Scaling of inefficiencies in the U.S. equity markets: Evidence from three market indices and more than 2900 securities*

David Rushing Dewhurst,[1, 2, 3, 4] Colin M. Van Oort,[1, 2, 3, 4] John H. Ring IV,[1, 2, 3, 4]
Tyler J. Gray,[2, 3] Christopher M. Danforth,[2, 3, 4, 5] and Brian F. Tivnan[1, 2, 3, 5]

[1] *The MITRE Corporation, McLean, VA*
[2] *MITRE-UVM Computational Finance Lab, Burlington, VT*
[3] *Vermont Complex Systems Center, University of Vermont, Burlington, VT*
[4] *Department of Computer Science, University of Vermont, Burlington, VT*
[5] *Department of Mathematics and Statistics, University of Vermont, Burlington, VT*

(Dated: April 18, 2019)

Using the most comprehensive, commercially-available dataset of trading activity in U.S. equity markets, we catalog and analyze dislocation segments and realized opportunity costs (ROC) incurred by market participants. We find that dislocation segments are common, observing a total of over 3.1 billion dislocation segments in the Russell 3000 during trading in 2016, or roughly 525 per second of trading. Up to 23% of observed trades may have contributed the the measured inefficiencies, leading to a ROC greater than $2 billion USD. A subset of the constituents of the S&P 500 index experience the greatest amount of ROC and may drive inefficiencies in other stocks. In addition, we identify fine structure and self-similarity in the intra-day distribution of dislocation segment start times. These results point to universal underlying market mechanisms arising from the physical structure of the U.S. National Market System.

I. INTRODUCTION

Securities markets utilize auction mechanisms to facilitate the valuation and trade of assets [1–4]. The operational structure of these markets including the auction mechanism, physical infrastructure through which the market is implemented, and endogenous information asymmetries are thus intrinsic factors in their efficiency [5, 6]. While some authors have largely neglected or minimized these factors in analyses of the efficiency of financial markets [7, 8], others have recognized that these so-called microstructure variables are central to the performance of modern-day markets [9–12].

A. Modern U.S. market

We focus our investigation on equities traded in the U.S. National Market System (NMS), a network of privately-owned and operated stock exchanges locate in the U.S., since it is the proverbial center of the world equity markets. In particular, we turn our attention to the roughly 3000 largest equities traded on the NMS, compiled by FTSE International Ltd. as the Russell 3000 index. These securities represent the vast majority of the securities traded in the U.S. and can serve as a nearly comprehensive cross-section of publicly-traded securities from which the observation and assessment of microstructure quantities can be made.

Using the most comprehensive commercially-available dataset of NMS messages available, we enumerate and describe dislocation segments and realized opportunity costs, defined in Section III, in Russell 3000 securities and a selected subset of exchange-traded funds (ETFs) during calendar year 2016. Extending earlier work by the present authors and others [13], we find that dislocation segments and differing trades resulting in realized opportunity cost occur frequently in the National Market System, with over 3×10^9 dislocation segments and a total realized opportunity cost of $2,051,916,739.66 in 2016 among securities studied. As shown in Table I, a significant portion of all trades studied here (23.71%) were trades for which there existed a different price elsewhere in the NMS at which the trade did not execute. These findings correspond with, and extend, the findings presented in [13].

B. Scaling in finance

There is widespread agreement that [14, 15] was one of the first to characterize the scaling properties of price returns in modern markets. The scaling of returns was later revisited [16] and formalized and extended [17, 18]. Beyond just price fluctuations (i.e., returns in price time series), additional financial variables have been found to display scaling properties. For some representative examples, market indices and foreign exchange rates [19] as well as share volume and number of trades [20] adhere to scaling properties.

With the dramatic increase in the number of securities under study and concomitant increase in the range of market capitalization, we examine scaling relationships between microstructure variables, such as dislocation segments and realized opportunity cost, and market capitalization and its derivative statistics. Dislocation seg-

* Corresponding authors: david.dewhurst@uvm.edu, btivnan@mitre.org

1	Realized Opportunity Cost	$2,051,916,739.66
2	SIP Opportunity Cost	$1,914,018,654.41
3	Direct Opportunity Cost	$137,898,085.25
4	Trades	4,745,033,119
5	Diff. Trades	1,124,814,017
6	Traded Value	$28,031,002,997,692.75
7	Diff. Traded Value	$7,077,357,462,641.67
8	Percent Diff. Trades	23.71
9	Percent Diff. Traded Value	25.25
10	Ratio of 9 / 8	1.0651

TABLE I. Summary statistics of the realized opportunity cost for all studied securities that traded in 2016. Realized opportunity cost across all equities in the year is over $2B USD. We discuss statistical characteristics of realized opportunities extensively in Section IV. We note that ratio of the percent of differing trades to the percent of differing traded values is discussed as a statistic of the relative value "on the table" per differing trade—a proxy for the potential profitability of latency arbitrage strategies.

Category	Duration Cond.	Magnitude Cond.	Count
Dow	-	-	120,355,462
Dow	$> 545\mu s$	-	65,073,196
Dow	$> 545\mu s$	$> \$0.01$	2,872,734
SPexDow	-	-	1,126,186,592
SPexDow	$> 545\mu s$	-	530,499,458
SPexDow	$> 545\mu s$	$> \$0.01$	51,187,430
RexSP	-	-	1,888,686,248
RexSP	$> 545\mu s$	-	704,416,718
RexSP	$> 545\mu s$	$> \$0.01$	110,447,787

TABLE II. Total number of dislocation segments in mutually-exclusive market categories. Number of opportunities is calculated unconditioned, conditioned on duration, and conditioned on both duration and magnitude.

ments occur in equities of all sizes. While they are more frequent in equities with large market capitalizations, the distributions of their qualities, such as their size (magnitude) and duration (how long they lasted), are more extreme among equities with smaller market capitalizations. The majority of realized opportunity cost is generated by equities in the S&P 500 that are not also in the Dow (termed the SPexDow). The SPexDow also Granger-causes realized opportunity costs in other mutually-exclusive market categories (Dow 30 and Russell 3000 less the S&P 500, or RexSP), pointing to its centrality in the U.S. equities market. When realized opportunity cost is analyzed at marketplace (all exchanges and the aggregation of dark pools or alternative trading facilities—ATFs—which is collectively known as FINRA) granularity, the marketplaces with highest volume also have the highest realized opportunity cost.

In the following sections, we first provide a brief overview of the U.S. National Market System for the unfamiliar reader. We then detail our data, the available and used fields, and summarize the equities studied. After describing statistics of dislocation segments, including distributions of start times and durations, we move to analysis of realized opportunity cost, providing summary statistics, comparisons across exchanges, and correlation and Granger-causality analyses. We close with a brief exploration of exchange-traded funds (ETFs), a discussion of results, and possibilities for future work.

II. MARKET OVERVIEW

We provide a brief overview of the U.S. equities market for the unfamiliar reader; a more comprehensive summary is given in [13], along with formal set of definitions that is used in this work. The U.S. equities market, known as the National Market System (NMS), is composed of 13 unique, privately-owned exchanges, platforms on which price of equity securities are discovered by market participants using a continuous or discrete auction mechanism. Another core component of the market infrastructure making up the NMS are on the order of 40 privately-owned alternative trading systems (ATSs) [21], also known as dark pools, on which market participants can trade but usually cannot participate in price discovery [22]. Each exchange and ATS keeps its own book of orders submitted by agents; trading facilities (exchanges and ATSs) attempt to execute orders at the "top" of the book (the highest-priced bid orders, or orders to buy securities, and the lowest-priced ask orders, or orders to sell securities) along with "market" buy and sell orders—orders that should be executed immediately, without regard for execution price. The top of the book at each exchange is termed the best bid and offer (BBO). All BBOs are aggregated over the entire NMS and the best bid and offer among all BBOs is calculated and termed the national best bid and offer (NBBO). By law, trades must execute at the NBBO, albeit with certain exceptions.[23]

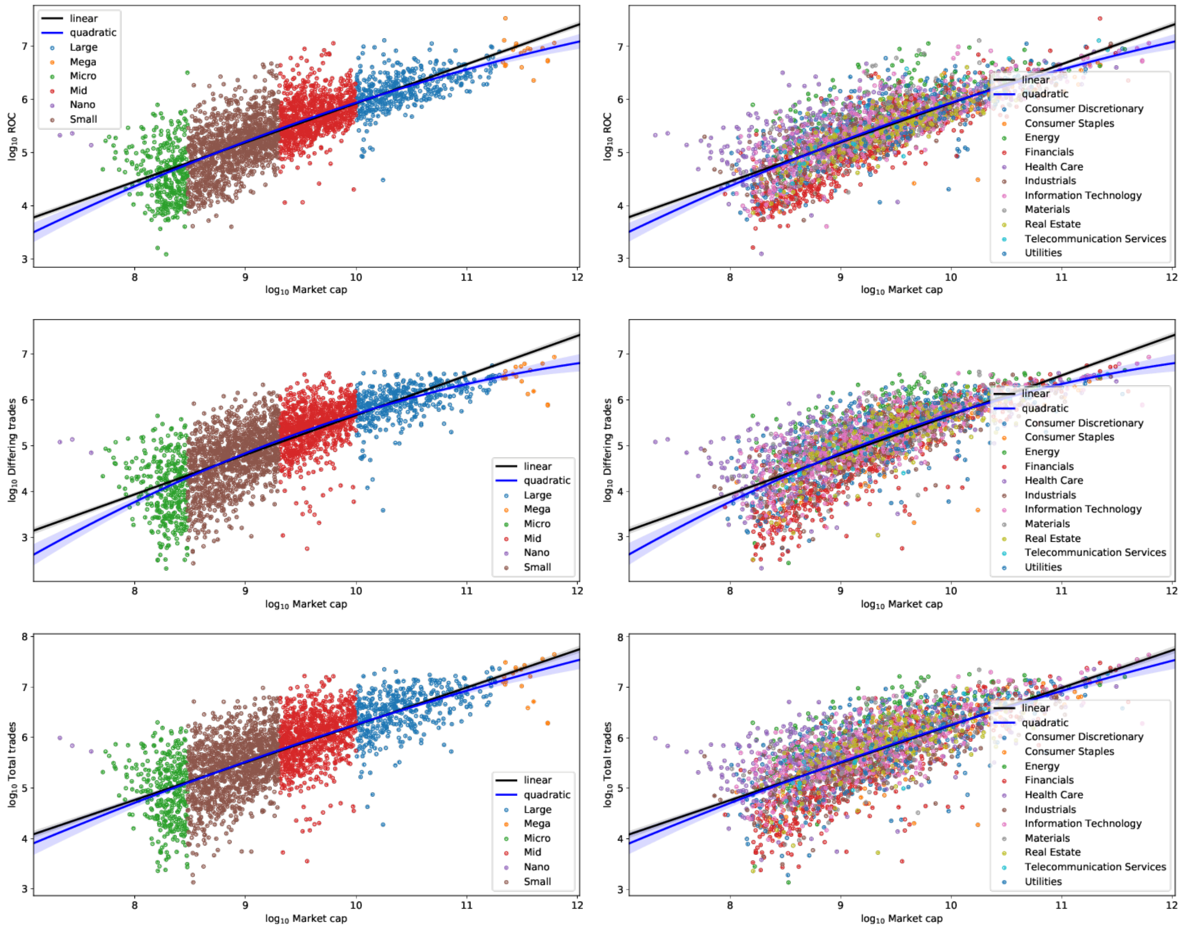

FIG. 1. Relationships between market capitalization and different dependent variables: ROC, differing trades, and total trades. It is immediately apparent that there is a strong positive relationship between market cap and realized opportunity cost. The data exhibits interesting nonlinearity and heteroskedasticity. Equities with smaller market capitalization have higher variability in each dependent variable, while equities with larger Market Capitalization (MC) have generally lower variability in each dependent variable. Also note the equities in the financial sector have a consistantly lower ROC relative to MC while equities in the energy sector have a consistantly higher ROC relative to MC. The data are well-fit by linear and quadratic functions in doubly-logarithmic space. Error ellipses are 95% confidence intervals for the true regression curve calculated using bootstrapping techniques.

There are different methods by which these trading facilities communicate with one another. The Securities Information Processor (SIP) provides a regulation-required communication system by which trades are reported and price information is disseminated. There also exist privately-owned direct information feeds that provide faster information updates than the SIP data feeds. This differential information quality can contribute to price discrepancies and dislocation segments that sometimes exist between trading facilities.

The NMS is regulated by the U.S. Securities and Exchange Commission (SEC), a federal government agency, and self-regulated by the Financial Industry Regulatory Authority (FINRA), a professional organization. FINRA self-polices its members and attempts to ensure they adhere to SEC rules and other professional guidelines, while SEC designs, implements, and enforces rules that are intended to promote market stability and economic efficiency.

The physical structure of the NMS, in conjunction with the existence and usage of information feeds of differing speeds (regulator-imposed Securities Information Processor (SIP) versus direct, proprietary feeds), leads to market inefficiencies [24] in the form of dislocation segments and realized opportunity costs. On Dow 30 equities, over 120 million dislocation segments and over \$160M USD in realized opportunity cost were cataloged during calendar year 2016 [13]. These are not the first violations of the so-called "law of one price" that have been detected in financial markets; multiple violations in international and domestic financial markets have been identified for more than a decade [25, 26]. However, our results here present one of the more egregious violations of this "law".

III. DATA AND METHODS

A. Data

Our data is the most comprehensive set of market data commercially available and is effectively identical to the SEC's MIDAS dataset [27]. It is available for purchase from Thesys Technologies [28] and is comprised of every message, quote, and trade that occurred on the SIP and of direct feeds in the NMS during the time period of study. Information on individual companies, such as the indexes into which companies fall, their Global Industry Classification Standard (GICS) sector, and market capitalization were gathered using the standard commercial Bloomberg Terminal.

The indices we consider are subject to change daily. In order to have consistent sets of companies to study we consider the Dow 30 and S&P 500 as they stood on Jan 1 2016. Additionally we consider the Russell 3000 2016 (i.e., the Russell 3000 constructed in June 2016) excluding components that were not publicly traded on a major exchange on Jan 1 2016.

Each index under study was curated to only include companies that survived as an individual, publicly traded entity on a major exchange for the entire calendar year of 2016. In other words, companies that were delisted from NYSE or NASDAQ for any reason (e.g. Chapter 11 bankruptcy or buyout) were removed. Companies that merged with another member of our dataset remained under study while those who merged with a firm outside of our dataset were removed from consideration. This curation process allows us to avoid several edge cases of market behavior including IPOs and de listings.

Many companies in our dataset changed their ticker symbol over the course of the calendar year and thus appear as a different entity in the data. To study a company over a long time period it is necessary to know all tickers it traded under and when the ticker changes occurred. As there is an absence of a consolidated public record of these symbol changes, we tracked them through an extensive manual review of press releases. These symbol changes were then validated by comparison with the Thesys data feed. Specifically, we confirmed that on the date of the change, trading activity ceased on the old ticker and began on the new symbol.

This curation reduced the Russell 3000 from 3005 companies to 2903 and the S&P 500 from 500 companies to 472. No companies were removed from the Dow 30 set. Note that the mutually exclusive indices are constructed using the curated sets and the curated sets are designated by appending a prime to the respective base index (e.g. Dow 30 $\rightarrow$ Dow 30$'$). Finally all companies in our dataset were classified by their market capitalization as it stood in the beginning of Q4 2016. There is no industry standard on either the range of market caps that define a class or the number of classes. Furthermore these definitions change over time as companies reach larger market caps with the overall growth of the economy. Our definitions for these classes are displayed in Table IV [29].

Our work here focuses on dislocations (and their constituent dislocation segments) and the related realized opportunity costs arising from price discrepancies between the SIP and direct feeds. Both concepts have been discussed throughout the empirical market microstructure literature [30–33], though definitions of these quantities are not entirely settled, we will abide by those set in [13]. We briefly review these definitions and detail the calculation of dislocations, dislocation segments, and realized opportunity cost from our data.

Suppose that there exist two sources of information I_1 and I_2 on market data displaying respective prices for an asset $p_1(t)$ and $p_2(t)$ observed from a single, well-defined location. We will term a dislocation between these sources of data as the amount of time during which the price information given by the information feeds differs. A related but distinct concept is that of a dislocation segment, defined as an amount of time during which the price information given by the information feeds differs in a single direction (i.e., we must have $\Delta p > 0$ or $\Delta p < 0$ during the entire time of the dislocation segment). The realized opportunity cost of using I_1 instead of I_2 over an amount of time $[0,\ T]$ is calculated as follows: for every trade that occurs during $[0,\ T]$, the difference in execution price $|p_1 - p_2|$ is calculated and multiplied by the number of shares in the trade n. The total realized opportunity cost over the interval $[0,\ T]$ is then given by the sum over the realized opportunity cost associated with each trade in $[0,\ T]$.

