

University of Vermont

ScholarWorks @ UVM

College of Engineering and Mathematical
Sciences Faculty Publications

College of Engineering and Mathematical
Sciences

1-1-2020

Fragmentation and inefficiencies in US equity markets: Evidence from the Dow 30

Brian F. Tivnan
MITRE Corporation

David Rushing Dewhurst
University of Vermont

Colin M. van Oort
MITRE Corporation

John H. Ring
MITRE Corporation

Tyler J. Gray
University of Vermont

See next page for additional authors

Follow this and additional works at: <https://scholarworks.uvm.edu/cemsfac>



Part of the [Community Health Commons](#), [Human Ecology Commons](#), [Nature and Society Relations Commons](#), [Place and Environment Commons](#), and the [Sustainability Commons](#)

Recommended Citation

Tivnan BF, Dewhurst DR, Van Oort CM, Ring IV JH, Gray TJ, Tivnan BF, Koehler MT, McMahon MT, Slater DM, Veneman JG, Danforth CM. Fragmentation and inefficiencies in US equity markets: Evidence from the Dow 30. *PloS one*. 2020 Jan 22;15(1):e0226968.

This Article is brought to you for free and open access by the College of Engineering and Mathematical Sciences at ScholarWorks @ UVM. It has been accepted for inclusion in College of Engineering and Mathematical Sciences Faculty Publications by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.

Authors

Brian F. Tivnan, David Rushing Dewhurst, Colin M. van Oort, John H. Ring, Tyler J. Gray, Brendan F. Tivnan, Matthew T.K. Koehler, Matthew T. McMahon, David M. Slater, Jason G. Veneman, and Christopher M. Danforth

RESEARCH ARTICLE

Fragmentation and inefficiencies in US equity markets: Evidence from the Dow 30

Brian F. Tivnan^{1,2,3,4*}, David Rushing Dewhurst^{2,4,5,6}, Colin M. Van Oort^{1,2,4,5,6}, John H. Ring IV^{1,2,4,5,6}, Tyler J. Gray^{2,3,4,5}, Brendan F. Tivnan^{4,7}, Matthew T. K. Koehler^{1,4}, Matthew T. McMahon^{1,4}, David M. Slater¹, Jason G. Veneman^{1,4}, Christopher M. Danforth^{2,3,4,5*}

1 The MITRE Corporation, McLean, VA, United States of America, **2** Vermont Complex Systems Center, University of Vermont, Burlington, VT, United States of America, **3** Department of Mathematics and Statistics, University of Vermont, Burlington, VT, United States of America, **4** Computational Finance Lab, Burlington, VT, United States of America, **5** Department of Computer Science, University of Vermont, Burlington, VT, United States of America, **6** Computational Story Lab, University of Vermont, Burlington, VT, United States of America, **7** School of Engineering, Tufts University, Medford, MA, United States of America

* btivnan@mitre.org (BFT); chris.danforth@uvm.edu (CMD)



OPEN ACCESS

Citation: Tivnan BF, Dewhurst DR, Van Oort CM, Ring JH, IV, Gray TJ, Tivnan BF, et al. (2020) Fragmentation and inefficiencies in US equity markets: Evidence from the Dow 30. *PLoS ONE* 15(1): e0226968. <https://doi.org/10.1371/journal.pone.0226968>

Editor: Stefan Cristian Gherghina, The Bucharest University of Economic Studies, ROMANIA

Received: April 17, 2019

Accepted: December 9, 2019

Published: January 22, 2020

Copyright: © 2020 The MITRE Corp., Tyler Gray, Brendan Tivnan, and Christopher Danforth. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The primary data for this study is a commercial product. As such, data for this study are publicly available to researchers, such that any researcher will be able to obtain the primary data in the same manner by which the authors obtained it. The terms of purchase for access to the commercial data preclude the authors from publicly posting or distributing the primary data. All data necessary to replicate the study are publicly available from <https://www.thesystech.com/>.

Abstract

Using the most comprehensive source of commercially available data on the US National Market System, we analyze all quotes and trades associated with Dow 30 stocks in calendar year 2016 from the vantage point of a single and fixed frame of reference. We find that inefficiencies created in part by the fragmentation of the equity marketplace are relatively common and persist for longer than what physical constraints may suggest. Information feeds reported different prices for the same equity more than 120 million times, with almost 64 million dislocation segments featuring meaningfully longer duration and higher magnitude. During this period, roughly 22% of all trades occurred while the SIP and aggregated direct feeds were dislocated. The current market configuration resulted in a realized opportunity cost totaling over \$160 million, a conservative estimate that does not take into account intra-day offsetting events.

1 Introduction

The Dow Jones Industrial Average, colloquially known as the Dow 30, is a group of 30 equity securities (stocks) selected by S&P Dow Jones Indices that is intended to reflect a broad cross-segment of the US economy (all industries except for utilities and transportation) [1]. The Dow 30 is one of the best known indices in the US and is broadly used as a barometer of the economy. Thus, while the group of securities that composes the Dow 30 is in some sense an arbitrary collection, it derives economic import from its ascribed characteristics. We study the behavior of these securities as traded in modern US equity markets, known as the National Market System (NMS). The NMS is comprised of 13 networked exchanges coupled by information feeds of differential quality and subordinated to national regulation. Adding another layer of complexity, the NMS supports a diverse ecosystem of market participants, ranging from small retail investors to institutional financial firms and designated market makers.

Funding: BFT, DRD, CMVO, JHR, JV were supported by Defense Advanced Research Project Agency award #W56KGU-17-C-0010 (<https://www.darpa.mil/>). BFT, MTTK, MTM, DS, and JV were supported by Homeland Security Advanced Research Projects Agency award #HSHQDC-14-D-00006 (<https://www.dhs.gov/science-and-technology/hsarpa>). CMD was supported by National Science Foundation grant #1447634 (<https://nsf.gov/>). CMD and TJG were supported by a gift from the Massachusetts Mutual Life Insurance Company (<https://www.massmutual.com/>). The funders played no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Competing interests: For completeness, we restate here that we declare no Competing Interests. Specifically, the commercial funder had no influence on the study. As for the employment affiliation of many of the co-authors, MITRE has no commercial interests in the study. MITRE is a not-for-profit research and development company solving problems for a safer world. More information on MITRE can be found here: <https://www.mitre.org/about/corporate-overview>. The authors place no restrictions (e.g., patents) on the artifacts of this study.

We do not attempt to unravel and attribute the activity of each of these actors here; several others have attempted to classify such activities with varying degrees of success in diverse markets [2–4]. We take a first-principles approach by compiling an exhaustive catalog of every dislocation, defined as a nonzero pairwise difference between the prices displayed by the National Best Bid and Offer (NBBO), as observed via the Securities Information Processor (SIP) feed, and Direct Best Bid and Offer (DBBO), as observed via the consolidation of all direct feeds.

The SIP and consolidation of all direct feeds are representative of the displayed quotes from the national exchanges (lit market). Additionally, we catalog every trade that occurred in the NMS among the Dow 30 in calendar year 2016, allowing an investigation of the relationship between trade execution and dislocations. We compile a dataset of all trades that may lead to a non-zero realized opportunity cost (ROC). We find that dislocations—times during which best bids and offers (BBO) reported on different information feeds observed at the same time from the point of view of a unified observer differ—and differing trades—trades that occur during dislocations—occur frequently. We measure more than 120 million dislocation segments, events derived from dislocations between the NBBO and DBBO, in the Dow 30 in 2016, summary statistics of which are displayed in Table 1. Approximately 65 million of those dislocation segments are what we term *actionable*, meaning that we estimate that there exists a nontrivial likelihood that an appropriately equipped market participant could realize arbitrage profits due to the existence of such a dislocation segment. (We discuss actionability in detail in Sec. 3.2 and the role that potential arbitrageurs play in the functioning of the NMS in Sec. 7.) Market participants incurred an estimated \$160 million USD in opportunity cost due to information asymmetry between the SIP and direct feed among the Dow 30 in 2016. We calculate the ROC using the NBBO price as the baseline. Deviations from this price contribute to the ROC with positive sign if the direct feed displays a worse price than the SIP, or with negative sign if the direct feed displays a better price than the SIP (from the perspective of a liquidity demanding market participant).

To characterize these phenomena, we use a publicly available dataset that features the most comprehensive view of the NMS (see Sec. 3.3 below) and is effectively identical to that used by the Securities and Exchange Commission’s (SEC) Market Information Data Analytics System (MIDAS). In addition to its comprehensive nature, this data was collected from the viewpoint of a unified observer: a single and fixed frame of reference co-located from within the Nasdaq data center in Carteret, N.J. We are unaware of any other source of public information (i.e.,

Table 1. The SIP feed consistently displayed worse prices than the aggregate direct feed for liquidity demanding market participants during periods of dislocation, with a \$84 million net difference in opportunity cost. Statistics 8–10 indicate that trades occurring during dislocations involve approximately 5% more value per trade on average than those that occur while feeds are synchronized. The values reported above are sums of daily observations, except for statistics 8–10, and are conservative estimates of the true, unobserved quantities since positive (favoring the SIP) and negative (favoring the direct feeds) ROC can cancel in summary calculations.

1	Total 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	Differing Trades	87,432,231
6	Traded Value	\$3,858,963,034,003.48
7	Differing Traded Value	\$900,535,924,961.72
8	Fraction of differing trades	0.2230
9	Fraction of differing notional	0.2334
10	Ratio of (9) over (8)	1.0465

<https://doi.org/10.1371/journal.pone.0226968.t001>

dataset available for purchase) or private information (e.g., available only to government agencies) that is collected using the viewpoint of a single, unified observer.

We demonstrate that the topological configuration of the NMS entails endogenous inefficiency. The fractured nature of the auction mechanism, continuous double auction operating on 13 heterogeneous exchanges and at least 35 Alternative Trading Systems (ATs) [5], is a consistent generator of dislocations and opportunity cost realized by market participants.

2 Literature review

2.1 Theory of market efficiency

The efficient markets hypothesis (EMH) as proposed by Fama [6] has left an indelible mark upon the theory of financial markets. Analysis of transaction data from the late 1960s and early 1970s strongly suggested that individual equity prices, and thus equity markets, fully incorporated all relevant publicly available information—the typical definition of market efficiency. A stronger version of the EMH proposes the incorporation of private information as well, via insider trading and other mechanisms. Previous studies have identified exceptions to this hypothesis [7], such as price characteristics of equities in emerging markets [8], the existence of momentum in the trajectories of equity prices [9], and speculative asset bubbles. Recent work by Fama and French has demonstrated that the EMH remains largely valid [9] when price time series are examined at timescales of at least 20 minutes and over a sufficiently long period of time. However, the NMS operates at speeds far beyond that of human cognition [10] and consists of fragmented exchanges [11] that may display different prices to the market. More permissive theories on market efficiency, such as the Adaptive Markets Hypothesis [12], allow for the existence of phenomena such as dislocations due to reaction delays, faulty heuristics, and information asymmetry [13]. In line with this, the Grossman-Stiglitz paradox [14] claims that markets cannot be perfectly efficient in reality, since market participants would have no incentive to obtain additional information. If market participants do not have an incentive to obtain additional information, then there is no mechanism by which market efficiency can improve. The proposition that markets are not perfectly efficient is supported by recent research. O'Hara [11], Bloomfeld [15], Budish [16], and others provide evidence that well-informed traders are able to consistently beat market returns as a result of both structural advantages and the actions of less-informed traders, so called “noise traders” [17]. This compendium of results points to a synthesis of the competing viewpoints of market efficiency. Specifically, that financial markets do seem to eventually incorporate all publicly available information, but deviations can occur at fine timescales due to market fragmentation and information asymmetries.

2.2 Empirical studies of market dislocations

Since the speed of information propagation is bounded above by the speed of light in a vacuum, it is not possible for information to propagate instantaneously across a fragmented market with spatially separated matching engines, such as the NMS. These physically-imposed information propagation delays lead us to expect some decoupling of BBOs across both matching engines and information feeds. Such divergences were found between quotes on NYSE and regional exchanges as long ago as the early 1990s [18], in NYSE securities writ large [19], in Dow 30 securities in particular [20], between NASDAQ broker-dealers and ATs as recently as 2008 [21, 22], and in NASDAQ listed securities as recently as 2012 [23]. U.S. equities markets have changed substantially in the intervening years, hence the motivation for our research. It is *a priori* unclear to what extent dislocations should persist within the NMS beyond the round-trip time of communication via fiber-optic cable. A first-pass analysis of latencies

between matching engines could conclude that, since information traveling at the theoretical speed of light between Mahwah and Secaucus would take approximately $372 \mu\text{s}$ to make a round trip between those locations, then dislocations of this length might be relatively common. However, a light-speed round trip between Secaucus and Mahwah takes approximately $230 \mu\text{s}$ and between Secaucus and Carteret takes approximately $174 \mu\text{s}$. Enterprising agents at Secaucus could rectify the differences in quotes between Mahwah and Carteret without direct interaction between agents in Carteret and agents in Mahwah.

