A study of persistence of price movement using High Frequency Financial Data

Dragos Bozdog[1], Ionuţ Florescu, Khaldoun Khashanah, Jim Wang

Stevens Institute of Technology, Hoboken, NJ, 07030 USA

ABSTRACT: We present a methodology to discover how often the price deviates from established levels without sound reason. We look at the distribution of these so called "rare events" during the trading day and analyze the speed of the price bouncing back to the levels established before the rare events. We provide methods to calibrate trading rules based on the detection of these events and we exemplify for a particular trading rule. In order to draw comprehensive conclusions we group the equity into classes depending on the volume of daily trades. We find the behavior these classes to be very different. We also find possible evidence of algorithmic trading related to these events.

---

1. INTRODUCTION

When studying price dynamics, the price-volume relationship is one of the most studied in the field of finance. Perhaps the oldest model used to study this relationship is the work of Osborne (1959) who models the price as a diffusion process with its variance dependent on the quantity of transaction at that particular moment. Subsequent relevant work can be found in Karpoff (1987), Gallant, Rossi, and Tauchen (1992), Bollerslev and Jubinski (1999), Lo and Wang (2002), and Sun (2003). In general this line of work studies the relationship between volume and some measure of variability of the stock price (e.g., the absolute deviation, the volatility, etc.). Most of these articles use models in time, they are tested with low frequency data and the main conclusion is that the price of a specific equity exhibits larger variability in response to increased volume of trades. We also mention the Autoregressive Conditional Duration (ACD) model of Engle & Russell (1998) which considers the time between trades as a variable related to both price and volume. In the current work we examine the relationship between change in price and volume. We study the exception of the conclusion presented in the earlier literature. In our study we do not consider models in time but rather make the change in price dependent on the volume directly.

The old Wall Street adage that "it takes volume to move prices" is verified in this empirical study. Indeed, this relationship was studied using market microstructure models and it was generally found true (Admati and Pfleiderer, 1988, Foster and Viswanathan, 1990, Llorente et. al., 2002). The advent of electronic trading using high frequency data, the increase in the trading volume and the recent research in automatic liquidation of large orders may lead to inconsistencies and temporary contradictions of this statement. For short time periods during trading we may encounter large price movements with small volume. However, if the claim is true then large price movements associated with small volume should be only temporary and the market should regain the momentum it had exhibited before the fleeting price movement.

This is the premise of the current study. We propose a methodology to detect outlying observations of the price-volume relationship. We may refer to these outliers as rare events in high frequency finance or rare micro-events to distinguish them from rare events for low frequency sampled data. In our context, due to the

---

[1] Corresponding author. Email: dbozdog@stevens.edu Tel (201) 216-5452, Fax (201) 216-8321
Email contact for the authors: ifloresc@stevens.edu, kkhashan@stevens.edu, Jim.Wang@stevens.edu

joint price-volume distribution, we may encounter two types of outliers. The first type occurs when the volume of traded shares is small but is associated with large price movement. The second type occurs when the volume of traded shares is large coupled with small price movement. Of the two types of rare events, we are only interested in the first type. The second type is evidence of unusually high trading activity which is normally accompanied with public information release (a well documented event as early as (Beaver, 1968)). We formulate the main objectives of this work as follows.

Objectives:

- Develop a method to detect rare events in real time where the movement of price is large with relatively small volume of shares traded
- Analyze the price behavior *after* these rare events and study the probability of price recovery. What is the expected return if a trade is placed at the detected observation?

The second objective is of particular interest to us. Recent research (Alfonsi, Schied and Schulz, 2007, Zhang, Russell and Tsay, 2008) analyze ways of liquidating a large order by splitting it into smaller orders to be spread over a certain period of time. There are several available strategies to achieve this objective. However, all strategies make one or several assumptions about the dynamic or structure of the limit order book. One specific assumption seems to be common in the literature and that is to assume a degree of elasticity/plasticity of the limit orders, i.e., the capability of the bid/ask orders to regain the previous levels after a large order has been executed. This elasticity degree is usually assumed as given but there are no methods which actually estimate the *current* nature of the market when the large order is executed, immediately *before* the liquidating strategy is being put into place. We believe that our second objective provides a way to estimate the current market conditions at the time when an outlying observation is detected. In particular, we believe that the frequency of these rare events relative to the market total trade volume sheds light about the current market condition as well as the particular equity being researched.

The article is structured as follows. In Section 2 we present the basic methodology for detecting and evaluating the rare events. Section 3 details results obtained applying the methodology to tick data collected over a period of five trading days in April, 2008. Section 4 presents the distribution of the trades and the rare events during the trading day. Section 5 presents conclusions drawn using our methodology.

2. METHODOLOGY

In this analysis we use tick-by-tick data of 5,369 equities traded on NYSE, NASDAQ and AMEX for a five day period. We need the most detailed possible dataset; however, since our discovery is limited to past trades we do not require the use of a more detailed level 2 order data. We perform model free statistical analysis on this multivariate dataset.

For any given equity in the dataset an observation represents a trade. Each trade records the price P of the transaction, the volume V of the shares traded and the time t at which the transaction takes place. In this study we are primarily interested in large price movement with small volume, thus for **any** two observations in the dataset we construct a four dimensional random vector $(\Delta P, \Delta V, \Delta N, \Delta t)$. Here ΔP is the change in price, ΔV is the change in volume, ΔN is the number of trades, and Δt is the period of time all variables calculated between the two trades. The number of trades elapsed between two observations is a variable that may be calculated using the given dataset.

The reason for considering any pair of trades and not only consecutive trades is that in general the price movement occurs over several consecutive trades. The main object of our study is the conditional distribution:

$$h(Max(\Delta P)|\, \Delta V < V_0)$$

i.e., the maximum price movement given the cumulative volume between two trades is less than a value V_0 specific to each equity. The study of this distribution will answer the specific questions asked in the beginning of this paper.

2.1 Justification of the method

1. Why restricting the distribution conditional on V_0?

According to our declared objective, we are interested in price movement corresponding to small volume. Therefore, by conditioning the distribution we are capable of providing answers while keeping the number of computations manageable.

2. Why should V_0 be constant in time and only depend on the equity?

Indeed, this is a very important question. There is no reason for V_0 to be constant other than practical reasons. A valid objection is that the dynamics of the equity change in time. A time changing model is beyond the scope of the current study, though in this work we investigate several (fixed) levels of this parameter.

3. Why not the more traditional approach of price and volume evolution in time?

First the price evolution in time will not answer the questions asked. Furthermore, the volume of traded shares changes predictably during the day. In general heightened trading activity may be observed at the beginning and the end of the trading day due to pre-market trading activity, rebalancing of portfolio positions and other factors. By tracking a window in volume we are unaffected by these changes in trading behavior. The net consequence is a change in time duration of the volume window which is irrelevant for our study.

2.2 Sampling method. Rare event detection

Consider the current trade S_n for a certain equity. Construct the sequence of consecutive trades $S_k, S_{k+1}, \dots, S_n$ and their associate volumes $v_k, v_{k+1}, \dots, v_n$, such that $v_k + v_{k+1} + \cdots + v_n < V_0$. Then let

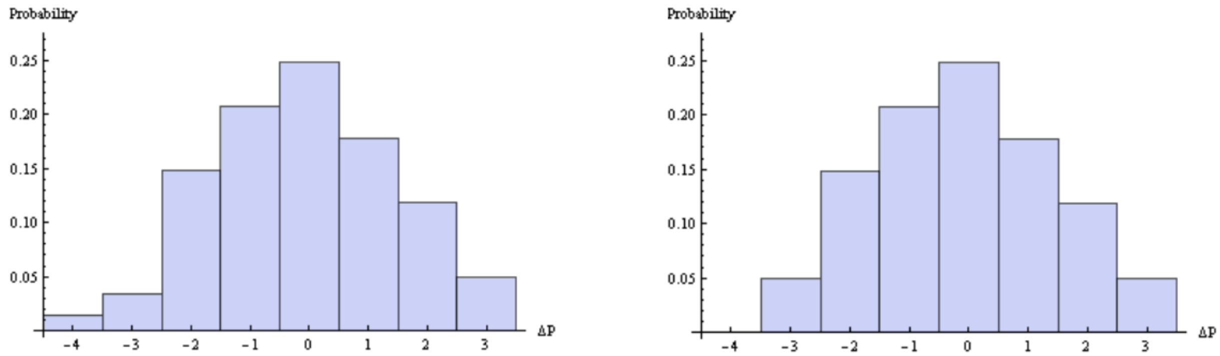
$$\Delta p_n = max\{S_n - S_k, S_n - S_{k+1}, \dots, S_n - S_{n-1}\}$$

