SYNTHETIC FINANCIAL DATA:
AN APPLICATION TO REGULATORY COMPLIANCE FOR BROKER-DEALERS

8 November 2019

Jan Hendrik Witte*
University College London (UCL)

*joint work with Basile Despond and J B Heaton

Part 1: Big Data

In the age of “big data”, those in the investment management industry have a “small data” problem.

In the age of “big data”, those in the investment management industry have a “small data” problem.

- Companies like Walmart, Amazon, PayPal, Facebook, and Google collect petabytes (one petabyte equals a million gigabytes) of data every hour.

- Financial market participants often use the small data of financial markets to generate and test investment strategies.

In the age of “big data”, those in the investment management industry have a “small data” problem.

- Companies like Walmart, Amazon, PayPal, Facebook, and Google collect petabytes (one petabyte equals a million gigabytes) of data every hour.

- Financial market participants often use the small data of financial markets to generate and test investment strategies.

This comes with two major problems.

In the age of “big data”, those in the investment management industry have a “small data” problem.

- Companies like Walmart, Amazon, PayPal, Facebook, and Google collect petabytes (one petabyte equals a million gigabytes) of data every hour.

- Financial market participants often use the small data of financial markets to generate and test investment strategies.

This comes with two major problems.

- Historical data from an earlier time may tell us little or nothing about future prices and returns.

- Inferences about the profitability of investment strategies may be sensitive to a handful of outliers.

Example 1:

Example 1:

- For decades, investment advisers and broker-dealers have assumed that the historical premium of equities over risk-free securities implied that stocks are a generally superior investment strategy for the “long term.”

- But the implication is highly fragile.

Example 1:

- For decades, investment advisers and broker-dealers have assumed that the historical premium of equities over risk-free securities implied that stocks are a generally superior investment strategy for the “long term.”

- But the implication is highly fragile.

In a pathbreaking work:

Example 1:

- For decades, investment advisers and broker-dealers have assumed that the historical premium of equities over risk-free securities implied that stocks are a generally superior investment strategy for the “long term.”

- But the implication is highly fragile.

In a pathbreaking work:

- Bessembinder (2018) finds that the majority of U.S. listed common stocks have returned (inclusive of dividends) less than the risk-free rate (that is, the one-month Treasury bill) over their lives as listed companies.

- Just 4% of listed U.S. companies account for all of the gains of the U.S. stock market from 1926 to 2016.

Example 2:

Example 2:

- Dogs generally look the same over a period of decades (even allowing for new hybrid breeds).

Example 2:

- Dogs generally look the same over a period of decades (even allowing for new hybrid breeds).

- This allows for successful image recognition algorithms.

Example 2:

- Dogs generally look the same over a period of decades (even allowing for new hybrid breeds).

- This allows for successful image recognition algorithms.

Financial markets present a far different problem.

Example 2:

- Dogs generally look the same over a period of decades (even allowing for new hybrid breeds).

- This allows for successful image recognition algorithms.

Financial markets present a far different problem.

- We know very little about the mechanisms that generate prices.

- Financial-market data is likely to be generated by mechanisms (interactions of traders using information) that are not stable through time.

- Even if a researcher finds a good model of price behavior in a particular period of time, there is little reason to believe that prices will behave today as they did 10 or 20, or even 5 years, ago.

Part 2: Financial Regulation

The New Compliance Risk

The New Compliance Risk

- In June 2019, the U.S. Securities and Exchange Commission (SEC) adopted Regulation Best Interest (RBI). The regulation requires broker-dealers to exercise reasonable diligence, care, and skill in making a recommendation to a retail customer. This is known as the “Care Obligation.”

The New Compliance Risk

- In June 2019, the U.S. Securities and Exchange Commission (SEC) adopted Regulation Best Interest (RBI). The regulation requires broker-dealers to exercise reasonable diligence, care, and skill in making a recommendation to a retail customer. This is known as the “Care Obligation.”

The SEC’s Final Rule states:

The New Compliance Risk

- In June 2019, the U.S. Securities and Exchange Commission (SEC) adopted Regulation Best Interest (RBI). The regulation requires broker-dealers to exercise reasonable diligence, care, and skill in making a recommendation to a retail customer. This is known as the "Care Obligation."

The SEC's Final Rule states:

- "Whether a broker-dealer's recommendation satisfies the Care Obligation will be an objective evaluation turning on the facts and circumstances of the particular recommendation and the particular retail customer."

- The care obligation requires that a broker-dealer understands "potential risks, rewards, and costs associated with the recommendation."

- "[Bad intent] will not be required to establish a violation of Regulation Best Interest."

How to reduce the risk of litigation and regulatory actions?

How to reduce the risk of litigation and regulatory actions?

- Given the known limitations of historical data, how can a broker or fiduciary gain confidence that an investment strategy **will not result in future regulatory action or litigation?**

- **What work would a broker-dealer or fiduciary want to show was done** to support its recommendations and actions if accused of basing advice on bad inferences from historical data?

Part 3: Possible Solutions

A Synthetic Data Approach

A Synthetic Data Approach

- We use a **fraud-detection approach to identify high-dimensional outliers** in the historical dataset.