Equities contained in our dataset represent approximately 98% of all publicly-traded equities in the U.S. by market capitalization [34]. Tables III - V provide summary statistics and distribution of these equities across business sector, market cap, and market category, for several indices.

B. Methods

We first compute basic summary statistics and qualitative descriptions of the distributions of dislocation segments and realized opportunity costs. In addition to computing summary statistics, the larger sample of equities (all equities traded on the National Market System during 2016) allows us to conduct a cross-sectional study of the effect of company "size" on these microstructure quantities. We quantify the notion of size of a company

Sector	Statistic	Russ 3K′	S&P 500′	Dow 30′	RexSP	SPexDow
Consumer Discretionary	% by #	14.92	16.95	13.33	14.52	17.19
	% by MC	12.97	11.92	8.98	16.40	13.10
	Count	433	80	4	353	76
	($) MC Min	95,330,024	1,244,719,232	84,654,022,656	95,330,024	1,244,719,232
	($) MC Max	356,313,137,152	356,313,137,152	165,862,064,128	89,539,158,016	356,313,137,152
Consumer Staples	% by #	4.03	7.20	10	3.41	7.01
	% by MC	8.54	9.99	10.74	3.83	9.69
	Count	117	34	3	83	31
	($) MC Min	114,570,432	9,794,159,616	178,815,287,296	114,570,432	9,794,159,616
	($) MC Max	224,997,457,920	224,997,457,920	224,997,457,920	17,508,790,272	150,058,582,016
Energy	% by #	5.20	7.63	6.67	4.73	7.69
	% by MC	6.57	7.14	10.40	4.71	5.83
	Count	151	36	2	115	34
	($) MC Min	160,502,160	2,427,903,232	222,190,436,352	160,502,160	2,427,903,232
	($) MC Max	374,280,552,448	374,280,552,448	374,280,552,448	27,468,929,024	116,800,331,776
Financials	% by #	17.81	12.50	13.33	18.84	12.44
	% by MC	15.17	13.07	8.91	21.99	14.73
	Count	517	59	4	458	55
	($) MC Min	89,903,488	3,021,111,552	34,774,474,752	89,903,488	3,021,111,552
	($) MC Max	401,644,421,120	308,768,440,320	308,768,440,320	401,644,421,120	276,779,139,072
Health Care	% by #	15.23	12.08	13.33	15.84	11.99
	% by MC	12.49	13.53	14.38	9.12	13.19
	Count	442	57	4	385	53
	($) MC Min	21,050,850	1,478,593,408	152,328,667,136	21,050,850	1,478,593,408
	($) MC Max	313,432,473,600	313,432,473,600	313,432,473,600	18,889,377,792	108,768,911,360
Industrials	% by #	13.47	13.77	16.67	13.41	13.57
	% by MC	10.40	10.20	10.94	11.03	9.91
	Count	391	65	5	326	60
	($) MC Min	58,695,636	2,821,674,240	54,259,630,080	58,695,636	2,821,674,240
	($) MC Max	279,545,937,920	279,545,937,920	279,545,937,920	13,281,452,032	100,041,220,096
Information Technology	% by #	14.40	13.35	20	14.60	12.90
	% by MC	21.40	23.74	30.74	13.81	20.93
	Count	418	63	6	355	57
	($) MC Min	114,370,240	3,334,570,240	151,697,113,088	114,370,240	3,334,570,240
	($) MC Max	617,588,457,472	617,588,457,472	617,588,457,472	32,402,583,552	538,572,161,024
Materials	% by #	4.55	5.30	3.33	4.40	5.43
	% by MC	3.26	2.47	1.11	5.83	3.02
	Count	132	25	1	107	24
	($) MC Min	103,733,456	2,823,849,728	63,809,703,936	103,733,456	2,823,849,728
	($) MC Max	69,704,540,160	63,809,703,936	63,809,703,936	69,704,540,160	46,132,944,896
Real Estate	% by #	6.61	4.66	0.00	6.99	4.98
	% by MC	3.89	2.41	0.00	8.67	3.38
	Count	192	22	0	170	22
	($) MC Min	161,591,616	7,130,559,488	0.00	161,591,616	7,130,559,488
	($) MC Max	55,830,577,152	55,830,577,152	0.00	24,264,243,200	55,830,577,152
Telecommunication Services	% by #	1.03	1.06	3.33	1.03	0.90
	% by MC	2.40	2.57	3.79	1.82	2.09
	Count	30	5	1	25	4
	($) MC Min	285,299,072	3,964,831,488	217,610,731,520	285,299,072	3,964,831,488
	($) MC Max	261,176,721,408	261,176,721,408	217,610,731,520	47,389,126,656	261,176,721,408
Utilities	% by #	2.76	5.51	0.00	2.22	5.88
	% by MC	2.91	2.95	0.00	2.78	4.13
	Count	80	26	0	54	26
	($) MC Min	141,720,064	3,867,331,328	0.00	141,720,064	3,867,331,328
	($) MC Max	57,253,351,424	57,253,351,424	0.00	12,880,323,584	57,253,351,424

TABLE III. Market Capitalization (MC) statistics of equities under study broken out by Global Industry Classification Standard (GICS) sector. The composition of various indexes is displayed by the percentage of index constituents that are a member of each given sector (% by #) and by the weighting of those constituents (% by MC). Additionally the MC of the smallest and largest constituent for each index in each category is displayed.

Class	Stat	Russ 3K′	S&P 500′	Dow 30′	RexSP	SPexDow
Nano	% by #	0.14	0.00	0.00	0.16	0.00
MC $\leq$ \$50M	% by MC	0.00	0.00	0.00	0.00	0.00
	Count	4	0	0	4	0
Micro	% by #	11.51	0.00	0.00	13.74	0.00
\$50M < MC $\leq$ \$300M	% by MC	0.26	0.00	0.00	1.09	0.00
	Count	334	0	0	334	0
Small	% by #	42.89	0.42	0.00	51.13	0.45
\$300M < MC $\leq$ \$2B	% by MC	4.37	0.01	0.00	18.50	0.02
	Count	1,245	2	0	1,243	2
Mid	% by #	30.35	20.76	0.00	32.21	22.17
\$2B < MC $\leq$ \$10B	% by MC	15.11	3.37	0.00	53.19	4.72
	Count	881	98	0	783	98
Large	% by #	14.50	75.21	66.67	2.71	75.79
\$10B < MC $\leq$ \$200B	% by MC	56.68	67.77	43.28	20.72	77.59
	Count	421	355	20	66	335
Mega	% by #	0.62	3.60	33.33	0.04	1.58
MC > \$200B	% by MC	23.58	28.85	56.72	6.50	17.67
	Count	18	17	10	1	7

TABLE IV. Composition of indexes under study by market capitalization (MC) classification. The composition of various indexes is displayed by the percentage of index constituents that are a member of each given index (% by #) and by the weighting of those constituents (% by MC).

	Russ 3K′	S&P 500′	Dow 30′	RexSP	SPexDow
Count	2,903	472	30	2,431	442
(\$) MC Sum	26,217,754,755,404	20,040,462,107,136	5,736,789,102,592	6,177,292,648,268	14,303,673,004,544
(\$) MC Min	21,050,850	1,244,719,232	34,774,474,752	21,050,850	1,244,719,232
(\$) MC Max	617,588,457,472	617,588,457,472	617,588,457,472	401,644,421,120	538,572,161,024

TABLE V. Makeup of market indexes by number of constituents. Additionally the Market Capitalization (MC) of the smallest and largest constituent for each indexed is displayed along with sum of all constituent MCs.

by both its market capitalization and its rank in relation to other companies. In addition, we create disjoint sets of equities and compute aggregate statistics across these sets. Since the S&P 500 is a strict superset of the Dow 30 and the Russell 3000 is a strict superset of the S&P 500, the natural division of the superset of all equities under study is into three distinct classes: the Dow 30, the S&P 500 excluding the Dow 30 (SPexDOW), and the Russell 3000 excluding the S&P 500 (RexSP). We compute measures of correlation between these disjoint subsets, and characterize the statistical properties of the time series of dislocation segments and realized opportunity cost across these disjoint categories. We further explore the relationship between these categories by conducting a Granger causality analysis of aggregated realized opportunity cost time series [35].

We also analyze the distribution of microstructure quantities across exchanges. Since different exchanges may have substantially different hardware (such as fiber-optic cables and computers) and software (such as matching engines), it is possible that dislocation segment and realized opportunity cost distributions vary significantly across exchanges; examples of the hardware and software used in exchanges vary in type and complexity [36–39]. We compute summary statistics and provide qualitative comparisons of the results.

The next section gives results on dislocation segments, including summary statistics overall and across mutually-exclusive market category, and regressions of dislocation segments against market cap. We then discuss structure in the intra-day distribution of dislocation segment start times and dislocation segment duration. Following this, we provide statistics of the realized opportunity cost across the market as a whole and again within mutually-exclusive market categories. We then explore statistical properties of the ROC time series and conduct a Granger-causality analysis of ROC by mutually-exclusive market category. We close with an overview of the statistics of ETF dislocation segments and realized opportunity costs, contrasting these with those of the market as a whole.

IV. RESULTS

A. Dislocation Segments

Dislocation segments can occur when prices propagated by distinct information feeds differ. We cataloged all dislocation segments occurring in the equities studied here and present summary statistics and qualitative comparisons of their distributions and higher-order moment statistics. Tables VIII - XVI display means of summary statistics of dislocation segments for each mutually-exclusive subset of the equities under study. We will use the notation $\langle f_A \rangle$ to denote an average of the quantity f conditioned on the condition A. These averages are interpreted as the quantity f conditioned on condition A averaged over all securities and all times of observation; defining the number of instances of the quantity f having condition A as N_A, we have

$$\langle f_A \rangle = \frac{1}{N_A} \sum_{\substack{1 \le n \le N_A \\ f \text{ has condition } A}} f_n. \tag{1}$$

On average, there were more dislocation segments in Dow 30 securities ($\langle N_{uncond} \rangle \simeq 4 \times 10^6$, $\langle N_{duration} \rangle \simeq 2 \times 10^6$, $\langle N_{duration,\ magnitude} \rangle \simeq 9 \times 10^4$) than in SPexDow ($\langle N_{uncond} \rangle \simeq 2 \times 10^6$, $\langle N_{duration} \rangle \simeq 1 \times 10^6$, $\langle N_{duration,\ magnitude} \rangle \simeq 1 \times 10^5$) or RexSP ($\langle N_{uncond} \rangle \simeq 7 \times 10^5$, $\langle N_{duration} \rangle \simeq 2 \times 10^5$, $\langle N_{duration,\ magnitude} \rangle \simeq 4 \times 10^4$) securities. However, the average maximum magnitude of latency arbitrage opportunities in the Dow30 in each conditioning class (unconditioned, conditioned only on duration, and conditioned on both duration and magnitude) is lower than those of the SPexDow, which in turn are lower than those of the RexSP. In particular, actionable dislocation segments (those with duration $> 545\mu s$) with magnitude $> \$0.01$ exhibit more extreme behavior in the SPexDow and RexSP than in the Dow. On average, the median maximum magnitude in the Dow 30 among actionable dislocation segments was $\langle \text{median max mag}_{duration,magnitude} \rangle \simeq \$0.023$, while in the SPexDow we observed $\langle \text{median max mag}_{duration,magnitude} \rangle \simeq \$0.034$ and in the RexSP $\langle \text{median max mag}_{duration,magnitude} \rangle \simeq \$0.045$, a roughly one-cent increase in the median maximum magnitude of a dislocation segment in each mutually-exclusive category. Examples of distributions of these quantities are given in Figure 2, where the distributions of the means of minimum magnitude, maximum magnitude, and duration are plotted for the RexSP.

Taken as a whole, these results provide evidence for the existence of a market capitalization scaling effect in dislocation segments: larger securities are on average traded more and hence have a higher average number of dislocation segments, but exhibit on average less extreme behavior in moment and quantile statistics of these opportunities; more frequent trading implies a lower probability that prices across differing information feeds will diverge by large magnitudes.

Since dislocation segments are not distributed evenly throughout the day in the Dow 30 [13], we examine their distribution in the SPexDow and the RexSP as well. Appendix B contains figures displaying the distribution of latency arbitrage opportunity start times plotted modulo day and aggregated over the year and figures displaying the distribution of dislocation segment durations for each mutually exclusive market category. Distributions are plotted both without conditioning and when conditioned on duration or magnitude (or both).

Distributions of start times display predictable structure. In all market categories, there are large peaks at the very beginning and end of the trading day (circa 9:30 AM and 4:00 PM), along with a noticeable and sudden increase in density around 2:00 PM. The peak in density that occurs at the end of the day is most noticeable when the distribution of start times is not conditioned on dislocation segment size; when the distribution is conditioned, this peak is present but the density is much lower. These observations correspond with the results found for the Dow 30 in [13]. However, along with these granular observations, there also exists structure on shorter timescales. The distribution exhibits self-similarity on the half-hour timescale, with large peaks every half-hour and decreasing density toward a sudden peak at the next half-hour. There is also structure at the five-minute timescale that is noticeable before the 2:00 PM spike in density but does not appear to be present after the spike. (Future work could statistically test for the presence of this structure and for its persistence across multiple timescales.) The structure on shorter timescales is present in all distributions but, again, is more pronounced in distributions not conditioned on magnitude.

Distributions of dislocation segment duration also exhibit definite structure, though it is realized in a different way. These distributions are plotted log-transformed, as they are leptokurtic. All distributions exhibit one or more peaks in the range $10^{-4}s \le \log_{10}$ duration $\le 10^{-3}s$, but there is also a distinct and much lower peak in the distribution near approximately one second in length.

As a visual aid to these results, we have included circle plots, as introduced in [13], to demonstrate the non-uniform distribution of dislocation segments in both time- and event-space that can occur. We construct these circle graphs for the equities with the smallest market cap in our sample (GALE), highest market cap in our sample (AAPL), and the highest-volume S&P 500 ETF (SPY) for comparison. Figure 3 displays circle graphs for the above-mentioned tickers for an arbitrary day (2016-01-07), while Figure 4 displays circle graphs for the same tickers but aggregated over a year (modulo day). There is substantial variation between the circle graphs of these tickers. AAPL displays intricate structure in event space, while GALE display similar structure but with much lower density, as it has far fewer dislocation segments. In time-space, AAPL and GALE display similar nonuniform density of events, with both exhibiting notable peaks near the beginning of the trading day. AAPL also exhibits large spikes in density near 12:30 PM and 2:30 PM. In contrast, SPY displays time-space event density that is far more uniform than either AAPL or

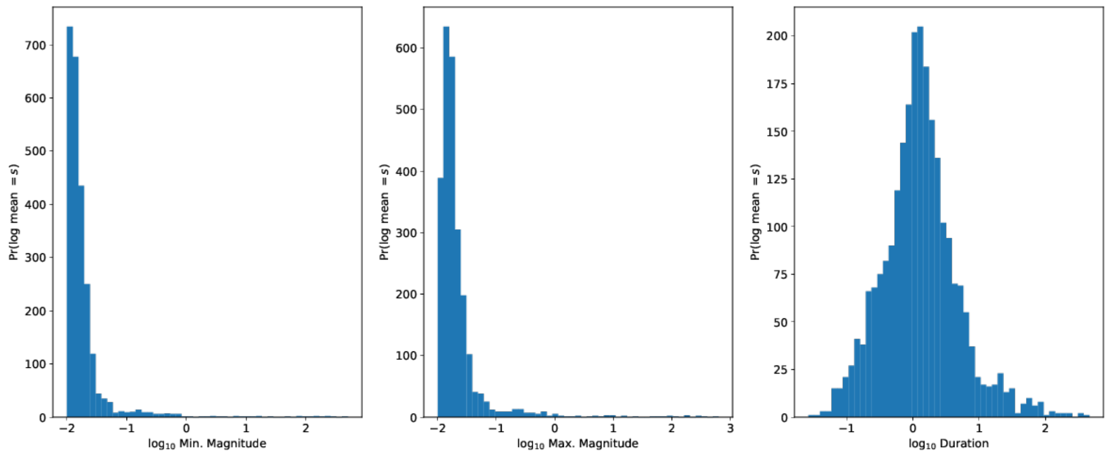

FIG. 2. Histograms of the base-10 logarithm of minimum magnitude, maximum magnitude, and duration of dislocation segments in the RexSP without conditioning on duration or magnitude. The distributions are leptokurtic, with the log-distributions of minimum and maximum magnitude presenting a long right tail and the distribution of log-duration displaying a rough bell-shape.

