

May 9, 2023

Sitting on a Gold Mine: Information Leakage at the DTCC

A Themis Trading White Paper

By Joe Saluzzi and Scott Wilson

Introduction

Information leakage takes many forms in our day-to-day life. For example, you click “Agree to terms” when downloading a “free” app like TikTok and you know you are granting the developer the right to sell your data.

In our business, we have come to understand that the major stock exchanges sell our data in their proprietary feeds. These data feeds reveal information on an order-by-order basis and are used by sophisticated market participants to help predict short-term stock price movements.

But what most people don’t expect is that their daily trades, after they are sent to the Depository Trust & Clearing Corporation (DTCC) for matching, clearing and settlement, are then packaged into data products and sold by the DTCC.

In this paper, we will cover:

1. What is the DTCC?
2. Two of the data products that the DTCC is selling.
3. Sample DTCC data analysis.
4. What can be done with the data?
5. Potential plans the DTCC may have for enhancements to the data products.
6. Why does this matter and should it be allowed?

1. DTCC

Many people are not that familiar with the most important company responsible for the smooth functioning of the US stock market. We are talking about the DTCC which is “the largest depository in the world and the primary infrastructure in the U.S. for clearing, settlement and asset servicing.”¹ DTCC is comprised of three clearing agencies: National

¹ DTCC website <https://www.dtcc.com/data-services-kinetics>

Securities Clearing Corporation (NSCC), Fixed Income Clearing Corporation (FICC) and The Depository Trust Company (DTC).

DTCC and their subsidiaries are essentially a monopoly that provides the critical service of clearing and settlements to the markets. They are so important that they have been designated as Systemically Important Financial Market Utilities² by the U.S. government. This means that they are subject to enhanced standards for risk management, operation and governance. In other words, if something happened to DTCC, the financial markets would be in deep trouble.

DTCC is not a public corporation but rather it is privately held and owned by numerous users including banks, broker/dealers and investment managers. DTCC is overseen by a Board of Directors composed of 21 industry participants that include executives from some of the biggest US banks and brokers.³

2. DTCC Data Products

DTCC offers a suite of data products known as DTCC Kinetics⁴ which are robust data feeds derived from trading information processed by the DTC and NSCC. These products are marketed as a way for investors to gain “market insights at the security level.”⁵ While the products are aggregated, anonymized, delayed and require a minimum level of participation, they contain unique and very sensitive trading information. We are particularly concerned about the information that is being distributed from two of these products: DTCC Equity Kinetics and DTCC Investor Kinetics.

DTCC Equity Kinetics

Introduced in August 2018, DTCC Equity Kinetics⁶ provides aggregated and anonymized trading data on all U.S. equities trading volume. The data, which is sourced from NSCC, includes trading information by security and transaction type and lists “buy activity and three categories of sell activity—sale, short sale, and short sale exempt.” The data are for prior day’s trades for all brokers, including ATS’s, and are available before the start of each trading day. An intraday snapshot is also available to be received by 3pm ET for current day trades. The data also show the total volume for the 10 most active brokers on an anonymized basis. According to DTCC, “using this information, users can

² DTCC website <https://www.dtcc.com/about/managing-risk/sifmu>

³ DTCC website <https://www.dtcc.com/about/leadership>

⁴ DTCC website <https://www.dtcc.com/data-services-kinetics>

⁵ Ibid.

⁶ DTCC website <https://www.dtcc.com/data-services/market-aggregates-and-analytics/equity-kinetics>

determine where trading is concentrated for given securities and assess the depth of market participation.”⁷

Since they are sourcing their information direct from NSCC, DTCC Equity Kinetics is the only database that we know of which has a complete set of trading information. For example, stock exchanges provide historical information on short sale activity but can only provide this information for their own stock exchange. DTCC, on the other hand, provides market-wide short sale and short sale exempt volumes and even supplies this information on an intraday basis.

DTCC Investor Kinetics

Introduced in February 2020, DTCC Investor Kinetics⁸ contains information about institutional trades matched with their brokers each day, and is delivered on a T+3 basis. According to DTCC:

“DTCC Investor Kinetics delivers current and historical time series information. The data includes volume, volume-weighted average price (VWAP), number of trades, number of unique investment managers, and concentration. The data is segmented by investor type, such as traditional asset managers, hedge funds and retail wealth managers. DTCC Investor Kinetics is a powerful source of historical data for enhancing users’ research and analytical capabilities. The service can be used for the back testing of models and to identify price and volume trends by investor type for alpha generation, risk modeling, and other purposes.”