Several other authors have considered the questions of calculating and quantifying the occurrence of dislocations or dislocation-like measures. In the aggregate, these studies conclude that price dislocations do not have a substantial effect on retail investors, as these investors tend to trade infrequently and in relatively small quantities, while conclusions differ on the effect of dislocations on investors who trade more frequently and/or in larger quantities, such as institutional investors and trading firms. Ding, Hanna, and Hendershot (DHH) [23] investigate dislocations between the SIP NBBO and a synthetic BBO created using direct feed data. Their study focuses on a smaller sample, 24 securities over 16 trading days, using data collected by an observer at Secaucus, rather than Carteret, and does not incorporate activity from the NYSE exchanges. They found that dislocations occur multiple times per second and tend to last between one and two milliseconds. In addition, DHH find that dislocations are associated with higher prices, volatility, and trading volume. Bartlett and McCrary [24] also attempted to quantify the frequency and magnitude of dislocations. However, Bartlett and McCrary did not use direct feed data, so the existence of dislocations was estimated using only Securities Information Processor (SIP) data, making it difficult to directly align their results to those presented here. A study by the TABB Group of trade execution quality on midpoint orders in ATs also noted the existence of latency between the SIP and direct data feeds, as well as the existence of intra-direct feed latency, due to differences in exchange and ATS software and other technical capabilities [25]. Wah [26] calculated the potential arbitrage opportunities generated by latency arbitrage on the S&P 500 in 2016 using data from the SEC's MIDAS platform [27]. Wah's study is of particular interest as it is the only other study of which we are aware that has used comprehensive data. Though similar in this respect, the quantities estimated in Wah's study differ substantially from those considered here. Wah located time intervals during which the highest buy price on one exchange was higher than the lowest sell price on another exchange, termed a "latency arbitrage opportunity" in that work, and examined the potential profit to be made by an infinitely-fast arbitrageur taking advantage of these price discrepancies. This idealized arbitrageur could have captured an estimated \$3:03B USD in latency arbitrage among S&P 500 tickers during 2014, which is on the same order of magnitude (on a per-ticker basis) as our approximately \$160M USD in realized opportunity cost among Dow 30 tickers during calendar year 2016.

Other authors have analyzed the effect of high-frequency trading (HFT) on market microstructure, which is at least tangentially related to our current work due to its reliance on low-latency, granular timescale data and phenomena. O'Hara [11] provides a high-level overview of the modern-day equity market and in doing so outlines the possibility of dislocation segments arising from differential information speed. Angel [28, 29] claims that price dislocations are relatively rare occurrences, while Carrion [30] provides evidence of high-frequency trading strategies' effectiveness in modern-day equity markets via successful, intra-day market timing. Budish [16] notes that high-frequency trading firms successfully perform statistical arbitrage (e.g., pairs trading) in the equities market, and ties this phenomenon to the continuous double auction mechanism that is omnipresent in the current market structure. Menkveld [31] analyzed the role of HFT in market making, finding that HFT market making activity correlates negatively with long-run price movements and providing some evidence that HFT market

making activity is associated with increasingly energetic price fluctuations. Kirilenko [2] provided an important classification of active trading strategies on the Chicago Mercantile Exchange E-mini futures market, which can be useful in creating statistical or agent-based models of market phenomena. Mackintosh noted the effects of both fragmented markets and differential information on financial agents with varying motives, such as high-frequency traders and long-term investors, in a series of Knight Capital Group white papers [32]. These papers provide at least three additional insights relevant to our study. The first is a comparison of SIP and direct-feed information, noting that “all data is stale” since, regardless of the source (i.e., SIP or direct feed), rates of data transmission are capped at the speed of light in a vacuum as discussed above. The second is that the SIP and the direct feeds are almost always synchronized. That is, for U.S. large cap stocks like the Dow 30 in 2016, synchronization between the SIP and direct feeds existed for 99.99% of the typical trading day. Stated another way, Mackintosh observed dislocations between quotes reported on the SIP and direct feeds for 0.01% of the trading day, or a sum total of 23 seconds distributed throughout the trading day. The third insight from the Mackintosh papers relevant to our study reflects the significance of dislocations. Mackintosh observed that 30% of daily value typically traded during these dislocations.

For a more comprehensive review of the literature on high frequency trading and modern market microstructure more generally, we refer the reader to Goldstein et al. [33] or Chordia et al. [34]. Arnuk and Saluzzi [35] provide a monograph-level overview of the subject from the viewpoint of industry practitioners.

3 Description of exchange network and data feeds

Here we provide a brief overview of the National Market System (NMS), including a description of infrastructure components and some varieties of market participants. In particular, we note the information asymmetry between participants informed by the Securities Information Processor and those informed by proprietary, direct information feeds.

3.1 Market participants

There are, broadly speaking, three classes of agents involved in the NMS: traders, of which there exist essentially four sub-classes (retail investors, institutional investors, brokers, and market-makers) that are not mutually exclusive; exchanges and ATs, to which orders are routed and on which trades are executed; and regulators, which oversee trades and attempt to ensure that the behavior of other market participants abides by market regulation. See [S3 Appendix](#) for an overview of select regulations. We note that Kirilenko *et al.* claim the existence of six classes of traders based on technical attributes of their trading activity [2]. This classification was derived from activity in the S&P 500 (E-mini) futures market, not the equities market, but is an established classification of trading activity. It is not possible to perform a similar study in the NMS since agent attribution is not publicly available. However, the Consolidated Audit Trail (CAT) is an SEC initiative (SEC Rule 613) that may provide such attribution in the future [36]. At the time of writing this framework was not yet constructed. Though the scope of this work does not encompass an analysis of various classes of financial agents, we describe some important agent archetypes in [S1 Appendix](#).

3.2 Physical considerations

Contrary to its moniker, “Wall Street” is actually centered around northern New Jersey. The matching engines for the three NYSE exchanges are located in Mahwah, NJ, while the matching engines for the three NASDAQ exchanges are located in Carteret, NJ. The other major exchange families base their matching engines at the Equinix data center, located in Secaucus,

Table 2. The speed of light is approximated by 186,000 mi/s (or 300,000 km/s) and fiber propagation delays are assumed to be 4.9 μ s/km. These propagation delays form the basis for estimates of the duration required for a dislocation segment to be considered actionable, though these figures do not account for any computing delays and thus are lower bounds for the definition of actionable.

NMS Propagation Delay Estimates				
	Carteret-Mahwah	Mahwah-Secaucus	Carteret-Secaucus	Secaucus-Weehawken
Straight-line Distance	34.55 mi	21.31 mi	16.22 mi	2.56 mi
	55.6 km	34.3 km	26.1 km	4.12 km
Light speed, one-way	185.75 μ s	114.57 μ s	87.2 μ s	13.76 μ s
Light speed, two-way	371.5 μ s	229.14 μ s	174.4 μ s	27.52 μ s
Fiber, one-way	272.44 μ s	168.07 μ s	127.89 μ s	20.19 μ s
Fiber, two-way	544.88 μ s	336.14 μ s	255.78 μ s	40.38 μ s
Hybrid laser, one-way	-	-	94.5 μ s	-
Hybrid laser, two-way	-	-	189 μ s	-

<https://doi.org/10.1371/journal.pone.0226968.t002>

NJ, except for IEX, which is based close to Secaucus in Weehawken, NJ. The location of individual ATSs is generally not public information. However, since there is a great incentive for ATSs to be located close to data centers (see sections 2 and 6), it is likely that many ATSs are located in or near the data centers that house the NMS exchanges. For example, Goldman Sachs's Sigma X² ATS has its matching engine located at the Equinix data center in Secaucus, NJ [37].

Since matching engines perform the work of matching buyers with sellers in the NMS, we hereafter refer to the locations of the exchanges by the geographic location of their matching engine. For example, IEX has its point of presence in Secaucus, but its matching engine is based in Weehawken; we locate IEX at Weehawken.

This geographic decentralization has a profound effect on the operation of the NMS. We calculate minimum propagation delays between exchanges and are displayed in Table 2. In constructing Table 2 we use estimates of propagation delays in fiber optic cables provided by M2 Optics [38] as well as data center locations, distances between data centers, and one-way hybrid laser propagation delays from Anova Technologies [39].

In reality, the time for a message to travel between exchanges will be strictly greater than these lower bounds, since light is slowed by transit through a fiber optic cable, and further slowed by any curvature in the cable itself. The two-way estimates in Table 2 give a lower bound on the minimum duration required for a dislocation segment to be "actionable" and a more realistic estimate derived by assuming propagation through a fiber optic cable with a refractive index of 1.47 [38]. These estimates do not account for computing delays, which may occur at either end of the communication lines, in order to avoid speculation. In practice such computing delays will also have a material effect on which dislocation segments are truly actionable and will depend heavily on the performance of available computing hardware.

Connecting the exchanges are two basic types of data feeds: SIP feeds, containing quotes, trades, limit-up / limit-down (LULD) messages, and other administrative messages compiled by the SIP; and direct data feeds, which contain quotes, trades, order-flow messages (add, modify, etc), and other administrative messages. The direct data feeds operate on privately-funded and installed fiber optic cables that may have differential information transmission ability from the fiber optic cables on which the SIP data feeds are transmitted. Latency in propagation of information on the SIP is also introduced by SIP-specific topology (SIP information must travel from a matching engine to a SIP processing node before being propagated from that node to other matching engines) and computation occurring at the SIP processing node. Due to the observed differential latency between the direct data feeds and the SIP data feed

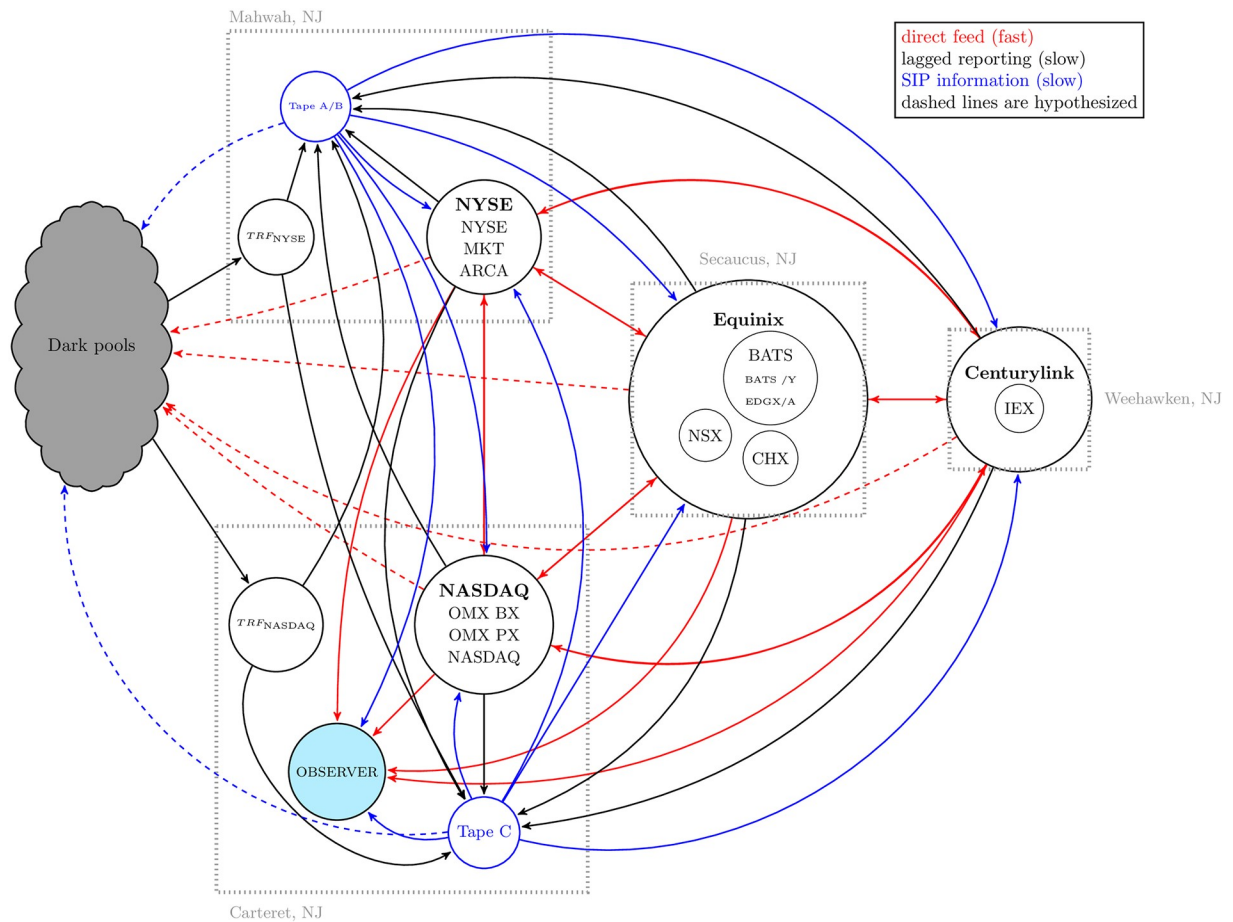


Fig 1. The NMS (lit market and ATs) as implied by the comprehensive market data. As we do not have the specifications of inter-market center communication mechanisms and have minimal knowledge of intra-market center communication mechanisms, we simply classify information as having high latency, as the SIP and lagged information heading to the SIP do, or low latency, as the information on the direct feeds does. Note the existence of the observer, located in Carteret NJ. Without a single, fixed observer it is difficult to clock synchronization issues and introduces an unknown amount of noise into measurements of dislocations and similar phenomena. Clock synchronization issues are avoided when using data collected from a single point of presence since all messages may be timestamped by a single clock, controlled by the observer.