GALE. Visualizations of many securities in this format can be found at the authors' webpage [40].

B. Market capitalization

Further evidence for scaling behavior arises from analysis of market capitalization. Tables III and V display market capitalization statistics broken down by industry sector and categorical size, e.g., micro-cap, mega-cap, etc. Market capitalization is significantly positively correlated with realized opportunity cost. Tables XXXII - XXXV display results from ordinary least squares regressions predicting realized opportunity cost using predictors including market capitalization. A linear fit predicting $\log_{10}$ ROC from $\log_{10}$ market capitalization, $\log_{10}$ total trades, and $\log_{10}$ differing trades gives $R^2 \simeq 0.908$, with a positive coefficient relating $\log_{10}$ ROC to $\log_{10}$ market capitalization; higher market capitalization is associated with higher ROC. A similar regression is computed including quadratic terms in $\log_{10}$ market capitalization, which has a significant, but weak, negative association with ROC. Similar relationships hold for both the linear and quadratic models when the dependent variable is instead chosen to be total or differing trades.

Though behavior of ROC as a function of market capitalization is generally similar when equities are stratified by sector, some sectors display lower average levels of ROC, differing trades, or total trades when market capitalization is held constant. Equities classified as being in the financial sector generally have a smaller amount of ROC, while equities classified as being in the energy sector exhibit a higher amount of ROC on average. However, there is no clear general trend linking sectors to market capitalization or to ROC.

C. Realized opportunity cost

As expected with an increase in the number of analyzed equities from 30 to more than 2900, the amount of realized opportunity cost rose substantially from the quantity reported in an earlier study by the present authors and others, from \$160M to \$2.05B USD. Realized opportunity cost (ROC) clearly displays sublinear scaling with number of studied equities; we do not observe a thousandfold increase in the amount of ROC with a thousandfold increase in the number of equities. The information advantage afforded traders with access to direct feed information is not uniform; though a vast majority of the ROC (\$1.91 B) favored the direct feeds in this way, a non-negligible amount of ROC (\$137 M) did favor the SIP feeds. Of all trades observed (trades that occurred at the SIP BBO), approximately a quarter (23.71%) of all trades were classified as differing trades, or trades that occurred at a time during which there existed a different price at which the trade could have executed. The fraction of "differing traded value"—the nominal market value of all differing trades—was slightly higher (25.25%) than the fraction of all trades that were differing trades. The fractional part of the ratio of these fractions (1.0651) can be interpreted as a rough estimate of a rate of profitability of latency arbitrage trading strategies.

SPexDow securities account for a plurality of all trades under study that resulted in realized opportunity cost, with the median number of differing trades per day that

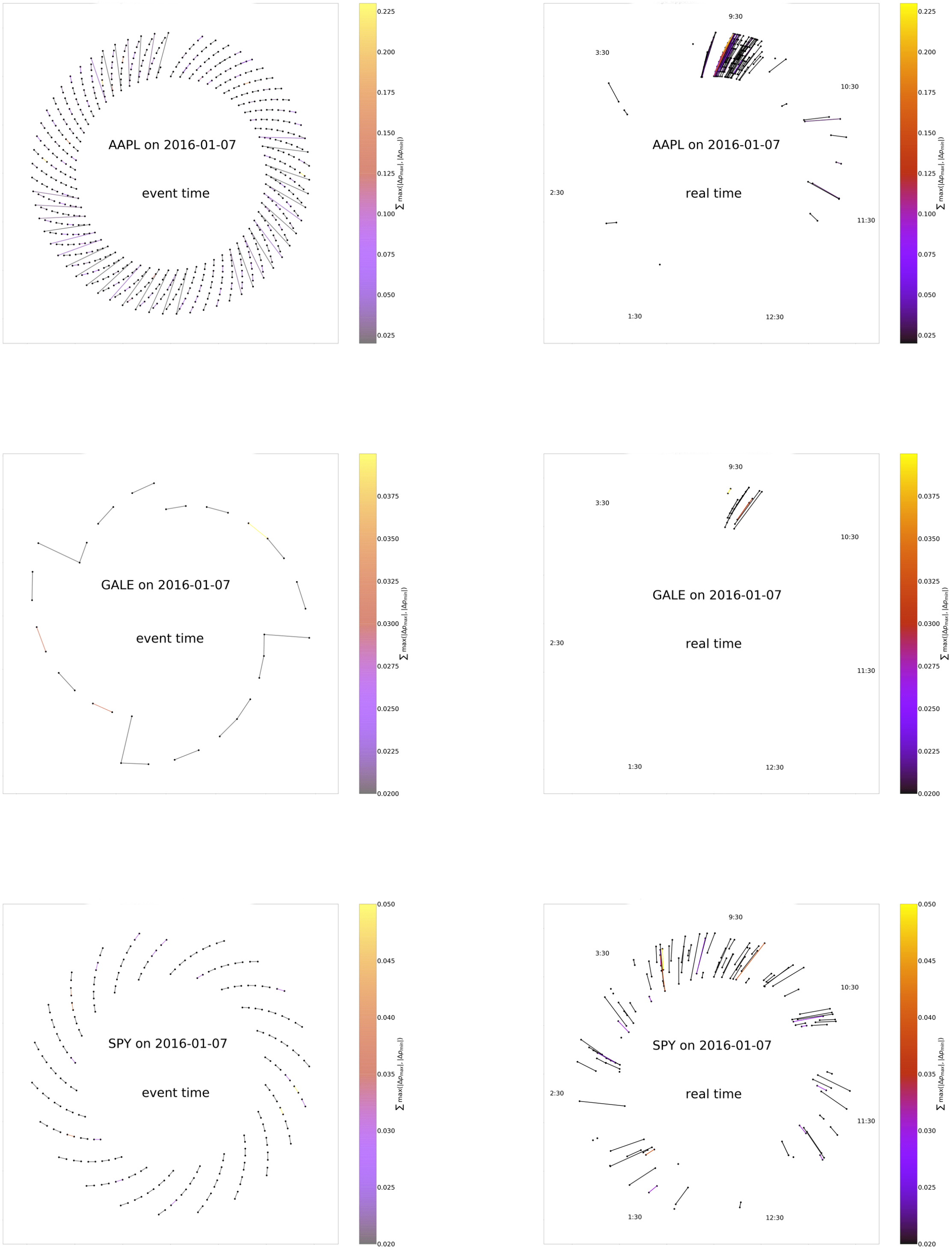

FIG. 3. Circle graphs constructed using the dislocation segments that occurred between 9:30am and 4:00pm on 2016-01-07. From top to bottom, where the left column is event-space and right column is time-space: AAPL, the equity with the highest market capitalization; GALE, the equity with the lowest market capitalization in our study; and SPY, the ETF that tracks the S&P 500.

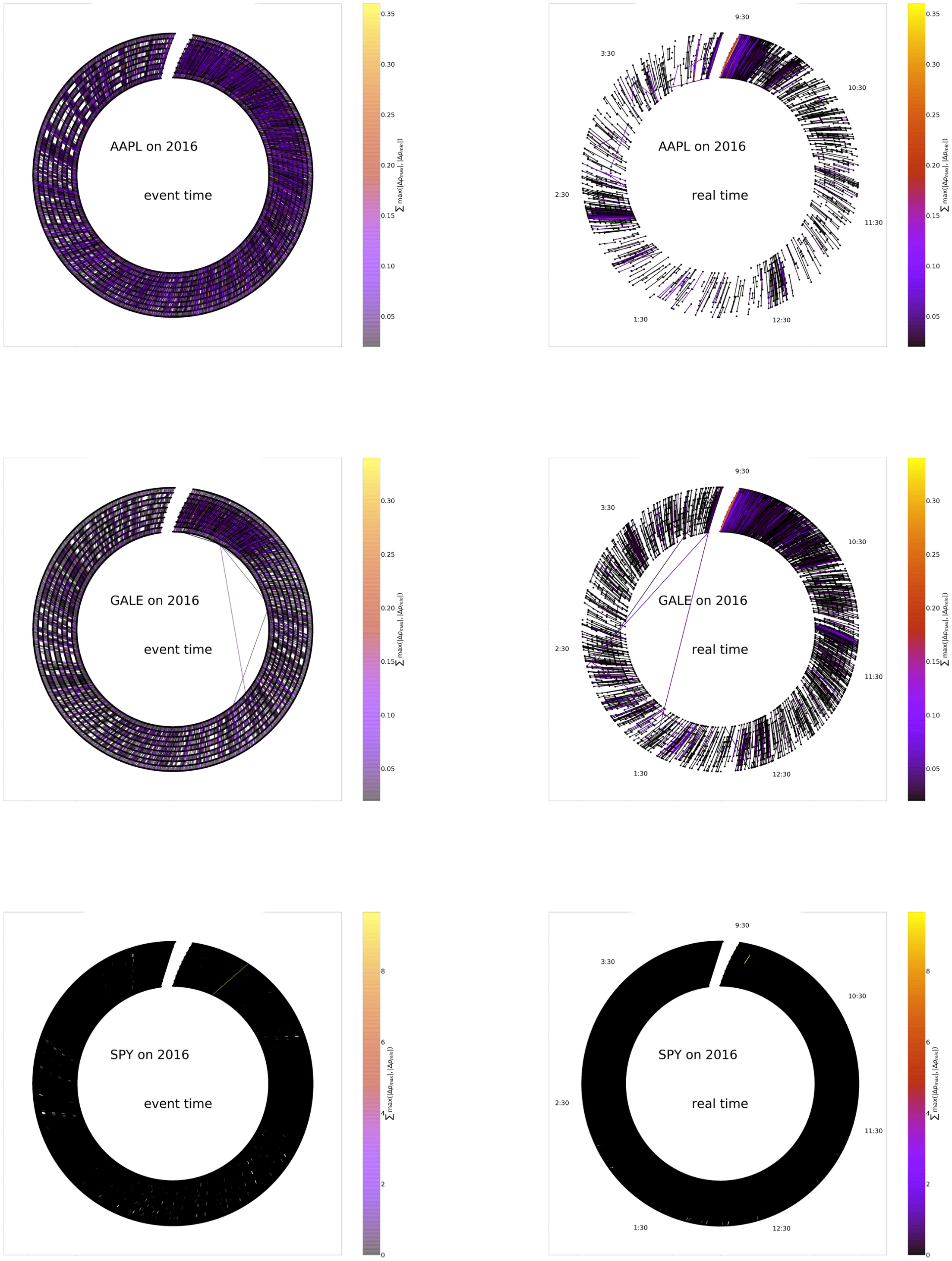

FIG. 4. Circle graphs constructed using the dislocation segments that occurred between 9:30am on 2016-01-01 and 4:00pm on 2016-12-31. From top to bottom, where the left column is event-space and right column is time-space: AAPL, the equity with the highest market capitalization; GALE, the equity with the lowest market capitalization in our study; and SPY, the ETF that tracks the S&P 500.

occurred in the SPexDow calculated to be 2,006,091, in contrast to the 309,158 in the Dow 30 or 1,921,121 in the RexSP. The median differing traded value per day in the SPexDow was also the highest among the three categories, totaling approximately \$14.07T versus the ReXSP's total of \$6.7T and the Dow's total of \$3.27T. Realized opportunity cost per share differed across the three categories, with median ROC per share per day on the Dow calculated to be \$0.011, on the SPexDow to be \$0.015, and on the RexSP to be \$0.021. This result is not surprising, as this particular measure of "market inefficiency" (ROC per share) increases with the decrease in average size of equity, with lowest ROC per share occurring in the Dow and highest ROC per share occurring in the set of equities ranked 501 and above. Median total ROC per day on the Dow amounted to \$514.8K, while median total ROC per day on the SPexDow totaled \$3.384M and on the RexSP amounted to \$3.564M. Summary statistics for distributions of ROC for each mutually-exclusive category are given in Tables XXV, XXIV and XXVI.

It is interesting to consider the distribution of both total ROC and ROC per share by both equity and mutually-exclusive category. Figure 6 displays ROC of the top 30 and bottom 30 of all securities under study when ranked by ROC. Included in this figure for comparison is the exchange-traded fund SPY, an ETF that tracks the S&P 500. (Selected ETFs are also treated separately in Section IV D.) It is notable that the equity with the largest ROC, Bank of America (BAC), has more than twice the ROC of the equity with the second-largest amount of ROC, Verizon (VZ). Though not an equity and not included in the rest of this study, it is also notable that SPY, one of the most heavily traded securities on the NMS along with BAC, is close to BAC in amount of ROC. Of the top 30 securities with most ROC, eight of the 30 are Dow 30 equities; only four out of 30 are RexSP equities, while the other 17 non-ETF securities are SPexDow equities. Since the S&P 500 appears to be the primary driver of ROC across all equities (c.f. below), we find the top 30 and bottom 30 S&P 500 securities ranked by ROC, including Dow 30 securities, and plot their ROC in Figure 7. Even in this subset, only 10 of the top 30 equities are Dow 30 securities. However, when the unit of analysis changes to ROC per share, as in Figure 8, we find that RexSP equities fill 27 out of 30 top ranks, which corresponds with the aggregated statistics reported in Table XXVI when compared with Tables XXIV and XXV.

Since there appear to be differences between the (stationary) summary statistics of the mutually-exclusive market categories, it is reasonable that there may be significant differences between the ROC statistics considered as time-dependent stochastic processes and simply considered as random variables decoupled from time. Within each category, the ROC was computed for all equities in that category for each day. Each ROC series is then normalized as $r_i \mapsto \frac{r_i - \langle r_i \rangle}{\sqrt{\mathrm{Var}(r_i)}}$, which allows direct comparison of the series. Figure 16 displays a quantile-quantile plot of the Dow, SPexDow, and RexSP ROC distributions. The Dow distribution is plotted as linear and the other two distributions are compared with it. It is immediately obvious that the left tails of the SPexDow and RexSP distributions are heavier than that of the Dow; this also appears to be the case for the right tails of the distributions, but there is little sampling in this region and so no conclusion can be drawn. This similarity of the SPexDow and RexSP distributions is also striking; when normalized they appear almost identical.

Figure 17 displays the time-dependent sample paths of ROC sampled at daily resolution. These processes are anti-autocorrelated—they display mean reversion—as evidenced by their detrended fluctuation analysis (DFA) [41] exponents of $\alpha_{\mathrm{Dow}} = 0.438$, $\alpha_{\mathrm{SPexDow}} = 0.242$, and $\alpha_{\mathrm{RexSP}} = 0.235$. All series exhibit rare large values from time to time, with the Dow ROC series exhibiting the largest rare values relative to its mean fluctuations and the SPexDow series exhibiting the smallest. We also note that, in accordance with the QQ plot of the time-decoupled distributions above, the DFA exponents of the SPexDow and RexSP—and thus their corresponding dynamical behavior—are closer than they are to the Dow DFA exponent.

A review of the above results points to the SPexDow as being the "dominant" mutually-exclusive category in some sense: it accounts for a plurality of differing trades, differing traded value, and total ROC, while also having a DFA exponent lower than that of the Dow and close in value to that of the RexSP, meaning that its time-series of ROC is strongly mean-reverting. The amalgamation of these facts can be interpreted as evidence that the SPexDow ROC time series is possibly least likely to be influenced by the other series of ROC. To test this hypothesis, we conduct a number of Granger causality tests on the time series of ROC. Granger causality is the notion that, beyond mere correlation in time series, past values of one time series may be useful in predicting current and future values of another time series [35]. A maximum lag of 40 days was set and four tests were calculated pairwise between each of the three mutually-exclusive categories: sum of squared residuals χ^2-test, a likelihood ratio test, sum of squared residuals F-test, and a Wald test. We consider there to be a significant Granger causality between series when all four tests indicate significant Granger causality at the $p = 0.05/N_{\mathrm{lags}}$ confidence level. The correction for multiple comparisons is done using the most conservative estimate, the Bonferroni correction, to minimize the probability of Type I error [42]. Figure 5 displays the results of these tests graphically as a directed network. The direction of edges denotes the direction of the Granger-causal relationship between the categories, while the weights on the edges denote the total number of lags for which the relationship was significant. The SPexDow is shown to significantly influence both the Dow and RexSP while not being significantly influenced by either category; this provides strong evidence to support our above hypothesis. We note that the SPY tracks the S&P 500, is one of the most heavily-traded securities, and has the second-highest amount of ROC of the securities under study here. The SPY's price dynamics and ROC may thus have a material effect on the relationships between the S&P 500's ROC and those of the other market categories, providing a partial confounding effect to the Granger-causal relationship determined here; there may be a mutually-causal relationship between the real S&P 500 and the ETF that tracks it. The RexSP and Dow have a mutually Granger-causal relationship, with the Dow exerting more influence on the RexSP than the other way around. This finding corresponds with the ranking of categories on a total shares traded per number of equities basis; this is not a surprising result. We also find that the SPexDow exerts far less influence on the RexSP than does the Dow (four total lags for the SPexDow versus 23 total lags for the Dow), a fact for which we do not have a ready explanation.