We repeat the process for every trade by calculating a corresponding maximum price movement within the last V_0 trades. Once we obtain these values for the entire sequence of trades we detect the extreme observations by applying a simple "quantile type" rule. Namely, for a fixed level α we select all the observations in the set:

$$Q_\alpha^+(x) = \{x: Prob(\Delta p < x) < \alpha \text{ or } Prob(\Delta p > x) > 1 - \alpha\} \quad (1)$$

The probability above is approximated using the constructed histogram of maximum price movements. We note that the rule above is different than the traditional quantile definition which uses non-strict inequalities. The modification above is imposed by the specific nature of the tick data under study (i.e., discrete data).

For illustration consider the two distributions of the price change Δp[cents] in Fig. 1. Suppose we are interested in rare events that occur with probability $\alpha = 0.015$. The rule in (1) will select the observations corresponding to $x = -4$ for distribution in Fig. 1a and no observation in Fig. 1b. A traditional quantile rule for any level $\alpha \leq 0.015$ no matter how small will indeed select the observations corresponding to $x = -4$ for the distribution in Fig. 1a however for the distribution in Fig. 1b will select all the observations at $x = -3$ and $x = 3$. Therefore, using a traditional quantile rule would force us to analyze points from distributions which lack extreme observations.



(a) Distribution with one interesting observation (b) Distribution lacking interesting observations

Fig. 1: Two price change distributions

Note: Using rule (1) with returns instead of change in price will be preferable in a trading environment. We use change in price (Δp) for clarity of exposition.

A discussion about the appropriateness of the rule of detecting rare events

Our rule is nonstandard and further discussion is necessary. We first note that due to the way Δp_n quantities are constructed they are not independent. Thus their histogram is only an approximation of the true probabilities of price movement. However, since we are only interested in extreme price movement, rule (1) will identify candidate rare events which may or may not correspond to the true probability level α. We may have a better depiction of the true histogram of the price movement by considering non-overlapping windows. There are two reasons why this is not feasible. First, by considering non-overlapping windows we may lose extreme price differences calculated using prices from these non-overlapping windows. Second, in a previous study (Mariani et. al., 2009) the authors have shown that returns calculated from tick data exhibit long memory behavior. Thus, even by considering non-overlapping windows one cannot guarantee that the observations are independent.

Furthermore, why do we use our rule and not a more traditional rule for detecting outliers such as 1.5xIQR rule or a parametric outlier test? A parametric detection rule does not make sense in our context since we do not want to hypothesize an underlying statistical model. The Inter-Quartile Range (IQR) rule is useful for outlier detection not rare events. It is essentially equivalent with our rule since it uses quantiles but it is a very rigid rule. In general, it does not find outliers very often for fat tailed distributions (such as the ones under study here).

2.3 Rare event analysis. Choosing the optimal level α

After we obtain the rare event candidates, we need to develop a systematic methodology to evaluate them. According to our assumption the movement in price is abnormal and the equity should recover and reverse its momentum. We assume that a trade is placed at the time when a rare event is discovered. We consider a limited volume window (called the *after-event window*) and we analyze the price behavior.

Definition 1: *We say that a favorable price movement occurs for a fixed rare event if either*

- *the price level within the after-event window raises above the event price for at least one trade if the event was generated by a negative value for rule (1), or*
- *the price level within the after-event window decreases below the event price level for at least one trade if the event was generated by a positive value for rule (1).*

This definition allows to estimate the probability of a favorable price movement for a specific level α. Specifically, if n is the total number of rare events detected by rule (1) and k is the number of favorable price movements among them then the probability desired is simply k/n. As we shall see this definition allows the optimal selection of the level α. As the level α increases the events will stop being rare and just plain events.

Definition 1 does not allow the selection of the optimal volume window size V_0 or the optimal after-event window size. To investigate this selection we consider the return on a trade. To this end we consider the following strategy:

- A trade is placed at every rare event, long or short according to the sign of the quantile detected
- An after-event window size is fixed at the moment of the trade
- We close the position either during the after-event window if a favorable price movement takes place or at the last trade of the after-event window if a favorable price movement does not take place

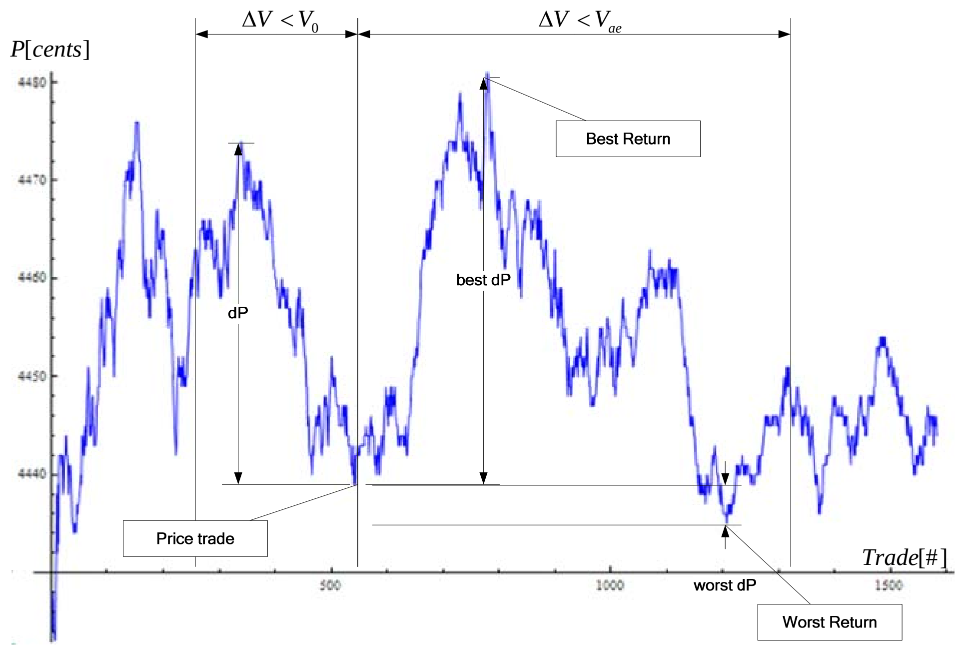


Fig. 2 Visual depiction of the quantities used in the study

The return of such a strategy depends on the price at which the position is closed during the after event window. To determine the optimal window size and optimal α level we use the following trading strategy.

Definition 2: *A position is opened at a point determined according to rule (1). The position is closed according to the following:*

- *If a favorable price movement takes place in the after-event window we close the position using the best return possible.*
- *If a favorable price movement does not take place within the after-event window we close the position using the worst return possible within the window.*

For a certain level α and an after event window size V_{ae} we calculate the expected return by averaging all the trade returns placed following the above strategy.

We note that we shall use the trading rule in Definition 2 only for determining optimal level α and window size. In practice, using back-testing and strategy calibration will determine a satisfactory favorable price movement and the position will be closed as soon as that level is reached.

2.4 Multi-scale volume classification

Econometric analysis traditionally distinguishes between results obtained for highly traded stocks versus less frequently traded equities. Most of the studies are focused on what are called large capitalization equities which are defined as having market capitalizations larger than a specified cutoff. This definition is often vague, varies over the years and, more importantly, does not necessarily have direct relevance to trading patterns. For example, an equity traditionally classified as a large-cap stock may have a small Average Daily Volume (ADV) and since the later is essential for us we use a different nomenclature based directly on ADV. The results obtained for a highly liquid equity do not necessarily hold true for less liquid stocks even if both belong to the same capitalization class. Herein, we analyze the change in price from the volume perspective; therefore, we recognize the need for classifying equities into classes based on the average daily traded volume. We refer to this classification as the *multi-scale volume classification*.

