

# SHORT IS LONG

Using dark pool short sales as a proxy for buying activity

SqueezeMetrics  
Research

March 2018

We're going to do three things in this paper.

First, we're going to tell you why a short sale in a U.S. exchange-traded stock is actually a meaningful indicator of *buying* activity. Then we're going to demonstrate this using only data that is freely available. Finally, we're going to invite readers to replicate our results (and extrapolate) using the same data, to which we will provide a URL.

## Why Short Is Long

A short sale is the sale of a security that the seller does not own. Traditionally, short sales are associated with speculation—traders betting that a stock will go down.

But according to the SEC, trades marked with the seller “short” comprise about 49% of equity share volume.\* In other words, about half of all selling volume in the market is *short* selling. Unless we actually believe that half of market volume is speculative short selling, this demands an explanation.

To understand, you need to know how market-makers (MMs) do business. Traditionally, market-makers make their money by “quoting a spread.” This means placing a bid at, say, \$19.95, and an offer at \$20.00. Since the MMs have no position in the stock, the offer at \$20.00 is necessarily entered as a short sale—they don't own it, so they can't actually sell it.†

Now imagine the common scenario where one investor *sells* to the MM at \$19.95 and another *buys* from the MM at \$20.00. Knowing that MMs place short sales at the offer, we know that only one of these trades will be reported as “short”—the one where the investor is *buying* stock at \$20.00.

\* See “Short Sale Position and Transaction Reporting” in the References.

† This is why MMs are exempt from short-selling restrictions.

Investor	Market-maker	Sale reported as
Buys	Sells	Short
Sells	Buys	Long (not short)

This means that whenever a market-maker fills an investor's buy order, the MM is facilitating the trade by shorting shares. Thus, short volume is actually representative of investor buying volume, and non-short volume is representative of investor selling volume. It's no coincidence that short volume is predictably *half* of total volume—short sales represent the buying half of the market, and long (non-short) sales represent the selling half.

Maybe you don't find this convincing. What about those speculative short sellers? Don't they throw the number off? (No. That volume is tiny.) What about passive sell orders from traders who *aren't* market-makers? Wouldn't their sales show up as *not* short? (Yes. But they'll still be selling to a MM, so we would correctly interpret them as sales.) The U.S. equities market is infamous for its complexity—how can we make such sweeping assertions?

Granted, it does seem that the market is too multifaceted, and opaque, for short volume to simply equate to investor buying volume, but there's a practical reason that today's market is more transparent than you might imagine. In a word, *rebates*. Here's the story.

*\* Later in 1997, the tightest spread actually changed from  $\frac{1}{8}$  to  $\frac{1}{16}$ , or from \$0.125 to \$0.0625.*

Back in 1997, when the U.S. market's tightest spread was 12.5 cents,\* the Island ECN (an off-exchange trading venue for NASDAQ stocks) started doing something different: They *paid* market-makers to do what market-makers already do—quote a spread. Whenever a trader “took” the liquidity that the MM posted at the bid or offer, that trader would pay a small fee, and most of that fee would be given, as a “rebate,” to the MM.

As a result, Island attracted all kinds of MMs, ending up with tighter spreads, and more depth, than their competitors. Naturally, customers liked this, and within two years, Island had gone from 3% to 13% market share of NASDAQ volume. The idea caught on and other exchanges, public and private, started using the “maker-taker” model and offering rebates. So spreads tightened everywhere.

By 2005, when the SEC's Regulation NMS officially made sub-penny pricing a reality, nearly everyone had adopted the maker-taker model. By then, MMs made more money by collecting rebates than their traditional business of quoting spreads. This was a new environment. There were tiny margins, and no margin for error—only the very fastest MMs would get that rebate.

