

Market Structure Primer

Get Started!

The purpose of this document is to provide an overview of the US equity trading landscape. It started as an onboarding document to help new hires who aren't as well versed in market structure. It can be read start to finish in order, or you can jump around. The visual Life Cycle of an Order also provides a handy view of the various events and entities involved in the process of a trade - from conception to completion. For those with little to no prior knowledge, the three core sections on market participants, mechanisms of communication between participants, and a birds-eye view of overall market activity should provide a working understanding of trading in the US equities market as it is today. Over time, we will be adding material that delves more deeply into further topics, especially into market structure history and the conflicts of interest that evolve along with market structure.

Birds-eye View of Overall Market Activity

How Much and How Often

Now that we've fleshed out the different kinds of entities in the US equities market and the logistics of communications between them, it's time to zoom out a bit and get a sense of scale and proportions. How much trading activity happens on a typical day? How it is distributed by size and time and place? How much variation is there intraday and interday? Here we will give you some basic answers to these questions. Obviously, there is a lot more structure and patterns underlying trading than the first order statistics here can capture, but our goal for now is just to get you started on building a basic intuition for the scale and nature of market activity.

Venues

There currently 3 large parent companies which each operate several stock exchanges. ICE operates five (this is often also referred to as the "NYSE family"): NYSE, NYSE Arca, NYSE American, NYSE National, and NYSE Chicago. CBOE operates four: EDGX, EDGA, BZX, and BYX. Nasdaq operates three: Nasdaq, BX, and PSX. The only exchange outside of the three families is IEX, which is operated as an independent company.

There are roughly 30 ATS's, which is down from around 40 in 2014.

Stocks and ETFs

Currently, there are roughly 8300 distinct securities traded in the US equities market. For each security, the market capitalization is calculated by taking the number of shares outstanding and multiplying by the current price per share. If we sum up the market capitalization in each security over all the publicly traded securities, we get the total market capitalization of the US public market, which passed 32 trillion in 2018. This represents about 38% of the global market.

Of the roughly 8300 securities, approximately 5300 of them are publicly listed companies. Most of the rest are ETFs, which stands for "Exchange-Traded Funds." More generally, there are ETPs ("Exchange-Traded Products"). An exchange-traded fund is typically designed to track the behavior of a particular collection of stocks and/or bonds, and it trades in the market throughout the trading day like an individual stock. When its price deviates from the underlying stocks/bonds it is referencing, arbitrageurs will use mechanisms that exchange shares of the fund for the underlying securities to bring the ETF's price back in line with what it is intended to track. ETNs ("Exchange-Traded Notes") similarly trade like individual stocks and relate to a collection of underlying securities, but in this case the underlying securities are unsecured debt securities.

In terms of scale, 8300 securities may not seem like a very big number, especially when we are talking about a market that functions electronically. 8300 things is certainly a number of things that computers can easily manage, right? Well, yes and no. 8300 rows in a database, or even an excel spreadsheet, is easy enough. But as a multiplier, 8300 can become a real problem. If a computation and data-intense analysis on one day's data for one stock takes 1 second to run, for example, running it on all stocks in sequence (one at a time) could take 8300 seconds, which is 138 minutes. And this is probably too optimistic - as the program is likely to save some information into memory as it goes, using up resources, and making each run of the computation potentially slower than the previous ones. Not to mention that the running time per stock is likely to vary widely based on how much data there is to churn through.


Though trading in different stocks may frequently exhibit correlation (e.g. multiple stocks impacted by common news or other market drivers), at a mechanical level trading in different stocks is separable, and can be implemented in parallel on different machines. To match a trade in Microsoft, for example, you only have to keep track of the orders and current price information for Microsoft, and you don't need to know anything about the other stock orders and prices. Stock exchanges will typically split the 8300 securities into several groups, and the trading for each group will be processed by its own machine. In the extreme case, you could imagine 8300 different machines, each acting in parallel to process the orders and produce trades for a single security. The same kind of trick can be used to speed up analytics or any other computational process that looks only at the data for one stock at a time. Parallelization is one tool that keeps things running quickly in the US equities market, mitigating the effect of the 8300 multiplier.

Trades and Sizes

On any given day, around 6.7 billion shares are traded in the US equities market. ETFs represent about 18-20% of these volumes. Trades typically happen in round "lots," which for most stocks are multiples of 100 shares. The average trade size is currently very small - roughly 200 shares per trade. This is not to say that large trades aren't also common. Block trades (trades of at least 10,000 shares) happen frequently. For example, looking at the [Nasdaq daily market summary](#) of all trading in Nasdaq-listed securities for June 7, 2019, we see that there were over 8000 block trades, accounting for close to 20% of the day's total trading volume. The 200 share average trade size hides a much more varied distribution, comprised of many large trades, many small trades of 200 or 100 shares, and some even smaller "odd lot" trades of less than 100 shares.

