

# A Systematic Trading Approach from Data Mining to Live Deployment

Documentation

Submitted by:

**Juri Stoffers**

Supervised by:

**Dr. Geoffrey Ostrin**

August 16, 2025

## **Abstract**

This document documents the mathematical and statistical background of the research and testing I conducted for my strategy development. Explination of the formulars I used and the different versions of my algorithms.

# Contents

<b>1</b>	<b>Outlier Detection Using Mean and Standard Deviation (Z-Score Based Outlier Detection)</b>	<b>4</b>
1.1	Orderbook Delta defenition . . . . .	4
1.2	Normal Range . . . . .	4
1.3	Outlier Condition . . . . .	5
1.4	My Hypothesis . . . . .	5
1.5	Bullish and Bearish Outliers version 1 . . . . .	5
1.5.1	. . . . .	5
1.5.2	. . . . .	5
1.6	Idea behind . . . . .	6
1.7	Future Plans . . . . .	6
<b>2</b>	<b>Measuring Volatility After Outlier Detection</b>	<b>7</b>
2.1	Dictionary of Terms . . . . .	7
2.2	In Python code . . . . .	7
2.3	Statistical evidence and results . . . . .	8
<b>3</b>	<b>Optimising for best Z-Score threshold for outliers</b>	<b>9</b>
<b>4</b>	<b>Measuring avearge return after price outlier detection</b>	<b>10</b>
4.1	Formulars . . . . .	10
4.2	Dictionary of Terms . . . . .	10
<b>5</b>	<b>Underlying strategy Bias</b>	<b>11</b>
5.0.1	Trend identification version 1 . . . . .	11
5.0.2	Trend identification version <i>v2</i> . . . . .	12
<b>6</b>	<b>Finding an edge</b>	<b>13</b>
6.1	Fees . . . . .	13
6.2	No Fees results . . . . .	13
6.3	With Fees results . . . . .	14
6.4	No clear path . . . . .	14
<b>7</b>	<b>Combining Indicators</b>	<b>15</b>
<b>8</b>	<b>From outliers to strategy</b>	<b>16</b>
8.1	Progress description . . . . .	16
8.2	Where did I start my research . . . . .	16
8.3	Finding an edge . . . . .	17
8.4	Monte Carlo Testing . . . . .	18
8.5	Backtesting framework . . . . .	20

<b>9</b>	<b>Developing a strategy</b>	<b>22</b>
9.1	Initial thoughts . . . . .	22
9.2	Reverse strategy . . . . .	22
9.3	First strategy test . . . . .	22
9.3.1	Code . . . . .	22
9.3.2	Strategy 1 results . . . . .	23
9.4	Learning from the first strategy . . . . .	23
9.5	Explaining figure . . . . .	24
9.6	Strategy 2 . . . . .	24
<b>10</b>	<b>Algorithms Implementation</b>	<b>25</b>
10.1	Outlier detection version 2 . . . . .	25
<b>11</b>	<b>Algorithmic infrastructure implementation</b>	<b>27</b>
11.1	Handeling trades asynchronously . . . . .	28
11.2	Strategy 3 . . . . .	29
<b>12</b>	<b>sources</b>	<b>30</b>
12.1	How good or random is your trading . . . . .	30
12.2	Tweets . . . . .	30
12.3	TRDR This platform allow you to use different kind of metrics on different Timeseries datasets (BTC/USD Price, orderbookdelta and Open Interest	30
12.4	Trend line automation . . . . .	30
12.5	Vectorbt github documentation respository. . . . .	30
12.6	Deep research by ChatGPT . . . . .	30
12.7	Markets in profile – James Dalton . . . . .	30
12.8	Do retail traders stande a chance? . . . . .	30

# 1 Outlier Detection Using Mean and Standard Deviation (Z-Score Based Outlier Detection)

## 1.1 Orderbook Delta definition

The orderbook is a real-time electronic list of all pending buy (bid) and sell (ask) orders for a specific asset, organized by price level. It represents the current market depth and shows:

- Bids: The prices at which buyers are willing to purchase an asset
- Asks: The prices at which sellers are willing to sell their asset
- The difference between the highest bid and lowest ask creates what we call the "spread"

Each price level in the orderbook shows:

- The price of the order
- The total volume (quantity) of orders at that price
- The number of individual orders at that price (on some exchanges)

The **orderbook delta** is calculate by the difference between the sum of the bid and ask orders at a certain depth. Formular:

$$\Delta_{x\%} = \sum_{i=1}^{x\%} \Delta_i$$

- $\Delta_{x\%}$  is the sum of the orderbook delta for the last  $x\%$  of the orderbook.
- $\Delta_i$  is the orderbook delta for the  $i$ -th level of the orderbook.