High-frequency trading (HFT) is primarily the story of the mad scramble for these rebates. If you submit an order to buy 100 shares of XYZ at the market, you *will* buy those shares from an HFT market-maker. If you later submit a passive limit order to sell 100 shares of XYZ at the offer, you *will* sell those shares to an HFT market-maker (once someone buys 100 shares from him at a higher price elsewhere). Call it “frontrunning” if you like, but competition by MMs to garner the rebate has actually been a boon for the average investor.†

*† While we don't deny the existence of some “predatory” HFTs, they are largely irrelevant.*

Here's why it matters to us: The competition to provide liquidity guarantees that the vast majority of orders go through MMs, because with spreads at a penny or less, they can't afford to let anyone else collect those rebates. Thanks to this state of affairs, there is a middle-man for just about every order that is executed on the market. Lucky for us, this middle-man *always* leaves a trail—he is compelled to sell short whenever he fills a buy order.

This decades-long struggle to collect a liquidity rebate, culminating in aggressive HFT market-making, is why short volume has become such a strong indicator of buying activity.

### Testing the Hypothesis

What we would hope to find from a test is that high short sale volume is positively correlated to intraday stock gains. Unfortunately, there is no real-time short sale reporting requirement, and so we cannot simply dig through the ticker tape to find the volume of short sales.\* Luckily, we *can* use a proxy, and even better, it's free.

\* More on tick-level short sale data in the Appendix.

Ever since 2010, FINRA has collected short sale volume data from their Trade Reporting Facilities (TRFs).† The TRFs receive data from exclusively *off-exchange*, or “dark,” venues. Some of these venues are Alternative Trading Systems (ATs), or “dark pools” and some are “internalizers.” Since our analysis doesn't need to distinguish between types of off-exchange venues, we'll refer to all data from the TRFs as “dark pool” data.

† This was pursuant to an SEC request as part of Regulation SHO. More on this in the Appendix.

The only thing that's “dark” about a dark pool is that there is no pre-trade data, i.e., there is no visible order book or quoting. The one thing that's particularly appealing about dark pools for the customer is that you can buy or sell stocks *between* the bid and ask prices, usually at the midpoint. This can dramatically reduce trading costs.

‡ The materials presented in Barclays' 2014 lawsuit made it clear that liquidity providers are a standard feature of dark pools.

For a market-maker in a dark pool (don't be fooled, dark pools have MMs)‡ this only changes one thing—midpoint trades mean spreads become even tighter (it's sub-penny). But for our purposes, this is irrelevant. The rebate is still king, and the MMs will still buy what's sold to them and *short sell* what's bought from them, just like they do on the lit exchange.

With all of this in mind, the only suitability concern about using short volume data from dark pools as a proxy for the whole market is that it just isn't enough data. Fortunately, off-exchange volume accounts for about *a third* of all trade volume. This should be more than enough to get a picture of the correlation between short volume and intraday stock gains—if there is one.

