Rare Events Analysis for High-Frequency Equity Data

Dragos Bozdog, Ionuţ Florescu, Khaldoun Khashanah, and Jim Wang
Stevens Institute of Technology, e-mail: ifloresc@stevens.edu

Abstract
In this work we present a methodology to detect rare events which are defined as large price movements relative to the volume traded. We analyze the behavior of equity after the detection of these rare events. We provide methods to calibrate trading rules based on the detection of these events and illustrate for a particular trading rule. We apply the methodology to tick data for thousands of equities over a period of five days. In order to draw comprehensive conclusions, we group the equities into classes and calculate probabilities of price recovery after these rare events for each class. The methodology that we have developed is based on non-parametric statistics and makes no assumption about the distribution of the random variables in the study.

Keywords
high-frequency trading, average daily volume, trading strategy

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 et al. (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 and 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 exceptions to the conclusions 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 brought into question 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; Podobnik et al., 2009). 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 extreme observations of the price–volume relationship. We may refer to the corresponding observations as *rare events in high-frequency finance* or simply *rare events*. Please note that these events are not necessarily catastrophic events and in fact, due to the high-frequency nature of the data in the study, they happen quite often.

In our context, due to the joint price–volume distribution, we encounter two types of rare event. The first type occurs when the volume of traded shares is small, coupled with a large price movement. The second type occurs when the volume of traded shares is large, coupled with a small price movement. Of the two types of rare event, we are only interested in the first type. The second type is evidence of unusually high trading activity, which is normally accompanied by 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 (Zhang et al., 2008; Alfonsi et al., 2010) analyzes 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 a rare event is observed.

This article is structured as follows. In Section 2, we present the basic methodology for detecting and evaluating rare events. Section 3 details results obtained by applying the methodology to tick data collected over a period of five trading days in April 2008. Section 4 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 over five days. 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 two-dimensional random vector $(\Delta P, \Delta V)$. Here ΔP is the change in price, ΔV is the change in volume.

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 objective of our study is the conditional distribution:

$$Prob(\max \Delta P \mid \Delta V < V_0)$$

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

2.1 Justification of the Method

1. Why restrict the distribution to be conditional on V_0?

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

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

Indeed, this is a very important question. There is no reason for V_0 to be constant other than for practical reasons. A valid objection is that the dynamics of the equity change over 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?

It is known from the literature that the duration defined as the time between consecutive trades has an impact on the price evolution. Tsay and Ting (2006) consider the distribution of price change conditional on duration and volume $(\Delta P \mid d, V)$ using only consecutive trades. As illustration they use tick data for four NYSE stocks. The authors conclude that the duration affects prices, but the effect is stronger when the volume is high. They find that at the intraday frequency the volume has an influence, even after controlling for duration. Similar results are documented in Manganelli (2005), who extends the ACD model of Engle. Generally, these types of model use consecutive trades or sampled consecutive trades. Rare events, as defined in our context, normally happen over several trades not consecutively. Because of this, we need to consider all trades within a moving window. We analyze over 5,000 equities. The large number of combinations makes the study of duration and volume difficult. Furthermore, according to the cited results, in our case (conditional volume is small) the influence of duration on price is weak. All these considerations led us to the study of the distribution $\Delta P \mid V$.

2.2 Sampling Method: Rare Event Detection

A thorough investigation of the distribution of price changes, conditional on cumulative trading window, would involve the evaluation of all observations for each equity $(S_n - S_j \mid v_k + v_{k+1} + \ldots + v_n < V_0)$ for $k \le j \le n$ where n runs through all the trades, S_n is the price, and v_n is the volume associated with trade n. Although the distribution would be accurate, such a task would involve significant computational effort considering the large database we use. Because we are interested only in the rare events, we chose to select the extreme observations in this sequence and subsequently analyze only a percentage of these observations at the tails of the distribution $\Delta P \mid V < V_0$.