<https://doi.org/10.1371/journal.pone.0226968.g001>

and the heterogeneous distance between exchanges, dislocation segments are created solely by the macro-level organization of the market system. We note that in the intervening years since data was collected for analysis, the SIP has been upgraded substantially to lower latency arising from computation at SIP processing nodes.

Our understanding of the physical layout of the NMS is depicted in Fig 1 at a relatively high level.

There are three basic types of information flow within the NMS:

1. Direct feed information, which flows to anyone who subscribes to it. Direct feed information is associated with non-trivial costs (on the order of \$130, 000 USD per month, see S2 Table for details) and so is used primarily by exchanges, large financial firms, and ATs. Direct feed information thus flows to and from the exchanges (and the major exchange participants). We hypothesize that direct feed information also flows to ATs, since they require some type of price signal in order for the market mechanism to function and may benefit from low latency data. This was the case for at least one major ATs, Goldman

Sachs's Sigma X², as of May 2019, so it is plausible that it is true for others [37]. The direct feeds provide the fastest means by which to acquire a price signal, and thus may provide the best economic value to traders dependent on frequent information updates; this provides the economic foundation for our hypothesis.

2. SIP information, which is considerably less expensive than direct feed information and exists by regulatory mandate. However, market participants may still subscribe to the SIP as a tool for use in arbitrage; see Section 2 for discussion of this possibility. Market participants that choose not to purchase the direct feed data might also choose to purchase the SIP data for use as a price signal and as a backup to the consolidated direct feeds. At least one ATS, Goldman Sachs's Sigma X², uses SIP data as a backup to direct feed data and combines both data sources to construct their local BBO [37].
3. Lagged reporting data that is not yet collated by the SIP. Regulation requires that exchanges report all local quote and trade activity, and that ATSS report all trade activity. This information is collected by the appropriate SIP tapes and then disseminated through the SIP data feeds. It is the responsibility of the exchanges to report their quote and trade information to the SIP, and of ATSS to report their trade information to FINRA Trade Reporting Facilities (TRF). Thus, though this information will be eventually visible to all subscribers to SIP or direct feed data, it differs qualitatively from that data due to its lagged nature. For example, suppose a trade occurs at NYSE MKT on a NASDAQ-listed security that updates the NBBO for that security. Since this trade occurs at Mahwah, it takes a non-negligible amount of time for the information to propagate to SIP Tape C, located in Carteret. However, traders located at Mahwah have access to this information more quickly, possibly allowing them an information advantage over their Carteret-based competitors.

3.3 Data

Our study utilizes all quotes and trades associate with Dow 30 stocks that occurred in calendar year 2016 (2016-01-01 through 2016-12-31), observed via the SIP and Direct feeds from a single point of presence in Carteret, NJ. This data is provided by Thesys Group Inc., formerly known as Tradeworx, who is the sole data provider for the SEC's MIDAS [27, 40]. MIDAS ingests more than one billion records daily—order flow, quote updates, and trade messages—from the direct feeds of all national exchanges. These records represent the exhaustive set of posted orders, quotes, order modifications, cancellations, trades, and administrative messages issued by national exchanges. Prior to awarding Thesys Group the MIDAS contract [41], the SEC conducted a sole source selection [42], thereby designating Thesys Group as the only current authoritative source for NMS data.

In addition to being the authoritative data source for the SEC's MIDAS program, another significant attribute of the Thesys data is that it is collected by a single observer from a consistent location in the NMS (the Nasdaq data center in Carteret, NJ) as depicted in Fig 1. The single observer not only allows the user to account for the relativistic effects described above but also to directly observe dislocation segments and realized opportunity cost instead of compiling estimates of these quantities as has been done in previous studies. At the NASDAQ data center, Thesys applies a new timestamp to each message received, including messages originating from the SIP feed or one of the direct feeds, that allows subscribers to observe information flow through the NMS in the same manner as a market participant located at the Carteret data center. In our analysis we use this "Thesys timestamp" to synchronize information from disparate data feeds and avoid issues that otherwise could arise from clock synchronization errors and relativistic effects. Since this timestamp is given at the time the data arrives at the server

from which the data is collected, any discrepancies in the clocks at different exchanges, ATSS, and the SIP do not affect our measurement procedures. This timestamping procedure is identical to that used in Ding, Hanna, and Hendershott [23]. Ideally, we would have data from four different unified observers—an observer located at each data center—so that we could compile the different states of the market that must exist depending on physical location of observation, but we do not believe that comprehensive consolidated data is available from the point of view of observers located anywhere but at Carteret, hence our selection of this location for observation.

4 Dislocations

We provide a brief definition of a dislocation segment as calculated and used in this work. Each dislocation segment can be represented by a 4-tuple:

$$v_n = (t_n^{\text{start}}, t_n^{\text{end}}, \min \Delta p, \max \Delta p). \quad (1)$$

The maximum (resp. minimum) value of the dislocation segment are simply the maximum (resp. minimum) difference in the prices that are generating the dislocation segment over the time period $[t_n^{\text{start}}, t_n^{\text{end}}]$. The time period $[t_n^{\text{start}}, t_n^{\text{end}}]$ is determined by identifying a contiguous period of time where $\Delta p > 0$ or $\Delta p < 0$. From the above quantities the duration of the dislocation segment can also be calculated. The quantity $\Delta p(t)$ is the difference in the price displayed by the information feeds at time t as measured and timestamped by our observer in Carteret. From the definitions of $\max \Delta p$ and $\min \Delta p$ the reader will note that dislocation segments will tend to feature $\min(|\min \Delta p|) \geq \0.01 , since the minimum tick size in the NMS is set at one penny for securities with a share price of at least \$1.00. In collating dislocation data, we record the maximum and minimum value of each dislocation segment rather than a time-weighted average of dislocation value or other statistic for the sake of simplicity. In much of our analysis we take the absolute values of the maximum and minimum values of each dislocation segment as the fundamental object of study as any dislocation, regardless of which feed is favored, presents an opportunity for market inefficiency.

See Fig 2 for a stylized depiction of two dislocation segments, along with annotations denoting their recorded attributes.

Based on the definition of dislocation segments given above, and fully specified in S2 Appendix, we may identify the necessary and sufficient conditions for a dislocation segment to occur. Specifically, the market state must include two or more distinct trading locations, two or more information feeds with differing latency, and a price discrepancy. These all follow directly from elements of the definition; such that a simple, null model configuration of a single exchange with a single data feed cannot support the existence of dislocation segments as specified here.

5 Realized opportunity cost

We used the following decision procedure to calculate realized opportunity cost: for each trade that occurred in the NMS we checked if a price discrepancy between the SIP and consolidated direct feeds was present at the time the trade executed, from the point of view of our observer in Carteret, and counted each as a *differing trade*. If the differing trade executed at a price displayed by the prevailing NBBO then a price difference was calculated, i.e. $p_{\text{SIP}} - p_{\text{direct}}$ if the liquidity-demanding order was a offer and $p_{\text{direct}} - p_{\text{SIP}}$ if the liquidity-demanding order was a bid, and a cost, termed the realized opportunity cost (ROC), was assigned to the trade using the number of shares multiplied by the price difference. Depth of book was not taken into account in this calculation. The sum total of all ROC occurrences

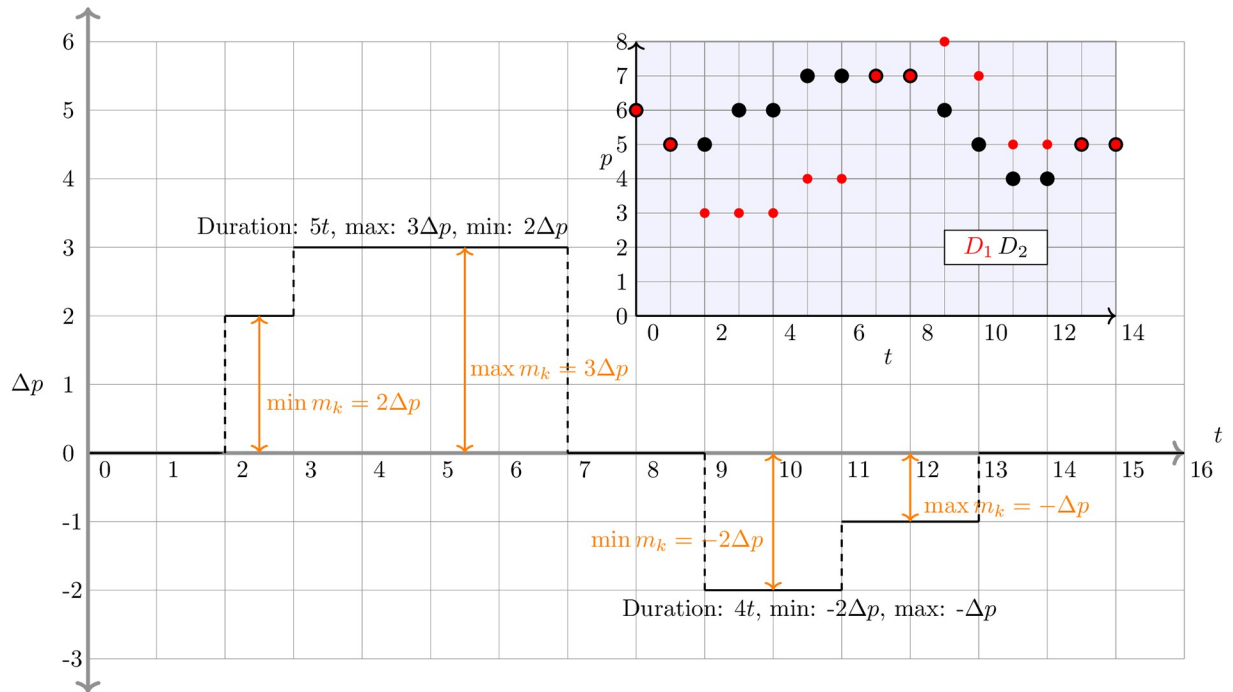


Fig 2. Diagram of two dislocation segments (DS). The inset plot shows the time series of best quotes that generate the DSs. Where the time series diverge from the same value, a DS occurs. We have deliberately not placed units on t , Δp , and p to indicate that DSs can occur in any market in which there are differing information feeds, not just in the NMS, though we do assume that these quantities are quantized. In the case of the NMS, we take t in units of μs and Δp in units of $\$0.01$. For the sake of simplicity this figure only displays one side of a hypothetical book. Marker size in the inset plot is used only for visual distinction.

<https://doi.org/10.1371/journal.pone.0226968.g002>

over a day was calculated and recorded. With this construction, positive opportunity costs indicate an incentive for liquidity demanding market participants to use the SIP feed while negative opportunity costs indicate an incentive to use the aggregated direct feeds. By ignoring the sign of the opportunity costs, and thus which feed is favored, an aggregate or total realized opportunity cost is constructed. Intra-day events can offset—e.g., a trade that resulted in ROC that disadvantaged direct data users and a trade that resulted in ROC that disadvantaged SIP data users could both occur on the same day, partially offsetting the total ROC due to opposite signs. Precise definitions of quantities described here are located in [S2 Appendix](#).