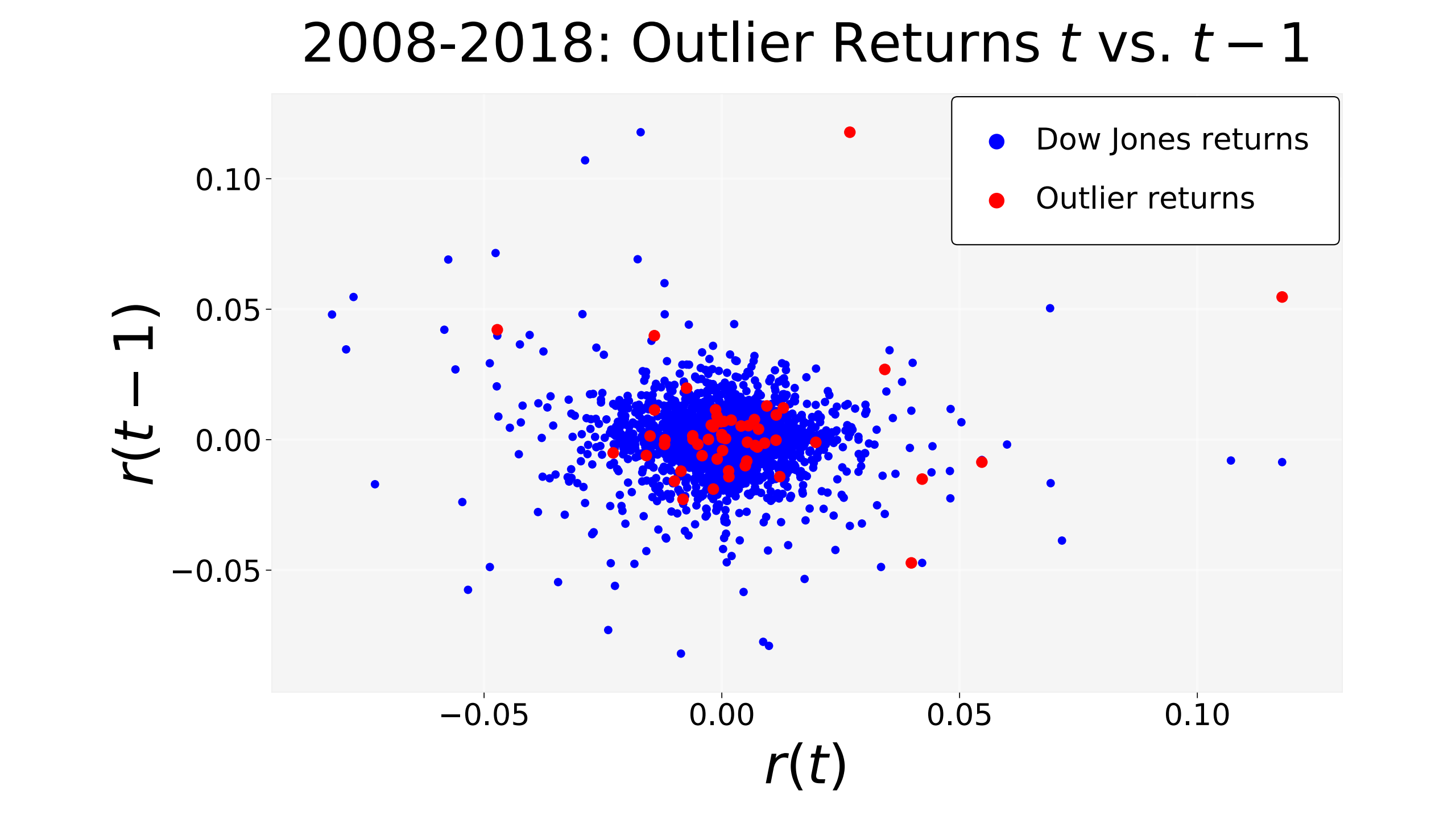
- We replace the outliers with a larger alternative dataset that reflects the different ways in which the joint prices might alternatively have been realized in the past.

