024	47
	Pr(+|+, +, −, −, −)	0.389	0.029	52
	Pr(+|+, +, −, −, +)	0.388	0.023	52
	Pr(+|+, +, −, +, −)	0.401	0.027	52
	Pr(+|+, +, −, +, +)	0.414	0.024	52
	Pr(+|+, +, +, −, −)	0.431	0.030	42
	Pr(+|+, +, +, −, +)	0.439	0.027	43
	Pr(+|+, +, +, +, −)	0.460	0.028	26
	Pr(+|+, +, +, +, +)	0.464	0.027	20

Table A.7
Conditional probabilities of market directional movements for 5 s timespan.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.498	0.003	n/a
1	Pr(+|−)	0.524	0.024	40
	Pr(+|+)	0.472	0.023	45
2	Pr(+|−, −)	0.494	0.028	36
	Pr(+|−, +)	0.551	0.021	48
	Pr(+|+, −)	0.444	0.020	51
	Pr(+|+, +)	0.502	0.025	33
3	Pr(+|−, −, −)	0.475	0.025	34
	Pr(+|−, −, +)	0.513	0.032	30
	Pr(+|−, +, −)	0.551	0.022	48
	Pr(+|−, +, +)	0.551	0.024	45
	Pr(+|+, −, −)	0.445	0.021	49
	Pr(+|+, −, +)	0.444	0.023	47
	Pr(+|+, +, −)	0.481	0.028	29
	Pr(+|+, +, +)	0.523	0.024	29
4	Pr(+|−, −, −, −)	0.466	0.026	30
	Pr(+|−, −, −, +)	0.486	0.029	17
	Pr(+|−, −, +, −)	0.514	0.033	26
	Pr(+|−, −, +, +)	0.511	0.037	20
	Pr(+|−, +, −, −)	0.541	0.024	35
	Pr(+|−, +, −, +)	0.560	0.025	46
	Pr(+|−, +, +, −)	0.558	0.024	41
	Pr(+|−, +, +, +)	0.544	0.028	36
	Pr(+|+, −, −, −)	0.451	0.025	39
	Pr(+|+, −, −, +)	0.439	0.022	48
	Pr(+|+, −, +, −)	0.439	0.024	47
	Pr(+|+, −, +, +)	0.450	0.026	41
	Pr(+|+, +, −, −)	0.482	0.030	23
	Pr(+|+, +, −, +)	0.481	0.030	23
	Pr(+|+, +, +, −)	0.516	0.026	15
	Pr(+|+, +, +, +)	0.529	0.025	29
5	Pr(+|−, −, −, −, −)	0.459	0.027	29
	Pr(+|−, −, −, −, +)	0.473	0.033	19
	Pr(+|−, −, −, +, −)	0.483	0.033	16
	Pr(+|−, −, −, +, +)	0.489	0.034	12
	Pr(+|−, −, +, −, −)	0.503	0.036	13
	Pr(+|−, −, +, −, +)	0.523	0.037	22
	Pr(+|−, −, +, +, −)	0.517	0.043	17
	Pr(+|−, −, +, +, +)	0.505	0.037	10
	Pr(+|−, +, −, −, −)	0.530	0.026	16
	Pr(+|−, +, −, −, +)	0.550	0.031	30
	Pr(+|−, +, −, +, −)	0.564	0.029	46
	Pr(+|−, +, −, +, +)	0.554	0.026	39
	Pr(+|−, +, +, −, −)	0.556	0.031	37
	Pr(+|−, +, +, −, +)	0.560	0.030	38
	Pr(+|−, +, +, +, −)	0.550	0.028	31
	Pr(+|−, +, +, +, +)	0.539	0.034	21
	Pr(+|+, −, −, −, −)	0.454	0.031	29
	Pr(+|+, −, −, −, +)	0.448	0.029	30
	Pr(+|+, −, −, +, −)	0.431	0.026	46
	Pr(+|+, −, −, +, +)	0.448	0.030	31
	Pr(+|+, −, +, −, −)	0.439	0.031	42
	Pr(+|+, −, +, −, +)	0.438	0.023	47
	Pr(+|+, −, +, +, −)	0.441	0.033	37
	Pr(+|+, −, +, +, +)	0.461	0.025	26
	Pr(+|+, +, −, −, −)	0.487	0.036	12
	Pr(+|+, +, −, −, +)	0.477	0.031	18