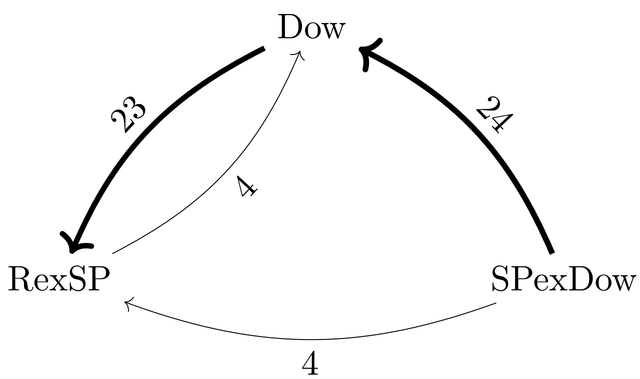

FIG. 5. Network of relationships between mutually-exclusive market categories implied by results of four Granger causality tests. The direction of the edges gives the direction of the Granger-causal relationship, while the weight on the edge is the total number of lags for which the relationship was significant at the $p = 0.05/N_{\text{lags}}$ level (the conservative Bonferroni correction). The maximum number of lags was chosen to be $N_{\text{lags}} = 40$. Thickness of the edge is proportional to edge weight and is plotted for emphasis in visualization.

Providing further evidence for the above hypothesis, we compute Pearson correlations between pairs of mutually exclusive categories for both ROC and ROC per share; these results are displayed in Table VI. ROC correlations are strongest between SPexDow and RexSP ($\rho = 0.72$) and SPexDow and Dow ($\rho = 0.45$), while the correlation between the RexSP and Dow is lower ($\rho = 0.31$). ROC per share correlations are universally lower than those for ROC, but the correlations between SPexDow and RexSP ($\rho = 0.41$) and SPexDow and Dow ($\rho = 0.10$) are still higher than that between RexSP and Dow ($\rho = -0.01$), which is actually negative.

Finally, we break ROC out by both exchange and (not mutually-exclusive) market category. Figures 9, 10, and 11 display the distributions of total ROC per day in 2016 by exchange; the vertical axis is transformed as $x \mapsto \log_{10}(|x| + 1)$ for ease of viewing. (We choose this transformation for its usefulness in the case where data has both a large range and many values near zero. Since $\log(|x| + 1) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{|x|^n}{n}$ for $|x| < 1$, near zero we have $\log(|x| + 1) \simeq |x|$, while at large values of $|x|$ it is essentially indistinguishable from $\log|x|$.) In all three categories, the exchanges with the most total ROC are NYSE and NASDAQ, followed by ARCA, BATS, and EDGX. Consistent with all summary statistics reported in previous work, CHX and NSX have by far the least ROC across all three categories [13]. Turning our analysis to ROC per share by exchange, Figures 12, 13, and 14 display daily ROC per share by exchange. The left panel of each figure shows the entire distribution, while the right panel displays the distribution truncated to within the 99th and 1st percentiles. The non-truncated distributions are almost-comically heavy-tailed—the excess kurtosis (kurtosis minus 3) of the EDGX - Russell distribution is $\kappa \simeq 55,652$—hence the truncation for the purposes of analysis; it seems likely that the kurtosis of the theoretical distributions do not exist, implying tail exponent $\gamma < 4$ in the distribution $\Pr(X > x) \sim x^{-(\gamma-1)}$.

(Table XXVII displays the kurtosis and skew for each non-truncated distribution.) The means of the truncated distributions are significantly above one cent per share favoring the direct feeds (i.e., significantly below $-\$0.01$) in all distributions. Among Dow 30 equities, CHX and NYSE have exceptionally long tails favoring the direct feeds. In the case of S&P 500 tickers, FINRA (alternative trading facilities / dark pools) have a distribution significantly more toward the direct feed than any exchange; its mean ROC per share is above \$0.20, while when considering the entire superset of all equities in the Russell 3000 NYSE, AMEX and FINRA are all leptokurtic favoring the direct feeds.

D. ETFs

Exchange traded funds (ETFs) are securities that trade on the NMS and are designed to mimic as closely as possible a particular portfolio of other securities. They are thus governed by the same price discovery mechanism as other securities that trade on the NMS, as opposed to the end-of-day price discovery mechanism to which mutual funds are subjected, but also allow investors to own a portion of potentially many underlying assets (or at least a simulacrum of such), similar to a mutual fund. Here, we briefly remark on the similarities and differences between ETFs designed to track subsets of the market and those subsets of the market themselves. We calculated statistics on the dislocation segments and realized opportunity cost attributed to ten ETFs, the descriptions of which are given in Table VII. We concentrate on measures of whole-market activity, with three ETFs that track the S&P 500 and Russell 3000 respectively. To iso-

ROC

	Dow	SPexDow	RexSP
Dow	1.000000	0.451072	0.319018
SPexDow	0.451072	1.000000	0.724903
RexSP	0.319018	0.724903	1.000000

ROC / Share

	Dow	SPexDow	RexSP
Dow	1.000000	0.103061	-0.019662
SPexDow	0.103067	1.000000	0.411443
RexSP	-0.019662	0.411443	1.000000

TABLE VI. Pearson correlation matrices of mutually-exclusive index subsets. For each index subset a daily resolution time series is constructed for the given statistic over all stocks in the index subset. For the ROC series the ROC generated for each stock on a particular trading day is summed, while in the ROC per share case the values are averaged. The correlation coefficients are then calculated between pairs of time series in order to construct the tables above. The left table displays ROC correlations, while the right table displays ROC per share correlations. The ROC per share statistic normalizes the number of traded shares, allowing for a fair comparison between the more heavily traded stocks in the Dow 30 or S&P 500 subset with the more lightly traded stocks in the Russell 3000 subset.

late dynamics among ETFs that track smaller equities, we also include three ETFs that track the Russell 2000, the smallest 2000 equities by market cap in the Russell 3000.

Table XXVIII summarizes ROC statistics for the ETFs under study. The fraction of differing trades and differing traded value are lower than for any of the indexes as a whole; in fact, the ratio of the fraction of differing traded value to the fraction of differing trades is less than one. Total realized opportunity cost incurred from trades in ETFs studied here totaled over $38 million in calendar year 2016. This statistic provides some evidence to suggest that ETFs have their own endogenous statistical behavior that differs from the behavior of the assets from which they are derived.

ETF	Description
SPY	SPDR S&P 500 ETF, one of the most heavily traded securities on the NMS (c.f. above)
VOO	Vanguard S&P 500 ETF
IVV	iShares S&P 500 ETF
THRK	SPDR Russell 3000 ETF
VTHR	Vanguard Russell 3000 ETF
IWV	iShares Russell 3000 ETF
TWOK	SPDR Russell 2000 ETF
VTWO	Vanguard Russell 2000 ETF
IWN	iShares Russell 2000 ETF

TABLE VII. Descriptions of the ETFs studied.

V. CONCLUSION

In sum, we have demonstrated that the existence of dislocation segments and realized opportunity cost is not restricted to Dow 30 securities. Furthermore, we have established that these microstructure quantities occur with non-negligible frequency and size; we show that total realized opportunity cost in Russell 3000 securities was well in excess of $2 billion USD during 2016. Compounding these results, we provide strong statistical evidence that the S&P 500 excluding Dow 30 securities, to which we refer as the SPexDow, is the primary driver of realized opportunity cost among the three mutually exclusive categories of equities (Dow 30, SPexDow, and Russell 3000 excluding S&P 500 securities, or the RexSP).

Compounding the above results, we find that structure in the distributions of dislocation segment start times and duration persist across the entire Russell 3000, indicating some broader microstructure-based proximate cause of this structure. Distributions of latency arbitrage duration exhibit a large peak between 10^{-4} and 10^{-3} seconds (100 microseconds to one millisecond), but also exhibit a second smaller, yet distinct, peak near one second. This separation of timescales in the distribution provide evidence for the existence of at least two distinct proximate causes of latency arbitrage opportunities. Distributions of dislocation segment start times display even more intricate structure, with large peaks at the beginning and end of the trading day, self-similarity on the half-hour and ten-minute timescales, and a large spike at 2:00 PM.

Realized opportunity cost was highest among SPexDow securities, but realized opportunity cost per share was highest among RexSP securities, which were also the most lightly-traded securities. All time series of realized opportunity cost exhibit behavior of anti-autocorrelation, meaning that they are mean-reverting. Realized opportunity costs in the SPexDow Granger-cause realized opportunity costs in the other market categories, but the converse is not true; while the Dow Granger-causes the RexSP, the RexSP only weakly Granger-causes the Dow and does not have any effect on the SPexDow. When considering realized opportunity cost by exchange (including the aggregation of all alternative trading facilities, denoted as FINRA), the largest exchanges by market volume (NYSE, INET, ARCA, BATS, and EDGX) capture a large majority of the realized opportunity cost. However, passing to analysis of realized opportunity cost per share, NYSE, CHX, and FINRA are clear outliers in the Dow, S&P. and Russell; AMEX is a clear outlier when considering the Russell superset.

Taken together, these results paint the picture of a NMS the physical structure of which generates effects that are persistent across size of equity and exchange. Amplifying these persistent effects is the apparent central role of the SPexDow; in number of dislocation segments, amount of realized opportunity cost, spectral properties of realized opportunity cost time series, and Granger-causal relationships, the story emerges of the SPexDow's characteristics being generated by largely-endogenous factors and subsequently influencing the characteristics of the Dow and RexSP. Future work could explore in more depth the extent to which microstructure effects arising first in the SPexDow then spread to other mutually exclusive market categories and propagate through time. This work could also explore the evolutionary dynamics of the modern NMS from its birth following the financial crisis of 2007/8 to the present day. The NMS may not have remained static, with a constant number of market centers and a stationary distribution of market agents and trading strategies, but rather may have experienced fluctuations in the number of exchanges, in the regulatory environment, and in strategy profiles of trading agents. Such an analysis could pave the way for more well-informed modelling efforts and development of financial economic theory.

VI. ACKNOWLEDGEMENTS

The authors gratefully acknowledge helpful discussions with Anshul Anand, Yosry Barsoum, Lashon Booker, David Bringle, Peter Carrigan, Gary Comparetto, Peter Sheridan Dodds, Bryanna Dienta, Jordan Feidler, Andre Frank, Bill Gibson, Frank Hatheway, Michelle Herd, Emily Hiner, Chuck Howell, Eric Hunsader, Neil Johnson, Matthew Koehler, Connie Lewis, Matthew McMahon, Wade Shen, David Slater, Jonathan Smith, Brendan Tivnan, Jason Veneman, and Thomas Wilk. All opinions and remaining errors are the sole responsibility of the authors and do not reflect the opinions nor perspectives of their affiliated institutions nor that of the funding agencies. D.R.D., C.M.V.O., J.H.R., and B.F.T. were supported by DARPA award #W56KGU-17-C-0010. C.M.D. was supported by NSF grant #1447634. C.M.D. was supported by a gift from Massachusetts Mutual Life Insurance Company. The authors declare no conflicts of interest. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

[1] Josef Penso de la Vega. *Confusion Des Confusiones*. Baker Library, Harvard Graduate School of Business Administration, 1688.

[2] Louis Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.

[3] Frank H Knight. Risk, uncertainty and profit. *New York: Hart, Schaffner and Marx*, 1921.

[4] Eugene F Fama. The behavior of stock-market prices. *The journal of Business*, 38(1):34–105, 1965.

[5] George A Akerlof. The market for lemons: Quality uncertainty and the market mechanism. In *Uncertainty in Economics*, pages 235–251. Elsevier, 1978.

[6] David Easley and Maureen O'hara. Price, trade size, and information in securities markets. *Journal of Financial economics*, 19(1):69–90, 1987.

[7] James J Angel, Lawrence E Harris, and Chester S Spatt. Equity trading in the 21st century. *The Quarterly Journal of Finance*, 1(01):1–53, 2011.

[8] James J Angel, Lawrence E Harris, and Chester S Spatt. Equity trading in the 21st century: An update. *The Quarterly Journal of Finance*, 5(01):1550002, 2015.

[9] Alex D Wissner-Gross and Cameron E Freer. Relativistic statistical arbitrage. *Physical Review E*, 82(5):056104, 2010.

[10] Shengwei Ding, John Hanna, and Terrence Hendershott. How slow is the nbbo? a comparison with direct exchange feeds. *Financial Review*, 49(2):313–332, 2014.

[11] Jacob Adrian. Informational inequality: How high frequency traders use premier access to information to prey on institutional investors. *Duke L. & Tech. Rev.*, 14:256, 2016.

[12] Phil Mackintosh. The need for speed. 2014.

[13] Brian F. Tivnan, David Rushing Dewhurst, Colin M. Van Oort, John H. Ring IV, Tyler J. Gray, Brendan F. Tivnan, Peter Sheridan Dodds, Christopher M. Danforth, Matthew T. K. Koehler, Matthew T. McMahon, David Slater, and Jason Veneman. Communication inefficiencies in the 2016 u.s. stock market: Evidence from the dow 30. 2018.

[14] Benoit B Mandelbrot. The variation of certain speculative prices. In *Fractals and scaling in finance*, pages 371–418. Springer, 1997.

[15] Benoit B Mandelbrot. *Fractals and scaling in finance: Discontinuity, concentration, risk. Selecta volume E*. Springer Science & Business Media, 2013.

[16] H Eugene Stanley and Vasiliki Plerou. Scaling and universality in economics: empirical results and theoretical interpretation. 2001.

[17] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. 2001.

[18] Felix Patzelt and Jean-Philippe Bouchaud. Universal scaling and nonlinearity of aggregate price impact in financial markets. *Physical Review E*, 97(1):012304, 2018.

[19] Tiziana Di Matteo. Multi-scaling in finance. *Quantitative finance*, 7(1):21–36, 2007.

[20] H Eugene Stanley, Vasiliki Plerou, and Xavier Gabaix. A statistical physics view of financial fluctuations: Evidence for scaling and universality. *Physica A: Statisti-*

cal Mechanics and its Applications, 387(15):3967–3981, 2008.
[21] Laura Tuttle. Alternative trading systems: Description of ats trading in national market system stocks, October 2013.
[22] For example, electronic communication networks occasionally offer incentives to trade on their platforms, including small price discounts; this may be considered as a form of price discovery.
[23] For example, execution of certain order types, such as an intermarket sweep order, which sweeps up as many shares as possible from across the market in a particular security, may result in some shares being purchased or sold at a non-NBBO price.
[24] Marc Schauten, Martijn J van den Assem, Dennie van Dolder, and Remco CJ Zwinkels. Can the market divide and multiply? a case of 807 percent mispricing in absence of arbitrage risk. 2018.
[25] Owen A Lamont and Richard H Thaler. Anomalies: The law of one price in financial markets. *Journal of Economic Perspectives*, 17(4):191–202, 2003.
[26] James J Choi, David Laibson, and Brigitte C Madrian. Why does the law of one price fail? an experiment on index mutual funds. *The Review of Financial Studies*, 23(4):1405–1432, 2009.
[27] Midas: Market information data analytics system, 2013.
[28] Thesys Technologies website: https://www.thesystech.com/products.html.
[29] Market capitalization classifications (e.g., small, mid, large) are adopted from https://www.investopedia.com/insights/understanding-small-and-big-cap-stocks/.
[30] Sal Arnuk and Joseph Saluzzi. Latency arbitrage: The real power behind predatory high frequency trading. *Themis Trading white paper*, 2009.
[31] Robert A Jarrow and Philip Protter. A dysfunctional role of high frequency trading in electronic markets. *International Journal of Theoretical and Applied Finance*, 15(03):1250022, 2012.
[32] Joel Hasbrouck and Gideon Saar. Low-latency trading. *Journal of Financial Markets*, 16(4):646–679, 2013.
[33] Maureen OHara. High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270, 2015.
[34] https://fred.stlouisfed.org/series/RU3000TR.
[35] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
[36] Walter D Braddock III. Automated stock exchange, October 25 1983. US Patent 4,412,287.
[37] Erik Brynjolfsson and Shinkyu Yang. The intangible costs and benefits of computer investments: Evidence from the financial markets. In *Atlanta, Georgia: Proceedings of the International Conference on Information Systems*, 1999.
[38] Ronald S Indeck, Ron Kaplan Cytron, Mark Allen Franklin, and Roger D Chamberlain. Method and apparatus for processing financial information at hardware speeds using fpga devices, November 29 2011. US Patent 8,069,102.
[39] John W Lockwood, Adwait Gupte, Nishit Mehta, Michaela Blott, Tom English, and Kees Vissers. A low-latency library in fpga hardware for high-frequency trading (hft). In *2012 IEEE 20th annual symposium on high-performance interconnects*, pages 9–16. IEEE, 2012.
[40] https://compfi.org.
[41] C-K Peng, Sergey V Buldyrev, Shlomo Havlin, Michael Simons, H Eugene Stanley, and Ary L Goldberger. Mosaic organization of dna nucleotides. *Physical review e*, 49(2):1685, 1994.
[42] C Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8:3–62, 1936.