The histogram in Figure 3 corresponds to the average daily trading volume (ADV) of the total universe of 5,369 equities considered in this study.

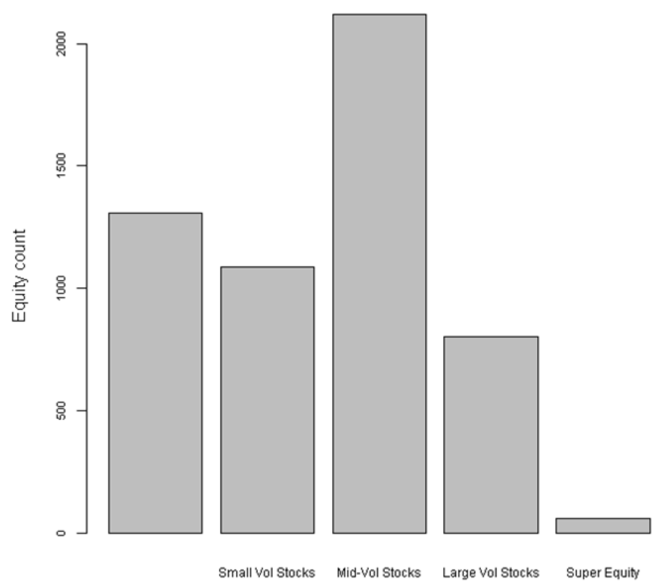


Fig. 3 Average daily volume distribution

The distribution of the average daily volume among the stocks is skewed to the right and our selection criterion follows certain features. As a preliminary step in our analysis, we need to eliminate all equities with average daily volume below 30,000 shares. The 30,000 volume cutoff value is not arbitrary, but it is found to be the minimum level required to perform our analysis. These stocks are grouped in class index 1 and are not used in any of the further analysis. The highest ADV values are concentrated around major indexes and large capitalization equities with more than 10 million shares traded daily. The three intermediary classes contain large, medium and small average daily volume stocks. The resulting five classes in our multi-scale volume classification are summarized in Table 1.

Table 1. Equities partitioned into 5 classes

	Class	Average daily volume (shares)	Number equities
1		$ADV \leq 30{,}000$	1,305
2	Small-Vol Stocks	$30{,}000 < ADV \leq 100{,}000$	1,088
3	Mid-Vol Stocks	$100{,}000 < ADV \leq 1{,}000{,}000$	2,117
4	Large-Vol Stocks	$1{,}000{,}000 < ADV \leq 10{,}000{,}000$	799
5	Super Equity	$10{,}000{,}000 < ADV$	60

3. RESULTS

The methodology described in Section 2 is applied to all the equity data within a class in a homogeneous way. For this purpose we combine all the outlying events detected according to rule (1) within each class. Table 2 presents the probabilities of a favorable price movement according to Definition 1.

We note that to calculate the probability of favorable price movement as in Definition 1 we need to specify a level α for the detection rule, a volume level V_0 as well as an after event volume size (V_{ae}). To analyze the optimal choices of these parameters, Table 2 presents the results obtained for a discrete set of parameters. Specifically, we look at $\alpha \in \{0.02, 0.015, 0.01, 0.005, 0.002, 0.0015, 0.001, 0.0005, 0.0002\}$, $V_0 \in \{3000, 5000, 10000\}$ and $V_{ae} = k * V_0$, where $k \in \{1, 2, 3\}$.

Table 2. Probability (%) of favorable price movement for equity classes for all days

Class	α level for rule (1)	$V_0 = 3{,}000$			$V_0 = 5{,}000$			$V_0 = 10{,}000$		
		V_{ae} (shares)			V_{ae} (shares)			V_{ae} (shares)		
		3,000	6,000	9,000	5,000	10,000	15,000	10,000	20,000	30,000
Small-Vol Stocks	0.02	84.13	88.97	91.05	89.00	92.68	94.06	93.80	95.73	96.36
	0.015	85.30	90.00	91.78	90.17	93.43	94.66	94.68	96.46	97.00
	0.01	86.68	91.23	92.66	91.83	94.62	95.68	95.02	96.83	97.30
	0.005	89.52	93.48	94.59	94.07	96.08	97.14	96.81	97.75	98.05
	0.002	92.68	95.59	96.63	96.33	98.16	98.82	98.22	98.57	98.75
	0.0015	93.72	96.03	96.86	95.53	97.65	98.82	98.77	99.08	99.08
	0.001	94.52	97.26	98.63	97.46	99.15	99.15	98.63	100.00	100.00
	0.0005	na	na	na	na	na	na	na	na	na
	0.0002	na	na	na	na	na	na	na	na	na
Mid-Vol Stocks	0.02	78.48	84.82	87.54	83.39	88.28	90.35	88.84	92.15	93.40
	0.015	78.85	85.09	87.71	83.70	88.55	90.54	89.28	92.42	93.62
	0.01	79.35	85.43	88.03	84.58	89.23	91.17	90.05	93.06	94.12
	0.005	81.24	86.95	89.28	86.37	90.44	92.31	91.76	94.27	95.03
	0.002	84.65	89.32	91.20	89.70	92.82	94.34	94.08	96.11	96.56
	0.0015	85.82	90.42	92.12	90.96	93.58	94.96	94.78	96.45	96.81
	0.001	86.98	91.25	92.73	91.71	94.07	95.36	95.56	97.07	97.38
	0.0005	88.91	92.78	93.88	93.21	94.93	96.11	96.46	97.65	97.87
	0.0002	88.87	92.23	93.49	94.28	95.65	97.25	97.58	98.07	98.07
Large-Vol Stocks	0.02	76.54	83.14	86.14	80.55	86.19	88.73	85.47	89.85	91.79
	0.015	76.82	83.36	86.29	80.99	86.49	88.99	85.80	90.12	92.04
	0.01	77.29	83.72	86.58	81.46	86.77	89.23	86.21	90.45	92.37
	0.005	78.31	84.46	87.06	82.40	87.49	89.78	86.99	91.09	92.90
	0.002	80.50	85.98	88.30	84.05	88.71	90.76	88.78	92.66	94.09
	0.0015	81.47	86.72	88.87	84.94	89.35	91.25	89.69	93.51	94.72
	0.001	82.69	87.74	89.82	86.20	90.32	92.17	91.26	94.52	95.43

	0.0005	85.42	89.62	91.58	89.28	92.55	94.09	93.05	95.49	96.18
	0.0002	88.23	92.01	93.64	92.67	95.17	96.27	95.17	96.66	96.93
Super Equity	0.02	71.75	79.76	83.52	77.36	83.99	87.05	81.49	86.93	89.21
	0.015	72.36	80.43	84.09	77.46	84.03	87.12	81.83	87.23	89.59
	0.01	74.10	81.90	85.28	78.00	84.57	87.68	83.03	88.04	90.29
	0.005	74.87	82.73	86.07	78.72	85.24	88.12	83.73	88.32	90.64
	0.002	76.27	83.25	86.76	80.53	86.77	89.50	86.08	90.16	91.78
	0.0015	76.44	83.25	86.86	80.96	87.23	90.05	86.21	90.04	91.69
	0.001	77.59	84.50	88.15	82.60	88.32	90.60	86.96	90.77	92.29
	0.0005	79.40	86.06	88.76	84.36	89.57	91.37	87.22	90.43	92.09
	0.0002	81.59	87.91	90.11	84.97	89.64	91.97	91.41	93.43	94.95

For a better visualization and interpretation of these numbers we construct probability surfaces for each class and we plot them with respect to the α level and volume V_{ae} in Figure 4.

According to the Definition 1 we expect the probabilities to increase as the α level becomes more selective, as well as the size of the after-event window volume to increase. Indeed, we observe this behavior in Figure 4, but it is remarkable that the surfaces are parallel and smooth. This seems to indicate that the probability has a similar behavior for each class. Furthermore, by using a simple translation in α and V_{ae} we may be able to map each surface into another. This translation is very important because once we decide on a optimal level for one class it automatically translates into optimal levels for the other classes.