As above, we provide a brief toy example of how realized opportunity cost can arise and a description of its' calculation. A minimal example involves two traders, each of which is in the market to buy the security XYZ. One trader places orders using the SIP NBO to determine the appropriate limit price and the other places orders using the best offer from a direct feed. If a trade for 100 shares of XYZ executes at $\$100.00$ per share, the current direct best offer, when the NBBO was a SIP quote of $\$100.01$ per share, a trader placing a bid informed by the SIP could receive an execution that resulted in a realized opportunity cost of $\$0.01$ per share, or $\$1.00$ in total. Because this opportunity cost favored the direct feed, this portion of ROC would be assigned a negative value. If, during another trade on the same day, another trade for 100 shares of XYZ executes when the direct best offer price is $\$101.02$ and the SIP NBO price is $\$101.00$ per share, the trader who places orders informed exclusively by the direct feeds could have experienced a realized opportunity cost of $\$0.02$ per share, or $\$2.00$ in

total, assuming that they may have been able to find counter-parties at the SIP NBO. This ROC is assigned a positive value because it favors the SIP feed. Summing these two together produces a net ROC of \$1.00, hence the conservative nature of our estimates. If, instead, our calculation summed the absolute value of each ROC-generating event, the figure above would instead be \$3.00. A more detailed example of ROC calculation from real trade data is located in [S4 Appendix](#).

6 Results

6.1 Dislocations and dislocation segments

We find that dislocations and dislocation segments are widespread, from the point of view of our observer in Carteret, and may have qualitative welfare effects on NMS participants, particularly large investors or investors that interact with the NMS directly on a frequent basis. There were a total of 120,355,462 dislocation segments among Dow 30 stocks in 2016. Now, let's assume a uniform distribution of dislocations throughout the trading day. On average, we therefore expect $\frac{120,355,462}{252 \times 6.5 \times 60^2} \approx 20.4$ dislocation segments per second. When restricting our attention to what we term *actionable* dislocation segments (those with a duration longer than 545 μ s), we find that there were 65,073,196 actionable dislocation segments, or on average, $\frac{65,073,196}{252 \times 6.5 \times 60^2} \approx 11$ actionable dislocation segments every second. Even when inspecting actionable dislocation segments with a minimum magnitude greater than 1 cent, we find that there were 2,872,734 instances of these dislocation segments, or on average, $\frac{2,872,734}{252 \times 6.5 \times 60^2} \approx 0.49$ dislocation segments per second, or almost one large and actionable dislocation segment every two seconds.

We focus much of our subsequent analysis on the dislocation segment distribution conditioned on both duration ($> 545\mu$ s) and magnitude ($> \$0.01$) From an academic point of view, dislocations with a minimum magnitude greater than one cent are more interesting, since one might expect many dislocations to feature a magnitude that corresponds with the price quantization—minimum tick size (\$0.01 in this case). There are several aspects of this conditional distribution that bear special notice. First, the distribution of each attribute is exceptionally heavy-tailed. In absolute value, the 75%-iles of the minimum and maximum magnitude are three cents—but the mean in absolute value of the minimum magnitude (resp. maximum magnitude) is 3.05 (resp. 8.23) cents. A similar phenomena is true for the duration distribution, displayed in [Fig 3](#), where the 75%-ile is 4231 μ s, while the mean is an astounding 0.389 seconds, almost two orders of magnitude longer. The max magnitude, min magnitude, and duration distributions are all highly skewed, while the distributions of the maximum and minimum magnitudes are nearly identical. Further summary statistics on dislocations with various conditioning are displayed in [Table 3](#).

[Fig 4](#) shows the distribution of dislocation segments modulo day, binned by minute. Intraday dislocation segment distributions are markedly nonuniform, with a majority of the probability mass concentrated toward the beginning of the trading day. There is also a notable spike in the number of dislocation segments occurring in mid-afternoon and at the very end of the trading day. Additionally, there seems to be a decaying cyclic pattern in the distribution, with spikes occurring with a 30 minute frequency.

We postulate that the mid-afternoon spike, which occurs at approximately 2:00pm, is associated with meetings of the Federal Open Market Committee (FOMC). These meetings release economically important information such as decisions regarding federal rate changes and economic forecasts, and their impact has been noted by several market participants, including analysts at NYSE [[43](#), [44](#)]. Note that the NYSE analysis of the impact of FOMC meetings is

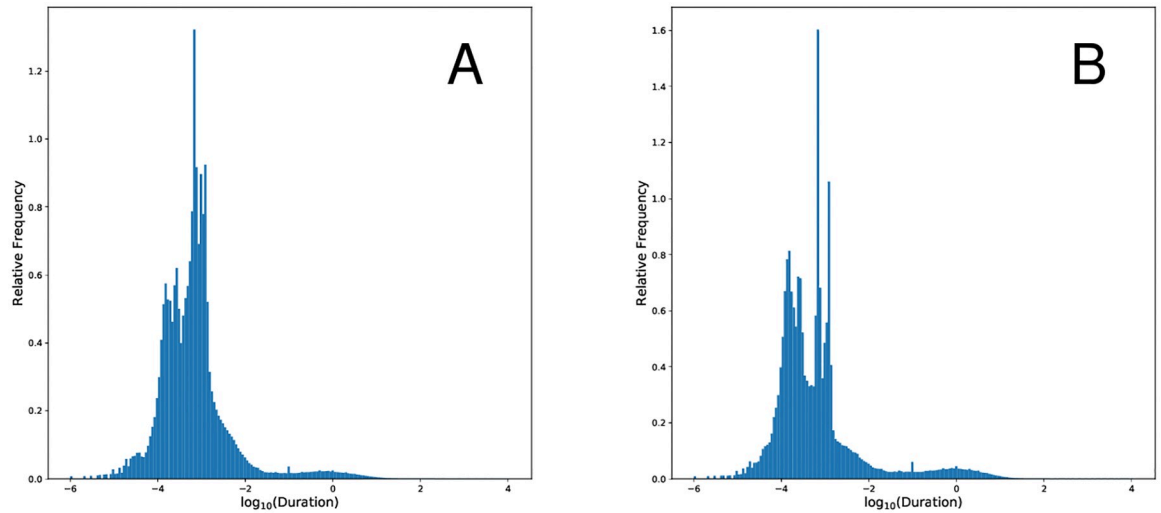


Fig 3. Panel A displays the distribution of dislocation segment (DS) durations. Panel B displays the distribution of DS durations with a magnitude greater than \$0.01. Both panels have a logged x-axis.

<https://doi.org/10.1371/journal.pone.0226968.g003>

Table 3. Dislocation segment (DS) attributes where the first section is unconditioned, the middle section is restricted to DSs with a duration longer than 545 μ s, and the final section is restricted to DSs with a duration longer than 545 μ s and a minimum magnitude greater than \$0.01. Of the approximately 120 million DSs observed, more than 54% of them have a duration that would allow them to be considered actionable, and about 2.4% of them are both actionable and feature a minimum magnitude greater than \$0.01. This makes the magnitude of the realized opportunity cost even more remarkable. Additionally, note that observed durations of “0” are the result of DSs that begin and end within the same microsecond, the maximum precision used for the majority of market data timestamps.

Filter	Statistic	Duration	Min. Value	Max. Value	Min. Mag.	Mean Mag.	Max. Mag.
None	count	120, 355, 462					
	mean	0.073712	-0.0012	0.0013	0.0112	0.0124	0.0137
	std	5.519033	0.1698	0.4815	0.0529	0.2581	0.5075
	min	0.000000	-141.49	-63.21	0.01	0.01	0.01
	25%	0.000216	-0.01	-0.01	0.01	0.01	0.01
	50%	0.000624	0.01	0.01	0.01	0.01	0.01
	75%	0.001190	0.01	0.01	0.01	0.01	0.01
	max	10,789.83	372.69	4,905.69	372.69	2,452.85	4,905.69
Duration > 545 μ s	count	65, 073, 196					
	mean	0.136142	-0.0020	0.0022	0.0109	0.0130	0.0151
	std	7.505197	0.2233	0.6511	0.0653	0.3474	0.6850
	min	0.000546	-141.49	-63.21	0.01	0.01	0.01
	25%	0.000751	-0.01	-0.01	0.01	0.01	0.01
	50%	0.001103	0.01	0.01	0.01	0.01	0.01
	75%	0.002391	0.01	0.01	0.01	0.01	0.01
	max	10,789.83	372.69	4,905.69	372.69	2,452.85	4,905.69
Duration > 545 μ s & Min. Mag. > \$0.01	count	2, 872, 734					
	mean	0.387866	-0.0250	0.0267	0.0305	0.0564	0.0823
	std	29.566716	0.9046	1.0021	0.3102	0.7116	1.3115
	min	0.000546	-141.49	-63.21	0.02	0.02	0.02
	25%	0.000724	-0.02	-0.02	0.02	0.02	0.02
	50%	0.001207	0.02	0.02	0.02	0.02	0.02
	75%	0.004231	0.02	0.02	0.03	0.03	0.03
	max	10,789.83	372.69	593.43	372.69	372.84	593.43

<https://doi.org/10.1371/journal.pone.0226968.t003>

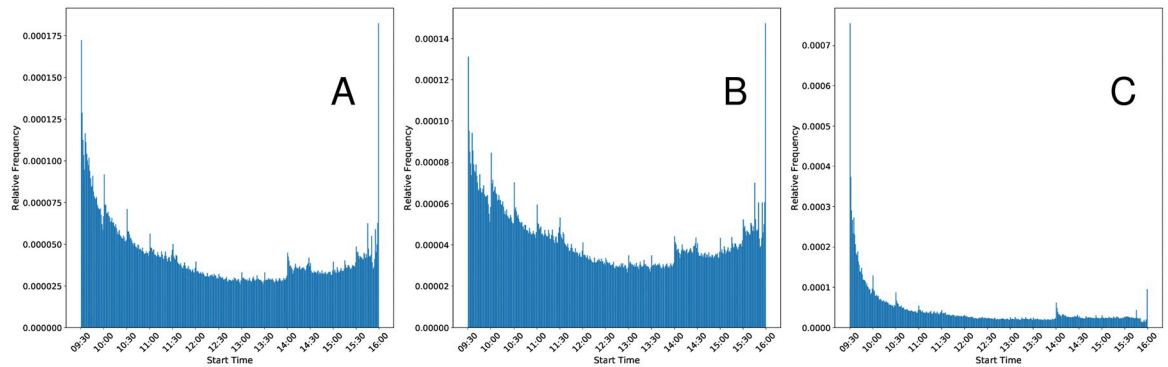


Fig 4. Panel A displays the distribution of dislocation segment (DS) start times binned by minute. Panel B displays the distribution of actionable DSs. Actionable DSs are those with a duration longer than 545µs. Panel C filters the actionable DSs to only include those with a minimum magnitude > \$0.01. Note that the distributions are heavily skewed right. A plurality of actionable DSs occur in the half-hour following the opening bell when compared to any other half-hour during the day. There is also a spike in the number of dislocation segments in the middle of the afternoon, which may be due to information events, such as press releases from meetings of the Federal Open Market Committee.

<https://doi.org/10.1371/journal.pone.0226968.g004>

based upon a quote volatility measure, which is conceptually quite similar to the dislocations discussed in our work. Regarding the cyclic pattern, it seems that most of this activity can be attributed to the aggregated effect of seemingly random market events. Investigating the data without aggregation reveals that almost no days exhibit this cyclic behavior for DS occurrence, though there are many days that seem to have one or more abnormal spikes in DS occurrence at seemingly random times. During aggregation, these potentially large spikes are not entirely smoothed out, leading to the cyclic pattern observed in Fig 4. Interested readers may investigate the dislocation segment occurrence distributions without aggregation by using the interactive application provided in our GitLab repository [45].

To further unpack the relationship between time of day, length, and magnitude of dislocation segments, we created a representation of dislocation segments modulo day as an ordered network, termed a circle plot. Fig 5 illustrates the construction of the circle plots from a few toy examples. Figs 6 and 7 depict circle plots for AAPL for an arbitrary day, whereas Figs 8 and 9 depict circle plots for AAPL for the entirety of 2016.