Appendix A: Tables

1. Dislocation Segments

2. Realized opportunity cost

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	4,011,848.7333	4,011,848.7333	4,011,848.7333
mean	0.0110	0.0136	0.0754
std	0.0391	0.2725	5.8295
min	0.0100	0.0100	0.0000
25%	0.0100	0.0100	0.0002
50%	0.0100	0.0100	0.0007
75%	0.0100	0.0103	0.0013
max	44.6933	279.2057	8,408.9315

TABLE VIII. Mean of summary statistics for the Dow 30 without conditioning on duration or magnitude.

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	2,169,106.5333	2,169,106.5333	2,169,106.5333
mean	0.0108	0.0149	0.1328
std	0.0436	0.3548	7.6454
min	0.0100	0.0100	0.0005
25%	0.0100	0.0100	0.0008
50%	0.0100	0.0100	0.0011
75%	0.0100	0.0107	0.0027
max	43.4150	279.1987	8,408.9315

TABLE IX. Mean of summary statistics for the Dow 30 conditioning on duration $> 545\mu s$

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	95,757.8000	95,757.8000	95,757.8000
mean	0.0427	0.2370	0.9557
std	0.3355	1.6130	48.2148
min	0.0200	0.0200	0.0005
25%	0.0200	0.0200	0.0007
50%	0.0200	0.0227	0.0011
75%	0.0307	0.0432	0.0036
max	43.4150	114.3480	7,186.8665

TABLE X. Mean of summary statistics for the Dow 30 conditioning on both duration $> 545\mu s$ and magnitude >$0.01.

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	2,525,082.0448	2,525,082.0448	2,525,082.0448
mean	0.0135	0.0168	0.2530
std	0.2801	0.3996	9.3252
min	0.0100	0.0100	0.0000
25%	0.0100	0.0100	0.0002
50%	0.0100	0.0101	0.0006
75%	0.0115	0.0136	0.0011
max	476.1177	522.6072	9,084.0401

TABLE XI. Mean of summary statistics for the SPexDow without conditioning on duration or magnitude

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	1,189,460.6682	1,189,460.6682	1,189,460.6682
mean	0.0134	0.0185	0.5558
std	0.4601	0.6076	13.0295
min	0.0100	0.0100	0.0005
25%	0.0100	0.0100	0.0008
50%	0.0102	0.0107	0.0011
75%	0.0117	0.0160	0.0082
max	471.7331	515.4222	9,084.0401

TABLE XII. Mean of summary statistics for the SPexDow conditioning on duration $> 545\mu s$

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	114,770.0224	114,770.0224	114,770.0224
mean	0.0557	0.1249	1.5915
std	1.9177	2.5050	54.0650
min	0.0200	0.0200	0.0005
25%	0.0202	0.0209	0.0007
50%	0.0228	0.0346	0.0012
75%	0.0375	0.0625	0.0278
max	471.7331	506.9715	6,943.1063

TABLE XIII. Mean of summary statistics for the SPexDow conditioning on both duration $> 545\mu s$ and magnitude >$0.01.

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	770,577.8246	770,577.8246	770,577.8246
mean	0.9734	1.1361	4.4132
std	34.0534	37.7472	50.0793
min	0.0100	0.0100	0.0000
25%	0.0116	0.0121	0.0002
50%	0.0139	0.0149	0.0010
75%	0.0225	0.0302	0.0138
max	2,238.1205	2,514.9617	8,796.9568

TABLE XIV. Mean of summary statistics for the RexSP without conditioning on either duration or magnitude.

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	287,399.7217	287,399.7217	287,399.7217
mean	1.2116	1.7162	12.7495
std	37.6277	46.3599	83.4650
min	0.0100	0.0100	0.0005
25%	0.0110	0.0118	0.0021
50%	0.0147	0.0188	0.0722
75%	0.0263	0.0408	0.9755
max	2,033.1633	2,302.4541	8,796.9568

TABLE XV. Mean of summary statistics for the RexSP conditioning on duration $> 545\mu s$

	min. magnitude ($)	max. magnitude ($)	duration (s)
count	45,062.3366	45,062.3366	45,062.3366
mean	2.1734	3.0486	13.1546
std	53.2211	66.0958	112.1013
min	0.0200	0.0200	0.0005
25%	0.0239	0.0272	0.0039
50%	0.0338	0.0449	0.0536
75%	0.0611	0.0806	0.7988
max	2,033.9931	2,295.6782	7,139.0753

TABLE XVI. Mean of summary statistics for the RexSP conditioning on both duration $> 545\mu s$ and magnitude >$0.01.

1	Realized Opportunity Cost	$2,013,458,668.87
2	SIP Opportunity Cost	$1,876,048,519.06
3	Direct Opportunity Cost	$137,410,149.76
4	Trades	4,658,307,833
5	Diff. Trades	1,105,201,803
6	Traded Value	$24,352,760,600,270.47
7	Diff. Traded Value	$6,272,439,590,589.91
8	Percent Diff. Trades	23.73
9	Percent Diff. Traded Value	25.76
10	Ratio of 9 / 8	1.0855

TABLE XVII. Summary statistics of realized opportunity cost for all equities (all securities except for ETFs) during 2016.

1	Realized Opportunity Cost	$160,213,922.95
2	SIP Opportunity Cost	$122,081,126.40
3	Direct Opportunity Cost	$38,132,796.55
4	Trades	392,101,579
5	Diff. Trades	87,432,231
6	Traded Value	$3,858,963,034,003.48
7	Diff. Traded Value	$900,535,924,961.72
8	Percent Diff. Trades	22.30
9	Percent Diff. Traded Value	23.34
10	Ratio of 9 / 8	1.0465

TABLE XVIII. Summary statistics of realized opportunity cost for Dow 30 securities during 2016.

1	Realized Opportunity Cost	$1,064,715,340.25
2	SIP Opportunity Cost	$964,098,388.26
3	Direct Opportunity Cost	$100,616,951.99
4	Trades	2,564,892,761
5	Diff. Trades	623,146,506
6	Traded Value	$18,429,250,470,003.83
7	Diff. Traded Value	$4,567,166,871,544.24
8	Percent Diff. Trades	24.30
9	Percent Diff. Traded Value	25.83
10	Ratio of 9 / 8	1.0631

TABLE XIX. Summary statistics of realized opportunity cost for S&P 500 securities during 2016.

1	Realized Opportunity Cost	$904,501,417.30
2	SIP Opportunity Cost	$842,017,261.86
3	Direct Opportunity Cost	$62,484,155.44
4	Trades	2,172,791,182
5	Diff. Trades	535,714,275
6	Traded Value	$13,824,440,155,934.76
7	Diff. Traded Value	$3,666,630,946,582.52
8	Percent Diff. Trades	24.66
9	Percent Diff. Traded Value	26.52
10	Ratio of 9 / 8	1.0757

TABLE XX. Summary statistics of realized opportunity cost for SPexDow equities during 2016.

1	Realized Opportunity Cost	$948,743,328.62
2	SIP Opportunity Cost	$911,950,130.85
3	Direct Opportunity Cost	$36,793,197.77
4	Trades	2,093,415,072
5	Diff. Trades	482,055,297
6	Traded Value	$6,669,357,410,332.23
7	Diff. Traded Value	$1,705,272,719,045.67
8	Percent Diff. Trades	23.03
9	Percent Diff. Traded Value	25.57
10	Ratio of 9 / 8	1.1104

TABLE XXI. Summary statistics of realized opportunity cost for RexSP equities during 2016.

	Trades	Traded Value ($)	Diff. Trades	Diff. Traded Value ($)	ROC ($)	ROC/Share ($/shares)
count	720,991	720,991	720,991	720,991	720,991	720,991
mean	6,460.98	33,776,788.61	1,532.89	8,699,747.42	2,792.63	0.020880
std	13,249.67	109,021,779.70	3,036.98	25,738,960.57	17,611.14	0.087810
min	0	0	0	0	0	0
25%	599	1,118,022.02	101	199,882.83	237.6100	0.009510
50%	2,020	5,316,322.22	450	1,246,241.41	826.6000	0.011448
75%	6,478	24,797,793.44	1,600	6,525,124.17	2,578.75	0.018836
max	517,270	8,280,915,338.59	103,885	1,596,912,962.05	6,798,041.07	19.3381

TABLE XXII. Purse statistics for all stocks under study in 2016. The data used to construct this table is aggregated by date and stock, resulting in 720,991 data points that correspond with the 731,556 combinations of 252 trading days in 2016 and 2903 stocks under study.

	Trades	Traded Value ($)	Diff. Trades	Diff. Traded Value ($)	ROC ($)	ROC/Share ($/shares)
count	252	252	252	252	252	252
mean	10,178,145.88	70,172,234,880.71	2,472,803.60	18,123,678,061.68	4,225,060.87	0.014624
std	2,406,751.15	14,303,150,882.94	775,201.38	4,760,162,875.50	1,531,548.30	0.002019
min	3,862,591	29,960,724,653.08	709,540	5,941,906,620.96	1,281,537.70	0.011127
25%	8,716,552.50	60,764,387,798.11	2,034,844.50	15,251,685,767.67	3,371,948.52	0.013502
50%	9,684,039	67,776,548,100.32	2,310,806	17,479,288,594.91	3,918,496.70	0.014407
75%	11,120,226.50	75,672,607,052.02	2,783,838.50	20,074,235,595.26	4,654,693.39	0.015434
max	19,505,364	128,057,685,223.37	5,715,448	37,114,729,300.67	14,335,072.09	0.031484

TABLE XXIII. Aggregated purse statistics for S&P 500 constituents in 2016. The data used to construct this table is aggregated by date, resulting in 252 data points that correspond with the 252 trading days in 2016.

	Trades	Traded Value ($)	Diff. Trades	Diff. Traded Value ($)	ROC ($)	ROC/Share ($/shares)
count	252	252	252	252	252	252
mean	8,622,187.23	54,858,889,507.68	2,125,850.30	14,550,122,803.90	3,589,291.34	0.014818
std	1,960,102.37	10,686,728,768.81	632,025.23	3,571,347,460.11	1,119,395.15	0.002029
min	3,283,385	23,296,053,599.93	619,976	4,906,051,591.25	1,136,332.05	0.011271
25%	7,398,970.25	48,123,050,130.46	1,762,152.75	12,329,749,894.94	2,915,802.29	0.013729
50%	8,237,387.50	53,383,376,977.72	2,006,091.50	14,073,439,429.50	3,384,654.11	0.014579
75%	9,405,905.75	59,188,646,444.18	2,398,085.25	15,973,362,072.81	4,050,343.31	0.015660
max	15,909,358	99,048,039,796.82	4,642,419	27,685,776,913.57	9,097,891.31	0.032760

TABLE XXIV. Aggregated purse statistics for S&P 500 constituents that were not also Dow 30 constituents in 2016. The data used to construct this table is aggregated by date, resulting in 252 data points that correspond with the 252 trading days in 2016.

	Trades	Traded Value ($)	Diff. Trades	Diff. Traded Value ($)	ROC ($)	ROC/Share ($/shares)
count	252	252	252	252	252	252
mean	1,555,958.65	15,313,345,373.03	346,953.30	3,573,555,257.78	635,769.54	0.011792
std	463,558.93	3,891,299,900.31	146,677.85	1,234,882,079.43	655,911.15	0.008071
min	579,206	6,664,671,053.15	89,564	1,035,855,029.71	145,205.65	0.008879
25%	1,278,813.25	12,915,031,172.08	262,209	2,804,569,367.64	417,485.73	0.009667
50%	1,429,062	14,431,597,662.01	309,158	3,274,390,601.60	514,856.64	0.010213
75%	1,715,351.25	16,829,521,684.38	387,772	3,993,470,514.97	666,268.27	0.011288
max	3,596,006	30,999,914,293.66	1,073,029	9,428,952,387.10	7,817,684.58	0.093108

TABLE XXV. Aggregated purse statistics for Dow 30 constituents in 2016. The data used to construct this table is aggregated by date, resulting in 252 data points that correspond with the 252 trading days in 2016.

	Trades	Traded Value ($)	Diff. Trades	Diff. Traded Value ($)	ROC ($)	ROC/Share ($/shares)
count	252	252	252	252	252	252
mean	8,307,202.67	26,465,704,009.25	1,912,917.85	6,766,955,234.31	3,764,854.48	0.022109
std	1,370,512.88	3,786,979,882.64	473,884.96	1,299,054,438.46	1,048,372.83	0.002874
min	3,183,224	11,363,776,182.38	487,500	2,268,729,995.29	1,436,093.46	0.017744
25%	7,528,810.25	24,222,297,224.76	1,648,499.25	6,053,458,251.52	3,182,173.91	0.020092
50%	8,175,352.50	26,166,834,634.22	1,921,121.50	6,779,433,456.68	3,564,482.05	0.021393
75%	9,061,096.50	28,685,877,060.20	2,161,350.50	7,599,965,429.85	4,206,538.80	0.023737
max	13,408,508	41,337,807,991.92	3,537,890	10,627,257,029.61	10,083,342.57	0.047415

TABLE XXVI. Purse statistics for Russell 3000 constituents that were not S&P 500 constituents in 2016. The data used to construct this table is aggregated by date, resulting in 252 data points that correspond with the 252 trading days in 2016.

		Kurtosis	Skew
AMEX	Dow	57.59	-3.16
	S&P 500	1265.74	-18.55
	Russell	565.18	-20.06
ARCA	Dow	7392.86	85.55
	S&P 500	80740.09	258.93
	Russell	27314.13	-118.14
BATS	Dow	7334.11	85.06
	S&P 500	37895.12	143.83
	Russell	19742.38	-107.57
BATSY	Dow	548.55	20.35
	S&P 500	543.06	3.32
	Russell	10735.67	-59.29
BX	Dow	592.30	21.83
	S&P 500	531.38	-5.23
	Russell	15141.20	-84.69
CHX	Dow	2409.74	48.36
	S&P 500	4372.12	63.08
	Russell	8728.17	84.94
EDGA	Dow	641.74	22.78
	S&P 500	322.96	-6.51
	Russell	15832.73	-25.59
EDGX	Dow	4136.75	57.38
	S&P 500	8940.48	45.80
	Russell	55652.52	-169.68
FINRA	Dow	1777.76	35.53
	S&P 500	10278.94	51.38
	Russell	7875.99	17.50
IEX	Dow	952.81	26.35
	S&P 500	5009.89	45.36
	Russell	5725.39	-45.48
INET	Dow	244.31	11.99
	S&P 500	5778.77	33.17
	Russell	22048.67	-119.37
NSX	Dow	71.78	4.32
	S&P 500	1440.79	27.10
	Russell	6001.35	62.79
NYSE	Dow	2223.37	43.78
	S&P 500	4250.17	30.79
	Russell	7836.95	-54.89
PSX	Dow	402.51	16.29
	S&P 500	620.01	-16.69
	Russell	12337.86	-74.91

TABLE XXVII. Kurtosis and skew for daily ROC by exchange. The distributions are remarkably heavy-tailed.

1	Realized Opportunity Cost	$38,458,070.79
2	SIP Opportunity Cost	$37,970,135.30
3	Direct Opportunity Cost	$487,935.49
4	Trades	86,725,286
5	Diff. Trades	19,612,214
6	Traded Value	$3,678,242,397,422.43
7	Diff. Traded Value	$804,917,872,051.93
8	Percent Diff. Trades	22.61
9	Percent Diff. Traded Value	21.88
10	Ratio of 9 / 8	0.9677

TABLE XXVIII. Summary statistics for realized opportunity cost observed in the ETFs under study. It is notable that, of all market subsets we study, only this small subset has a ratio of the fraction of differing traded value to fraction of differing trades with value below unity. On a per-trade basis, this means that there is on average less potential for realized opportunity cost.