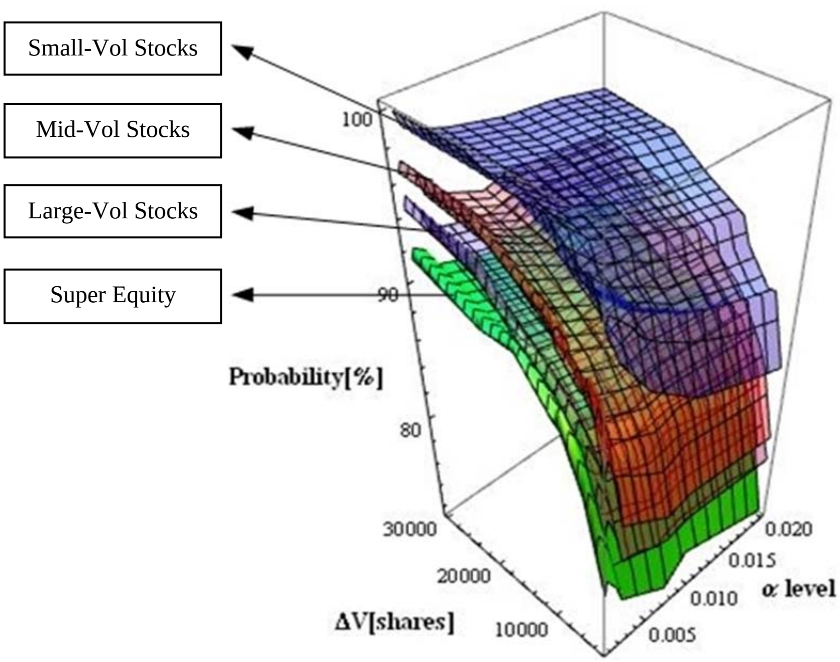


Fig. 4 Probability surfaces for equity classes.

To determine the optimal level for each class we calculate the expected return of trades according to the Definition 2. Specifically, for fixed levels of α and V_{ae}, we average all the returns within each class and present the results in Table 3. We also construct the corresponding surfaces in Figure 5.

Table 3. Expected return (%) for equity classes for all days

$$r = r_{fav} \cdot P + r_{unfav} \cdot (1 - P)$$

Class	α level for rule (1)	$V_0 = 3,000$			$V_0 = 5,000$			$V_0 = 10,000$		
		V_{ae} (shares)			V_{ae} (shares)			V_{ae} (shares)		
		3,000	6,000	9,000	5,000	10,000	15,000	10,000	20,000	30,000
Small-Vol Stocks	0.02	0.6119	0.8473	0.9963	0.8030	1.0562	1.2097	1.0696	1.3143	1.4380
	0.015	0.6570	0.8976	1.0513	0.8507	1.1065	1.2626	1.1268	1.3781	1.5093
	0.01	0.7026	0.9620	1.1199	0.9189	1.1847	1.3475	1.1751	1.4485	1.5832
	0.005	0.7900	1.0784	1.2490	1.0585	1.3585	1.5390	1.2947	1.5884	1.7266
	0.002	0.8072	1.0755	1.2359	1.0108	1.2934	1.4649	1.2195	1.5165	1.6656
	0.0015	0.7844	1.0292	1.1670	0.9820	1.2802	1.4746	1.2184	1.5356	1.6997
	0.001	0.7030	0.9506	1.0579	0.9299	1.2352	1.3898	1.0963	1.5649	1.7453
	0.0005	na	na	na	na	na	na	na	na	na
	0.0002	na	na	na	na	na	na	na	na	na
Mid-Vol Stocks	0.02	0.2396	0.3745	0.4643	0.3309	0.4821	0.5809	0.4685	0.6393	0.7437
	0.015	0.2529	0.3916	0.4819	0.3467	0.5014	0.6008	0.4871	0.6587	0.7647
	0.01	0.2693	0.4118	0.5051	0.3719	0.5294	0.6302	0.5182	0.6922	0.8008
	0.005	0.3111	0.4633	0.5603	0.4218	0.5861	0.6927	0.5794	0.7587	0.8686
	0.002	0.3732	0.5348	0.6350	0.4931	0.6697	0.7832	0.6581	0.8441	0.9541
	0.0015	0.3949	0.5591	0.6626	0.5066	0.6835	0.7944	0.6652	0.8518	0.9622
	0.001	0.4034	0.5639	0.6694	0.5085	0.6856	0.7959	0.6706	0.8582	0.9665
	0.0005	0.3829	0.5345	0.6368	0.4720	0.6392	0.7454	0.6160	0.7962	0.8951
	0.0002	0.2988	0.4230	0.5112	0.4053	0.5548	0.6495	0.5240	0.6733	0.7506
Large-Vol Stocks	0.02	0.0906	0.1385	0.1750	0.1171	0.1795	0.2263	0.1742	0.2596	0.3191
	0.015	0.0953	0.1461	0.1840	0.1248	0.1904	0.2387	0.1830	0.2707	0.3312
	0.01	0.1039	0.1593	0.1992	0.1362	0.2056	0.2552	0.1963	0.2867	0.3490
	0.005	0.1198	0.1824	0.2263	0.1570	0.2317	0.2850	0.2191	0.3144	0.3777
	0.002	0.1458	0.2151	0.2619	0.1843	0.2643	0.3190	0.2533	0.3598	0.4256
	0.0015	0.1578	0.2291	0.2761	0.1956	0.2784	0.3335	0.2672	0.3764	0.4442
	0.001	0.1694	0.2418	0.2894	0.2117	0.2971	0.3547	0.2895	0.4011	0.4696
	0.0005	0.1978	0.2771	0.3274	0.2474	0.3367	0.3966	0.3234	0.4387	0.5091
	0.0002	0.2289	0.3189	0.3771	0.2742	0.3761	0.4386	0.3595	0.4821	0.5581
Super Equity	0.02	0.0543	0.0721	0.0859	0.0666	0.0899	0.1081	0.0819	0.1130	0.1378
	0.015	0.0565	0.0762	0.0910	0.0659	0.0897	0.1072	0.0839	0.1178	0.1431
	0.01	0.0607	0.0828	0.0984	0.0646	0.0902	0.1083	0.0897	0.1272	0.1539
	0.005	0.0601	0.0833	0.1014	0.0636	0.0942	0.1143	0.0965	0.1334	0.1617
	0.002	0.0596	0.0829	0.1054	0.0708	0.1117	0.1343	0.1073	0.1508	0.1773
	0.0015	0.0615	0.0851	0.1092	0.0727	0.1167	0.1406	0.1149	0.1578	0.1865
	0.001	0.0659	0.0912	0.1172	0.0794	0.1263	0.1527	0.1175	0.1653	0.1942
	0.0005	0.0768	0.1111	0.1381	0.0811	0.1285	0.1555	0.1319	0.1807	0.2153
	0.0002	0.0881	0.1185	0.1502	0.0877	0.1436	0.1679	0.1490	0.1964	0.2470