	Pr(+|+, +, −, +, −)	0.475	0.034	18
	Pr(+|+, +, −, +, +)	0.489	0.033	12
	Pr(+|+, +, +, −, −)	0.516	0.029	11
	Pr(+|+, +, +, −, +)	0.517	0.029	12
	Pr(+|+, +, +, +, −)	0.526	0.023	16
	Pr(+|+, +, +, +, +)	0.532	0.031	26

Table A.8
Conditional probabilities of market directional movements for 10 s timespan.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.496	0.005	n/a
1	Pr(+|−)	0.463	0.022	43
	Pr(+|+)	0.529	0.020	41
2	Pr(+|−, −)	0.451	0.021	47
	Pr(+|−, +)	0.477	0.028	33
	Pr(+|+, −)	0.515	0.023	21
	Pr(+|+, +)	0.542	0.020	46
3	Pr(+|−, −, −)	0.445	0.020	49
	Pr(+|−, −, +)	0.457	0.027	38
	Pr(+|−, +, −)	0.485	0.028	17
	Pr(+|−, +, +)	0.469	0.030	32
	Pr(+|+, −, −)	0.524	0.025	22
	Pr(+|+, −, +)	0.506	0.024	14
	Pr(+|+, +, −)	0.535	0.024	30
	Pr(+|+, +, +)	0.548	0.020	44
4	Pr(+|−, −, −, −)	0.445	0.021	44
	Pr(+|−, −, −, +)	0.446	0.027	39
	Pr(+|−, −, +, −)	0.460	0.032	29
	Pr(+|−, −, +, +)	0.454	0.030	32
	Pr(+|−, +, −, −)	0.482	0.027	12
	Pr(+|−, +, −, +)	0.488	0.035	14
	Pr(+|−, +, +, −)	0.485	0.035	13
	Pr(+|−, +, +, +)	0.455	0.032	32
	Pr(+|+, −, −, −)	0.530	0.027	21
	Pr(+|+, −, −, +)	0.517	0.033	11
	Pr(+|+, −, +, −)	0.499	0.030	10
	Pr(+|+, −, +, +)	0.512	0.029	8
	Pr(+|+, +, −, −)	0.536	0.027	22
	Pr(+|+, +, −, +)	0.535	0.026	19
	Pr(+|+, +, +, −)	0.544	0.028	30
	Pr(+|+, +, +, +)	0.552	0.020	41
5	Pr(+|−, −, −, −, −)	0.443	0.027	37
	Pr(+|−, −, −, −, +)	0.447	0.030	30
	Pr(+|−, −, −, +, −)	0.444	0.034	29
	Pr(+|−, −, −, +, +)	0.447	0.034	30
	Pr(+|−, −, +, −, −)	0.457	0.038	22
	Pr(+|−, −, +, −, +)	0.463	0.042	15
	Pr(+|−, −, +, +, −)	0.460	0.040	18
	Pr(+|−, −, +, +, +)	0.449	0.031	26
	Pr(+|−, +, −, −, −)	0.483	0.034	8
	Pr(+|−, +, −, −, +)	0.481	0.031	8
	Pr(+|−, +, −, +, −)	0.492	0.049	12
	Pr(+|−, +, −, +, +)	0.484	0.033	3
	Pr(+|−, +, +, −, −)	0.482	0.044	14
	Pr(+|−, +, +, −, +)	0.489	0.042	9
	Pr(+|−, +, +, +, −)	0.463	0.043	17
	Pr(+|−, +, +, +, +)	0.447	0.033	31
	Pr(+|+, −, −, −, −)	0.537	0.030	17
	Pr(+|+, −, −, −, +)	0.522	0.038	10
	Pr(+|+, −, −, +, −)	0.509	0.039	3
	Pr(+|+, −, −, +, +)	0.523	0.036	9
	Pr(+|+, −, +, −, −)	0.504	0.035	5
	Pr(+|+, −, +, −, +)	0.494	0.043	10
	Pr(+|+, −, +, +, −)	0.511	0.039	6
	Pr(+|+, −, +, +, +)	0.513	0.032	4
	Pr(+|+, +, −, −, −)	0.538	0.032	17
	Pr(+|+, +, −, −, +)	0.533	0.037	15
	Pr(+|+, +, −, +, −)	0.532	0.033	11
	Pr(+|+, +, −, +, +)	0.537	0.031	16