Specifically, we construct the sequence of consecutive trades $S_k, S_{k+1}, \ldots, S_n$ and their associated volumes $v_k, v_{k+1}, \ldots, v_n$ such that $v_k + v_{k+1} + \ldots + v_n < V_0$ and we consider

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

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 univariate "quantile-type" rule. Namely, for a fixed level α we select all the observations in the set:

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

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

Note: The Δp_n quantities are not independent. Thus, the empirical distribution is only an approximation of the true probabilities of price movement. However, rule (1) will identify candidates for rare events which may or may not correspond to the true probability level α. If we use non-overlapping windows, the resulting rare events will correspond better to the level α. There are two reasons why this is not feasible. First, by considering non-overlapping windows we can lose extreme price differences calculated using prices from the 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.

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 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 rises 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 us 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 $\frac{k}{n}$. 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 become 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.

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 of V_{ae} we calculate the expected return by averaging all the trade returns placed following the above strategy.

A depiction of the quantities involved and the trading strategy chosen may be observed in Figure 1. We note that we shall use the trading rule in Definition 2 only for determining the 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 stocks. Most of the studies are focused on what are called "large cap" 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). In our case, ADV is essential and 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 to classify equities into classes based on the average daily traded volume. We refer to this classification as the *multi-scale volume classification*.

Figure 1: Visual depiction of the quantities used in the study.

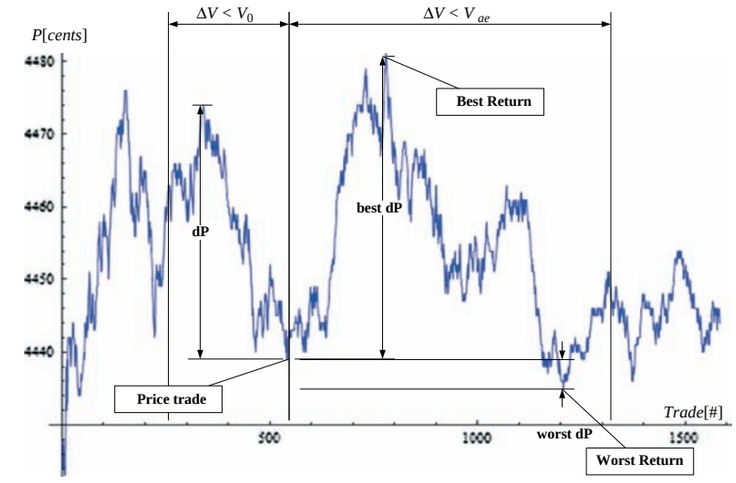
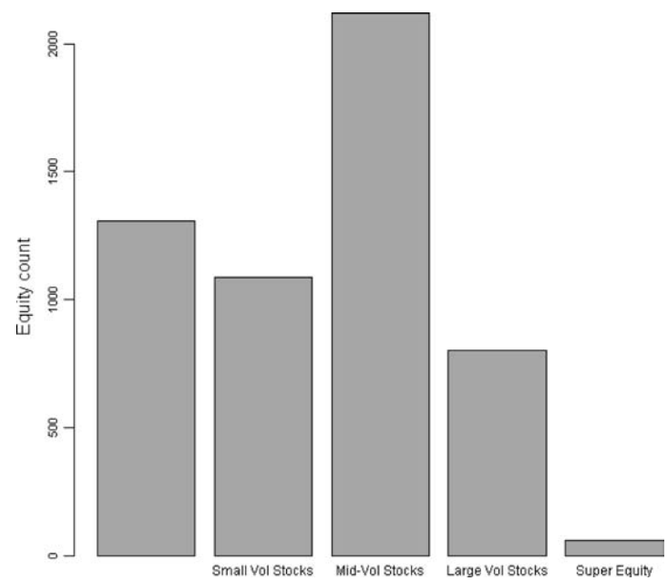


Figure 2: Average daily volume distribution.



The histogram in Figure 2 corresponds to the ADV of the total universe of 5,369 equities considered in this study.

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 analyses. The highest ADV values are concentrated around major

Table 1: Equities partitioned into five classes.

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

Table 2: Probability (%) of favorable price movement by equity class.

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

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.