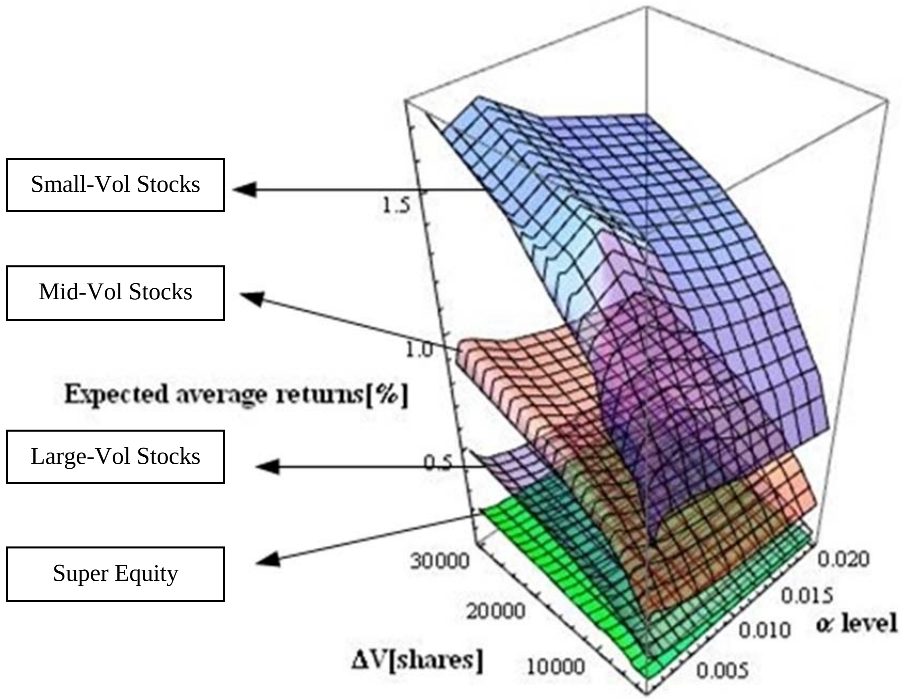


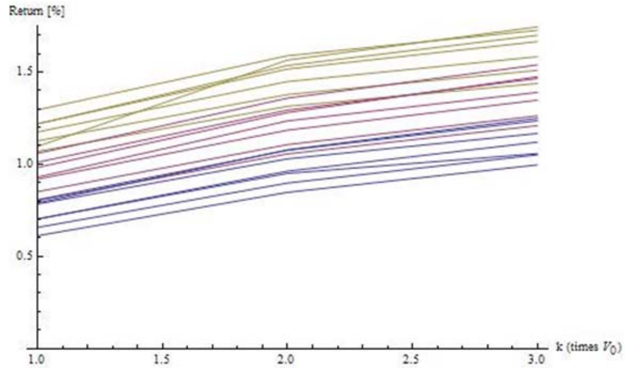
Fig. 5 Expected return surfaces for stock classes

Unlike the probability plots, the surfaces in Figure 5 have different curvatures. For each class surface we identify the α level which produces maximum return for each V_{ae}. First, unlike the probability surfaces which were decreasing in α the return surfaces have a maximum for each V_{ae}. Remarkably, within each class the maximum return is obtained for the same α level regardless of the V_{ae} value. The corresponding α level is thus construed as optimal. The following list presents these values.

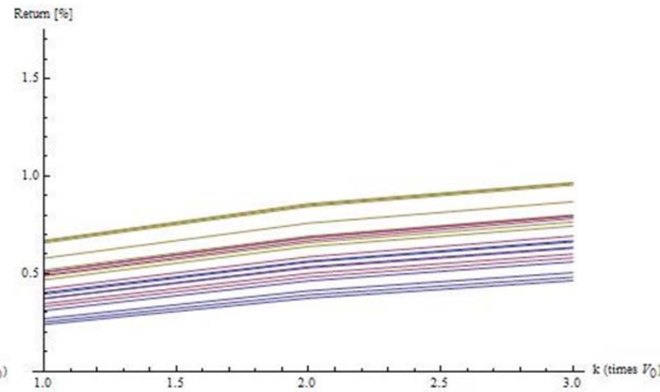
Class	Optimal level α
Small-Vol Stocks	0.0025
Mid-Vol Stocks	0.0005
Large-Vol Stocks	0.0001
Super Equity	less than 0.0001

The optimum α level is different for each surface and in general decreases as we consider larger ADV equities.

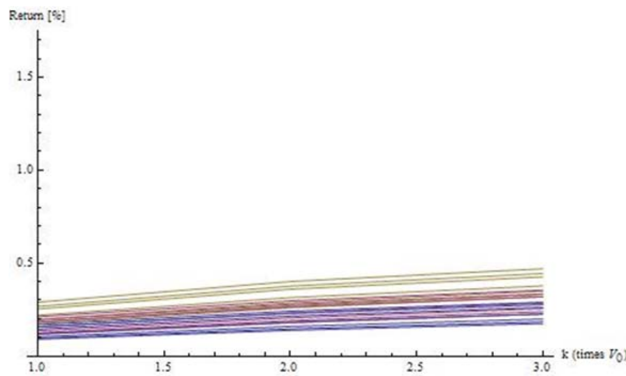
Once we have the optimal level α we analyze the 3D plot in more detail to determine the optimal V_0 and V_{ae} levels. The numbers in Table 3 tell us that in general the more we wait, the better the expected return. This however is an artifact due to the way we calculate the expected return (by taking the highest favorable value within the window). To calculate optimal values we consider projections of the 3D plot in Figures 6 and 7.