	Pr(+|+, +, +, −, −)	0.539	0.030	17
	Pr(+|+, +, +, −, +)	0.549	0.036	22
	Pr(+|+, +, +, +, −)	0.550	0.028	25
	Pr(+|+, +, +, +, +)	0.553	0.027	32

Table A.9
Conditional probabilities of market directional movements for 20 s timespan.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.495	0.007	n/a
1	Pr(+|−)	0.431	0.015	52
	Pr(+|+)	0.561	0.013	52
2	Pr(+|−, −)	0.435	0.013	52
	Pr(+|−, +)	0.425	0.024	50
	Pr(+|+, −)	0.569	0.021	49
	Pr(+|+, +)	0.554	0.014	51
3	Pr(+|−, −, −)	0.436	0.019	50
	Pr(+|−, −, +)	0.434	0.019	48
	Pr(+|−, +, −)	0.437	0.032	37
	Pr(+|−, +, +)	0.417	0.026	49
	Pr(+|+, −, −)	0.572	0.023	48
	Pr(+|+, −, +)	0.565	0.028	39
	Pr(+|+, +, −)	0.552	0.024	33
	Pr(+|+, +, +)	0.556	0.018	46
4	Pr(+|−, −, −, −)	0.437	0.024	39
	Pr(+|−, −, −, +)	0.435	0.029	38
	Pr(+|−, −, +, −)	0.434	0.027	36
	Pr(+|−, −, +, +)	0.432	0.025	42
	Pr(+|−, +, −, −)	0.439	0.037	28
	Pr(+|−, +, −, +)	0.433	0.042	24
	Pr(+|−, +, +, −)	0.429	0.035	35
	Pr(+|−, +, +, +)	0.407	0.03	48
	Pr(+|+, −, −, −)	0.578	0.029	44
	Pr(+|+, −, −, +)	0.563	0.035	30
	Pr(+|+, −, +, −)	0.572	0.038	34
	Pr(+|+, −, +, +)	0.56	0.033	28
	Pr(+|+, +, −, −)	0.551	0.028	28
	Pr(+|+, +, −, +)	0.552	0.031	19
	Pr(+|+, +, +, −)	0.558	0.024	30
	Pr(+|+, +, +, +)	0.555	0.023	33
5	Pr(+|−, −, −, −, −)	0.439	0.033	29
	Pr(+|−, −, −, −, +)	0.434	0.031	26
	Pr(+|−, −, −, +, −)	0.431	0.039	22
	Pr(+|−, −, −, +, +)	0.439	0.038	25
	Pr(+|−, −, +, −, −)	0.431	0.037	22
	Pr(+|−, −, +, −, +)	0.438	0.043	17
	Pr(+|−, −, +, +, −)	0.434	0.042	23
	Pr(+|−, −, +, +, +)	0.431	0.03	27
	Pr(+|−, +, −, −, −)	0.442	0.042	15
	Pr(+|−, +, −, −, +)	0.435	0.054	21
	Pr(+|−, +, −, +, −)	0.436	0.053	12
	Pr(+|−, +, −, +, +)	0.431	0.049	19
	Pr(+|−, +, +, −, −)	0.434	0.043	20
	Pr(+|−, +, +, −, +)	0.422	0.044	24
	Pr(+|−, +, +, +, −)	0.416	0.042	34
	Pr(+|−, +, +, +, +)	0.4	0.038	41
	Pr(+|+, −, −, −, −)	0.577	0.034	32
	Pr(+|+, −, −, −, +)	0.58	0.035	29
	Pr(+|+, −, −, +, −)	0.567	0.049	19
	Pr(+|+, −, −, +, +)	0.56	0.042	19
	Pr(+|+, −, +, −, −)	0.576	0.046	21
	Pr(+|+, −, +, −, +)	0.566	0.062	17
	Pr(+|+, −, +, +, −)	0.566	0.048	20
	Pr(+|+, −, +, +, +)	0.554	0.041	16
	Pr(+|+, +, −, −, −)	0.552	0.035	22
	Pr(+|+, +, −, −, +)	0.551	0.039	12
	Pr(+|+, +, −, +, −)	0.552	0.052	12
	Pr(+|+, +, −, +, +)	0.552	0.047	11
	Pr(+|+, +, +, −, −)	0.562	0.033	23
	Pr(+|+, +, +, −, +)	0.553	0.031	9