Since there are roughly 6.7 billion shares traded each day with an average trade size of roughly 200 shares, we can infer that the number of trades per day is approximately 34 million. These are not spaced evenly throughout the trading day. There are often bursts of activity in one or many stocks, followed by stretches of relative inactivity. These bursts are often at speeds a human observer cannot keep pace with - frequently events are following within less than 1 millisecond of each other. Trading is also typically heavy in the closing auction: for NYSE-listed securities, for example, the closing auctions represent about 7% of the volume in those securities.

Orders, Quotes, and Ticks

The number of orders that are transmitted to a venue is typically much higher than the number of resulting trades. This can be seen, for example, in [the data that the SEC provides](#)  concerning the ratio of order cancelations to trades at each exchange.

The ratio of cancelations to trades varies widely by date and by venue, but something like 20/1 is somewhat typical. This gives us a general sense that there is on average a lot of nimble maneuvering of quotes around each trade.

Over the course of a trading day across all listed securities, there are tens of millions of times when an NBO price or NBB price changes. NBO stands for "National Best Offer," and it refers to the lowest price advertised by a would-be seller across exchanges. NBB stands for the "National Best Bid," and it refers to the highest price advertised by a would-be buyer across exchanges. The "NBBO" is a shorthand often used to refer to the pair of the NBB and NBO prices. Naturally, the NBB/NBO are time-dependent constructs, and practically speaking they are also latency-dependent constructs, as information about price changes takes time to travel from one market participant to another.

The number of NBBO changes per day is roughly on the same order of magnitude as the number of trades. Note that this does not count times when the total *size* available at the best bid or best offer price changes, but the prices themselves do not change. Nor does it count the times when displayed orders away from these best prices are changed or canceled.

A fairly high percentage of NBBO price changes occur in rapid succession: nearly 40% of them occur within 1 millisecond of another NBBO price change. To put this in context, we can calculate the number of milliseconds (ms) in the trading day as: 6.5 hours x 60 minutes/hr x 60 seconds/minute x 1000 ms/s = 23,400,000. If we approximate the total number of NBBO price changes across all roughly 8300 listed securities as 20 million, then we have on average about 2400 NBBO price changes per trading day per stock. So this is about 1 NBBO price change every 10,000 milliseconds. So the fact that they often occur within 1 ms of each

other reflects a rather extreme amount of clustering, suggesting that we should conceptualize price changes as isolated bursts of activity, punctuating long stretches of relative calm.

Market Share

The market share of a venue is calculated as the amount of trading volume that occurs on that venue on a given day, divided by the total volume across the US equities market. Thus, the sum of the market share over all venues will be equal to 1. Roughly 62% of trading volume happens on exchanges, with the remaining 38% happening elsewhere (dark pools or internalizers). Within the exchange landscape, each of the 3 major exchange families, CBOE, Nasdaq, and NYSE, represents about 20% of trading volume (summing up all of the individual exchanges within each family). The remaining 2.5% market share belongs to IEX.

Indices

There are several indices which are commonly used as heuristic indicators of overall market health. The behavior of indices is a frequent headline of financial reporting (e.g. "The Dow Crossed 20,000 Today!"), presumably because indices are quick and easy to report, whereas more detailed statements like "these 1345 stocks trended up by small amounts today, whereas these 1642 stocks trended down by small amounts today, these 587 stocks wiggled a lot but ultimately landed in roughly the same place, and these 763 stocks..." well, you get the idea.

So how do indices condense the more detailed activity into a single number? Each index represents a hypothetical portfolio of holdings that an investor could choose to maintain among all of the publicly listed stocks. To form a hypothetical portfolio, one must choose which stocks to include, and in what proportions. Once that is decided, the "value" of the portfolio can be calculated by multiplying the current prices of the stocks in the portfolio and summing these with weightings by the specified proportions. Naturally this gives you a number that changes dynamically as prices change. The underlying unit (whether you imagine you are holding 100 shares of Microsoft and 200 shares of Intel vs. 200 shares of Microsoft and 400 shares of Intel, etc.) doesn't really matter, as long as it is consistent. It is the relative behavior of the index over time that tells you something about the evolution of stock market over time.

Sometimes what it tells you is pretty limited or difficult to interpret. There are many degrees of freedom in defining an index, as there are many ways to decide which stocks to include, and many ways to decide the relative proportions. Some indices are clearly more meaningful than others. An "index" that is comprised of simply one stock, for example, presumably tells you fairly little about the behavior of other stocks. The price movements in a single stock will perhaps reflect general trends across the entire market and more granular trends across the relevant sector (e.g. energy, financials, utilities, etc.), but will also be driven by the specifics of that particular company. If we want the meaning of an index to transcend the details of any

one particular company or small subset of companies, we need to choose a reasonably "representative" sample of stocks and proportions to define our index.

S&P 500

A very commonly used example of this is the S&P 500 index, which is comprised of the roughly 500 top companies, primarily chosen by market capitalization (recall: market cap = price per share times number of outstanding shares). The weighting for the S&P 500 Index corresponds to market cap, meaning that the relative weight given to a particular company is proportional to its market cap. This allows a rather clean interpretation of changes in the index: if the index increases by 10%, for example, that means the total market capitalization of all the included companies has increased by 10%.