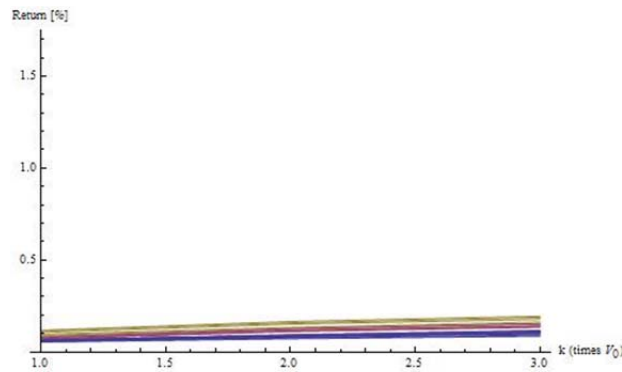
(a) Small cap stocks



(b) Medium cap Stocks

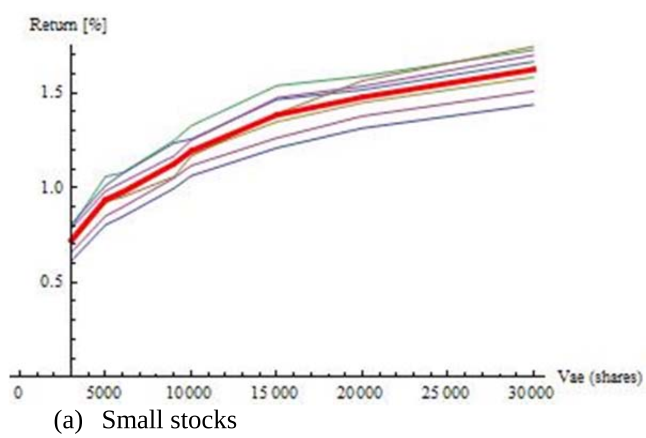


(c) Large cap stocks

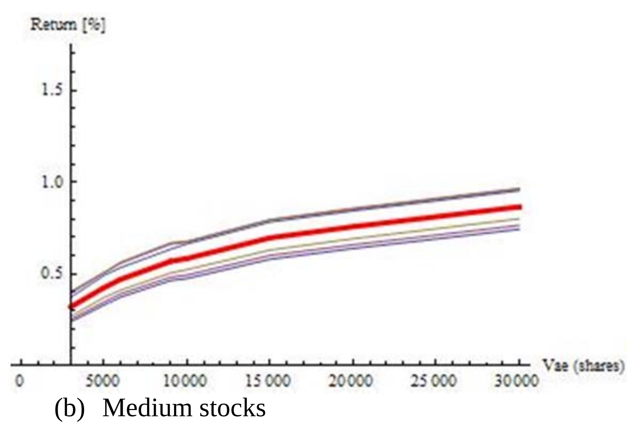


(d) Super Equity

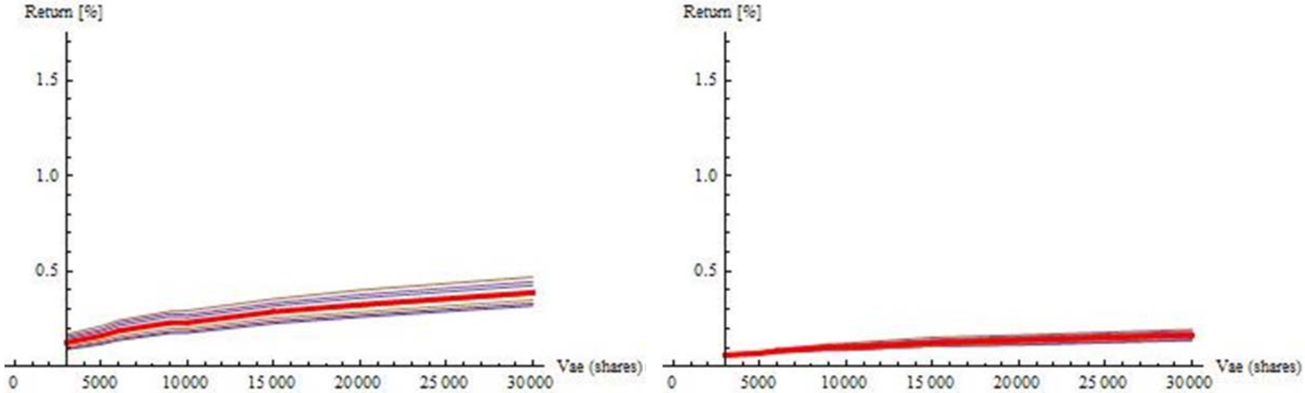
Fig. 6 Sectional 2D plots of the surfaces in Fig 5 for each of the quantile levels considered. Each subfigure represents one surface from Fig 5. The *x* axis is the proportion of after-event window size with respect to the before-event window size, and lines of the same color represent the three original window sizes chosen (blue for $V_0 = 3{,}000$, red for $V_0 = 5{,}000$, and yellow for $V_0 = 10{,}000$).



(a) Small stocks



(b) Medium stocks



(c) Large stocks (d) Super Equity

Fig. 7 Sectional 2D plots of the surfaces in Fig 5 for each of the quantile levels considered. The *x* axis is the value of *VAE* (in units of 100 shares). Each line represents a specific α level. The thicker red line is an average of all the returns for all α levels.

The analysis we need to perform is similar with a standard three way ANOVA. However, we avoid giving numerical values for the tests of interaction. The correlation between observations would cast a doubt on the validity of these numbers. Instead we prefer a graphical depiction of the values.

In Figure 6 we project the 3D plot onto the ratio V_{ae}/V_0. Each line represents a specific α level. With one exception the lines do not intersect. This means that there is little to no interaction between the α levels and the window sizes. Furthermore, using the same graphs we may determine if there is a significant increase in return as the ratio V_{ae}/V_0increases.

From this figure we deduce that for Small and Medium stocks it may pay to wait longer (after-event window size two or three times larger than the original). In contrast for Large and especially for Super Large equity the expected return does not appear to increase significantly by enlarging the after event window size (V_{ae}) with respect to the original window size V_0. In other words for highly traded stocks either the price bounces back very quickly or not at all.

Since the interaction was not found significant we proceed with Figure 7 where we plot return vs. V_{ae}. This figure provides an indication about the optimal after-event window size to use for each class of equity. Specifically, we look for points where the increase in return becomes negligible. Once again the lines are parallel (the level α factor and the windows size factor do not interact) thus we look at the average return for all quantile levels (the thicker red line in the image). Combining the results in Fig. 6 and 7 we may provide the following list of optimal values.

Class	Before event window size	After event window size
Small-Vol Stocks	5,000	15,000
Mid-Vol Stocks	5,000	15,000
Large-Vol Stocks	5,000	10,000
Super Equity	10,000	10,000

We emphasize that we give these values only as an example for these particular days and choice of classes. We observe that for small and medium volume stocks it takes a longer after event window for the price to recover. In contrast for the large volume stock and especially super-equity the price bounces back much faster.

4. RARE EVENTS DISTRIBUTION

In this section we analyze time distribution of the rare events during the trading day. Recall that the way we define the rare events may be viewed as a quantile of the two dimensional price-volume distribution. Since most of the trades today are small, it is natural to ask: are these rare events just a percentage of the trades or are they concentrated at certain periods during the day?

First we look at the distribution of trades within each minute for a full trading day. In Fig. 8 we present these distributions for small, mid, large volume, and super equity. They are constructed using the entire 5 days trading activity. The U-shape of these distributions is well documented in literature (Foster and Viswanathan, 1993) and (Gerety MS & Mulherin JH, 1992). However, in addition to these early studies we note the presence of high trading activity around 30 minutes after the market opening. This spike becomes more prevalent for large volume and super equity. As already noted, the trading activity is increasing towards the end of market trading hours but again we see that early market activity compared to end market activity is stronger when looking at the large and super-equity.

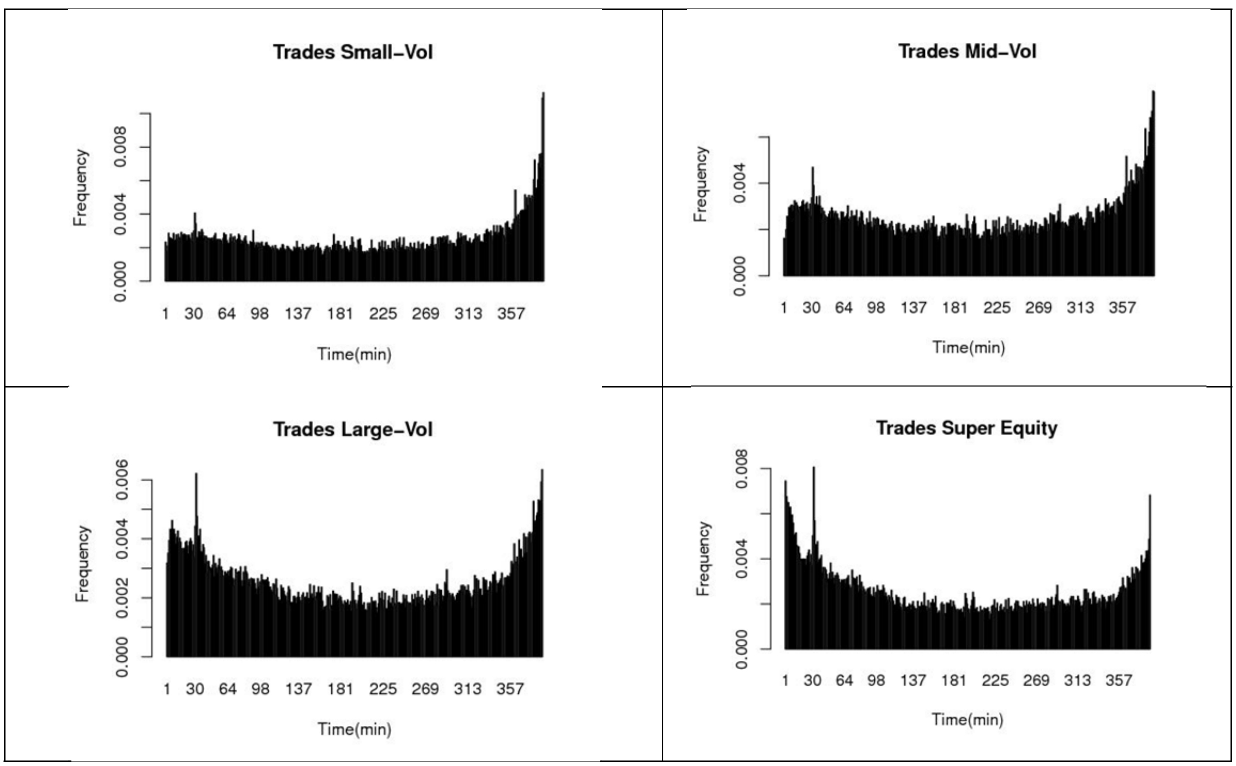


Fig. 8 Distribution of trades during the day for small, mid, large-vol, and super equity classes.

Next we look at the rare events corresponding to the level $\alpha = 0.02$. We construct histograms for the rare events distribution during the day for each of the V_0 levels considered and for each equity class. Fig. 9 presents these distributions. On each histogram we also represent the corresponding percentage level (0.02) of trades with a red line.