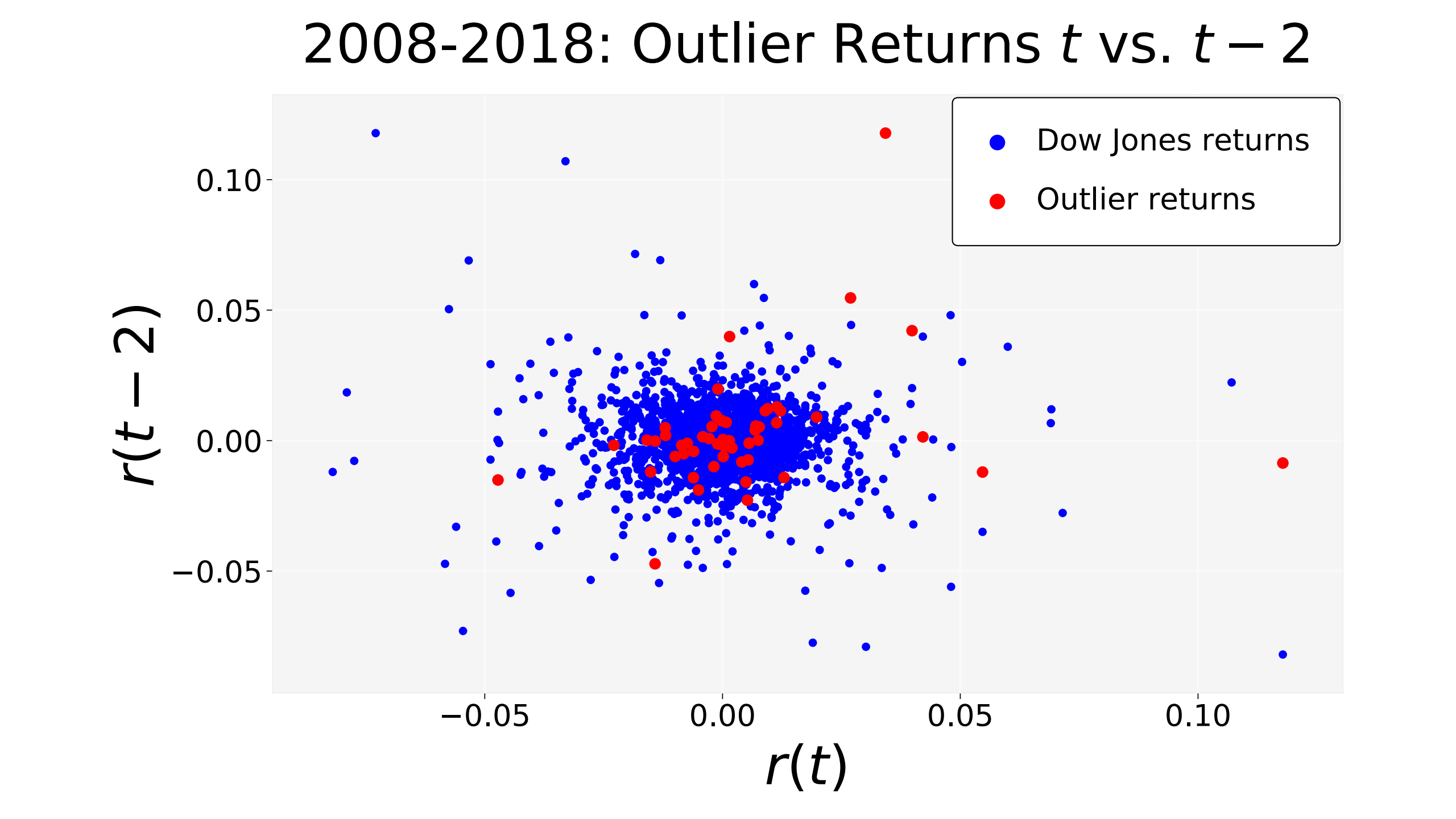
- The resulting synthetic datasets have **little to no dependence on historical outliers** while maintaining all other characteristics with a high degree of accuracy.

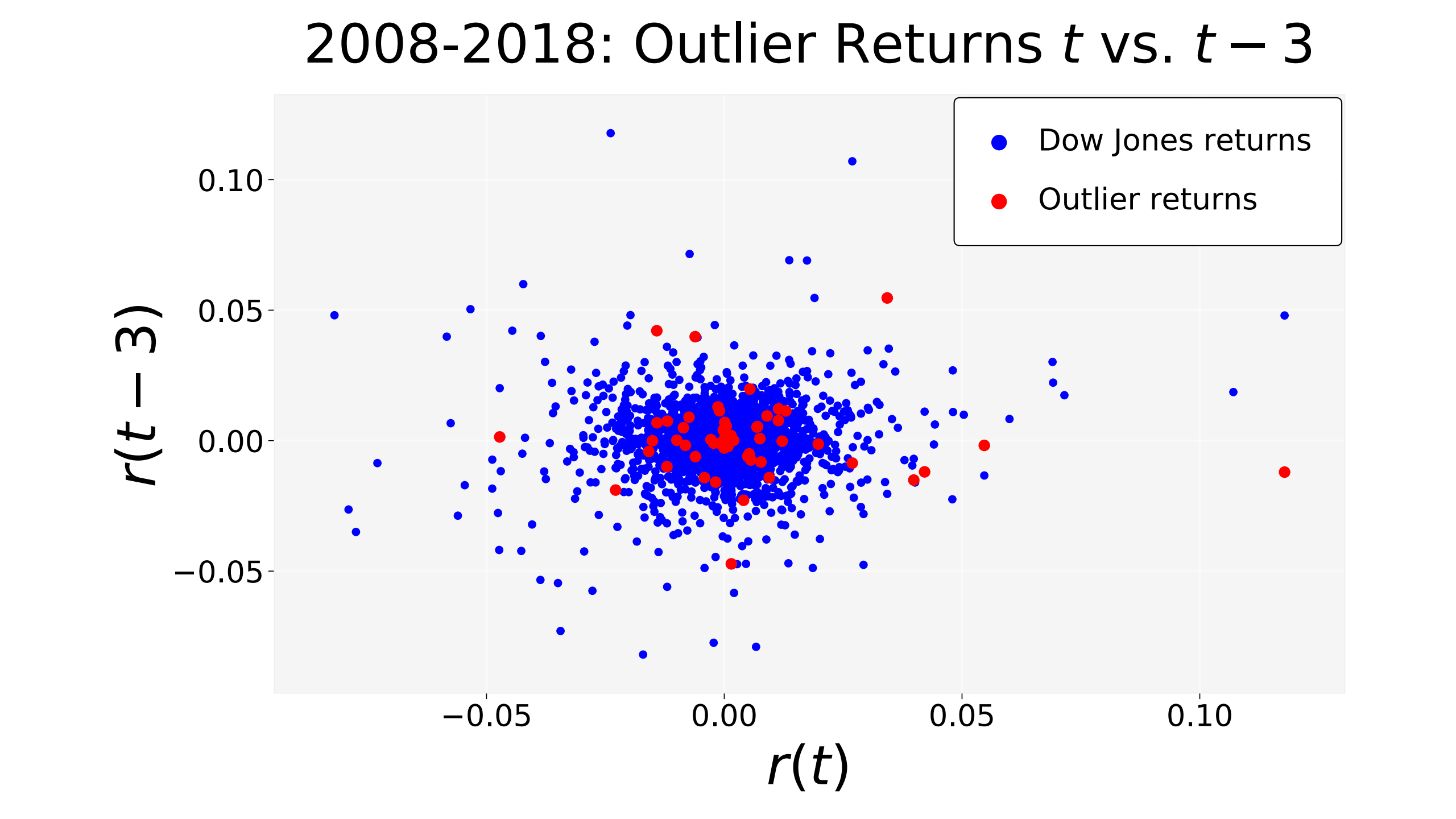
Example:

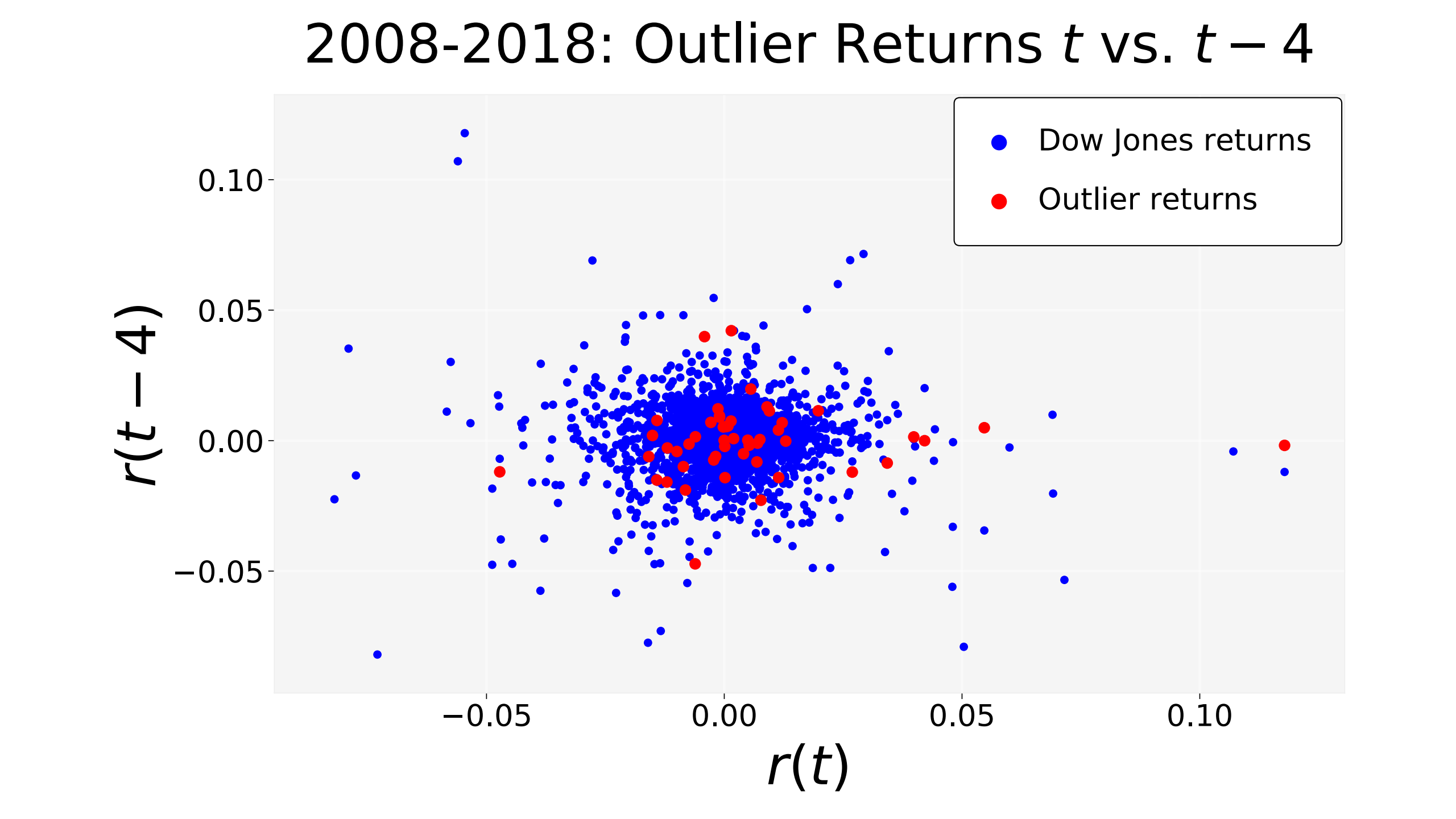
Example:

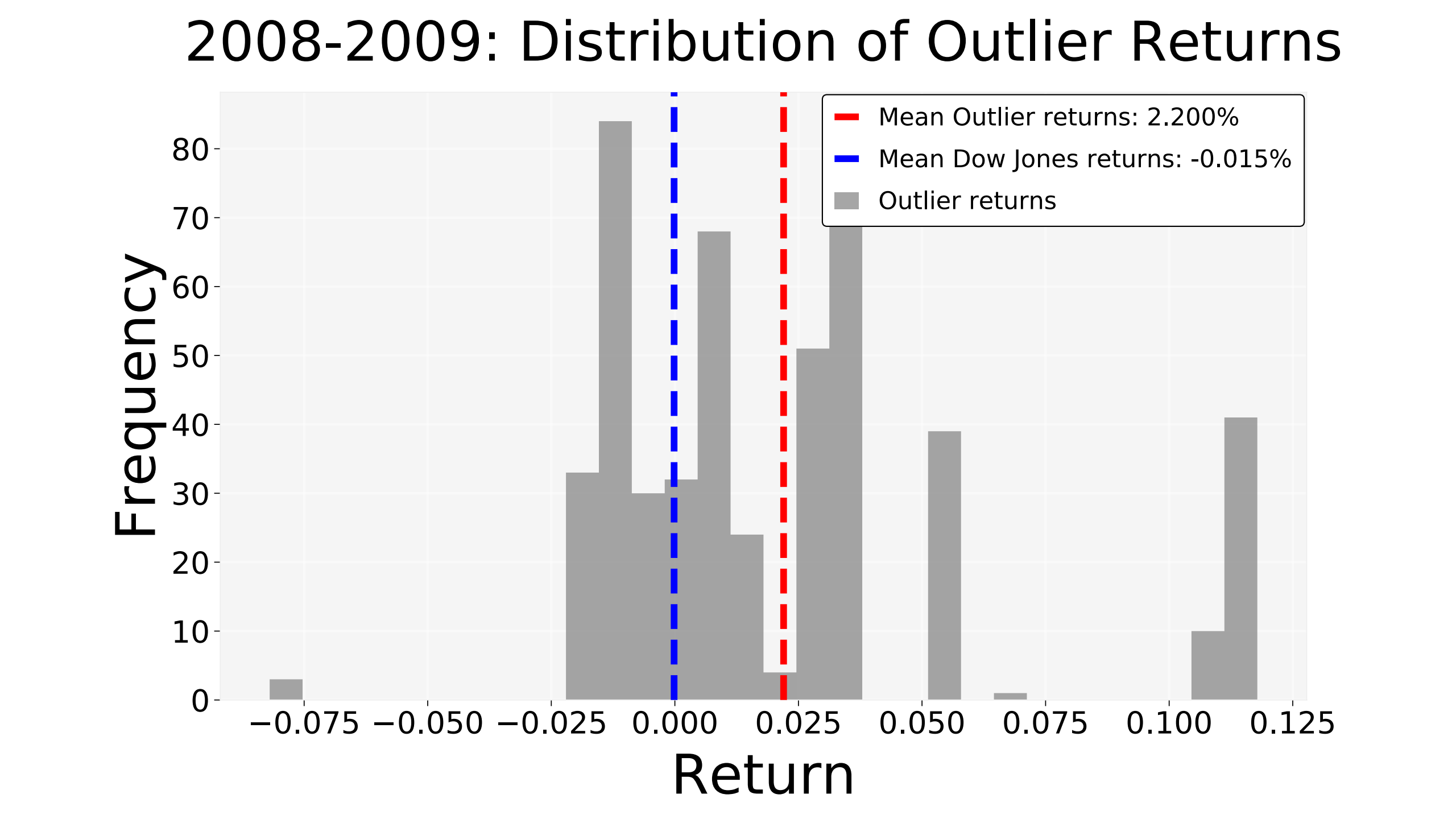
- Consider daily closing prices for Dow Jones Industrial Average (DJIA) from 2 January 2008 to 22 May 2019.


2008-2018: Outlier Returns t vs. $t-1$
Dow Jones returns
Outlier returns
0.10
0.05
0.00
−0.05
$r(t-1)$
−0.05
0.00
0.05
0.10
$r(t)$