Circle plots are constructed using the following algorithm. Starts and stops of dislocation segments at time t (as measured and timestamped by our observer in Carteret) are termed events $v(t)$ and denoted by black nodes. More than one event can occur at each time t ; all events are represented by the same node. Events $v_i(t)$ and $v_j(s)$ where $t < s$ are connected by an edge e_{ij} when a dislocation segment starts at $v_i(t)$ and ends at $v_j(s)$. It is not necessarily the case that dislocation segments start and stop in order as seen above; for example consider two dislocation segments, the first starting at v_i , and the second starting at v_j . The first dislocation segment could end at v_k , and the second could end at v_e . When N events occur “out of order” in this way, we identify the events as a single component (even though, as in the above example, the component decomposes into two two-tuples of events) and term it an N -component for reasons we state below; the above example is a 4-component. Nodes are plotted in rays that spread outward from the geometric center of the plot in a modulo 10 relation. Edges between nodes v_i and v_j are weighted according to the quantity

$$\sum_{(v_i, v_j)} \max(|\Delta p_{\max}|, |\Delta p_{\min}|), \tag{2}$$

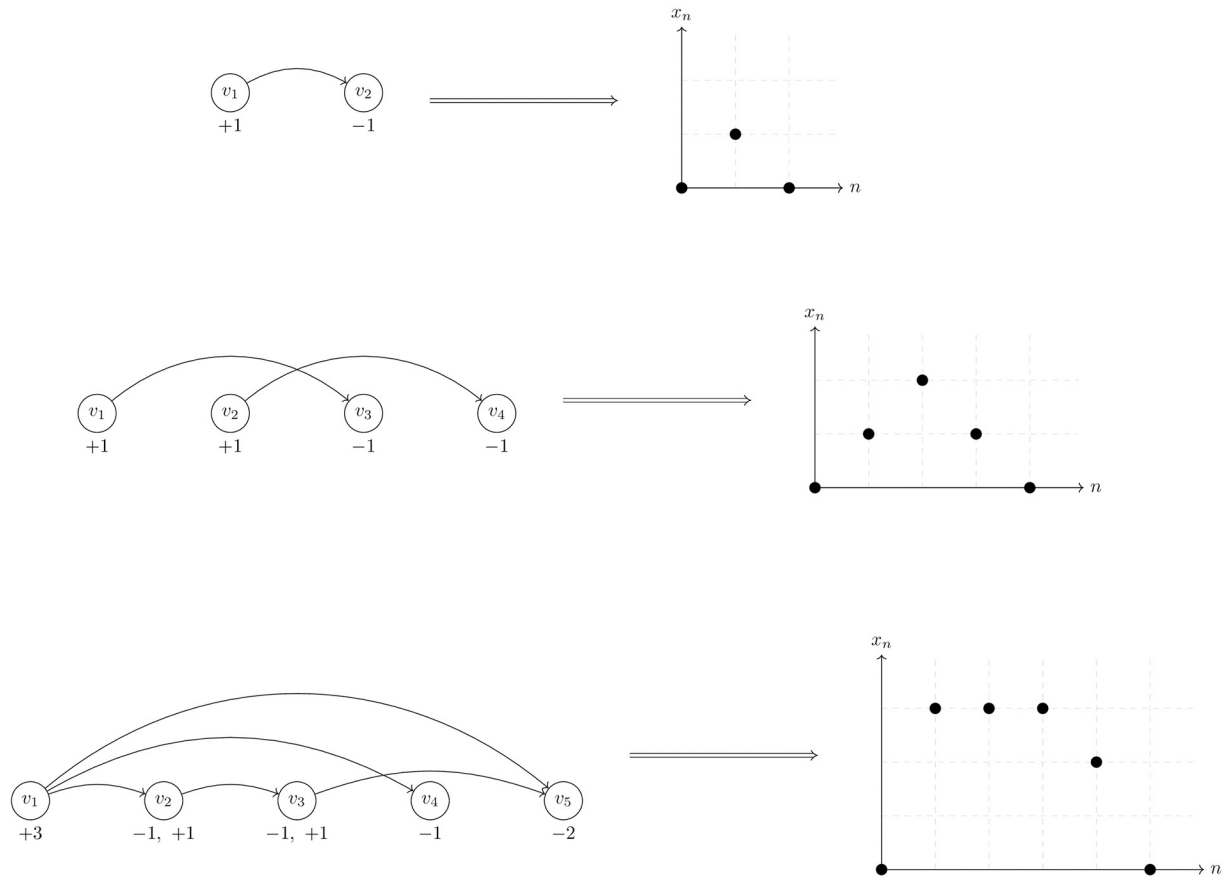


Fig 5. A depiction of the injection mapping from an N -component in an ordered network to a tied positive random walk of length $N + 1$. The injection is given by j outgoing edges $\cong j$ steps up and likewise k incoming edges $\cong k$ steps down. The total number of steps up or down is given by $x_{n+1} - x_n = \#$ of steps up + $\#$ of steps down. The top row displays a simple 2-component, where an equity begins a dislocation at time t_i and ends it at time t_{i+1} . The corresponding walk on the line starts at zero, moves up a step, and then moves down. The second row displays a 4-component identical to that described in the text of the article. This 4-component, which is separable into two disconnected pieces, demonstrates the geometric nature of the ordered network. Since an ordering is imposed on the nodes, the crossing of the edges implies the staggered starts and stops of the two dislocations.

<https://doi.org/10.1371/journal.pone.0226968.g005>

where the sum is taken over all events that started at node v_i and ended at node v_j and Δp_{\max} and Δp_{\min} are the largest positive (resp. smallest negative) change in value that occurred during each event. Fig 9 displays the ordered network for AAPL aggregated (modulo day) over the entire trading year. There is high event density near the beginning of the day and there is another spike in density near noon-12:30 PM. This clustering can make interpretation of the fine event structure difficult to discern, so we conduct a re-normalization into event space with a simple method: consecutive events $v_i(t)$ and $v_j(s)$ are plotted in order, but at a uniform distance so that the measure on the graph becomes a Stieltjes-type instead of a Lebesgue-type measure. In other words, in the case of the real time representation, an event represented by a node on a fixed but arbitrary circle of the graph occurred at a multiple of $10\mu s$ from all other events represented by nodes on the ring; in the case of the event-time representation, an event represented by a node on a fixed but arbitrary circle of the graph and another event represented by a node on the same circle are separated by an integer multiple of events that occurred between them. Fig 6 displays the ordered network in this re-normalized space, where it is easier to see that the usual behavior of dislocation segments is a regular cyclic, on-off (start-stop)

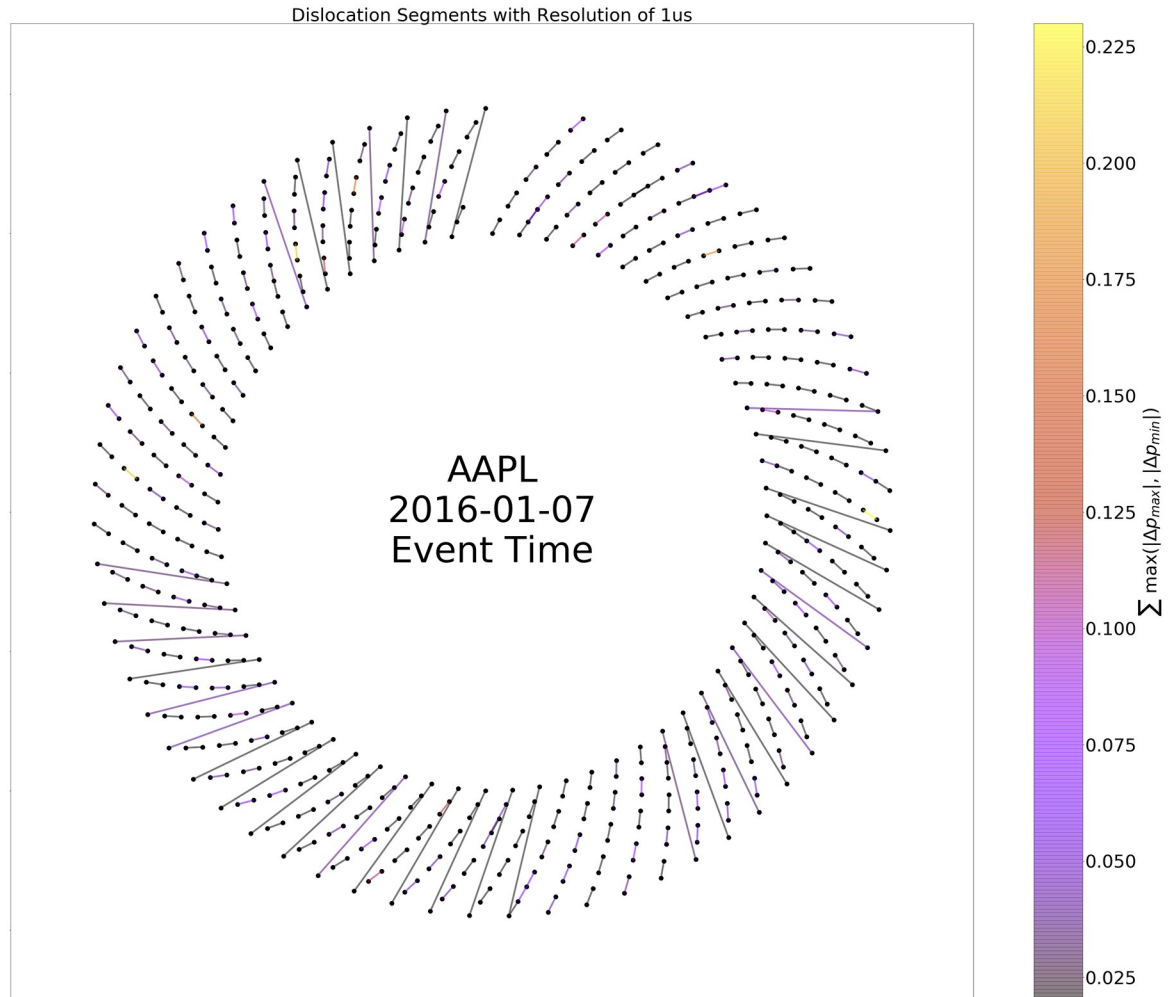


Fig 6. Distribution of dislocation segments (DS) with minimum magnitude greater than \$0.01 and duration longer than 545 μ s for AAPL on 2016-01-07 visualized with a time re-normalization procedure. Nodes are placed in rings modulo 10; nodes zero through 9 are in the first ray from the origin, then the angle in the plot is incremented and nodes 10 through 19 are in the second ray, etc. A link e_{ij} connects two nodes, v_i and v_j , if a dislocation segment starts at v_i and stops at v_j . This view of the dislocation segment network preserves time ordering while defining a nonlinear transformation between uniform time ordering, as shown below in Fig 7, and uniform event-space ordering, as shown here. As noted in the text, it is not necessary for only one dislocation segment to exist at the same point in time t . For example, there are many instances of new dislocation segments starting while another is still ongoing—the first starts at v_i and then another starts at v_j and ends at v_k , followed by the first dislocation segment ending at v_l . Irregular behavior such as this generates the banding of the edge distribution. Interested readers may wish to have some more context for the selected date. For AAPL, 2016-01-07 ranked 8th out of 252 trading days when considering ROC. \$106,990.23 in ROC was accumulated, which lies between the minimum of \$2,773.35 and the maximum of \$138,331.08. This day of AAPL also ranked 15th when considering the number of DSs. A total of 108,843 occurred, falling between the minimum of 9,256 and the maximum of 188,656.

<https://doi.org/10.1371/journal.pone.0226968.g006>

pattern. However, there are multiple deviations from this pattern—any component other than a 2-component is structurally different from a purely sequential pattern. In fact, there is an injection from an N -component and a tied, non-negative sequence $\{x_n\}_{n=0}^N$, $x_0 = x_{N+1} = 0$, $x_n \geq 0$ for all n . This injection is defined by the relationships “start of k events $\cong k$ steps up” and “end of k events $\cong k$ steps down”.