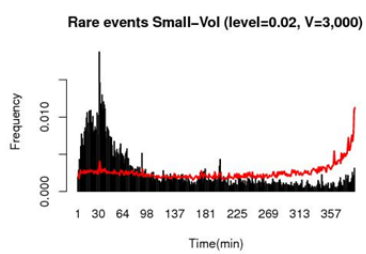
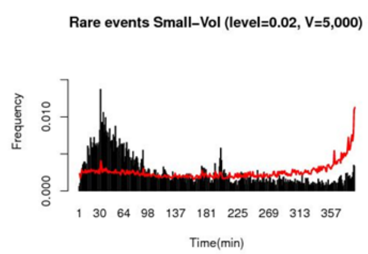
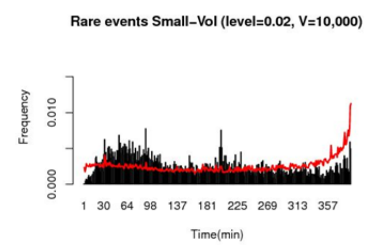
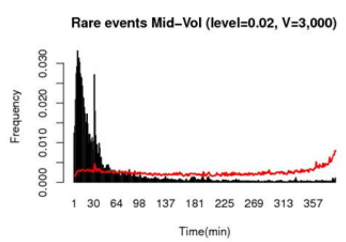
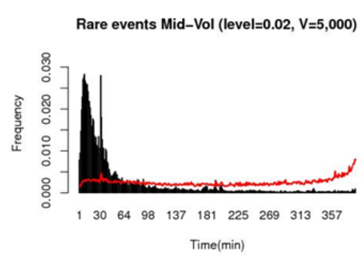
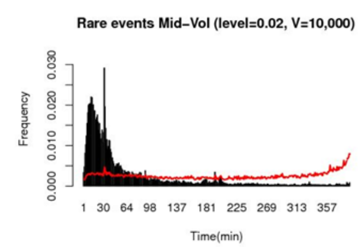
We first note that if the distribution of rare events would occur at a percentage rate of the trades, then the rare events distribution during the day would have had to follow approximately the red line. From the Figure it is clear that the rare events do not follow the profile of the trades. Furthermore, they are concentrated in a region close to the opening.

Second, we note the similarity of these distributions for a specific equity class (small vol, mid vol etc.). The skewness of the distributions decreases as V_0 increases and it is more significant for small and mid-vol stocks. This is easy to explain since this type of equities are more rarely traded and it takes a longer time period to detect the changes in equity using a 10,000 shares window versus a 3,000 shares window.

Third, we remark the presence of a peak in the distribution of rare events during the day after about 30 minutes of trading across all classes. This does correspond to the previously observed peak in trading activity (Figure 9) at about the same time. We hypothesize that the peak in rare events may be caused by the activation of various trading strategies after the stabilization of the market following the opening. Recall that the histogram presents the rare events detection for ALL equity within a class. This may be evidence of algorithmic trading starting at about the same time, reaching about the same conclusion, placing similar limit orders, and therefore pulling the market in the same direction with relatively little volume. We do underline however that this does not destabilizes the market. This much is evident by the ensuing pattern of rare events which follows the same trend as before the spike.

Four, we notice the presence of a significant number of rare events concentrated around noon for small and mid-vol equity. This is not evident in the trades distribution (Figure 9) and they correspond typically to the lowest level of volume traded during the day. We do not notice this concentration for the super-equity class. Somehow this period coincides with the end of lunch time (around 1:00) during the real time and this may show increased trader activity after regrouping and using the information accumulated during lunch. This hypothesis may actually be strengthen by the absence of activity in the large and supper equity since the human factor is much less present in this type of equity (plus trading in these equities comes from around the globe thus lunch time is meaningless).

If we look at the duration of time when the rare events are in excess of the red line we find periods of about 90 min for small-vol class, 60 min for mid-vol class, 40 min for large-vol class, and about 35 min for super-equity class.

<table>
<tr><th>Class</th><th>$V_0 = 3{,}000$</th><th>$V_0 = 5{,}000$</th><th>$V_0 = 10{,}000$</th></tr>
<tr><td>Small-Vol Stocks</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Mid-Vol Stocks</td><td>
</td><td>
</td><td>
</td></tr>
</table>

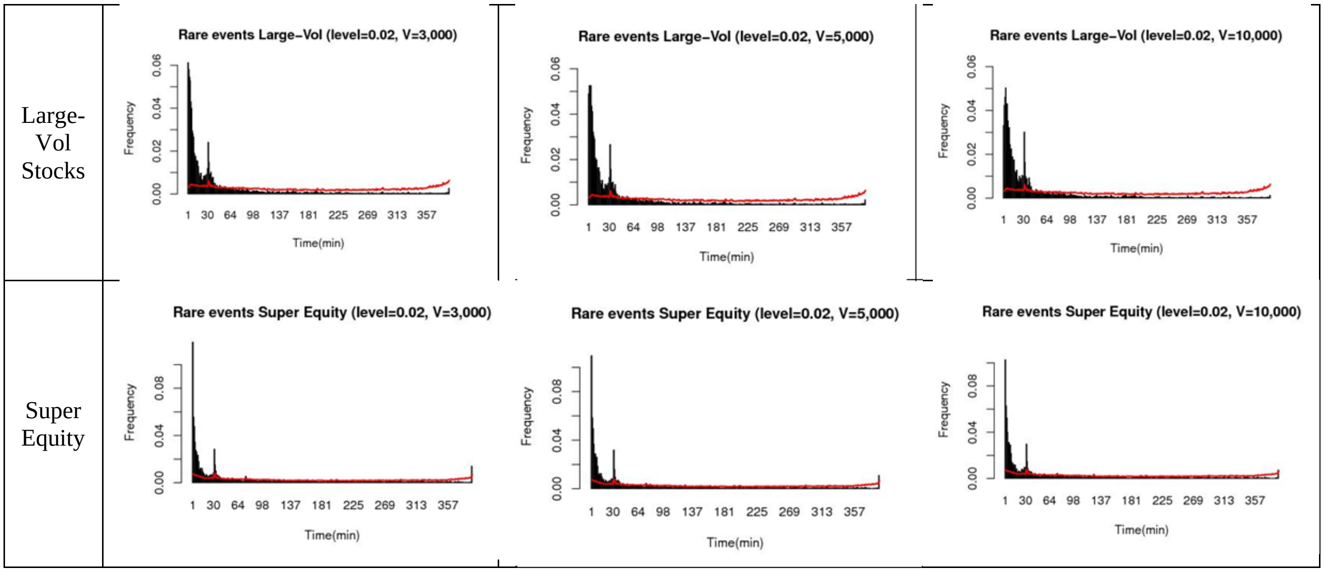
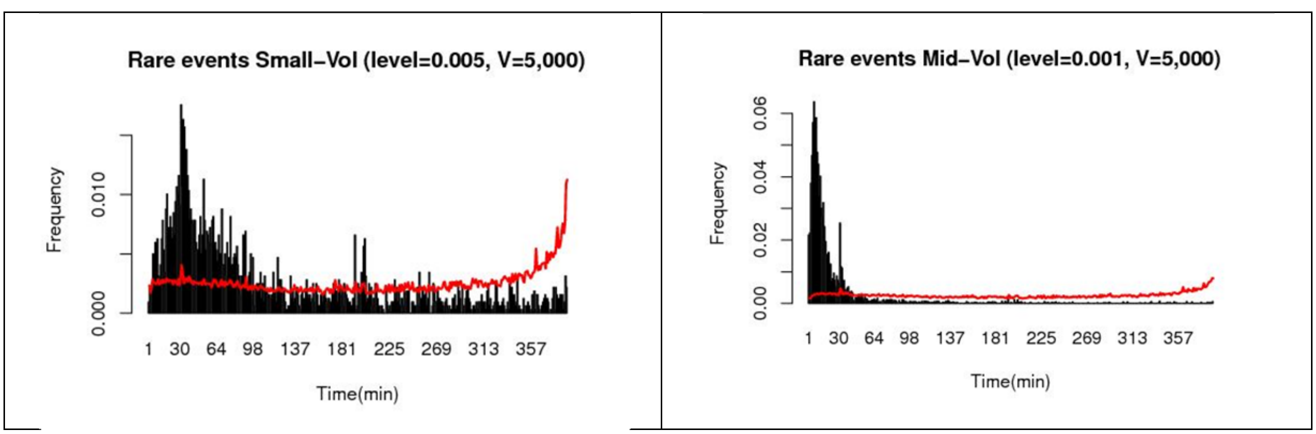


Fig. 9 Distribution of rare events (level $\alpha = 0.02$) during the day for each type of equity and window volume considered.