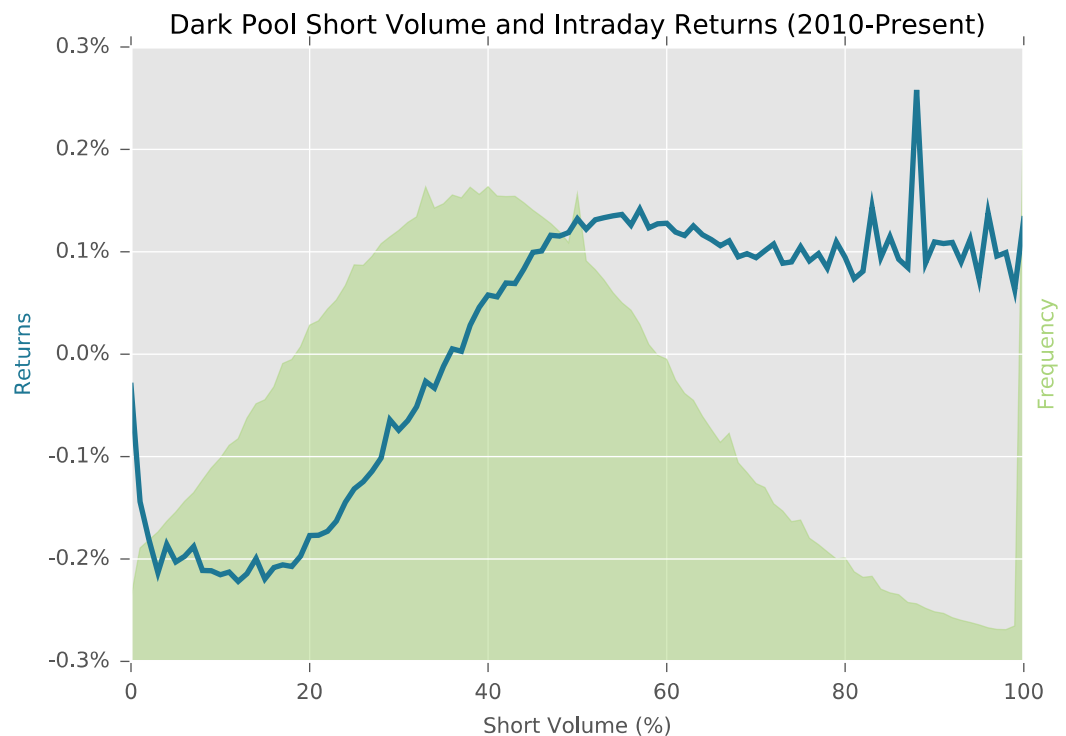
Our approach to test short volume is simple: First, find the average intraday return (open to close) of all exchange-listed U.S. stocks, preferreds, CEFs, ETFs, ETNs, etc., from 2010 to present. This is the baseline return. Then, break down those intraday returns according to the ratio of dark pool *short* volume to *long* (i.e., not short) volume on any given day. Naturally, we expect

to see that a higher short volume—representative of buying activity—is correlated to higher intraday returns. Here's what we found:

Across a universe of 11,254 securities, listed between 2010 and the present, we are able to sample a total of 12.74 million discrete days of returns. The mean intraday return of this data is **0.003%**—an average gain of three thousandths of a percent from open to close.

When we subsequently halve the data according to dark pool short volume composition, the lower quantile being 0–49% short and the upper being 50–100% short, we find that the mean intraday return of the lower quantile is **-0.0593%**; the upper, **0.1184%**.

While this is already supportive of our hypothesis, we can see much more detail when we average returns by percent, dividing the data into 101 parts, 0 to 100%. In the following figure, intraday returns are plotted by their respective dark pool short volume percentages (blue) alongside data frequency (green). The data is unfiltered and unedited.



We think that it is difficult to dispute that higher short volume correlates to greater intraday returns. As short volume rises above 35%, average intraday returns are uniformly positive. Similarly, below 35%, returns grow steadily more negative, and remain so.

Especially if we focus on that part of the figure where data is most abundant (and thus the section of data that is most relevant), starting around 20% and ending at 60%, we see a nearly linear relationship between short volume and intraday returns.

*\* A few large block orders originating in dark pools is likely the cause of lower-than-average (~40%) aggregate dark pool short volume.*

Outside of this band of data, we encounter what could mostly be characterized as “unusual” volume composition. Exceptionally *low* short volume will often be linked to days with large block trades between two parties with no market-making intermediary.\* Exceptionally *high* short volume will often be linked to either large buy orders facilitated by a broker or a high volume of speculative short sales.†

But, a few outliers aside, we expect that the figure above is able to speak sufficiently about the basic tendency of the data. For more information on the source and nature of the data, see the Appendix.

### Extrapolating

Naturally, we view this market phenomenon with more than academic interest, and we expect that readers are similarly interested in practical applications. To that end, adventurous souls can retrieve the short volume data used in this analysis directly from FINRA. A URL is provided in the appendix.

The data is, of course, provided without warranty, and requires some wrangling. For advice or assistance in replicating the data we've used here, message us at the email address in the footer of this page.