	Pr(+|+, +, +, +, −)	0.56	0.037	23
	Pr(+|+, +, +, +, +)	0.552	0.026	20

Table A.10
Conditional probabilities of market directional movements for 30 s timespan.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.495	0.009	n/a
1	Pr(+|−)	0.422	0.013	52
	Pr(+|+)	0.570	0.013	52
2	Pr(+|−, −)	0.434	0.017	51
	Pr(+|−, +)	0.407	0.023	52
	Pr(+|+, −)	0.587	0.019	52
	Pr(+|+, +)	0.557	0.015	48
3	Pr(+|−, −, −)	0.437	0.021	45
	Pr(+|−, −, +)	0.429	0.024	42
	Pr(+|−, +, −)	0.414	0.031	43
	Pr(+|−, +, +)	0.402	0.026	50
	Pr(+|+, −, −)	0.587	0.024	49
	Pr(+|+, −, +)	0.586	0.029	45
	Pr(+|+, +, −)	0.558	0.023	40
	Pr(+|+, +, +)	0.556	0.021	42
4	Pr(+|−, −, −, −)	0.440	0.029	30
	Pr(+|−, −, −, +)	0.434	0.031	33
	Pr(+|−, −, +, −)	0.433	0.036	24
	Pr(+|−, −, +, +)	0.426	0.031	34
	Pr(+|−, +, −, −)	0.414	0.036	36
	Pr(+|−, +, −, +)	0.413	0.045	28
	Pr(+|−, +, +, −)	0.409	0.041	38
	Pr(+|−, +, +, +)	0.395	0.029	49
	Pr(+|+, −, −, −)	0.593	0.034	42
	Pr(+|+, −, −, +)	0.580	0.029	34
	Pr(+|+, −, +, −)	0.594	0.045	32
	Pr(+|+, −, +, +)	0.581	0.035	34
	Pr(+|+, +, −, −)	0.561	0.026	27
	Pr(+|+, +, −, +)	0.554	0.035	16
	Pr(+|+, +, +, −)	0.560	0.031	28
	Pr(+|+, +, +, +)	0.553	0.028	26
5	Pr(+|−, −, −, −, −)	0.441	0.034	19
	Pr(+|−, −, −, −, +)	0.438	0.043	18
	Pr(+|−, −, −, +, −)	0.437	0.047	13
	Pr(+|−, −, −, +, +)	0.431	0.041	24
	Pr(+|−, −, +, −, −)	0.438	0.052	18
	Pr(+|−, −, +, −, +)	0.429	0.054	10
	Pr(+|−, −, +, +, −)	0.426	0.039	21
	Pr(+|−, −, +, +, +)	0.426	0.049	24
	Pr(+|−, +, −, −, −)	0.427	0.050	17
	Pr(+|−, +, −, −, +)	0.397	0.050	21
	Pr(+|−, +, −, +, −)	0.414	0.067	14
	Pr(+|−, +, −, +, +)	0.414	0.060	13
	Pr(+|−, +, +, −, −)	0.423	0.054	21
	Pr(+|−, +, +, −, +)	0.389	0.062	31
	Pr(+|−, +, +, +, −)	0.406	0.047	27
	Pr(+|−, +, +, +, +)	0.387	0.040	41
	Pr(+|+, −, −, −, −)	0.599	0.039	34
	Pr(+|+, −, −, −, +)	0.585	0.052	24
	Pr(+|+, −, −, +, −)	0.589	0.044	20
	Pr(+|+, −, −, +, +)	0.574	0.041	23
	Pr(+|+, −, +, −, −)	0.607	0.060	29
	Pr(+|+, −, +, −, +)	0.575	0.059	12
	Pr(+|+, −, +, +, −)	0.583	0.058	21
	Pr(+|+, −, +, +, +)	0.579	0.045	20
	Pr(+|+, +, −, −, −)	0.557	0.035	13
	Pr(+|+, +, −, −, +)	0.565	0.045	12
	Pr(+|+, +, −, +, −)	0.558	0.055	12
	Pr(+|+, +, −, +, +)	0.551	0.049	11
	Pr(+|+, +, +, −, −)	0.555	0.037	11
	Pr(+|+, +, +, −, +)	0.567	0.043	13
	Pr(+|+, +, +, +, −)	0.562	0.046	17
	Pr(+|+, +, +, +, +)	0.546	0.040	15

Table A.11