Based on the previous analysis of the probabilities for price reversal and the expected return of a trading rule and the rare event distribution, we note significant market inefficiencies for small and mid-vol classes and an increase in market efficiencies for the other two high volume classes. This is to be expected since higher volume means inefficiencies will disappear faster, but the histograms provide a time-frame for the expected duration of these rare events in excess of the expectation.

In Fig. 10 we exemplify the rare events distribution for the optimum α level and V_0 determined in the prior analysis. This particularization for the optimal set provides a calibration method for optimal execution time and activation of a trading rule.



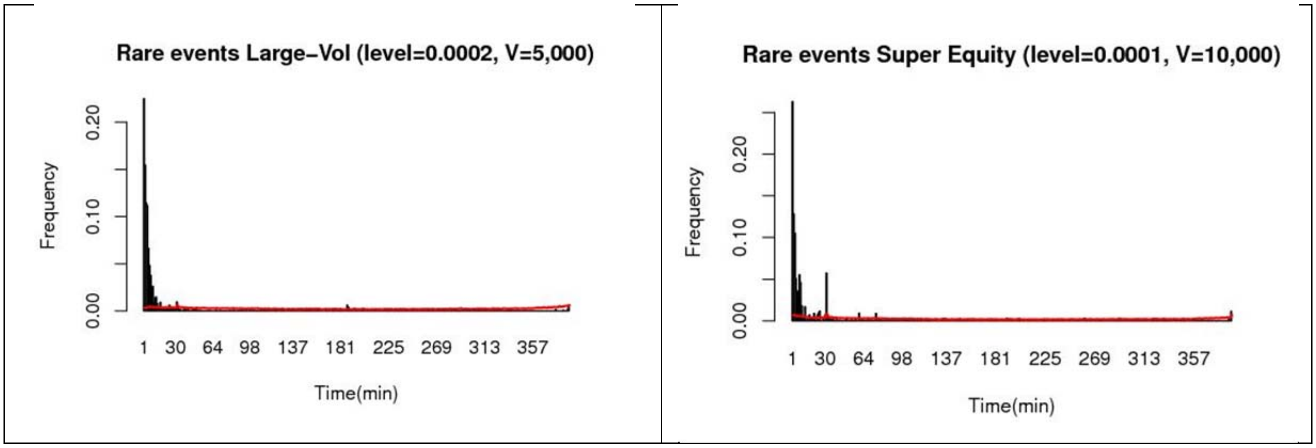


Fig. 10 Rare events distribution relative to the frequency of the trade activity with optimal parameters determined in the analysis.

5. CONCLUSIONS

This article presents a simple methodology of detecting and evaluating unusual price movements defined as large change in price corresponding to small volume of trades. We classify these events as "rare" and we show that the behavior of the equity price in the neighborhood of a rare event exhibits an increase in the probability of price recovery. The use of an arbitrary trading rule designed to take advantage of this observation indicates that the returns associated with such movements are significant. We therefore confirm the old Wall Street adage that "it takes volume to move prices" even in the presence of high frequency trading.

We present a way to calibrate and find optimal trading parameters for the specific trading strategy considered. The methods presented herein may be easily extended to any trading strategy based on rare events detection. The equity behavior is consistent throughout the equity classes considered in this work. The trading rule we consider provides positive returns when considering the entire universe of equities and neglecting transaction costs.

The classification of equity based on average daily volume (*ADV*) allows us to draw more specific inference about the rebound behavior of the equity. We confirm that it takes a larger volume window to observe a rare event for super equity (e.g., SPY, JPM, MSFT, etc.) than for less traded equity. Furthermore, the price recovery after a rare event is much faster for highly traded stocks than for low volume stocks.

We look at the distribution of these rare events during the trading day. We show that they are not simply a percentage of the trades and we show that they accumulate at the beginning of the day. We observe an increase frequency around 30 min across equities and another at about the middle of the trading day for equity which is not traded very frequently. These may be explained and we formulate hypotheses about their appearance.

Essentially, the methodology we present measures the reaction of the market to abnormal price movements. Notably, a possible application of this methodology may involve the development of forensic tools for market trading activity. The delimitation between rare events and suspicious events is rather thin and additional market data regarding the origination of the trades recorded would be useful in identification of irregular trades.

References

Admati A. R. and P. Pfleiderer (1988). A theory of Intraday Patterns: Volume and price variability, *The Review of Financial Studies*, vol.1, nr. 1, pp3-40

Alfonsi, A., A. Schied and A. Schultz (2007). Optimal execution strategies in limit order books with general shape functions. http://www.citebase.org/abstract?id=oai:arXiv.org:0708.1756.

Beaver W.H. (1968). The information content of annual earnings announcements. *Empirical research in Accounting: Selected Studies; suppl to Journal of Accounting Research*, vol. 6, pp 67-92

Bollerslev, T. and D. Jubinski (1999). Equity trading volume and volatility: latent information arrivals and common long-run dependencies. *Journal of Business & Economic Statistics* 17, pp. 9–21.

Engle, R.F. and J.R. Russell (1998), Autoregressive Conditional Duration: a New Model for Irregularly Spaced Transaction Data, *Econometrica*, 66, 1127-1162.

Foster, F. D. and S. Viswanathan (1990) A Theory of the Interday variations in volume, variance, and trading costs in securities markets, *The Review of Financial Studies*, vol. 3, nr. 4, pp. 593-624

Foster F.D. and S. Viswanathan (1993) Variations in trading volume returns volatility and trading costs: evidence on recent price formation models. *Journal of Finance* 48(1): 187-211.

Gallant, A.R., P.E. Rossi and G.E. Tauchen (1992), Stock prices and volume. *The Review of Financial Studies*, 5, pp. 199–242.

Gerety M.S. and J.H. Mulherin (1992) Trading halts and market activity: an analysis of volume at the open and the close. *Journal of Finance* 47(5): 1765-1784.

Karpoff, J. (1987). The relation between price change and trading volume: A survey, *Journal of Financial and Quantitative Analysis*, 22, March, pp 109-126.

Llorente G, R. Michaely, G. Saar and J. Wang (2002) Dynamic Volume-Return relation of individual stocks, *The Review of Financial Studies*, vol. 15, nr. 4, pp. 1005-1047

Lo, A.W., H. Mamaysky, and J. Wang (2000). Foundation of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of Finance* 55 (4), pp 1705-1765.

Lo, A.W and J. Wang. (2002) Trading volume: Implications of an intertemporal capital asset price model. *Advances in Economic Theory: Eighth World Congress*, pp 1-23

Mariani M.C., I. Florescu, M.P. Beccar Varela and E. Ncheuguim (2009): Long correlations and Levy Models applied to the study of Memory effects in high frequency (tick) data, *Physica A*, 388(8), April, p. 1659-1664

Osborne M.F.M. (1959) Brownian motion in the stock market, *Operations Research*, 7(2), pp 145-173.

Sun W. (2003) Relationship between Trading Volume and Security Prices and Returns, *MIT LIDS Technical Report 2638*, February 2003 Area Exam.

Tsay, A. S. and C. Ting (2006, January). Intraday stock prices, volume, and duration: a nonparametric conditional density analysis. *Empirical Economics* 30 (4), pp 253-268.

Tsay, R. (2005). *Analysis of Financial Time Series*. Wiley-Interscience.

Zhang, M. Y., J.R. Russell and R.S. Tsay (2008). Determinants of bid and ask quotes and implications for the cost of trading. *Journal of Empirical Finance*. 15 (4), pp 656-678.