Our own research suggests that there is, at least, longer-term predictive value in short volume data when aggregated. The Dark Index™ (DIX™) has chronicled, in dollar-weighted terms, dark pool short volume across all components of the S&P 500® since 2011.‡ Because the index is dollar-weighted, volume is multiplied by share value, giving more weight to larger and more frequently traded stocks.

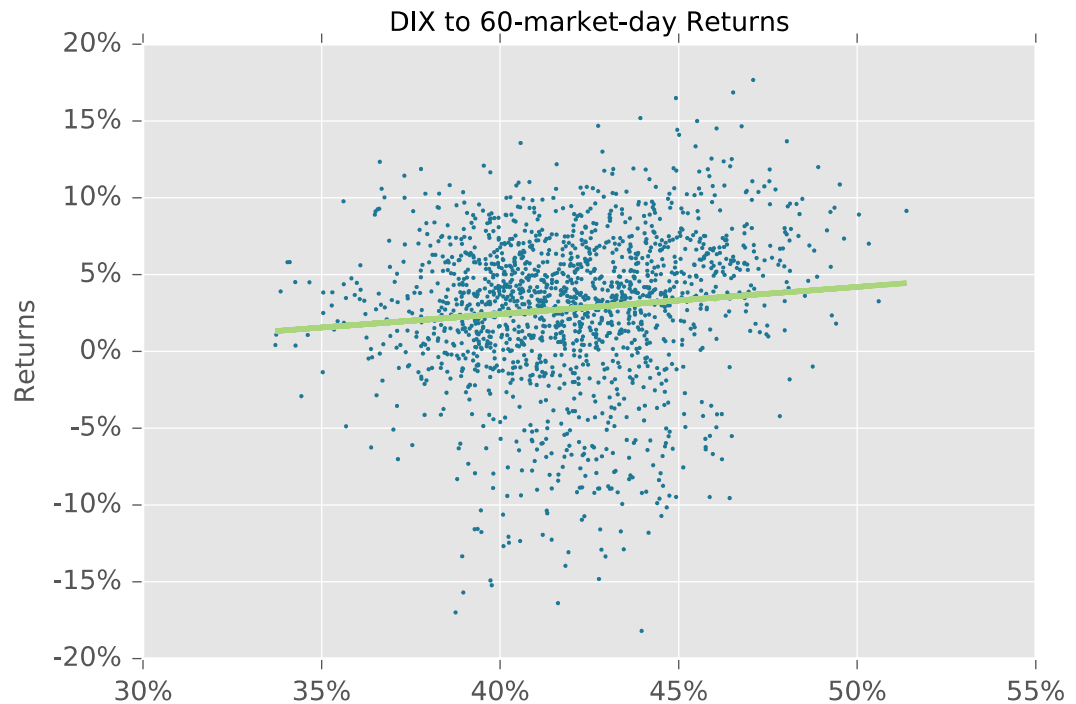
Very high relative percentages ( $\geq 45\%$ ) of dollar-weighted short volume are associated with mean 60-market-day returns of **5.3%**, as compared to a mean of **2.8%** across the whole dataset.

Since DIX tends to rise into corrections, we are left to believe that it reflects a broad willingness of investors to accumulate S&P 500 component stocks at lowered valuations, and that high levels of short volume correspond to positive medium-to-long term investor outlooks. This has certainly been reflected in the data.

The following scatter plot illustrates this relationship between DIX and S&P 500 returns.

*† There is a high incidence of data where 100% of volume is short. This is likely a broker-dealer shorting shares into a customer account as one leg of a transaction in a very illiquid stock.*

*‡ As of this writing, DIX data is available at <http://dix.sqzme.co>.*



It is unfortunate that we have less than 10 years of data to perform these tests, but the SEC has been reluctant to make more, or more detailed, data available. Among cost feasibility concerns, the SEC worries that providing too much data to the public may facilitate strategies disruptive to liquidity providers (MMs).\*

\* See “Short Sale Position and Transaction Reporting” (2015) in References.