DTCC is gathering the data for Investor Kinetics from DTC’s CTM automated trade-matching service and it is estimated that CTM represents approximately 70% of the institutional buy-side trade activity worldwide.⁹ DTCC segments this data into three categories that describe the nature of the institutions: traditional asset managers, hedge funds and retail wealth managers.

⁷ DTCC Kinetics Services presentation, 6/10/21, slide 6

https://communications.dtcc.com/rs/669-QIL-921/images/DTCC_Kinetics_Event_JUN21.pdf

⁸ DTCC Website <https://www.dtcc.com/data-services/market-aggregates-and-analytics/investor-kinetics>

⁹ DTCC Kinetics Services presentation, 6/10/21, slide 16

https://communications.dtcc.com/rs/669-QIL-921/images/DTCC_Kinetics_Event_JUN21.pdf

DTCC knows that this data is very powerful and has marketed¹⁰ it as a way to generate alpha:

It Can be Powerful for Quant Traders and Hedge Fund Managers

DTCC Investor Kinetics can be used for alpha generation, risk modelling, and more. The data can assist investors in understanding and identifying pricing and volume trends across key attributes.

The DTCC data could also be used with other data sources like stock exchange order-by-order proprietary data feeds to create an even clearer picture of what market participants are doing. DTCC acknowledged this possibility in their marketing piece:¹¹

It Can be Part of a Quant's Larger "Mosaic"

Quant users often create complicated algorithms made from many different data sources. DTCC Investor Kinetics can be one key piece of that larger mosaic. It's a service that can inform and enhance a user's larger overall strategy.

3. Sample DTCC Data Analysis

DTCC offers a one-day sample of the Investor Kinetics and Equity Kinetics data. A brief review of this sample revealed the potential predictive power of the data.

Since Investor Kinetics segments the data into different market participant categories, it is possible to determine the type of investors that are active in a security on a given day and over time.

- For example, suppose a stock traded 500,000 shares on a particular day. According to data from Investor Kinetics, it is possible to ascertain that four buyers were responsible for a total of 200,000 shares. The data also might show that all four of these participants were labeled "Hedge Fund-Alternative Managers" (HFA). This would mean that four HFA's were responsible for 40% of the day's volume.
- Another example, if a stock traded 700,000 shares on a particular day, the Investor Kinetics data might show that five market participants were responsible for 300,000 shares that were sold that day. The data might show that all five of these participants were labeled "Hedge Fund-Alternative Managers." Thus, 43%

¹⁰ A New Data Source for Quants by DTCC Connection Staff, 2/10/20, <https://www.dtcc.com/dtcc-connection/articles/2020/february/10/a-new-data-source-for-quants>

¹¹ Ibid.

of the sales were from a small group of HFA's. Although the data does not indicate if the sales were long or short, DTCC's broker-level product, Equity Kinetics, does offer long vs short sale detail on a security level basis. Combining the two data sets could indicate an increased probability that the institutional sellers were shorting the stock.

4. What can be done with the data?

The unique data supplied by DTCC Data Services can be used to create alpha-generating trading signals. One firm, CloudQuant, which is a partner of DTCC, evaluated the Kinetics datasets. They published two papers titled "Evaluation of Investor Kinetics"¹² and "Evaluation of Equity Kinetics"¹³ where they developed signals based on the data that they received from DTCC. The results were astonishing.

Equity Kinetics – CloudQuant claims that signals they developed from this data produced "an impressive dollar neutral portfolio return of 17.37% per annum." They claim the signals they produced were "novel and valuable" and that their analysis "highlights significant trades by highly informed institutional investors."

Investor Kinetics - CloudQuant identified two novel signals from the Investor Kinetics data: fund flow signals and concentration signals. CloudQuant claims these signals that they developed from the DTCC data "produced a dollar neutral portfolio return of 35.68% per annum" which "outperformed the S&P 500 by an average of 21.0% per annum over 10 years." If true, this is a remarkable outperformance and indicates that the DTCC data is truly unique and valuable.

CloudQuant boasted that the DTCC data is literally a gold mine: "The benefit of flow data is clear and it provides insight into short-term market momentum. It may also be a great catalyst risk monitoring tool by highlighting large institutional position movements. All that is left to answer is how to dig into this gold mine and how well has it performed."¹⁴

¹² How CloudQuant Researchers Derived Signals Out Of DTCC Data Services' Kinetics Datasets <https://www.cloudquant.com/how-cloudquant-researchers-derived-signals-out-of-dtcc-data-services-kinetics-datasets/>

¹³ Ibid.