	Trades	Traded Value ($)	Diff. Trades	Diff. Traded Value ($)	ROC ($)	ROC/Share ($/shares)
count	2,266	2,266	2,266	2,266	2,266	2,266
mean	38,272.41	1,623,231,419.87	8,654.99	355,215,300.99	16,971.79	0.016651
std	106,159.45	4,657,139,733.49	23,868.59	1,032,161,545.93	48,416.03	0.032486
min	0	0	0	0	0	0
25%	14	252,922.28	3	46,585.01	34.2000	0.008215
50%	676	14,513,478.58	175.5000	3,358,814.83	437.4200	0.009910
75%	12,101.50	282,371,508.76	4,123.25	93,738,485.93	6,009.93	0.013664
max	974,888	40,617,035,891.21	251,657	11,028,368,359.92	499,906.77	1.0200

TABLE XXIX. Aggregated purse statistics for the ETFs under study. The data used to construct this table is aggregated by date and instrument, resulting in 2,266 data points that correspond with the 2,268 combinations of 252 trading days in 2016 and 9 ETFs under study.

	Trades	Traded Value ($)	Diff. Trades	Diff. Traded Value ($)	ROC ($)	ROC/Share ($/shares)
count	252	252	252	252	252	252
mean	344,147.96	14,596,199,989.77	77,826.25	3,194,118,539.89	152,611.39	0.189762
std	157,107.76	6,043,079,696.41	45,179.00	1,675,731,349.39	85,509.19	0.118446
min	113,860	5,018,912,183.01	14,610	703,559,994.91	30,989.52	0.054358
25%	237,021.25	10,471,387,904.01	47,237.50	2,052,459,478.17	94,488.20	0.106098
50%	308,705	13,005,695,875.47	66,509	2,780,132,908	131,084.42	0.169572
75%	394,822.25	16,641,275,220.96	94,108	3,799,483,257.76	186,174.78	0.256871
max	1,177,148	44,900,644,748.00	339,480	12,945,336,256.63	616,859.86	1.0963

TABLE XXX. Aggregated purse statistics for the ETFs under study. The data used to construct this table is aggregated by date, resulting in 252 data points that correspond with the 252 trading days in 2016.

Appendix B: Figures

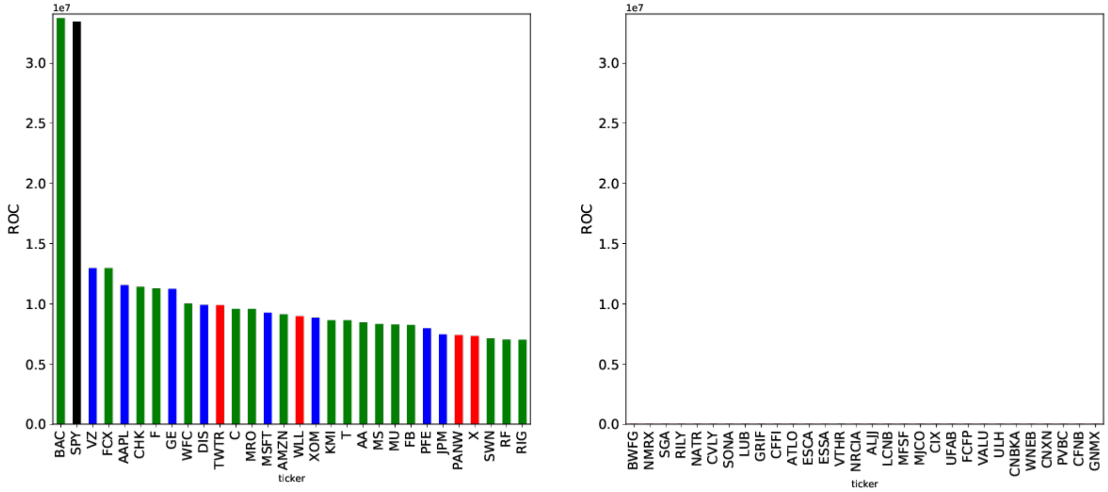

FIG. 6. ROC by ticker ($) for the top 30 (left panel) and bottom 30 (right panel) of all securities under study, ranked by ROC. Constituents of the Dow 30 are shown in blue, constituents of the S&P 500 (excluding the Dow 30) are shown in green, constituents of the Russell 3000 (excluding the S&P 500) are shown in red, and ETFs are shown in black.

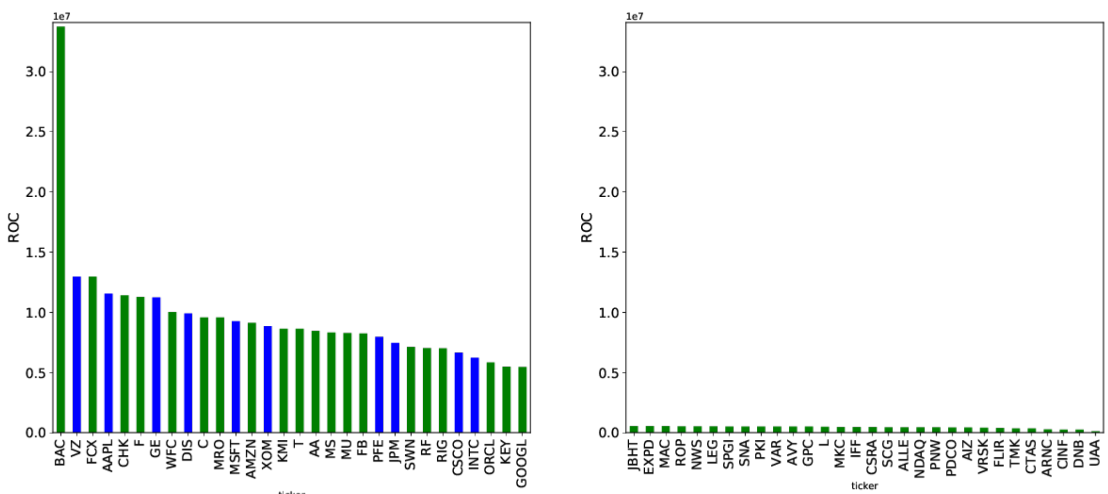

FIG. 7. ROC by ticker ($) for the top 30 (left panel) and bottom 30 (right panel) of S&P 500 securities, ranked by ROC. Constituents of the Dow 30 are shown in blue, while those belonging to the S&P 500 (excluding the Dow 30) are shown in green.

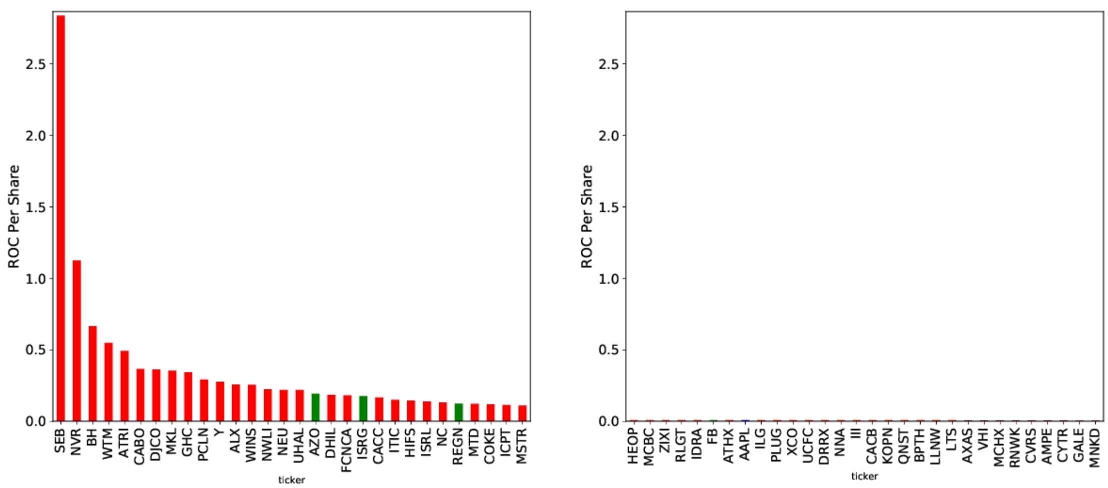

FIG. 8. ROC per share ($ / share) by ticker for the top 30 (left panel) and bottom 30 (right panel) of all securities under study, ranked by ROC. Constituents of the Dow 30 are shown in blue, constituents of the S&P 500 (excluding the Dow 30) are shown in green, and constituents of the Russell 3000 (excluding the S&P 500) are shown in red.

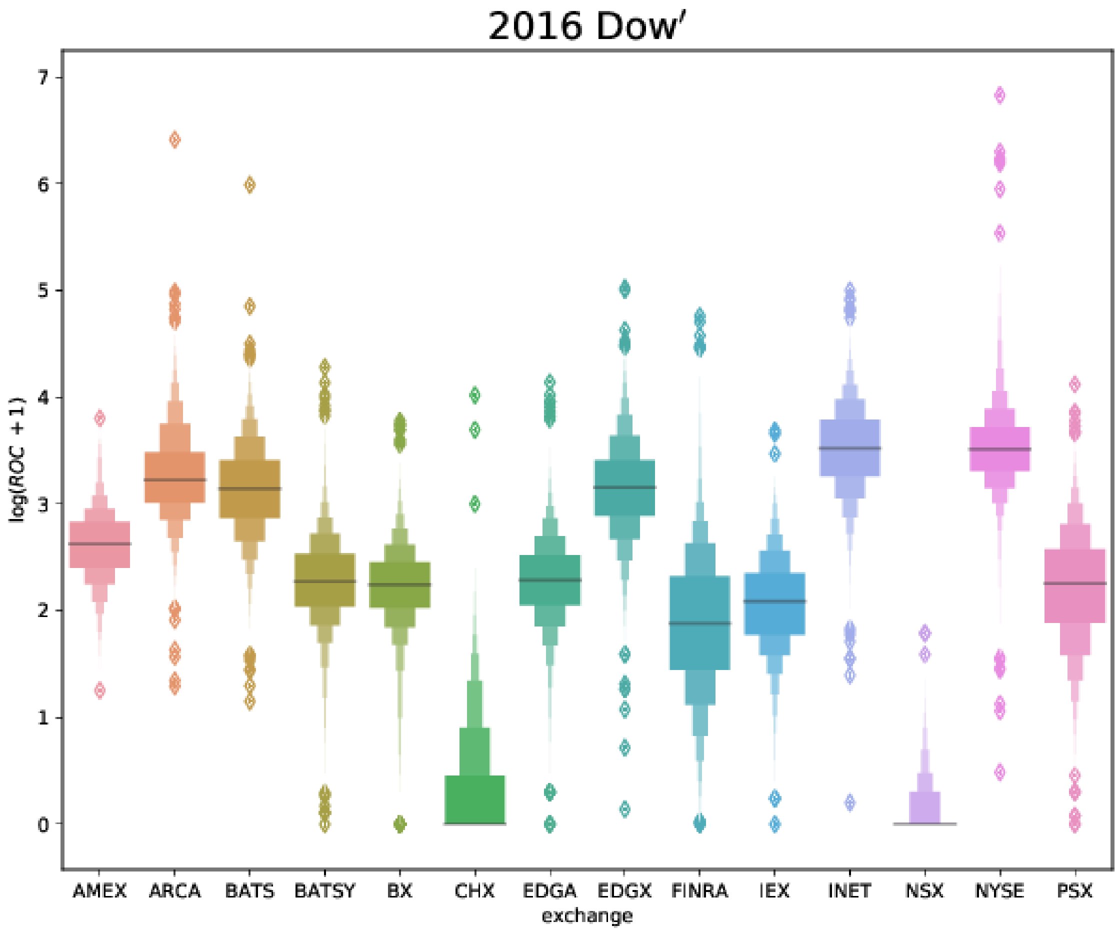

FIG. 9. Total ROC per day ($ / day) by exchange over Dow 30 equities. The vertical axis is transformed by $x \mapsto \log_{10}(x+1)$. The label INET refers to the NASDAQ exchange proper, while FINRA refers to the aggregation of all ATSs.

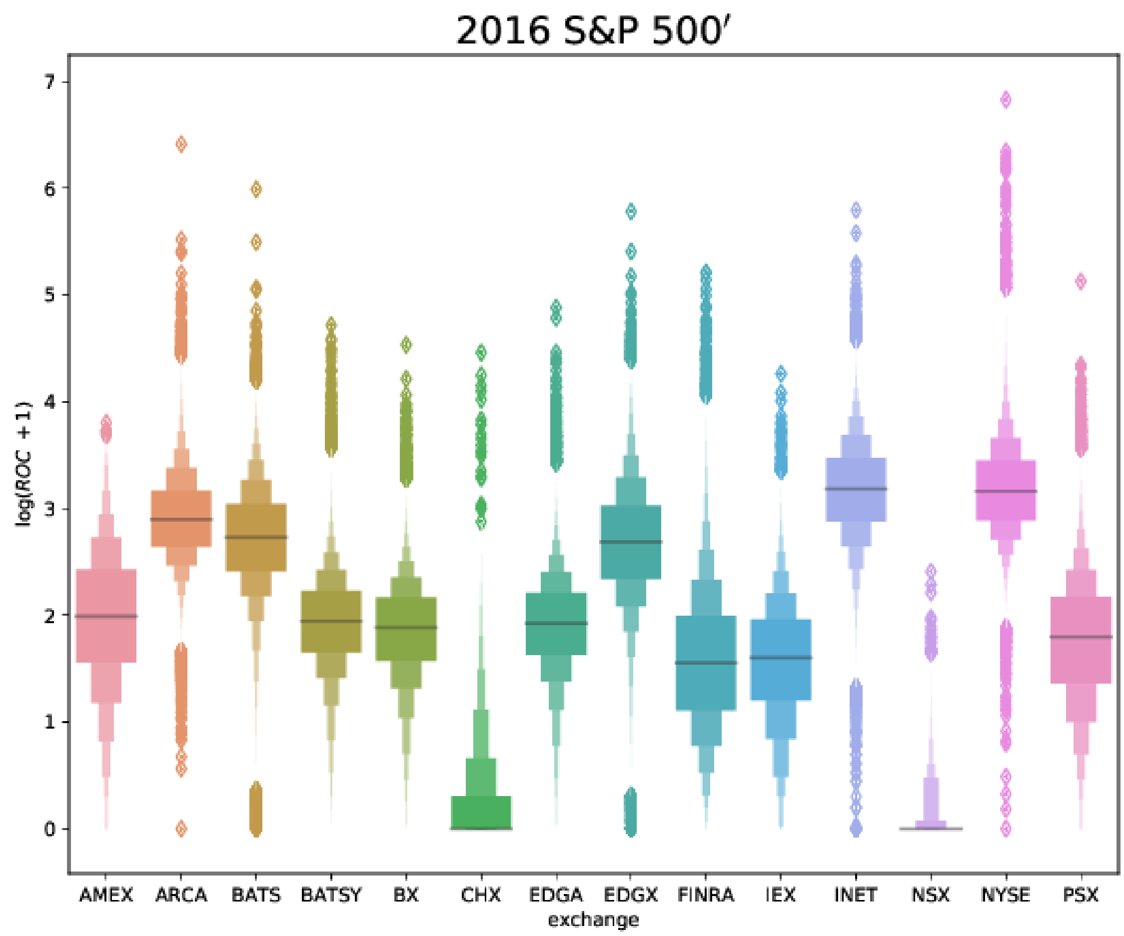

FIG. 10. Total ROC per day ($ / day) by exchange over S&P 500 equities. The vertical axis is transformed by $x \mapsto \log_{10}(x+1)$. The label INET refers to the NASDAQ exchange proper, while FINRA refers to the aggregation of all ATSs.