As a concrete example, the 4-component described above maps to the sequence steps $\{1, 1, -1, -1\}$, with values $x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 1, x_4 = 0$. Fig 5 displays a toy example of the

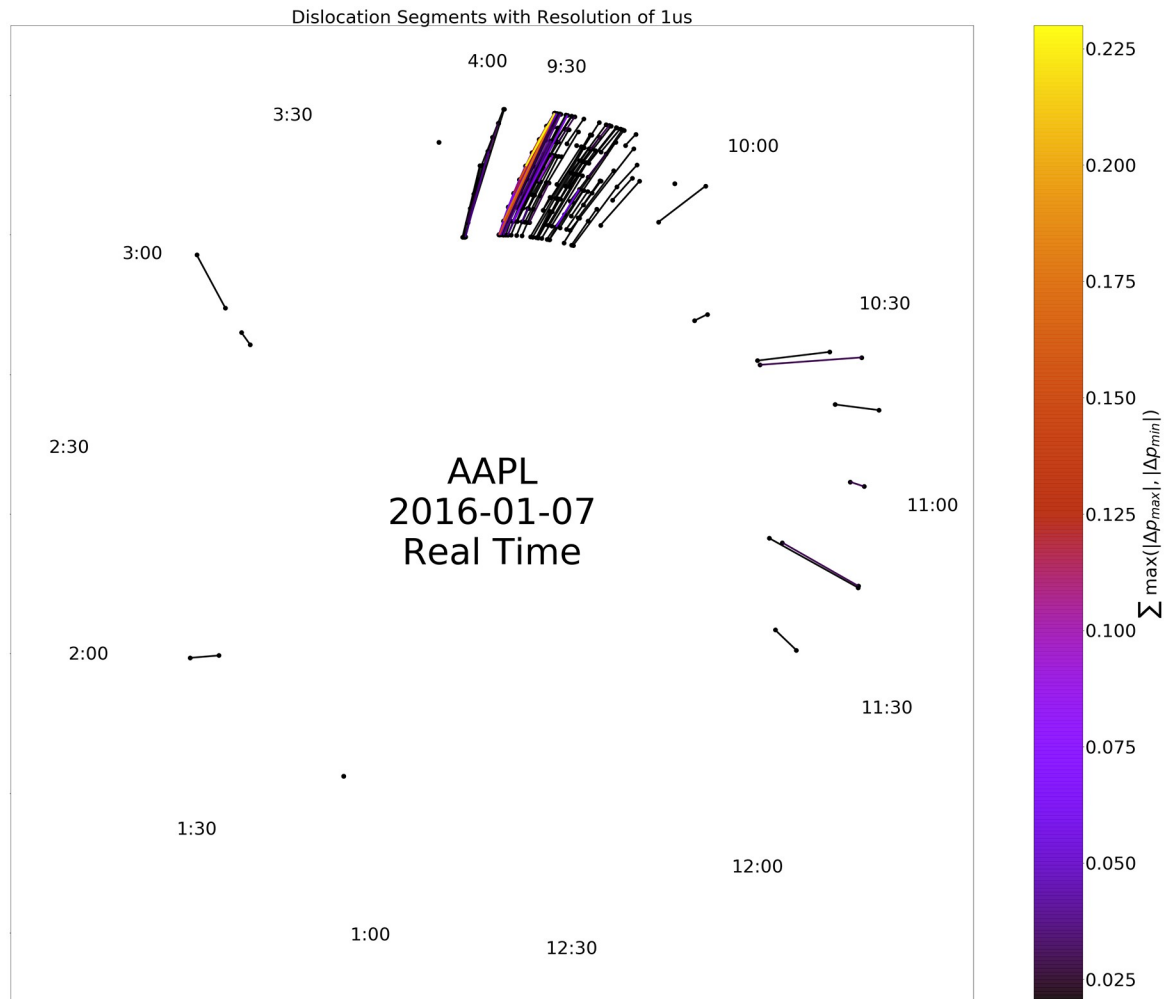


Fig 7. Dislocation segments in AAPL on 2016-01-07 without time re-normalization. The characteristic structure in the occurrence of dislocations segments is clearly displayed, with the majority occurring near the beginning of the trading day.

<https://doi.org/10.1371/journal.pone.0226968.g007>

injection between N -components in an ordered network and a tied positive sequence, as outlined above.

When aggregated over all trading days, evidence of persistent nontrivial structure in the event-space density of N -tuples emerges. As stated above, Figs 8 and 9 display the aggregate of events in AAPL modulo day. Visualizations of all Dow 30 securities in this format are at the authors' webpage (<https://compfi.org>).

6.2 Realized opportunity cost

The large number of actionable dislocation segments likely has a direct effect on the opportunity cost market participants may incur by using one information source over the other. The aggregate of this realized opportunity cost can be estimated by cataloging the quantity and characteristics (average price difference, etc.) of differing trades. Table 1 summarizes many of these findings. In the time period studied (01-01-2016 through 31-12-2016) there were a total of 392,101,579 trades of stocks in the Dow 30, with a traded value of \$3,858,963,034,003.48 USD. Of those trades, we classified 87,432,231 trades, or 22.3% of the total number of trades,

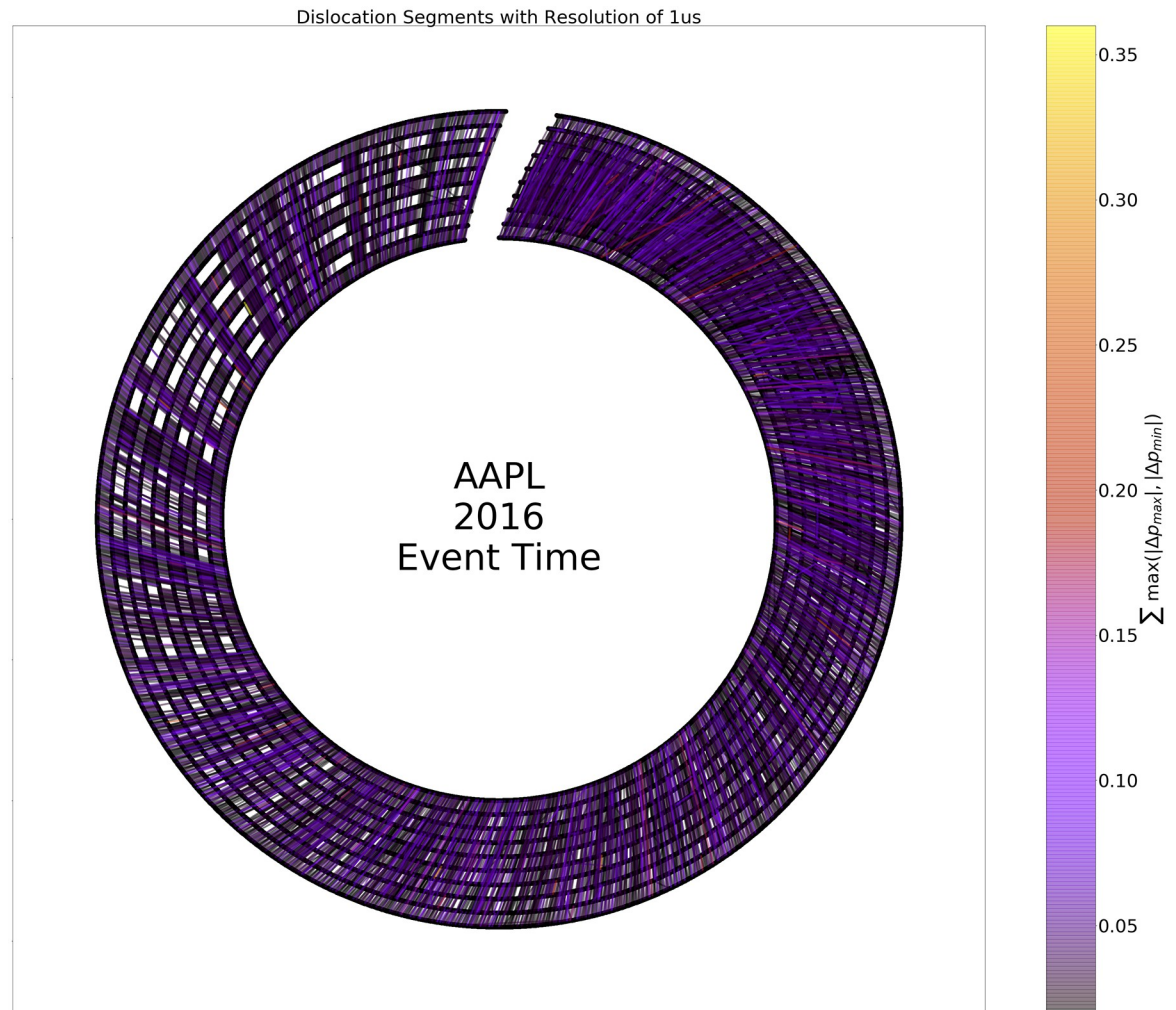


Fig 8. Dislocation segments (DS) aggregated over an entire year (modulo trading day). Investigating structures at the trading day timescale is of interest as this is likely the longest timescale over which HFT strategies are used. Here DSs are plotted in event space, where density is uniform between events. Note the presence of irregular structure even here, evidence of higher-order structure in the ordering of starts and stops of DSs.

<https://doi.org/10.1371/journal.pone.0226968.g008>

as *differing* trades, defined as follows: if the trade is on the buy side, it is a differing trade if the SIP bid is not equal to the direct bid; if the trade is on the sell side, it is a differing trade if the SIP offer is not equal to the direct offer. These differing trades had a traded value of \$900,535,924,961.72 USD, or 23.34% of the total traded value. More optimal use of information presented by the SIP and direct feeds could have saved market participants a total of \$160,213,922.95 USD in ROC. This opportunity cost was distributed unevenly, with traders informed by NBBO prices suffering \$122,081,126.40 USD in ROC, while traders informed by DBBO prices only accumulated \$38,132,796.55 USD in ROC.

Fig 10 provides insight into the joint distribution of total and differing trades. While we might expect that the ratio of total to differing trades would have a linear relationship, this is not observed empirically. Fig 11 displays the daily net opportunity cost aggregated over all tickers in our sample, showing some of the dynamics present in the occurrence of ROC over the period of study. Table 4 provides an aggregated summary that describes ROC and related

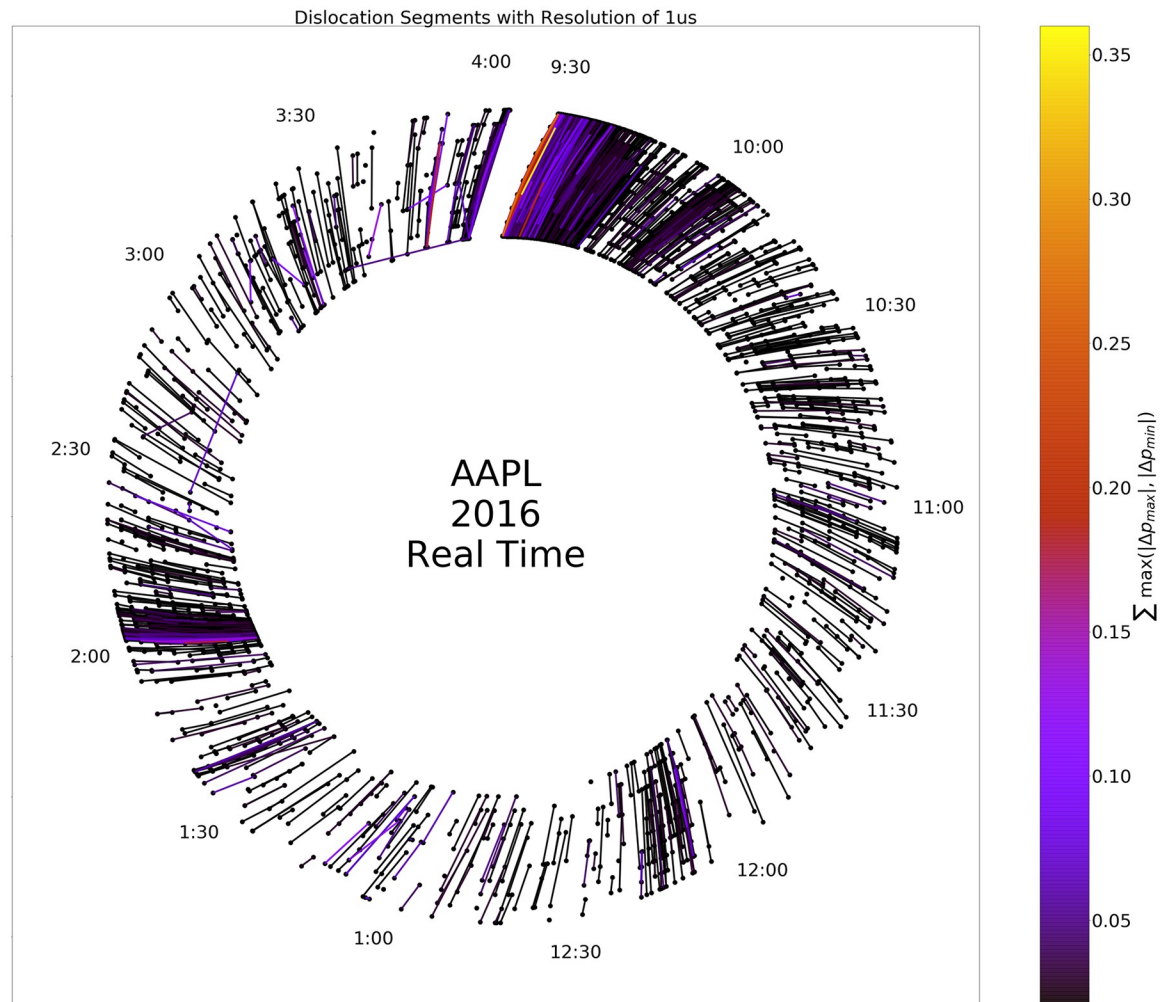


Fig 9. Dislocation segments (DS) aggregated over an entire year (modulo trading day), as above, but not transformed to event space. The high density of dislocation segments at the beginning of the trading day, near 12:10, and near 2:00 is readily apparent.