Conditional probabilities of market directional movements for 1 min timespan.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.494	0.015	n/a
1	Pr(+|−)	0.422	0.019	51
	Pr(+|+)	0.568	0.017	50
2	Pr(+|−, −)	0.444	0.022	38
	Pr(+|−, +)	0.391	0.030	51
	Pr(+|+, −)	0.599	0.027	52
	Pr(+|+, +)	0.544	0.022	30
3	Pr(+|−, −, −)	0.449	0.028	21
	Pr(+|−, −, +)	0.438	0.030	28
	Pr(+|−, +, −)	0.398	0.042	36
	Pr(+|−, +, +)	0.386	0.037	46
	Pr(+|+, −, −)	0.608	0.031	48
	Pr(+|+, −, +)	0.584	0.039	27
	Pr(+|+, +, −)	0.547	0.037	18
	Pr(+|+, +, +)	0.542	0.027	16
4	Pr(+|−, −, −, −)	0.459	0.037	15
	Pr(+|−, −, −, +)	0.435	0.043	16
	Pr(+|−, −, +, −)	0.435	0.054	13
	Pr(+|−, −, +, +)	0.440	0.039	17
	Pr(+|−, +, −, −)	0.407	0.050	28
	Pr(+|−, +, −, +)	0.387	0.069	24
	Pr(+|−, +, +, −)	0.404	0.049	31
	Pr(+|−, +, +, +)	0.371	0.041	44
	Pr(+|+, −, −, −)	0.614	0.038	41
	Pr(+|+, −, −, +)	0.600	0.047	26
	Pr(+|+, −, +, −)	0.604	0.071	20
	Pr(+|+, −, +, +)	0.570	0.049	14
	Pr(+|+, +, −, −)	0.544	0.044	13
	Pr(+|+, +, −, +)	0.550	0.057	11
	Pr(+|+, +, +, −)	0.563	0.041	13
	Pr(+|+, +, +, +)	0.524	0.039	6
5	Pr(+|−, −, −, −, −)	0.465	0.050	6
	Pr(+|−, −, −, −, +)	0.452	0.063	11
	Pr(+|−, −, −, +, −)	0.443	0.062	6
	Pr(+|−, −, −, +, +)	0.430	0.057	17
	Pr(+|−, −, +, −, −)	0.449	0.064	6
	Pr(+|−, −, +, −, +)	0.418	0.094	0
	Pr(+|−, −, +, +, −)	0.450	0.064	9
	Pr(+|−, −, +, +, +)	0.433	0.057	11
	Pr(+|−, +, −, −, −)	0.415	0.067	14
	Pr(+|−, +, −, −, +)	0.397	0.067	11
	Pr(+|−, +, −, +, −)	0.392	0.109	10
	Pr(+|−, +, −, +, +)	0.383	0.084	21
	Pr(+|−, +, +, −, −)	0.415	0.070	19
	Pr(+|−, +, +, −, +)	0.387	0.073	17
	Pr(+|−, +, +, +, −)	0.387	0.070	23
	Pr(+|−, +, +, +, +)	0.359	0.052	39
	Pr(+|+, −, −, −, −)	0.629	0.054	31
	Pr(+|+, −, −, −, +)	0.596	0.050	14
	Pr(+|+, −, −, +, −)	0.612	0.085	17
	Pr(+|+, −, −, +, +)	0.591	0.055	16
	Pr(+|+, −, +, −, −)	0.618	0.090	20
	Pr(+|+, −, +, −, +)	0.585	0.115	9
	Pr(+|+, −, +, +, −)	0.585	0.076	11
	Pr(+|+, −, +, +, +)	0.558	0.071	8
	Pr(+|+, +, −, −, −)	0.544	0.059	10
	Pr(+|+, +, −, −, +)	0.546	0.062	5
	Pr(+|+, +, −, +, −)	0.575	0.098	13
	Pr(+|+, +, −, +, +)	0.533	0.068	4
	Pr(+|+, +, +, −, −)	0.566	0.051	12
	Pr(+|+, +, +, −, +)	0.560	0.073	9
	Pr(+|+, +, +, +, −)	0.538	0.067	8
	Pr(+|+, +, +, +, +)	0.513	0.055	2

Table A.12
Conditional probabilities of market directional movements for 5 min timespan.