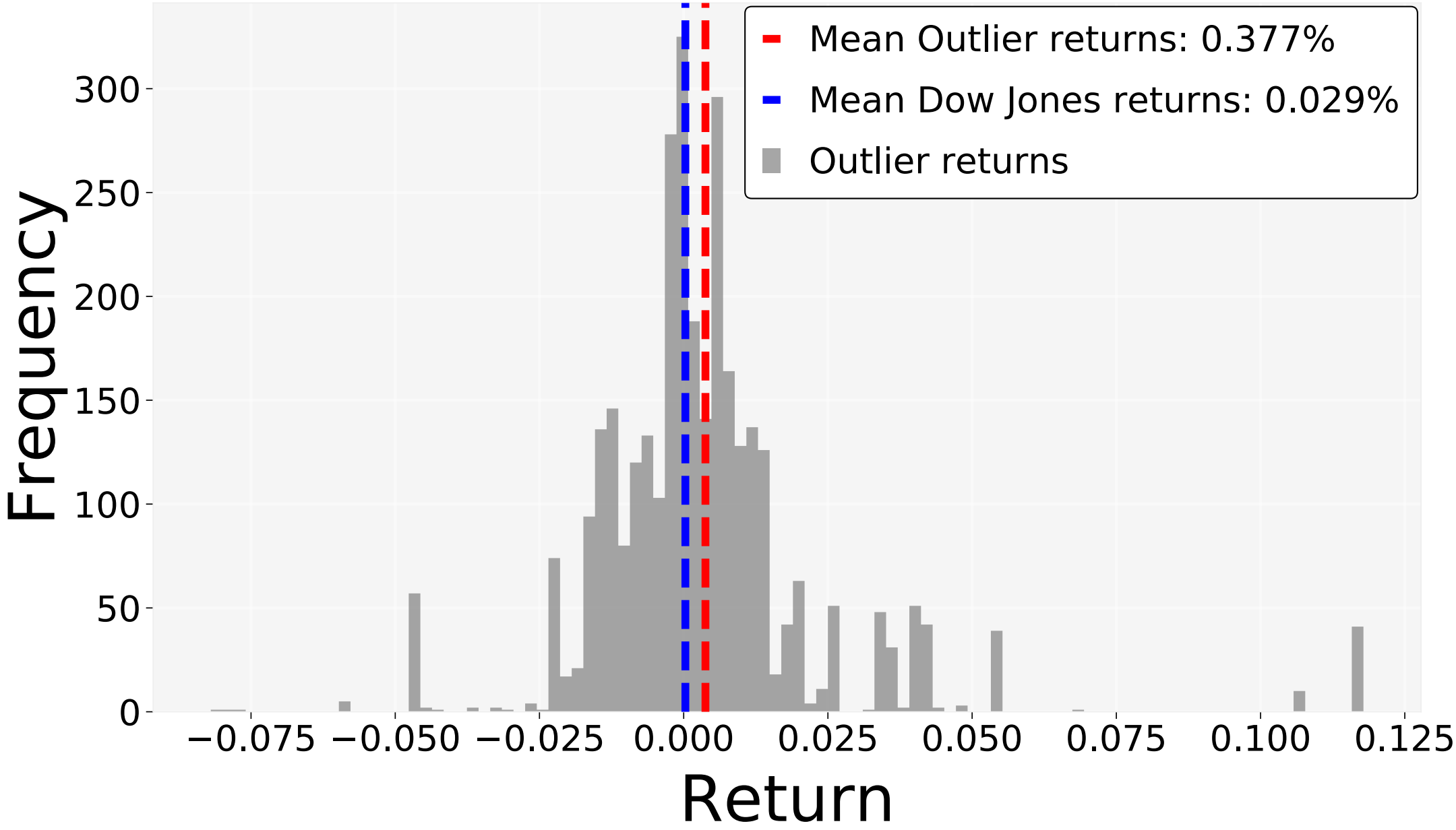

2008-2018: Outlier Returns t vs. $t-2$
Dow Jones returns
Outlier returns
0.10
0.05
0.00
−0.05
$r(t-2)$
−0.05
0.00
0.05
0.10
$r(t)$