<https://doi.org/10.1371/journal.pone.0226968.g009>

statistics over the tickers and trading days in our sample. [S1 Table](#) gives additional details of these statistics for each ticker in our study. Though our observer was located in Carteret while many securities (all but four during 2016) in the Dow 30 are listed on NYSE, located in Mahwah, consultation with [S1 Table](#) demonstrates that mean ROC per ticker does not differ significantly by listing venue (one-way ANOVA: $F(4, 20) = 1.35$, $p = 0.25$; Kruskal-Wallis H -test: $H = 0.84$, $p = 0.35$).

7 Concluding remarks

Using the most comprehensive set of NMS data publicly available, we have shown that market inefficiencies in the form of dislocations and realized opportunity cost were common in the Dow 30 in 2016 as measured by our observer in the NASDAQ data center in Carteret, NJ. We find that inefficiencies due to the physical fragmentation of the market are widespread, totaling over \$160M USD in realized opportunity cost and 2,872,734 dislocations of magnitude $> \$0.01$ and duration $> 545\mu\text{s}$. These figures correspond well with those reported in other bodies of work [23, 26]. Additionally, we found that the average trade that occurred during a

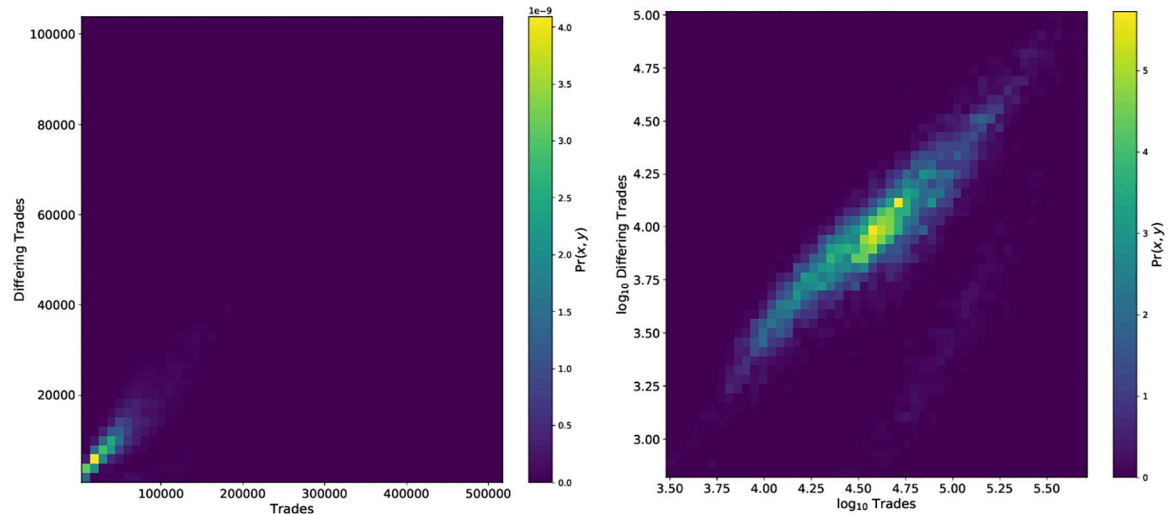


Fig 10. Left: A bivariate empirical distribution function for total trades and number of differing trades. Right: The same distribution, but with logged axes. We might expect *a priori* that they are related by a constant proportion and hence should observe a fit $\log_{10} \text{ total trades} = c + \log_{10} \text{ differing trades}$, where $c < 0$. Though there is good evidence of this linear relationship, we see there is a non-negligible area of higher total trades with markedly sub-linear scaling of differing trades.

<https://doi.org/10.1371/journal.pone.0226968.g010>

dislocation moved approximately 5% more value than the average trade that occurred when the NBBO and DBBO were synchronized (see Table 1 row 10). In the fifth Need for Speed report [32], Mackintosh and Chen indicate that 29% of traded value executes within a small window around quote changes, closely aligning with rows 8 and 9 from Table 1. This may indicate that market participants could be more heavily impacted by the existence of dislocation segments than previous analyses suggest.

Beyond our empirical results, S2 Table contains estimates of some costs associated with the usage of direct feeds, highlighting the stark cost difference between SIP data and direct feed data.

Though our work is empirical, our results do have implications for theoretical results on nuances of financial market efficiency. The discovery of systematically-different prices as measured in geographically-distinct locations that can be routinely observed by agents with access to higher-speed information flows—and cannot be routinely observed by agents without this access—has a logical bearing on questions of distributional effects of asymmetric information and market design. This feature of fragmented market structure can be viewed as a modern-day example of the Grossman-Stiglitz paradox [14]. Trading agents who are able to act at higher speeds may be rewarded for their investment, effort, and risk-taking behavior by executing on trading opportunities that exist for very short time intervals. In fact, without competition among traders to reduce processing time and infrastructure providers to implement faster communications protocols and networking equipment, dislocations and associated inefficiencies would likely be more prevalent. Opportunity cost realized by market participants (in the form of ROC as detailed above) is ultimately attributable to the physically- and topologically-fragmented nature of the NMS. Despite this fact, we believe that the current market configuration offers many benefits over alternative configurations, such as the null model defined in Section 4. These results should not be considered as evidence for or against a specific market configuration since, as stated above, the observed phenomena may incentivize the participation of certain kinds of market actors.

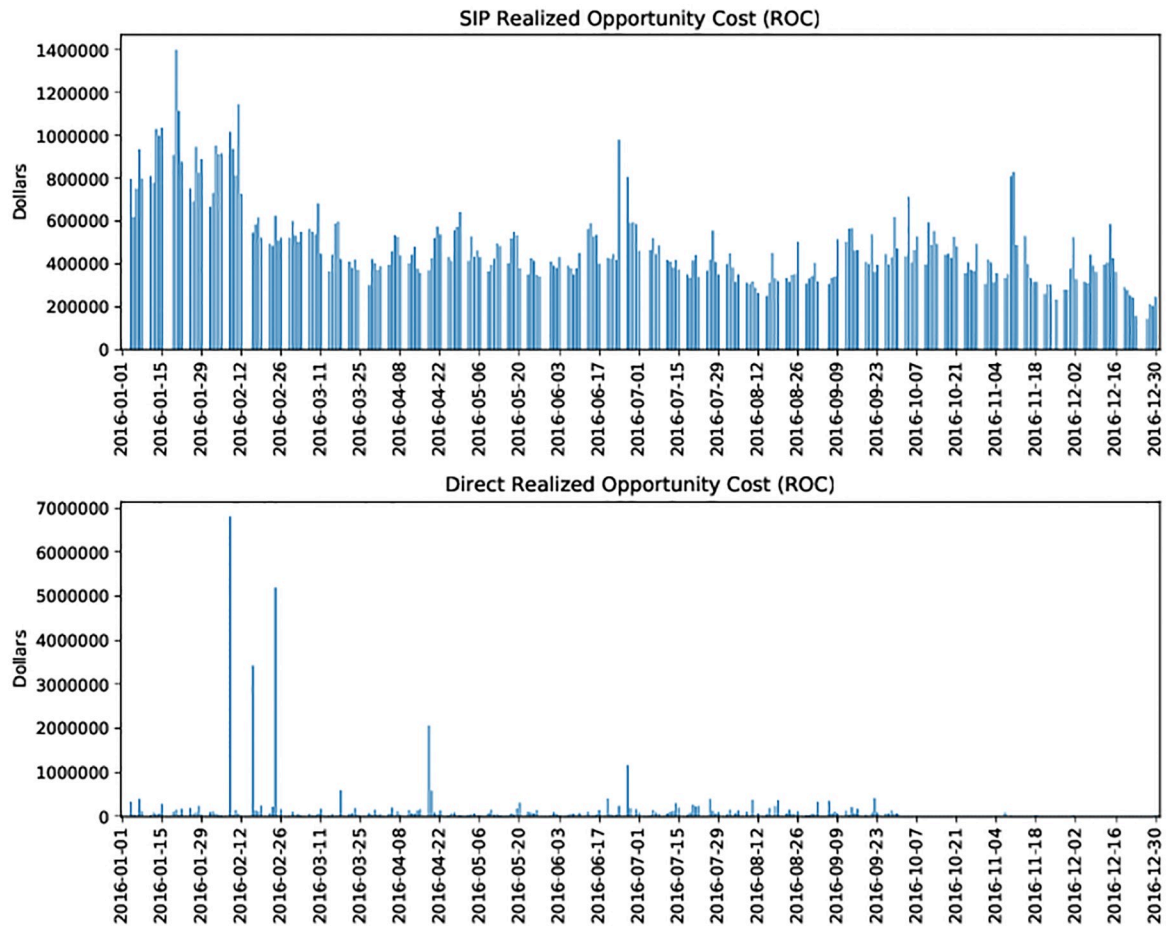


Fig 11. Daily ROC during calendar year 2016 aggregated across all tickers. A large majority of days favored the direct data feeds. Both Direct and SIP ROC time series show signs of decay across 2016, which may be due to infrastructure improvements.

<https://doi.org/10.1371/journal.pone.0226968.g011>

We focused our attention on the Dow 30 during calendar year 2016 in order to provide a strong, but tractable baseline. Future work should investigate longer time periods, larger groups of equities, and other exchange traded products such as Exchange Traded Funds (ETF). For example, an extension of the current work to larger groups of equities, such as the S&P 500 or the Russell 3000 would provide greater context for how fragmentation effects different portions of the equities market. While a time series analysis of dislocation segments and realized opportunity cost series over several years could provide useful information about how

Table 4. Summary statistics of realized opportunity cost and related statistics for Dow 30 stocks, aggregated over the 252 trading days in 2016.

	Trades	Traded Value	Diff. Trades	Diff. Traded Value	ROC	ROC/Share
mean	1,555,958.65	15,313,345,373.03	346,953.30	3,573,555,257.78	635,769.54	0.011804
std	463,558.93	3,891,299,900.31	146,677.85	1,234,882,079.43	655,911.15	0.008592
min	579,206	6,664,671,053.15	89,564	1,035,855,029.71	145,205.65	0.008848
25%	1,278,813.25	12,915,031,172.08	262,209	2,804,569,367.64	417,485.73	0.009613
50%	1,429,062	14,431,597,662.02	309,158	3,274,390,601.60	514,856.64	0.010154
75%	1,715,351.25	16,829,521,684.38	387,772	3,993,470,514.97	666,268.27	0.011213
max	3,596,006	30,999,914,293.66	1,073,029	9,428,952,387.10	7,817,684.58	0.098303

<https://doi.org/10.1371/journal.pone.0226968.t004>

fragmentation effects have evolved due to changes in regulation, technology, and market participant behavior.

Supporting information

S1 Appendix. Market participants. Describes several classes of market participants and how they interact.

(PDF)

S2 Appendix. Glossary. Provides formal definitions for many of the terms used in the study.

(PDF)

S3 Appendix. Regulation National Market System. A high level summary of some regulations that impact the implementation of the National Market System in the US.

(PDF)

S4 Appendix. Dislocations and ROC. Additional details on how we calculate Realized Opportunity Cost and Dislocations. Also provides some example calculations and discussion regarding the connections between ROC and Dislocations.

(PDF)

S1 Table. Summary ROC statistics for Dow 30 Stocks. Aggregated by day and trading symbol.