Depth	Event	Mean	SD	# weeks that are stat. sig. from Pr(+)
0	Pr(+)	0.493	0.030	n/a
1	Pr(+|−)	0.443	0.039	21
	Pr(+|+)	0.546	0.030	9
2	Pr(+|−, −)	0.471	0.054	6
	Pr(+|−, +)	0.408	0.054	19
	Pr(+|+, −)	0.591	0.051	19
	Pr(+|+, +)	0.508	0.044	2
3	Pr(+|−, −, −)	0.493	0.072	3
	Pr(+|−, −, +)	0.446	0.085	12
	Pr(+|−, +, −)	0.429	0.086	10
	Pr(+|−, +, +)	0.392	0.073	16
	Pr(+|+, −, −)	0.581	0.069	10
	Pr(+|+, −, +)	0.609	0.092	12
	Pr(+|+, +, −)	0.523	0.077	4
	Pr(+|+, +, +)	0.496	0.056	1
4	Pr(+|−, −, −, −)	0.515	0.091	2
	Pr(+|−, −, −, +)	0.468	0.102	4
	Pr(+|−, −, +, −)	0.479	0.125	4
	Pr(+|−, −, +, +)	0.425	0.104	10
	Pr(+|−, +, −, −)	0.450	0.119	6
	Pr(+|−, +, −, +)	0.392	0.142	11
	Pr(+|−, +, +, −)	0.395	0.123	10
	Pr(+|−, +, +, +)	0.384	0.096	15
	Pr(+|+, −, −, −)	0.582	0.103	9
	Pr(+|+, −, −, +)	0.577	0.100	7
	Pr(+|+, −, +, −)	0.637	0.139	13
	Pr(+|+, −, +, +)	0.583	0.126	8
	Pr(+|+, +, −, −)	0.547	0.105	4
	Pr(+|+, +, −, +)	0.490	0.104	2
	Pr(+|+, +, +, −)	0.538	0.085	2
	Pr(+|+, +, +, +)	0.455	0.081	1
5	Pr(+|−, −, −, −, −)	0.535	0.134	2
	Pr(+|−, −, −, −, +)	0.498	0.118	3
	Pr(+|−, −, −, +, −)	0.474	0.211	13
	Pr(+|−, −, −, +, +)	0.466	0.130	5
	Pr(+|−, −, +, −, −)	0.506	0.150	5
	Pr(+|−, −, +, −, +)	0.427	0.204	9
	Pr(+|−, −, +, +, −)	0.414	0.136	5
	Pr(+|−, −, +, +, +)	0.432	0.139	10
	Pr(+|−, +, −, −, −)	0.459	0.154	7
	Pr(+|−, +, −, −, +)	0.453	0.220	11
	Pr(+|−, +, −, +, −)	0.429	0.258	12
	Pr(+|−, +, −, +, +)	0.370	0.172	9
	Pr(+|−, +, +, −, −)	0.406	0.148	10
	Pr(+|−, +, +, −, +)	0.380	0.166	11
	Pr(+|−, +, +, +, −)	0.393	0.147	9
	Pr(+|−, +, +, +, +)	0.371	0.143	13
	Pr(+|+, −, −, −, −)	0.586	0.142	7
	Pr(+|+, −, −, −, +)	0.575	0.149	7
	Pr(+|+, −, −, +, −)	0.616	0.166	11
	Pr(+|+, −, −, +, +)	0.545	0.142	4
	Pr(+|+, −, +, −, −)	0.621	0.179	13
	Pr(+|+, −, +, −, +)	0.636	0.288	22
	Pr(+|+, −, +, +, −)	0.596	0.210	14
	Pr(+|+, −, +, +, +)	0.580	0.149	5
	Pr(+|+, +, −, −, −)	0.545	0.146	5
	Pr(+|+, +, −, −, +)	0.547	0.159	8
	Pr(+|+, +, −, +, −)	0.502	0.141	1
	Pr(+|+, +, −, +, +)	0.483	0.159	6
	Pr(+|+, +, +, −, −)	0.536	0.122	2
	Pr(+|+, +, +, −, +)	0.538	0.160	4
	Pr(+|+, +, +, +, −)	0.498	0.105	0
	Pr(+|+, +, +, +, +)	0.396	0.150	7

References

[1] E.F. Fama, The behavior of stock-market prices, Journal of Business 38 (1) (1965).