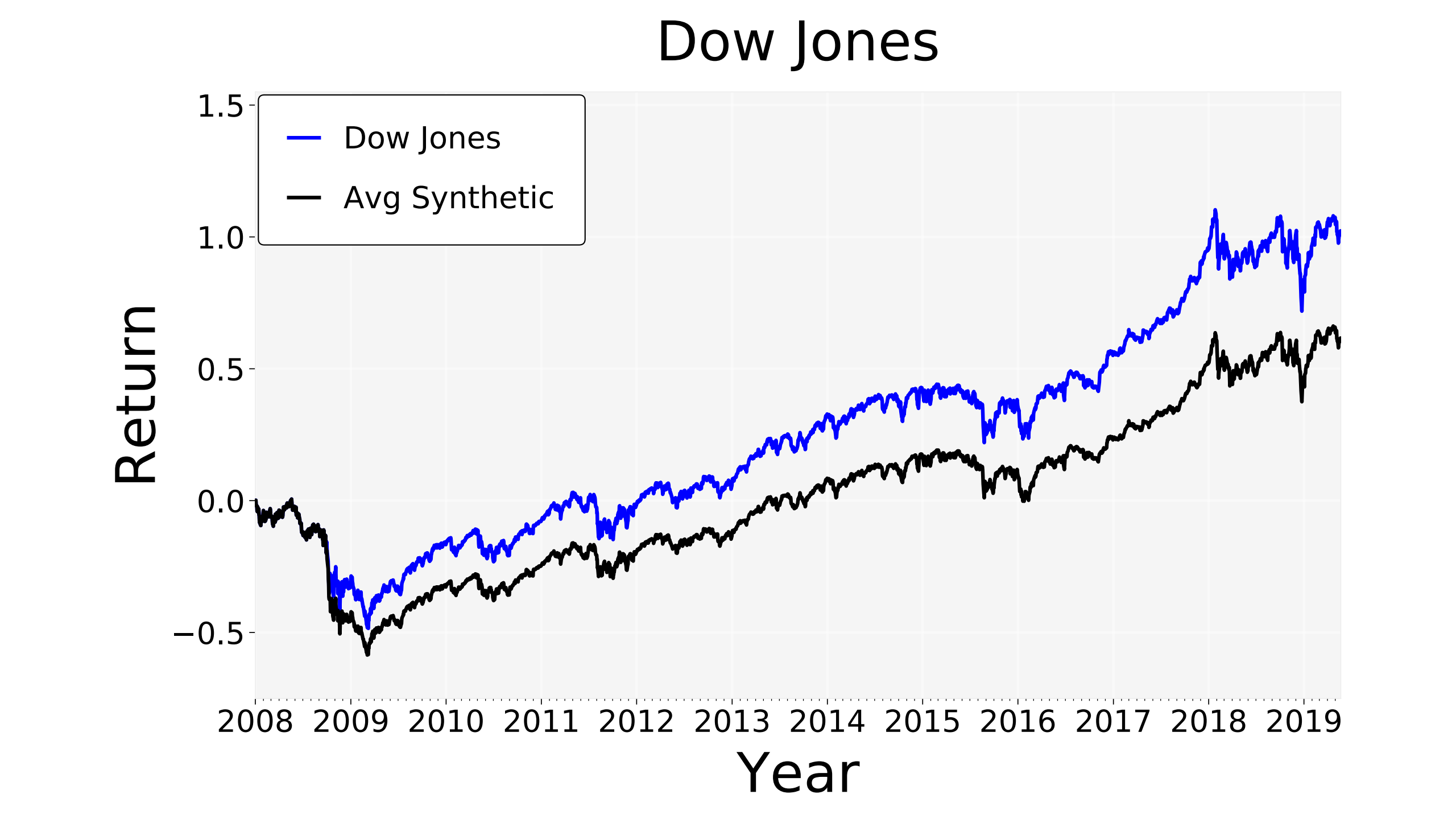

2008-2018: Outlier Returns t vs. $t-3$
Dow Jones returns
Outlier returns
0.10
0.05
0.00
−0.05
$r(t-3)$
−0.05
0.00
0.05
0.10
$r(t)$

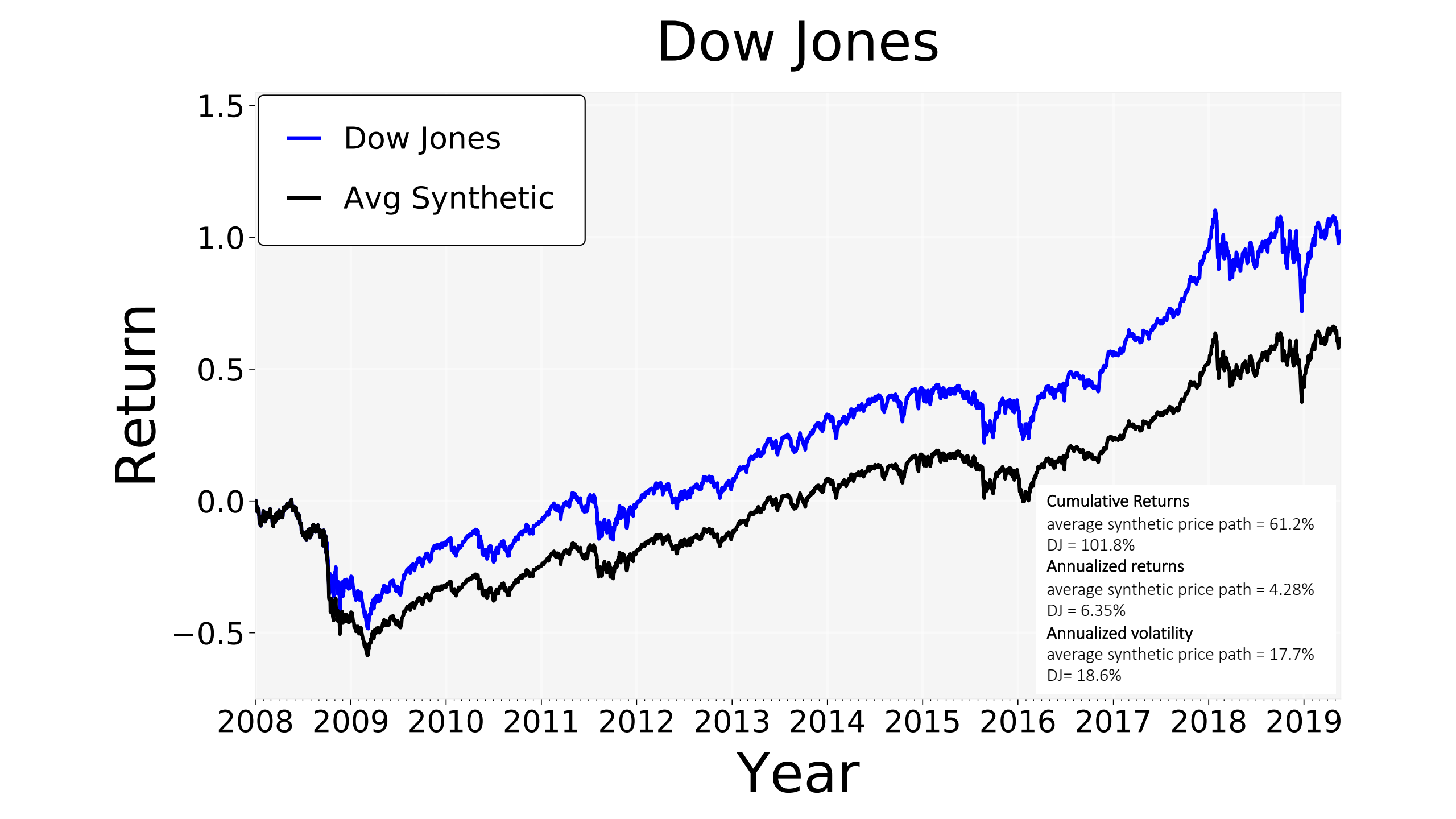

2008-2018: Outlier Returns t vs. $t-4$
Dow Jones returns
Outlier returns
0.10
0.05
0.00
−0.05
$r(t-4)$
−0.05
0.00
0.05
0.10
$r(t)$