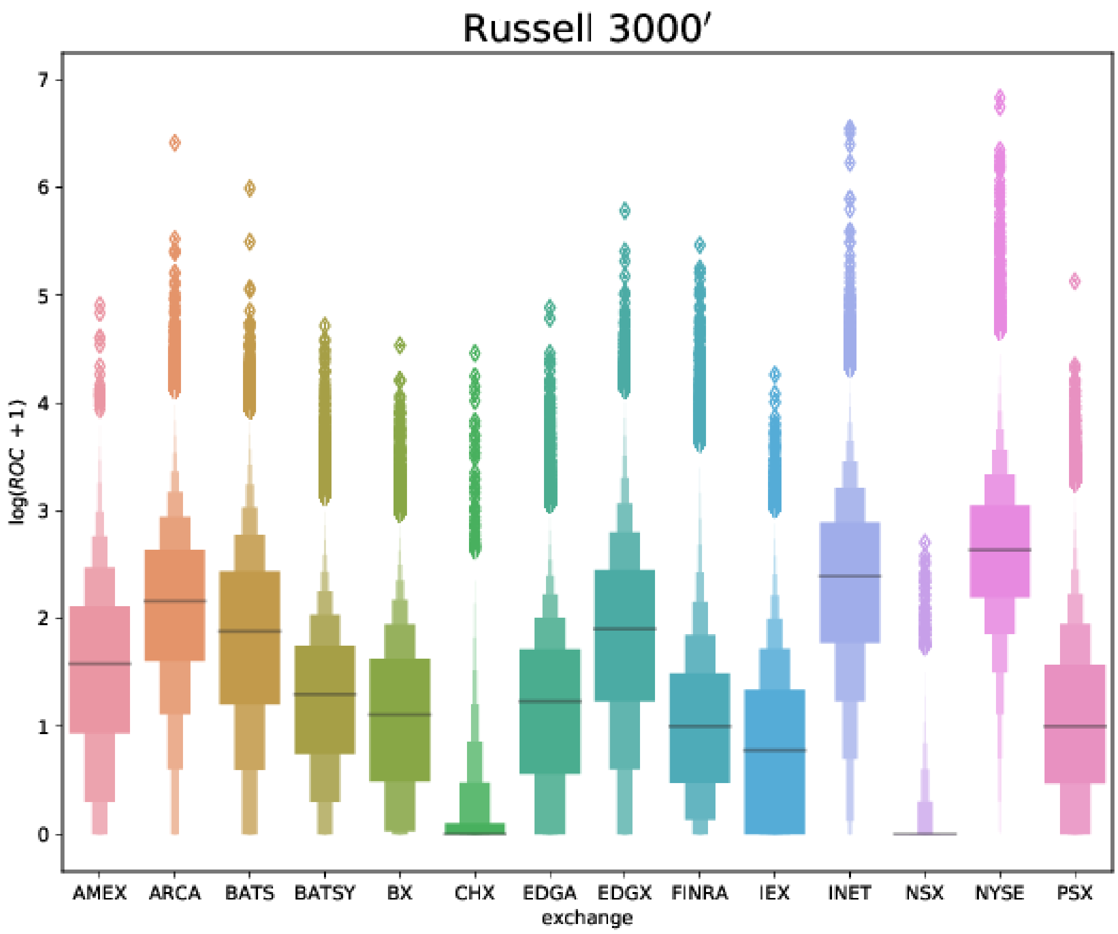

FIG. 11. Total ROC per day (\$ / day) by exchange over Russell 3000 equities. The vertical axis is transformed by $x \mapsto \log_{10}(x+1)$. The label INET refers to the NASDAQ exchange proper, while FINRA refers to the aggregation of all ATSs.

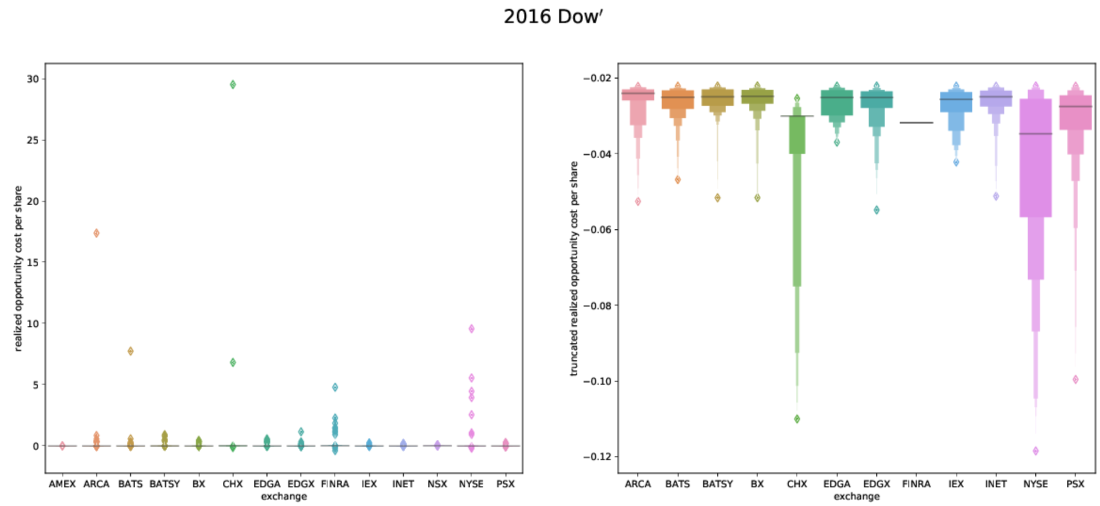

FIG. 12. ROC per share per day ($ / day) by exchange over Dow 30 equities. The label INET refers to the NASDAQ exchange proper, while FINRA refers to the aggregation of all ATSs.

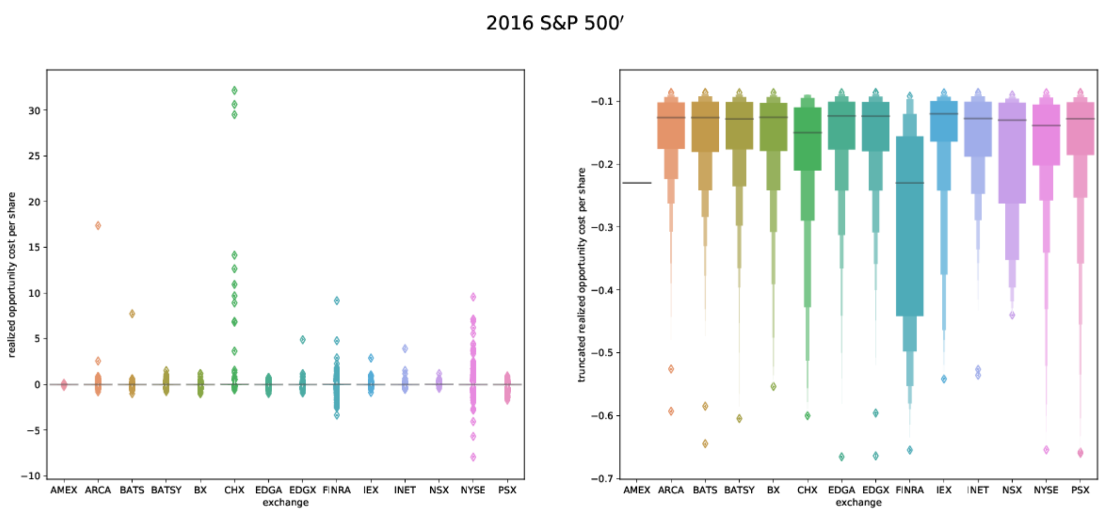

FIG. 13. ROC per share per day ($ / day) by exchange over S&P 500 equities. The label INET refers to the NASDAQ exchange proper, while FINRA refers to the aggregation of all ATSs.

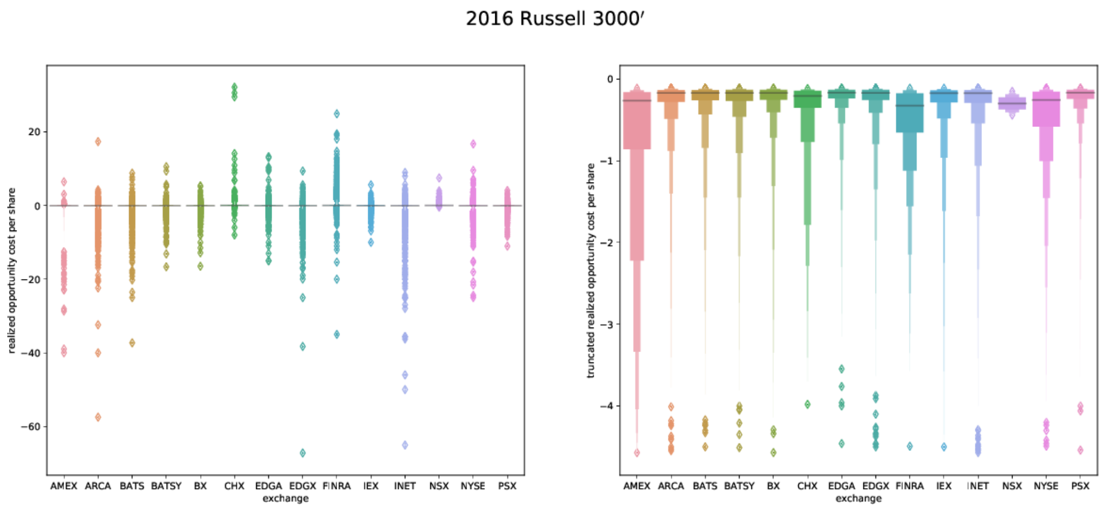

FIG. 14. ROC per share per day ($ / day) by exchange over Russell 3000 equities. The label INET refers to the NASDAQ exchange proper, while FINRA refers to the aggregation of all ATSs.

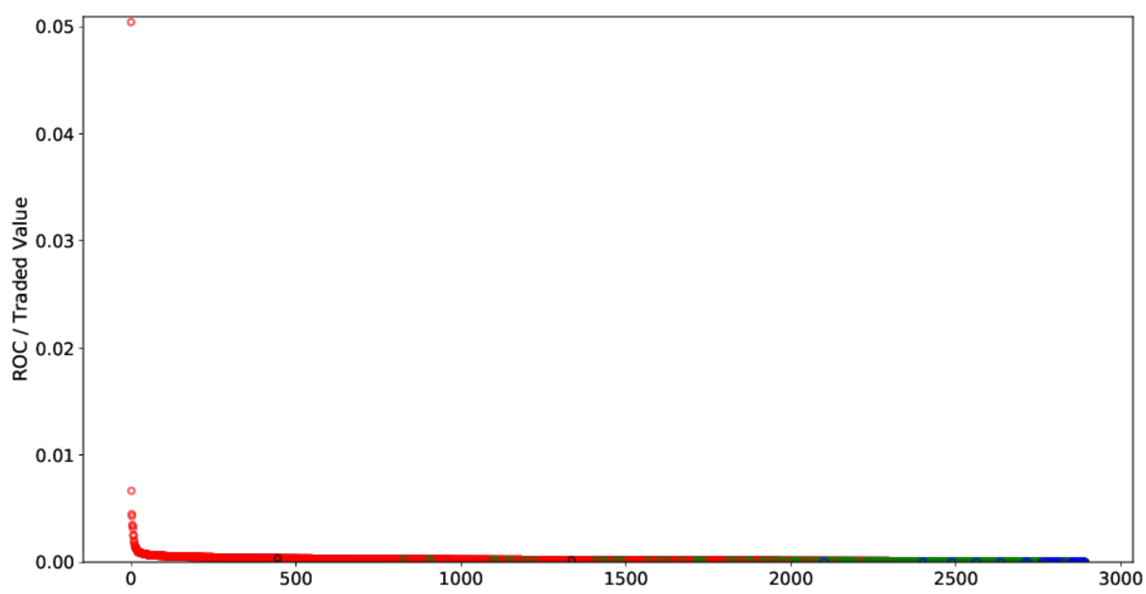

FIG. 15. Equities are plotted in rank-order of ROC per traded value; the 0-th equity has highest ROC per traded value. The first over-100 top equities are in the RexSP, which is unsurprising due to their combination of generally lower liquidity and lower share prices. Blue markers are associated with constituents of the Dow 30, green markers with constituents of the S&P 500 (excluding the Dow 30), red markers with constituents of the Russell 3000 (excluding the S&P 500), and black markers with ETFs.

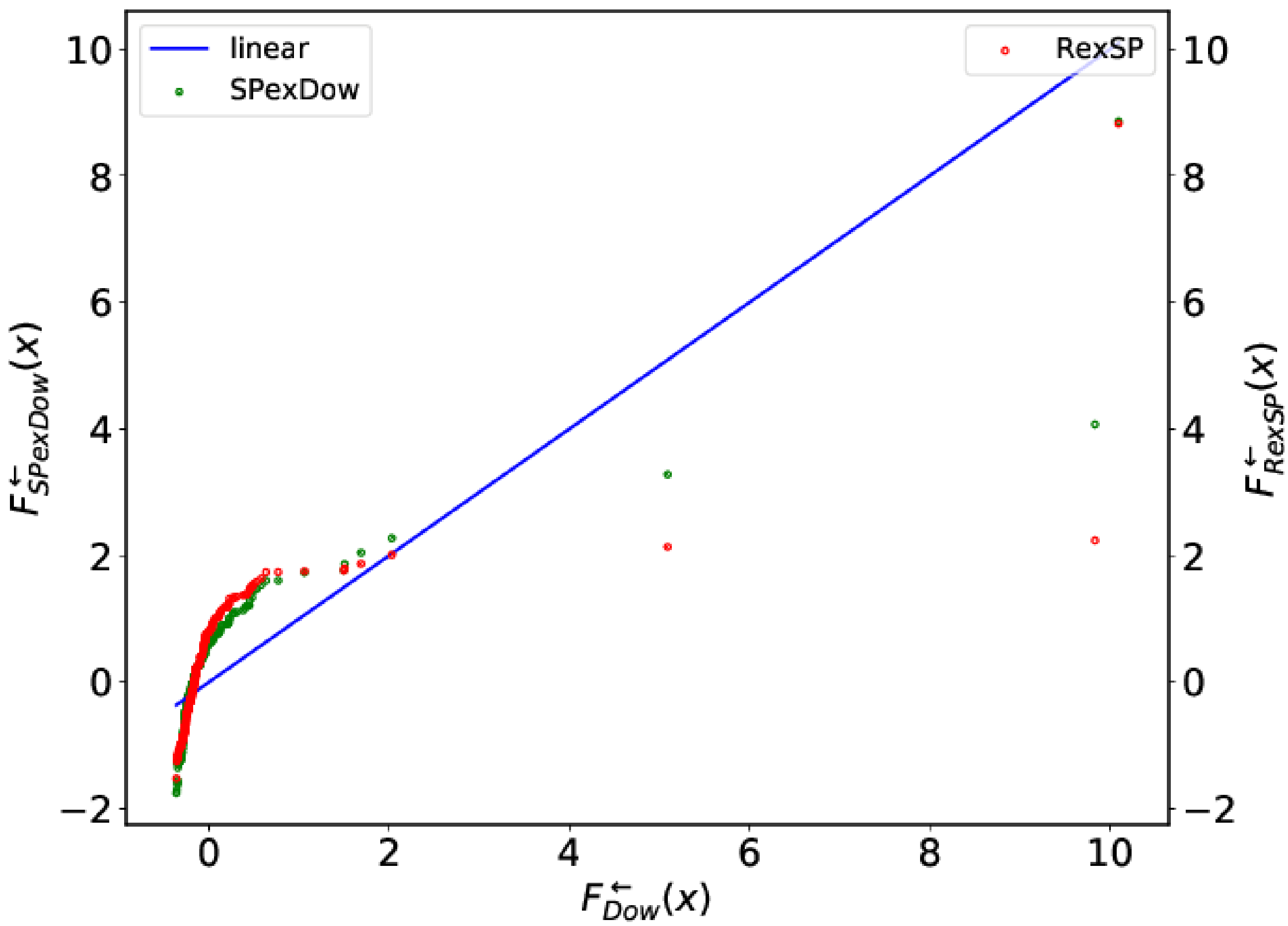

FIG. 16. Empirical quantile-quantile (QQ) plot for the normalized ROC per share processes. It is clear that the distribution of the SPexDow and RexSP processes are similar, and both are markedly different from the Dow process (blue line).