The S&P 500 is often used as a proxy for general market health, which is quite reasonable, as the included companies cover about 80% of the total value of the US stock market. This tells us that the stock market is pretty top heavy: total market capitalization is fairly concentrated in the top 500 companies among the many thousands of total companies. Looking at the S&P 500 over time gives us a broad sense of how the market has historically trended up and down.

DOW

The DOW Jones Industrial Average is comprised of 30 stocks of large companies. The included companies change over time, as individual companies may rise and fall in influence. When the index was created by Charles Dow and Edward Jones in the late 1800s, it included 12 companies, and it was calculated by simply adding the price of 1 share of each of the companies and dividing by 12. It is still calculated by adding the prices of 1 share of stock for each company in the index and dividing by a single number, but that number is now called the "divisor" and has historically been adjusted to make the index consistent across stock splits and mergers. In between adjustments of the divisor, we can think of the DOW as an average of the 30 stock prices with equal weight on each. This means, for example, that a 10% relative increase in the price of one of the higher priced stocks will cause a greater relative increase in the DOW than a 10% relative increase in the price of one

of the lower priced stocks. In contrast, a \$1 absolute price increase in any one of the component stock prices will have an indential effect on the index.

The companies included in the DOW index represent about a quarter of the value in the total US stock market. But because of the smaller sample size and the fact that the index is not weighted by market cap, the DOW is not really a great proxy for overall market health. It remains a popular data point mostly because of its historical longevity and broad familiarity.

Russell Indices

The Russell Indices, e.g. the Russell 3000 Index, the Russell 2000 Index, and the Russell 1000 Index, etc., pick out different subsets of stocks based on market capitilization. For example, the Russell 3000 Index includes the largest 3000 companies by market capitalization, while the Russell 1000 includes only the top 1000 companies and the Russell 2000 includes the next 2000. Once the subset of stocks is determined, their individual contributions to the index are weighted proportionally to market cap. The Russell indices can be used as proxies for the cummulative performance of different categories of stocks: large cap, small cap, etc.

Index Funds

Since an index corresponds to a hypothetical portfolio of stocks, it is possible to maintain a portfolio matching the hypothetical one, and turn the hypothetical behavior into realized returns. This process isn't quite perfect, since sometimes the stocks or weights in an index change, and adjusting a real portfolio to keep in sync with the hypothetical one involves buying and selling that incurs additional cost and risk. But to a reasonable extent, one can approximate the performance of an index with a real portfolio, and this is what index funds are designed to do. Investing in an index fund that is properly managed should allow you to achieve *approximately* the returns corresponding to the hypothetical index it is implementing. This is why index funds that simulate the S&P 500 index, for example, are popular investments for investors who just want to generally benefit from stock market growth in an overall sense, without entangling their returns too closely with the performance of a smaller number of individual stocks.

When the constitution of an index changes, a lot of market activity will result as index funds seek to adjust their holdings accordingly. The Russell indices, for example, are reconstituted on a scheduled yearly basis. This is to adjust for changes to market capitalization that have occurred over the preceding year. Trading on and around such scheduled events will be markedly heavier (and different in distribution) from ordinary trading activity.

Volatility and Variation

Over time, there is a lot of variation in the kind of daily stats we have quoted above. Some variation in trading behavior is driven by structural things: e.g. index rebalancing or the fact that stock options have certain standardized expiration dates (following the third Friday of a given month). But even aside from these anticipated structural events, within a day or hour or minute, etc. there can be wild changes in activity and prices within a single stock, multiple stocks, or broadly across the market. The number of NBBO changes in a day for instance, can vary by a multiplicative factor of 3 or 4 (or more) when we compare a relatively calm trading day to a relatively turbulent one.

In a technical sense, "volatility" in the context of stock trading usually refers to the standard deviation or variance of prices for a particular security or set of securities. Roughly, this is a measure of how far prices tend to drift above and below their average over a specified unit of time. This gives additional color around price trends that mere averages or medians do not. For example, if we say "the average price of Microsoft over the trading day today was X," we mean that we took all the trades of Microsoft over the day and averaged their prices (presumably weighted by their sizes). But this alone tells us nothing about whether we had, say, many trades at price exactly X, or trades that swung widely between lows of $X/2$ and highs of $2X$, averaging out to X overall. Volatility disambiguates these disparate situations. Low volatility means that we had trades mostly clustered tightly around the average price, while high volatility means that prices oscillated considerably.

Information about expected future volatility can be inferred by looking at option prices. Since the value of an option depends on the probability that a stock price will go above or go below a specified strike price in the future, the price of an option right now can tell us something about "the market's" estimate of the likelihood of particular future price movements. The CBOE Volatility Index, or VIX, measures this kind of "implied volatility" by using SPX option trades to infer the market's expectation of 30-day volatility in the S&P 500 index.

Further reading:

[SIFMA Insights: US equity Market Structure Primer, July 2018](#)  (*this was a source for many of the stats above)