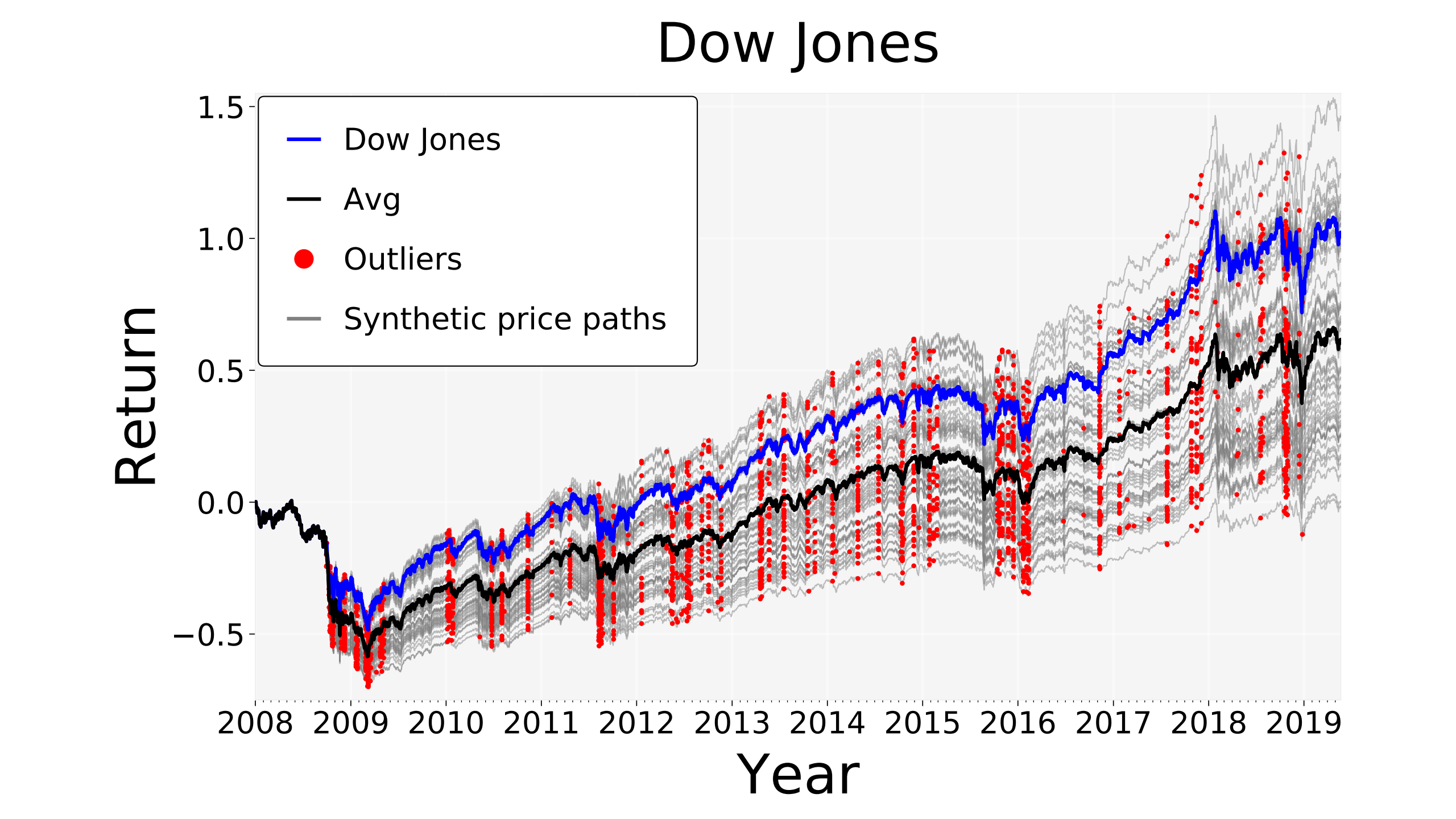

2008-2009: Distribution of Outlier Returns
Mean Outlier returns: 2.200%
Mean Dow Jones returns: -0.015%
Outlier returns
Frequency
0
10
20
30
40
50
60
70
80
−0.075
−0.050
−0.025
0.000
0.025
0.050
0.075
0.100
0.125
Return

2008-2018: Distribution of Outlier Returns




Dow Jones
Dow Jones
Avg Synthetic
1.5
1.0
0.5
0.0
−0.5
Return
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Year


Dow Jones
Dow Jones
Avg Synthetic
1.5
1.0
0.5
0.0
−0.5
Return
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Year
Cumulative Returns
average synthetic price path = 61.2%
DJ = 101.8%
Annualized returns
average synthetic price path = 4.28%
DJ = 6.35%
Annualized volatility
average synthetic price path = 17.7%
DJ= 18.6%


Dow Jones
Dow Jones
Avg
Outliers
Synthetic price paths
Return
1.5
1.0
0.5
0.0
−0.5
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Year

Part 4: Summary

Summary:

Summary:

- Finance has a small data problem.

Summary:

- Finance has a small data problem.
- Regulation Best Interest means broker-dealers have a Care Obligation. Bad intent is not required for regulatory violation.

Summary:

- Finance has a small data problem.
- Regulation Best Interest means broker-dealers have a Care Obligation. Bad intent is not required for regulatory violation.
- Working with historical data is misleading.

Summary:

- Finance has a small data problem.
- Regulation Best Interest means broker-dealers have a Care Obligation. Bad intent is not required for regulatory violation.
- Working with historical data is misleading.
- Synthetic data is one way of meeting the Care Obligation.