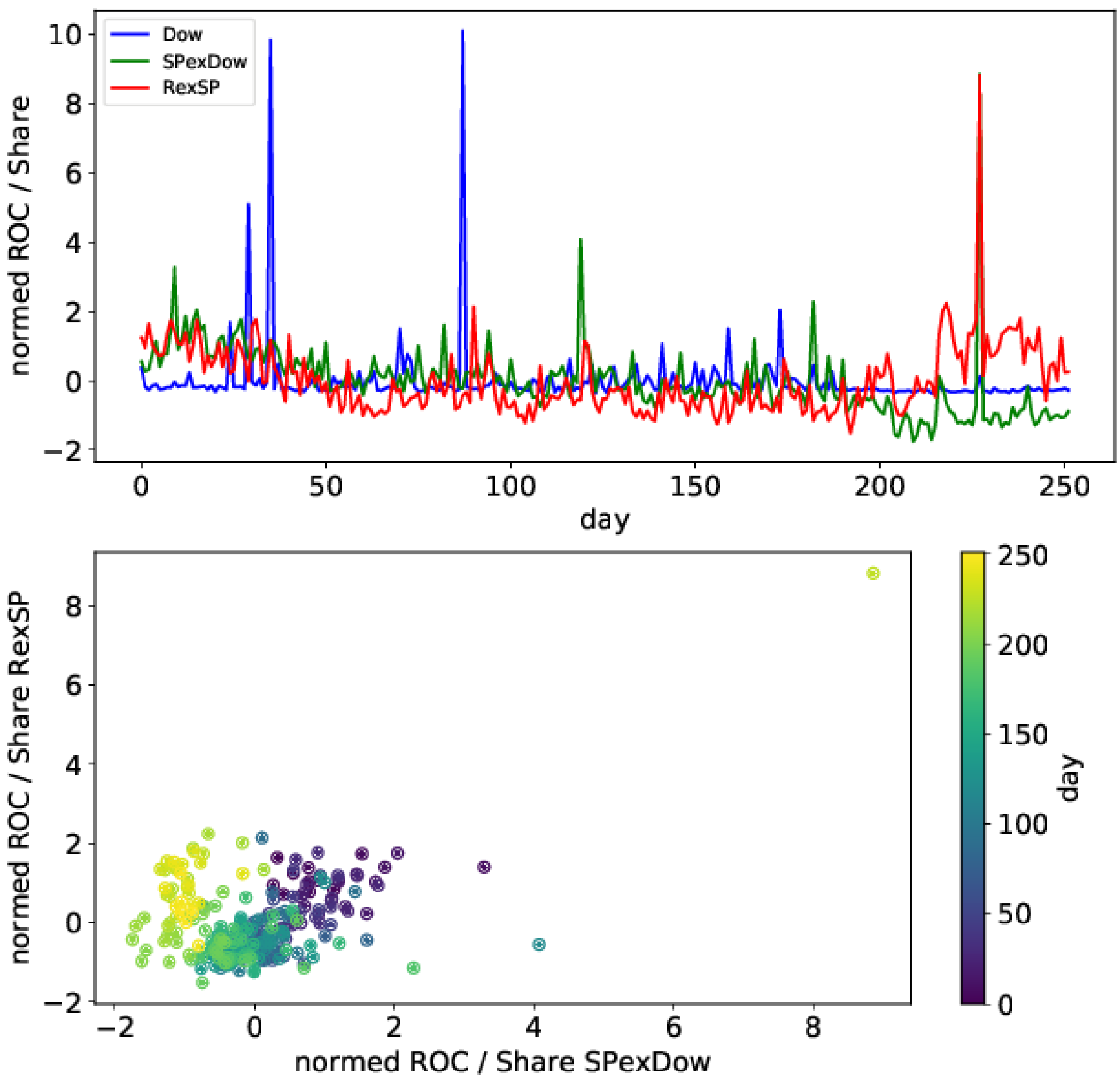

FIG. 17. Normalized ROC per share processes. There is one observation per day for a total of 252 observations in the process. These processes are anti-autocorrelated (Dow DFA exponent $\alpha = 0.434$, SPexDow DFA exponent $\alpha = 0.226$, RexSP DFA exponent $\alpha = 0.301$) and exhibit rare large values. The lower panel provides evidence for nonlinear cross-correlation between the SPexDow and RexSP ROC per share processes.

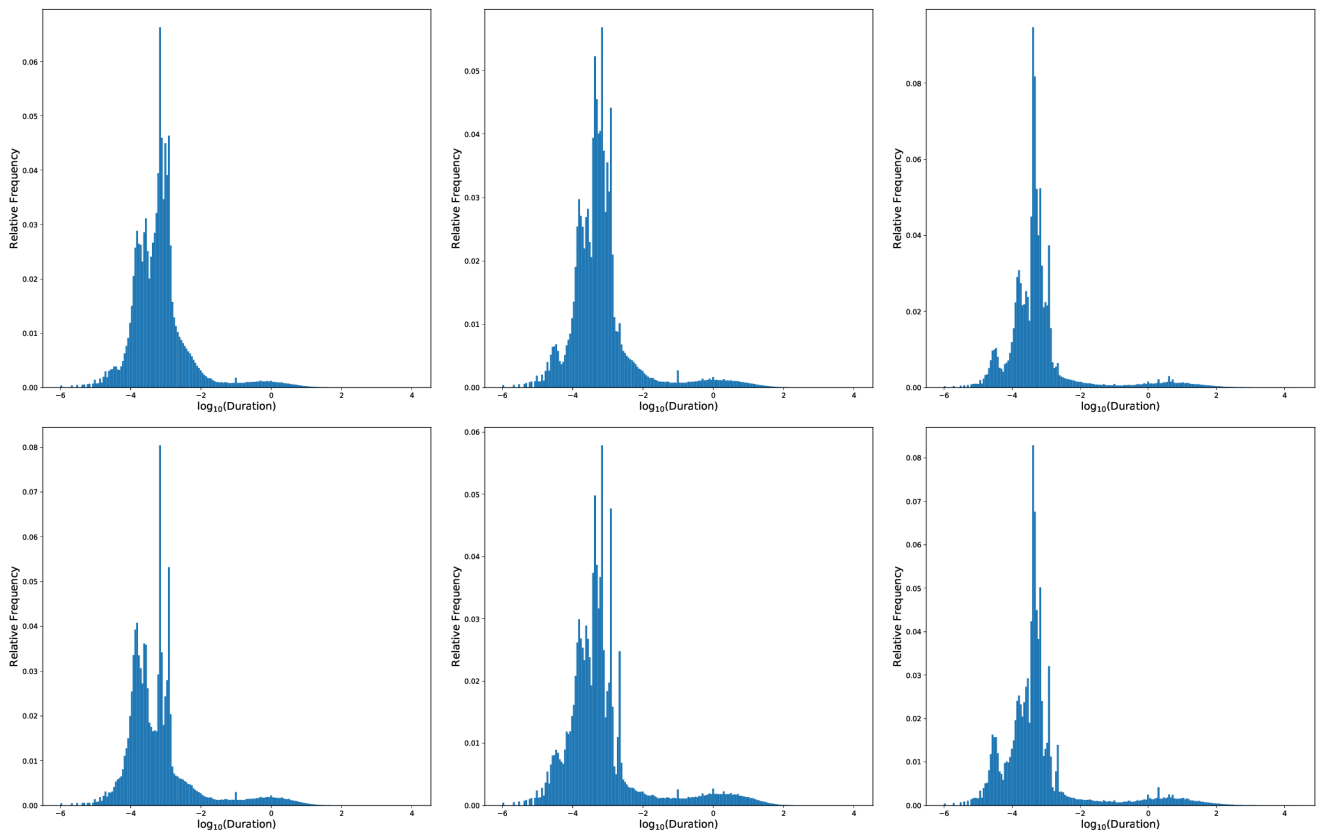

FIG. 18. Distributions of latency arbitrage duration. Columns are associated with an index (left to right: Dow 30, S&P 500 excluding the Dow 30, Russell 3000 excluding the S&P 500) and rows are associated with conditioning strategies (top to bottom: no conditioning, magnitude greater than $0.01).

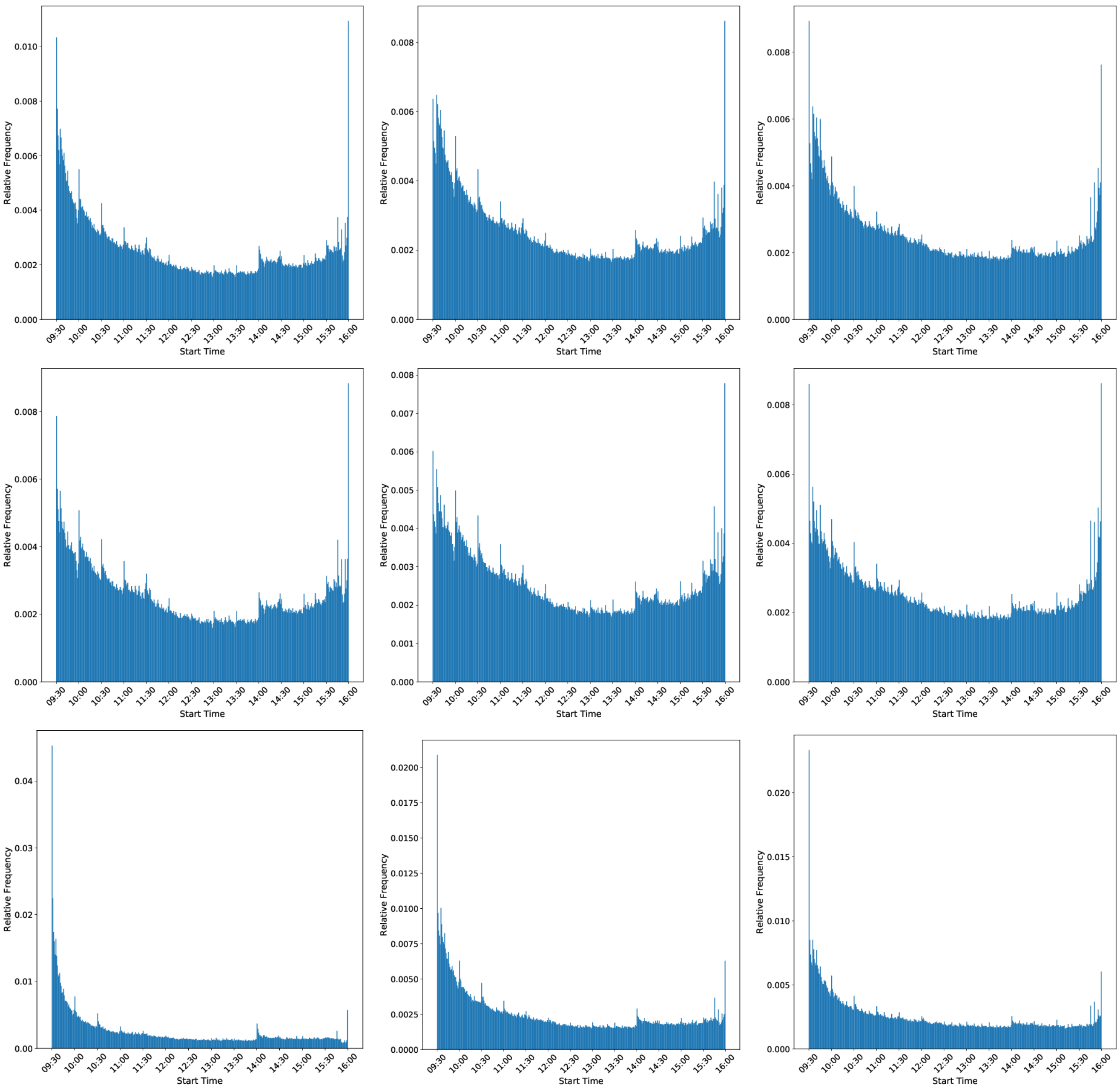

FIG. 19. Distributions of dislocation segment start time. Columns are associated with an index (left to right: Dow 30, S&P 500 excluding the Dow 30, Russell 3000 excluding the S&P 500) and rows are associated with conditioning strategies (top to bottom: no conditioning, duration greater than 545 μs, duration greater than 545 μs and magnitude greater than \$0.01).

Appendix C: Statistics

Ordered pair	Lags
Dow → RexSP	2,..., 4, 13,...,15,20, 22,...,37
Dow → SPexDow	
RexSP → Dow	1, 3, 35, 36
RexSP → SPexDow	
SPexDow → Dow	1, ..., 10, 15,...,24, 26, 30,...,34
SPexDow → RexSP	1, ..., 4

TABLE XXXI. Granger causality results for pairwise combinations of mutually-exclusive subsets of the equities under study. In order to qualify as a significant result, all four Granger causality tests (parameter F-test, sum of squared residuals F-test, likelihood-ratio test, χ^2-test) had to indicate causality at a $p = 0.05/N_{\text{lags}}$ level. The maximum number of lags was chosen to be $N_{\text{lags}} = 40$.

Dep. Variable:	$\log_{10}$ ROC	**R-squared:**	0.908
Model:	OLS	**Adj. R-squared:**	0.908
Method:	Least Squares	**F-statistic:**	7179.
Date:	Wed, 12 Sep 2018	**Prob (F-statistic):**	0.00
Time:	12:43:35	**Log-Likelihood:**	551.07
No. Observations:	2884	**AIC:**	-1094.
Df Residuals:	2880	**BIC:**	-1070.
Df Model:	3		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	1.0052	0.091	11.050	0.000	0.827	1.183
l_MarketCap	0.1183	0.011	10.675	0.000	0.097	0.140
l_total_trades	-0.2203	0.043	-5.127	0.000	-0.304	-0.136
l_differing_trades	0.9023	0.040	22.286	0.000	0.823	0.982

Omnibus:	1630.431	**Durbin-Watson:**	2.007
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	23812.396
Skew:	2.375	**Prob(JB):**	0.00
Kurtosis:	16.252	**Cond. No.**	259.

TABLE XXXII. Ordinary least squares regression predicting realized opportunity cost using market capitalization, differing trades, and total trades.

Dep. Variable:	$\log_{10}$ ROC	**R-squared:**	0.925
Model:	OLS	**Adj. R-squared:**	0.925
Method:	Least Squares	**F-statistic:**	5970.
Date:	Wed, 12 Sep 2018	**Prob (F-statistic):**	0.00
Time:	12:43:35	**Log-Likelihood:**	846.73
No. Observations:	2884	**AIC:**	-1679.
Df Residuals:	2877	**BIC:**	-1638.
Df Model:	6		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	7.8666	0.802	9.811	0.000	6.295	9.438
l_MarketCap	-0.0738	0.149	-0.496	0.620	-0.365	0.218
l_total_trades	-4.1661	0.432	-9.638	0.000	-5.013	-3.319
l_differing_trades	3.0804	0.338	9.103	0.000	2.417	3.744
np.power(l_MarketCap, 2)	0.0067	0.008	0.837	0.402	-0.009	0.022
np.power(l_total_trades, 2)	0.3385	0.038	8.936	0.000	0.264	0.413
np.power(l_differing_trades, 2)	-0.2042	0.034	-6.002	0.000	-0.271	-0.138

Omnibus:	1952.210	**Durbin-Watson:**	1.988
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	50808.169
Skew:	2.831	**Prob(JB):**	0.00
Kurtosis:	22.768	**Cond. No.**	1.70e+04

TABLE XXXIII. Ordinary least squares regression predicting realized opportunity cost using market capitalization, differing trades, and total trades. Quadratic terms are included.

Dep. Variable:	$\log_{10}$ ROC	**R-squared:**	0.600
Model:	OLS	**Adj. R-squared:**	0.600
Method:	Least Squares	**F-statistic:**	4280.
Date:	Wed, 12 Sep 2018	**Prob (F-statistic):**	0.00
Time:	12:43:30	**Log-Likelihood:**	-1574.9
No. Observations:	2884	**AIC:**	3154.
Df Residuals:	2882	**BIC:**	3166.
Df Model:	1		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-1.4415	0.108	-13.398	0.000	-1.652	-1.231
l_MarketCap	0.7368	0.011	65.422	0.000	0.715	0.759

Omnibus:	52.492	**Durbin-Watson:**	1.933
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	76.592
Skew:	0.199	**Prob(JB):**	2.34e-17
Kurtosis:	3.692	**Cond. No.**	126.

TABLE XXXIV. Ordinary least squares regression predicting realized opportunity cost using only market capitalization.

Dep. Variable:	$\log_{10}$ ROC	**R-squared:**	0.603
Model:	OLS	**Adj. R-squared:**	0.603
Method:	Least Squares	**F-statistic:**	2904.
Date:	Wed, 12 Sep 2018	**Prob (F-statistic):**	0.00
Time:	12:43:30	**Log-Likelihood:**	-1564.7
No. Observations:	2884	**AIC:**	3135.
Df Residuals:	2881	**BIC:**	3153.
Df Model:	2		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-6.2441	1.286	-4.857	0.000	-8.764	-3.724
l_MarketCap	1.7575	0.266	6.598	0.000	1.235	2.280
np.power(l_MarketCap, 2)	-0.0539	0.014	-3.927	0.000	-0.081	-0.027

Omnibus:	67.584	**Durbin-Watson:**	1.935
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	100.782
Skew:	0.242	**Prob(JB):**	1.30e-22
Kurtosis:	3.777	**Cond. No.**	1.24e+04

TABLE XXXV. Ordinary least squares regression predicting realized opportunity cost using only market capitalization. Quadratic terms are included.