[2] S.S. Alexander, Price movement in speculative markets: trend or random walks, No. 2, in: The Random Character of Stock Market Prices, MIT Press, Cambridge, MA, 1964, pp. 338–372.
[3] Burton G. Malkiel, A Random Walk Down Wall Street, second ed., W. W. Norton & Company, 2003.
[4] A. Timmermann, C.W.J. Granger, Efficient market hypothesis and forecasting, International Journal of Forecasting 20 (1) (2004) 15–27.
[5] C. Neely, P. Weller, R. Dittmar, Is technical analysis in the foreign exchange market profitable? A genetic programming approach, Journal of Financial and Quantitative Analysis 32 (04) (1997) 405–426.
[6] P.H.K Chang, C.L. Osler, Methodical madness: technical analysis and the irrationality of exchange rate forecasts, The Economic Journal 109 (458) (1999) 636–661.
[7] Richard M. Levich, Lee R. Thomas, The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach, Journal of International Money and Finance 12 (5) (1993) 451–474.
[8] Richard J. Sweeney, Some new filter rule tests: methods and results, The Journal of Financial and Quantitative Analysis 23 (3) (1988) 285–300.
[9] A. Timmermann, Elusive return predictability, International Journal of Forecasting 24 (1) (2008) 1–18.
[10] Toru Ohira, Naoya Sazuka, Kouhei Marumo, Tokiko Shimizu, Misako Takayasu, Hideki Takayasu, Predictability of currency market exchange, Physica A: Statistical Mechanics and its Applications 308 (1–4) (2002) 368–374.
[11] M. Tanaka-Yamawaki, Stability of Markovian structure observed in high frequency foreign exchange data, Annals of the Institute of Statistical Mathematics 55 (2) (2003) 437–446.
[12] N. Sazuka, Analysis of binarized high frequency financial data, The European Physical Journal B. Condensed Matter and Complex Systems 50 (1) (2006) 129–131.
[13] Y. Hashimoto, T. Ito, T. Ohnishi, M. Takayasu, H. Takayasu, T. Watanabe, Random Walk or a Run: Market Microstructure Analysis of the Foreign Exchange Fate Movements Based on Conditional Probability, National Bureau of Economic Research, Cambridge, Mass., USA, 2008.
[14] V. Niederhoffer, M.F.M. Osborne, Market making and reversal on the stock exchange, Journal of the American Statistical Association 61 (316) (1966) 897–916.
[15] H. Bessembinder, Bid-Ask spreads in the interbank foreign exchange markets, Journal of Financial Economics 35 (3) (1994) 317–348.
[16] Juan Reboredo, José Matías, Raquel Garcia-Rubio, Nonlinearity in forecasting of high-frequency stock returns, Computational Economics 40 (3) (2012) 245–264.
[17] Tao Wang, Jian Yang, Nonlinearity and intraday efficiency tests on energy futures markets, Energy Economics 32 (2) (2010) 496–503.
[18] Richard Roll, A simple implicit measure of the effective Bid–Ask spread in an efficient market, The Journal of Finance 39 (4) (1984) 1127–1139.
[19] R.C.A. Oomen, Properties of realized variance under alternative sampling schemes, Journal of Business and Economic Statistics 24 (2) (2006) 219–237.
[20] R.F. Engle, J.R. Russell, Autoregressive conditional duration: a new model for irregularly spaced transaction data, Econometrica (1998) 1127–1162.
[21] C.M.C. Lee, M.J. Ready, Inferring trade direction from intraday data, Journal of Finance (1991) 733–746.
[22] E. Rahm, H.H. Do, Data Cleaning: Problems and Current Approaches, Bulletin of the Technical Committee, 2000, p. 3.