(PDF)

S2 Table. Direct feed and historical data pricing. The pricing presented in this table assumes a single consumer with an academic use case aiming to construct a dataset similar to what was used in this analysis. It is also assumed that non-display fees do not apply. Historical data costs assume a 12 month period of interest, i.e. calendar year 2016. Strictly speaking, historical data may sufficient for replicating the analysis presented in this paper, making subscription to live feeds unnecessary. However, utilizing historical data provided by each exchange excludes the possibility of collecting data from a single point of observation, reintroducing the issues of clock synchronization and relativity. Additionally, highlighting the monthly cost for comprehensive direct feed access shines a light on one of the reasons for the lack of academic participation in the analysis of modern U.S. stock markets. This does not include costs which may be incurred while curating the data, fulfilling potential co-location requirements, ISP/networking costs, computing hardware acquisition and maintenance, etc. DoB indicates that a product contains full Depth of Book information (adds, mod, and cancel messages), while ToB indicates that a product contains only Top of Book information (trade and quote messages). The NYSE Historical ToB product, also called NYSE Daily TAQ, is frequently used in academic studies due to its relatively low cost and broad coverage (e.g. [24] use this product). Historical data from CHX is not directly available, and the live feeds are transitioning to NYSE technology, thus historical CHX data must be purchased from a third party. This list is not guaranteed to be comprehensive, additional fees/costs may exist. * Access to UTP data and NASDAQ direct feed data may granted freely to academic institutions, see UTP Feed Pricing and NASDAQ Academic Waiver Policy for more info. ** Historical data purchased from NYSE only covers 5/21/2018—present for NYSE National, thus an alternative data provider is required in order to obtain historical data from 2016. The sources used to construct this table include [CTA feed pricing](#), [UTP feed pricing](#) via the Data Policies document, [NYSE feed pricing](#), [NYSE historical data pricing](#), [NASDAQ feed pricing](#), [BATS/DirectEdge feed pricing](#), and [CHX feed pricing](#).

(PDF)

S3 Table. Example AAPL trades. Trades that occurred during a dislocation in AAPL on 2016-01-07 at approximately 9:48am, more than three minutes after the trading “guardrails” are enforced. The “Delta” column indicates the difference between the Thesys timestamp and the SIP publication timestamp (in microseconds). For trade 0, Thesys received the trade at 9:48:55.396951 and the SIP timestamp was 9:48:55.396696. The “Extra” column contains additional deltas related to the timestamps added in the 2015 SIP changes, see [24] for additional details. In particular, this column contains the difference (in microseconds) between the Thesys timestamp and the exchange timestamp. For trade 0, Thesys received the trade at 9:48:55.396951 and the exchange timestamp was 9:48:55.397602, an example of the timestamp inversion seen in [24], which is generally caused by clock synchronization issues.
(PDF)

S4 Table. Example AAPL trades with positive ROC. A subset of the trades from [S3 Table](#) that resulted in positive ROC. Positive ROC indicates that these trades received favorable prices that were aligned with the SIP NBBO.
(PDF)

S5 Table. Example AAPL trades with negative ROC. A subset of the trades from [S3 Table](#) that resulted in negative ROC. Negative ROC indicates that these trades executed at less favorable prices than what was offered by the DBBO.
(PDF)

Acknowledgments

The authors gratefully acknowledge helpful discussions with Anshul Anand, Yosry Barsoum, Aaron Blow, Lashon Booker, Robert Books, David Bringle, Richard Byrne, Peter Carrigan, Gary Comparetto, Bryanna Dienta, Peter Sheridan Dodds, Jordan Feidler, Andre Frank, Chase Garwood, Bill Gibson, Frank Hatheway, Emily Hiner, Chuck Howell, Eric Hunsader, Robert Jackson, Neil Johnson, Connie Lewis, Phil Mackintosh, Rishi Narang, Joseph Saluzzi, Wade Shen, Jonathan Smith, Thomas Wilk, the participants of the “Flash Mob Research in Computational Finance” hosted by the Complex Systems Center at the University of Vermont, and the participants of Swarm Fest 2016 hosted by the University of Vermont. 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. B.F.T., D.R.D., C.M.V.O., J.H.R., and J.G.V. were supported by DARPA award #W56KGU-17-C-0010. B.F.T., M.T.T.K., M.T.M., D.M.S., and J.G.V. were supported by HSARPA award #HSHQDC-14-D-00006. C.M.D. was supported by NSF grant #1447634. C.M.D., and T.J.G. were 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.

Author Contributions

Conceptualization: Brian F. Tivnan, David Rushing Dewhurst, Colin M. Van Oort, John H. Ring, IV, Tyler J. Gray, Brendan F. Tivnan, Matthew T. K. Koehler, Matthew T. McMahon, David M. Slater, Jason G. Veneman, Christopher M. Danforth.

Data curation: David Rushing Dewhurst, Colin M. Van Oort, John H. Ring, IV, Matthew T. K. Koehler, Matthew T. McMahon, David M. Slater, Jason G. Veneman.

Formal analysis: David Rushing Dewhurst, Colin M. Van Oort.

Funding acquisition: Brian F. Tivnan, Matthew T. K. Koehler, Matthew T. McMahon, David M. Slater, Jason G. Veneman, Christopher M. Danforth.

Investigation: David Rushing Dewhurst, Colin M. Van Oort, John H. Ring, IV, Tyler J. Gray, Brendan F. Tivnan, Matthew T. K. Koehler, Matthew T. McMahon, David M. Slater, Jason G. Veneman, Christopher M. Danforth.

Methodology: Brian F. Tivnan, David Rushing Dewhurst, Colin M. Van Oort, Christopher M. Danforth.

Project administration: Brian F. Tivnan, Matthew T. K. Koehler, Matthew T. McMahon, David M. Slater, Jason G. Veneman, Christopher M. Danforth.

Resources: Matthew T. K. Koehler, Matthew T. McMahon, David M. Slater, Jason G. Veneman, Christopher M. Danforth.

Software: David Rushing Dewhurst, Colin M. Van Oort, John H. Ring, IV, Matthew T. K. Koehler, Matthew T. McMahon, David M. Slater, Jason G. Veneman.

Supervision: Brian F. Tivnan, David Rushing Dewhurst, Christopher M. Danforth.

Validation: Colin M. Van Oort, John H. Ring, IV, Tyler J. Gray, Brendan F. Tivnan, Jason G. Veneman, Christopher M. Danforth.

Visualization: David Rushing Dewhurst, Colin M. Van Oort, John H. Ring, IV.

Writing – original draft: David Rushing Dewhurst.

Writing – review & editing: Brian F. Tivnan, David Rushing Dewhurst, Colin M. Van Oort, John H. Ring, IV, Tyler J. Gray, Brendan F. Tivnan, Christopher M. Danforth.

References

1. Indices SDJ. Dow Jones Industrial Average; 2018.
2. Kirilenko A, Kyle AS, Samadi M, Tuzun T. The flash crash: The impact of high frequency trading on an electronic market. Available at SSRN. 2011;1686004. <https://doi.org/10.2139/ssrn.1686004>
3. Goldstein MA, Kavajecz KA. Trading strategies during circuit breakers and extreme market movements. *Journal of Financial Markets*. 2004; 7(3):301–333. <https://doi.org/10.1016/j.finmar.2003.11.003>
4. Grinblatt M, Keloharju M. The investment behavior and performance of various investor types: a study of Finland's unique data set. *Journal of financial economics*. 2000; 55(1):43–67. [https://doi.org/10.1016/S0304-405X\(99\)00044-6](https://doi.org/10.1016/S0304-405X(99)00044-6)
5. FINRA. ATS Transparency Data Quarterly Statistics;.
6. Fama EF. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*. 1970; 25(2):383–417. <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>
7. Bouchaud JP. Econophysics: Still fringe after 30 years? arXiv preprint arXiv:190103691. 2019;.
8. Foye J, Mramor D, Pahor M. The Persistence of Pricing Inefficiencies in the Stock Markets of the Eastern European EU Nations. 2013.
9. Fama EF, French KR. Size, value, and momentum in international stock returns. *Journal of financial economics*. 2012; 105(3):457–472. <https://doi.org/10.1016/j.jfineco.2012.05.011>
10. Johnson N, Zhao G, Hunsader E, Qi H, Johnson N, Meng J, et al. Abrupt rise of new machine ecology beyond human response time. *Scientific reports*. 2013; 3:2627. <https://doi.org/10.1038/srep02627> PMID: 24022120
11. O'Hara M. High frequency market microstructure. *Journal of Financial Economics*. 2015; 116(2):257–270. <https://doi.org/10.1016/j.jfineco.2015.01.003>
12. Lo AW. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. 2004.
13. Akerlof GA. The market for “lemons”: Quality uncertainty and the market mechanism. In: *Uncertainty in Economics*. Elsevier; 1978. p. 235–251.

14. Grossman SJ, Stiglitz JE. On the impossibility of informationally efficient markets. *The American economic review*. 1980; 70(3):393–408.
15. Bloomfield R, O'hara M, Saar G. How noise trading affects markets: An experimental analysis. *The Review of Financial Studies*. 2009; 22(6):2275–2302. <https://doi.org/10.1093/rfs/hhn102>
16. Budish E, Cramton P, Shim J. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*. 2015; 130(4):1547–1621. <https://doi.org/10.1093/qje/qjv027>
17. Black F. Noise. *The Journal of finance*. 1986; 41(3):528–543. <https://doi.org/10.1111/j.1540-6261.1986.tb04513.x>
18. Blume ME, Goldstein MA. Differences in Execution Prices among the NYSE, the Regionals, and the NASD. Available at SSRN 979072. 1991;.
19. Lee CM. Market integration and price execution for NYSE-listed securities. *The Journal of Finance*. 1993; 48(3):1009–1038. <https://doi.org/10.1111/j.1540-6261.1993.tb04028.x>
20. Hasbrouck J. One security, many markets: Determining the contributions to price discovery. *The journal of Finance*. 1995; 50(4):1175–1199. <https://doi.org/10.1111/j.1540-6261.1995.tb04054.x>
21. Barclay MJ, Hendershott T, McCormick DT. Competition among trading venues: Information and trading on electronic communications networks. *The Journal of Finance*. 2003; 58(6):2637–2665. <https://doi.org/10.1046/j.1540-6261.2003.00618.x>
22. Shkilko AV, Van Ness BF, Van Ness RA. Locked and crossed markets on NASDAQ and the NYSE. *Journal of Financial Markets*. 2008; 11(3):308–337. <https://doi.org/10.1016/j.finmar.2007.02.001>
23. Ding S, Hanna J, Hendershott T. How slow is the NBBO? A comparison with direct exchange feeds. *Financial Review*. 2014; 49(2):313–332. <https://doi.org/10.1111/fire.12037>
24. Bartlett RP, McCrary J. How rigged are stock markets? Evidence from microsecond timestamps. *Journal of Financial Markets*. 2019;. <https://doi.org/10.1016/j.finmar.2019.06.003>
25. Alexander J, Giordano L, Brooks D. Dark Pool Execution Quality: A Quantitative View. <http://blogthemistrading.com/wp-content/uploads/2015/08/Dark-Pool-Execution-Quality-Short-Finalpdf>. 2015;.
26. Wah E. How Prevalent and Profitable are Latency Arbitrage Opportunities on US Stock Exchanges? 2016.
27. Securities US, Commission E. MIDAS: Market Information Data Analytics System; 2013.
28. Angel JJ, Harris LE, Spatt CS. Equity trading in the 21st century. *The Quarterly Journal of Finance*. 2011; 1(01):1–53. <https://doi.org/10.1142/S2010139211000067>
29. Angel JJ, Harris LE, Spatt CS. Equity trading in the 21st century: An update. *The Quarterly Journal of Finance*. 2015; 5(01):1550002. <https://doi.org/10.1142/S2010139215500020>
30. Carrion A. Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets*. 2013; 16(4):680–711. <https://doi.org/10.1016/j.finmar.2013.06.005>
31. Menkveld AJ. High frequency trading and the new market makers. *Journal of Financial Markets*. 2013; 16(4):712–740. <https://doi.org/10.1016/j.finmar.2013.06.006>
32. Mackintosh P, Herrick J, Chen KW. The Need for Speed Reports 1-5. 2014-2016;.
33. Goldstein MA, Kumar P, Graves FC. Computerized and High-Frequency Trading. *Financial Review*. 2014; 49(2):177–202. <https://doi.org/10.1111/fire.12030>
34. Chordia T, Goyal A, Lehmann BN, Saar G. High-frequency trading. 2013;.
35. Arnuk S, Saluzzi J. Broken markets: how high frequency trading and predatory practices on Wall Street are destroying investor confidence and your portfolio. FT Press; 2012.
36. Securities US, Commission E. Consolidated Audit Trail; 2012.
37. Sachs G. Sigma X2 Form ATS-N;.
38. Miller K. Calculating Optical Fiber Latency;.
39. Technologies A. Anova Technologies Network Map; 2018.
40. Inc TG. Thesys Group Inc.; 2018.
41. Opportunities FB. MIDAS Contract Award Notice; 2012.
42. Popper N, Protess B. To Regulate Rapid Traders, S.E.C. Turns to One of Them; 2012.
43. Insights NE. Fed Rate Decisions and Quote Volatility;.
44. Insights NE. Market Reactions After Fed Rate Cut;.
45. Lab CF. Dislocation Visualizer. <https://compfi.org/visualizer.php>.