3 Results

The methodology described in Section 2 is applied to each equity data within a class. Then, we combine all the rare 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 a 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 $k \in \{1,2,3\}$.

For a better visualization and interpretation of these numbers we construct probability surfaces for each class and plot them with respect to the α level and volume V_{ae} in Figure 3.

According to 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 increasing. Indeed, we observe this behavior in Figure 3, 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 an optimal level for one class it automatically translates into optimal levels for the other classes.

To determine the optimal level for each class we calculate the expected return of trades according to Definition 2. Specifically, for fixed levels of α

Figure 3: Probability surfaces for equity classes.

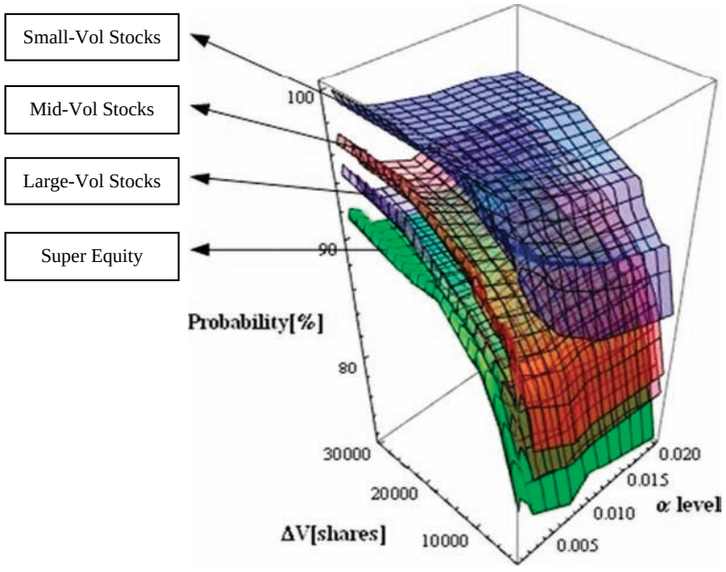
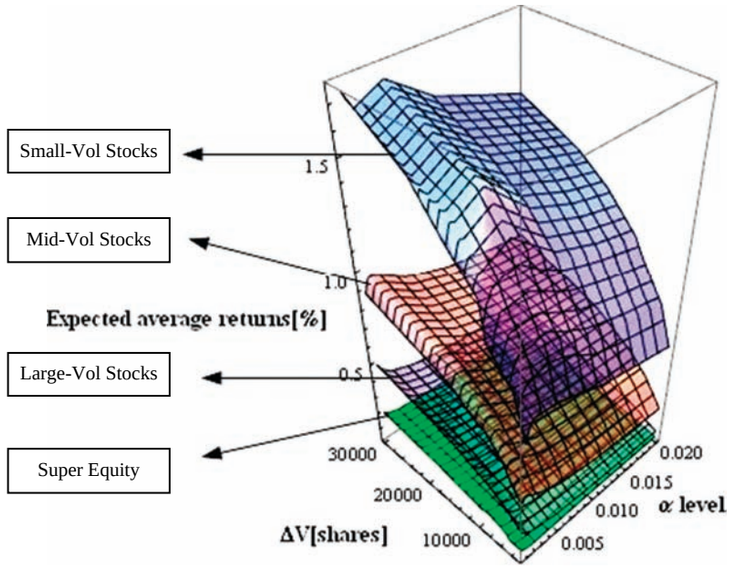


Figure 4: Expected return surfaces for stock classes.



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 4.

Unlike the probability plots, the surfaces in Figure 4 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 that 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 5 and 6.

In Figure 5 we project the 3D plot onto the ratio $\frac{V_{ae}}{V_0}$. 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 $\frac{V_{ae}}{V_0}$ increases.

