

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.