Furthermore, they go on to say that:

Market participants, including issuers and investors, do not appear to widely monitor or use this data, and data vendors informed the Division [of Economic and Risk Analysis] that they had not created products utilizing this data. The Division is unaware of the transaction-level data being widely used by any group other than academics.†

† Ibid, pp. 19–20. More on “transaction-level data” in Appendix.

Given what we see as potentially valuable information, this may come as a surprise. How can this data have been overlooked? (Academics don't count.) Our suspicion is that the explanation is simple: The conceptual link between short sales and investor buying activity—the key to understanding what “short volume” really means—has never been clearly drawn before.

And so, armed with this fresh understanding, we hope that the reader might feel compelled to explore this new avenue of market data. And perhaps to be more receptive, generally, to the value of unconventional data and methods in the investment process.

## Appendix

The data used to test intraday short volume to stock returns is FINRA's "Reg SHO Daily Files." The most recent year of this data can be found at the following URL.

<http://regsho.finra.org/regsho-Index.html>

The Reg SHO Daily Files come from two separate TRFs, NYSE and NASDAQ. There is no meaningful distinction between the two (reporting is rather arbitrary), and they should be added together to generate total off-exchange short volume and total volume numbers. We ignore "short-exempt" volume for our analysis.

Even less known than the daily files are FINRA's "Monthly Short Sale Transaction Files," which provide detailed trade activity (including timestamps) of all off-exchange short sale transactions. The data can be found at the following URL.

<http://www.finra.org/industry/trf/trf-regulation-sho-2018>

A more thorough evaluation of short volume data would include data from lit exchanges. While the CBOE (formerly BATS) exchanges have made their short volume data freely available, NASDAQ and NYSE charge exorbitant fees. CBOE data can be found at the following URLs. These links include both the broader daily files as well as the detailed transaction files.

[https://markets.cboe.com/us/equities/market\\_statistics/short\\_sale/?mkt=bzx](https://markets.cboe.com/us/equities/market_statistics/short_sale/?mkt=bzx)

[https://markets.cboe.com/us/equities/market\\_statistics/short\\_sale/?mkt=byx](https://markets.cboe.com/us/equities/market_statistics/short_sale/?mkt=byx)

[https://markets.cboe.com/us/equities/market\\_statistics/short\\_sale/?mkt=edga](https://markets.cboe.com/us/equities/market_statistics/short_sale/?mkt=edga)

[https://markets.cboe.com/us/equities/market\\_statistics/short\\_sale/?mkt=edgx](https://markets.cboe.com/us/equities/market_statistics/short_sale/?mkt=edgx)

Of interest to would-be data-wranglers is that lit exchanges will tend to have fewer large institutional cross trades and thus higher short volume. Taken together, they will add up to around the SEC's aforementioned 49% figure.

For clarification regarding any of this data, reach out to our research team at [research@sqzme.co](mailto:research@sqzme.co).

## References

Bartlett, Robert P. and McCrary, Justin, Dark Trading at the Midpoint: Pricing Rules, Order Flow and High Frequency Liquidity Provision (June 10, 2015).

[https://eml.berkeley.edu/~jmccrary/Bartlett\\_McCrary\\_June2015.pdf](https://eml.berkeley.edu/~jmccrary/Bartlett_McCrary_June2015.pdf)

Cardella, Laura and Hao, Jia and Kalcheva, Ivalina, Liquidity-based Trading Fees and Exchange Volume (August 1, 2017). <https://ssrn.com/abstract=2149302> or <http://dx.doi.org/10.2139/ssrn.2149302>

SEC, Key Points About Regulation SHO (April 18, 2015).

<https://www.sec.gov/investor/pubs/regsho.htm>

SEC DERA, Short Sale Position and Transaction Reporting (June 15, 2014).

<https://www.sec.gov/files/short-sale-position-and-transaction-reporting%2C0.pdf>

Ye, Linlin, Understanding the Impacts of Dark Pools on Price Discovery (October 1, 2016).

<https://ssrn.com/abstract=2874957>

© SqueezeMetrics  
*All rights reserved.*