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

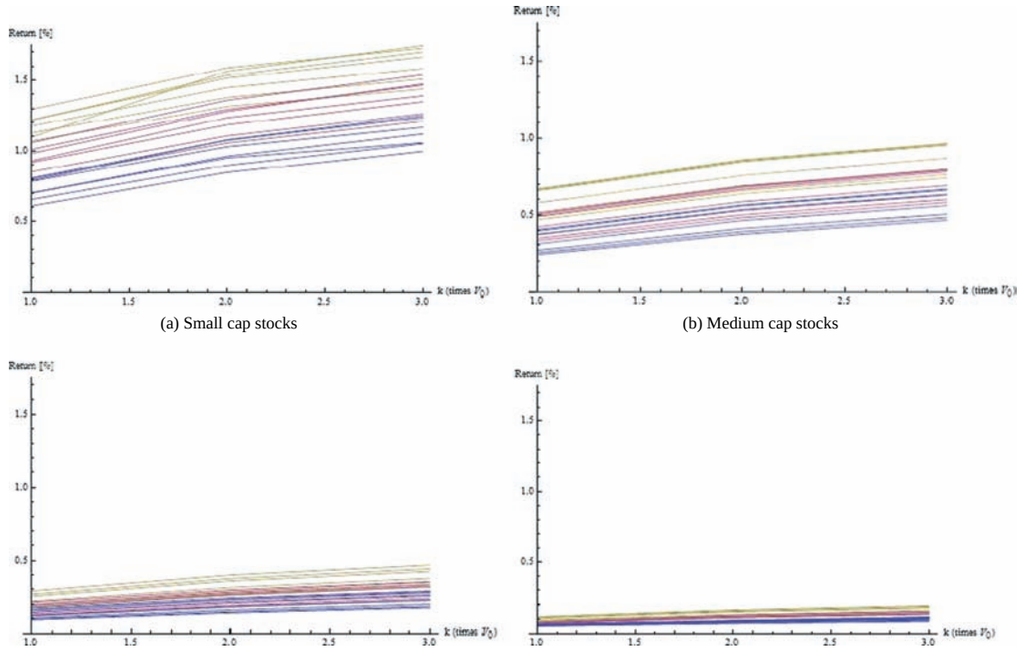
Table 3: Expected return (%) for equity classes for all days $r = r_{fav} \cdot P + r_{unfav} \cdot (1 - P)$.

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
	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
Small-Vol Stocks	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
	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
Mid-Vol Stocks	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
	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
Large-Vol Stocks	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
	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
Super Equity	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

original). In contrast, for large and especially for super large equities, 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 6 where we plot return vs. V_{ae}. This figure provides an indication of 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 Figures 5 and 6 we may provide the following list of optimal values:

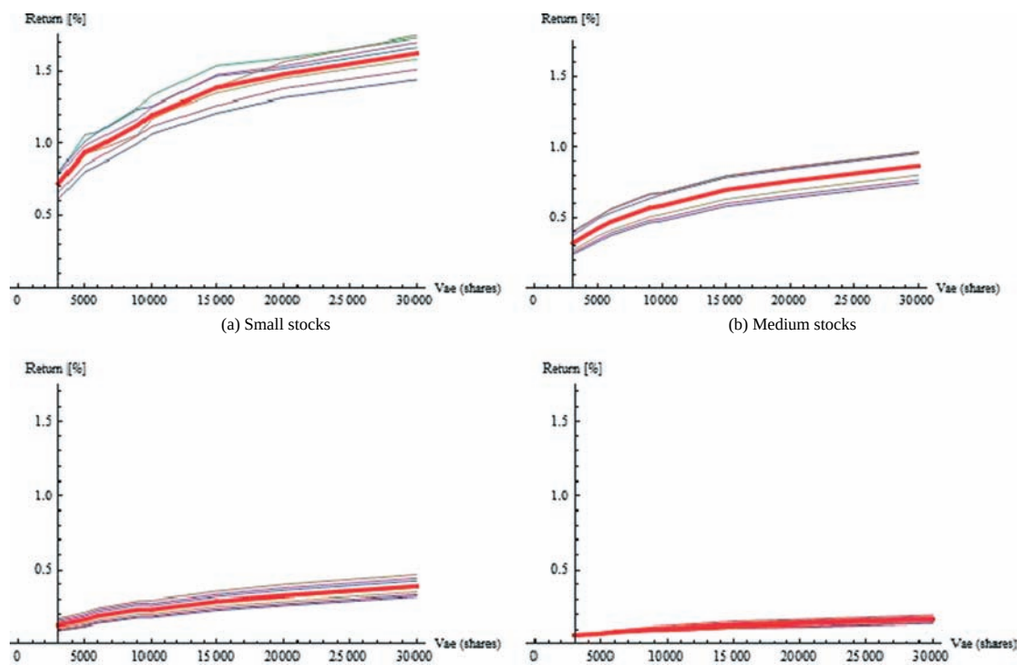
Figure 5: Sectional 2D plots of the surfaces in Figure 4 for each of the quantile levels considered. Each sub-figure represents one surface from Figure 4. The *x*-axis is the proportion of after-event window size with respect to the before-event window size, and lines of the same color represent the three original window sizes chosen (blue for V_0 = 3,000, red for V_0 = 5,000, yellow for V_0 = 10,000).