¹⁴ Ibid.

5. Potential plans the DTCC may have for enhancements to the data products

We have already described in this paper an egregious lack of awareness by the DTCC concerning the sensitivity of investor's data. Unfortunately, according to a recent conference presentation¹⁵, it appears that DTCC may have plans to:

- Distribute even more granular data.
- Shorten the delivery time to T+1.
- Refine the investor segmentation categories that the Investor Kinetics product uses.

According to the DTCC executive speaking at the conference:

"We're going to create new sub-aggregations that show the data in a very different way...these are going to be activity-based measures now. How diverse is someone's portfolio? If they trade lots of securities, let's group them together, if they trade only a few, let's put them in a separate bucket. How much turnover do they have in their portfolio?"

Clearly, the DTCC doesn't seem too concerned about revealing investors' trading data. We wonder what other products they might have planned for mining the value of your trading flows? The DTCC executive indicated that more enhancements could be in the works and said at the conference:

"It's a new pivot on other data. The good news is we can't run out these pivots. Math is pretty unlimited."

6. Why does this matter and should it be allowed?

Institutional investors spend enormous sums of time and money to generate alpha for their clients. It's their business and their competitive advantage. Why would they allow the central clearing agency to sell data about ongoing trade implementation to those seeking to profit from that alpha? Even if the information is aggregated and anonymized, the CloudQuant analysis producing a 35% annual return clearly demonstrates the signals that can be generated are of tremendous value.

Essentially, DTCC is using institutions' trading activity to help those who purchase the data feeds gain a competitive advantage. Should DTCC be allowed to distribute this

¹⁵ BattleFin Discovery Days Miami, DTCC Kinetics In Action, 1/27/22, <https://vimeo.com/683625095?s=09>

information? Under Section 17A of the Exchange Act¹⁶, the SEC is responsible for the supervision of clearing agencies which include the DTCC. Is the DTCC required to file their plans with the SEC for the Kinetics data products? We searched the Federal Register and could not find any proposals or filings related to DTCC Equity Kinetics or DTCC Investor Kinetics. Considering that stock exchanges need to file with the SEC when they launch new products, we would think that the DTCC would also need to alert the SEC about new data products.

Finally, are clients of DTCC aware that their data is being packaged and sold as part of a product that is arguably designed to compromise the alpha they work so hard to generate for their clients? Is there an agreement in place that states DTCC has a right to do whatever they like with the data? We would imagine that the DTCC lawyers checked this box and are comfortable with the legality of their product. Regardless, doesn't the central clearing authority have a bigger obligation to protect the data of their clients? Is the DTCC choosing to increase their revenue at the expense of leaking valuable client trade information?

Conclusion

Many institutional investors are unaware that the private trading information that they entrust to DTCC is being packaged and sold as part of an aggregated data product. Up until a few years ago, this data was kept inside the DTCC and not available for public consumption. But it appears that DTCC realized that they were sitting on a gold mine of information and decided to profit from it. Analysis of the DTCC 2022 and 2021 financial statements¹⁷ shows that "Data and Other Services" revenue has been growing at a faster rate than overall revenue (15% vs 5% in 2022, 20% vs 9% in 2021).

We would still be outraged if this was the end of the story. But what happens next? DTCC has indicated that they may further parse the data so that the consumers of the data feed can learn even more about market participant activities and potentially, ongoing intentions.

DTCC seems to think that they own the data and can do whatever they like with it. They might be thinking that they are providing the data to help with valuation, investment decisions and risk management. They might be thinking that the data they are selling is benign and only helps with high-level market direction calls. But what they are actually

¹⁶ SEC Final Rule: Clearing Agency Standards

<https://www.sec.gov/rules/final/2012/34-68080.pdf>

¹⁷ The Depository Trust & Clearing Corporation Consolidated Financial Statements for the years end December 31, 2022 and 2021 <https://www.dtcc.com/-/media/Files/Downloads/legal/financials/2023/DTCC-Annual-Financial-Statements-2022-and-2021.pdf>

doing is leaking sensitive trading data. DTCC stands for the Depository Trust & Clearing Corporation. By creating products like Equity and Investor Kinetics, the DTCC is violating the “Trust” in their name.

We believe that DTCC should immediately cease distribution of DTCC Equity and Investor Kinetics. If they do not to choose to do so voluntarily, then the SEC should enforce their supervisory authority under Section 17a of the Exchange Act and require the DTCC to eliminate these products.