(a) Small cap stocks (b) Medium cap stocks

(c) Large cap stocks (d) Super equity

Figure 6: Sectional 2D plots of the surfaces in Figure 4 for each of the quantile levels considered. The *x*-axis is the value of V_{ae} (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.



(a) Small stocks (b) Medium stocks

(c) Large stocks (d) Super equity

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 Conclusion

This article presents a simple methodology of detecting and evaluating unusual price movements defined as a large change in price corresponding to a small volume of trades. We classify these events as "rare" and 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 equities 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 a super equity (e.g., SPY, JPM, MSFT, etc.) than for a less traded equity. Furthermore, the price recovery after a rare event is much faster for highly traded stocks than for less liquid stocks.

Essentially, the methodology 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.

Dragos Bozdog works in both industry and academia with research interests in financial and engineering applications. He received his Ph.D. in engineering from The University of Toledo and an M.S. in financial engineering from Stevens Institute of Technology. He was a Post-Doctoral Fellow at Rutgers Center for Operations Research.

Ionuţ Florescu was graduated in 2004 with a Ph.D. in statistics from Purdue University. He has been an Assistant Professor at Stevens Institute of Technology since that time. His research interest lies in applied problems related to probability and stochastic processes.

Khaldoun Khashanah earned his Ph.D. in applied mathematics from the University of Delaware in 1994. He is a Distinguished Service Professor and Director of Financial Engineering Program at Stevens Institute of Technology. He is a member of Committee to Establish the National Institute of Finance, IAFE, and INCOSE.

Jim Wang is an Adjunct Professor of Financial Engineering Program at Stevens Institute of Technology. He earned an M.S. degree in computer science and a Ph.D. degree in biomedical engineering from Rutgers University, and an M.B.A. degree from New York University. He holds a CFA designation from the CFA Institute. He has his own business specializing in high-frequency trading in equity, FX, futures and options markets.

REFERENCES

Admati, A.R. and P. Pfleiderer (1988). A theory of intraday patterns: volume and price variability. *The Review of Financial Studies*, **1**(1), 3–40.

Alfonsi, A., A. Fruth, and A. Schied (2010). Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, **10**(2), 143–157.

Beaver, W.H. (1968). The information content of annual earnings announcements. *Empirical Research in Accounting: Selected Studies (Supplement to Journal of Accounting Research)*, **6**, 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**, 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*, **3**(4), 593–624.

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

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

Llorente, G., R. Michaely, G. Saar, and J. Wang (2002). Dynamic volume–return relation of individual stocks. *The Review of Financial Studies*, **15**(4), 1005–1047.

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.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, **8**(4), 377–399.

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), 1659–1664.

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

Podobnik, B., D. Horvatic, A. Petersen, and E. Stanley (2009). Cross-correlations between volume change and price change. *Proceedings of the National Academy of Sciences*, **106**(52), 22079–22084.

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). Intraday stock prices, volume, and duration: a nonparametric conditional density analysis. *Empirical Economics*, **30**(4), 253–268.

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), 656–678.