## 1.2 Normal Range

What I want to test is how price reacts to anomalous orderbook delta movements, particularly in scenarios where unrealistic or clearly outlying values are detected. In cryptocurrency markets, such inefficiencies can be caused by various events, one example is liquidation <sup>1</sup> events that interact with passive demand order stacked zones <sup>2</sup> During these events, the orderbook delta exhibits significant increases, providing a clear signal of market stress. This research will focus on understanding the relationship between rapid delta movements and how price reacts after these events.

---

<sup>1</sup>A liquidation event in cryptot is when traders who borrowed money from an exchange to open a so called leveraged position are forced to sell or buy at a loss.

<sup>2</sup>A passive demand order stacked zone is a zone where a lot of orders are stacked at a certain price level.

### 1.3 Outlier Condition

A outlier is defined when a the Orderbook  $\Delta_t$  is outside of the normal range. I define the normal range as the mean  $\mu(\Delta)$  plus or minus 2 times the standard deviation  $\sigma(\Delta)$ . For mean and standard deviation I use a rolling window of 1440  $\Delta$  values. On a timeseries dataset with a interval of 1 minute that is equal to one day.

Equation for outlier defenition:

$$\Delta < \mu(\Delta) - 2\sigma(\Delta) \quad \text{or} \quad \Delta > \mu(\Delta) + 2\sigma(\Delta) \quad (1)$$

### 1.4 My Hypothesis

- I expext realized volatility to increase after an outlier is detected.
- I expect that I can get a directional bias based on outliers if we combine the outlier signal with a underlying bias.

### 1.5 Bullish and Bearish Outliers version 1

What does bullish and bearish even mean? A bullish signal is a signal where we anticipate price will go up. A bearish signal is a signal where think price will go down.

Now I defined a a bullish outlier as bullish if he has a certain z-score value and a bearish outlier as a bearish if he has a certain z-score value.

The mean is calculated is calculate by making a rolling window of the last 1440  $\Delta$  values. (Basically a moving average of the last 1440  $\Delta$  values) Same period is applied for the standard deviation.

Python code:

```
df['mean'] = df['delta'].rolling(window=1440).mean()
df['std'] = df['delta'].rolling(window=1440).std()
```

The z-score is calculated by the following formular:

$$z = \frac{\Delta - \mu(\Delta)}{\sigma(\Delta)} \quad (2)$$

and shows how many standard deviations  $\Delta_t$  is away from current mean

#### 1.5.1

#### 1.5.2

## 1.6 Idea behind

- This method assumes data is roughly normally distributed.
- Using  $2\sigma$  captures approximately 95% of data points under a normal distribution.
- You can adjust the multiplier (e.g.,  $3\sigma$ ) for stricter or looser thresholds.

## 1.7 Future Plans

- Test on more data
- use rolling windows (e.g. 1 day or 1 week) for local context.
- Compare sensitivity with  $\pm 1.5\sigma$  or  $\pm 2.5\sigma \rightarrow$  optimize for best results

## 2 Measuring Volatility After Outlier Detection

The first Idea I had was to measure the volatility of the price after an outlier is detected. Volatility is a measure of how much price changes in either direction over a certain period of time. Using the following formular:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (3)$$

### 2.1 Dictionary of Terms

- $P_t$  Asset price at time  $t$ .
- $r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$  – 1-minute price return at time  $t$ .
- $\sigma_t^{(15)}$  – Realized volatility: the standard deviation of the next 15 one-minute returns,

$$\sigma_t^{(15)} = \sqrt{\frac{1}{14} \sum_{i=1}^{15} (r_{t+i} - \bar{r}_t)^2}, \quad \bar{r}_t = \frac{1}{15} \sum_{i=1}^{15} r_{t+i}. \quad (4)$$

aligned so that at time  $t$  it measures volatility over  $t + 1$  to  $t + 15$ .

### 2.2 In Python code

```
import pandas as pd
df = pd.read_csv(file_path)
df.set_index('timestamp', inplace=True)
#Compute 1-min return of delta_5

df['r_t'] = df['price'].pct_change().fillna(0)

#compute rolling std of the future 15 min window
window = 15

#rolling on r_t, then shift forward so index t hold vol of t+1...t+15
df['future_vol_15'] = (
    df['r_t']
    .rolling(window=window)
    .std()
    .shift(-window)
)
```



## 2.3 Statistical evidence and results

Once an outlier is detected (1) inside of the Orderbook  $\Delta$ , we calculate the 15-minute ahead realized volatility using Equation: (4)

if a  $\Delta_t$  values is flagged as an outlier (1) we record

$$\sigma_t^{(15)} = \sqrt{\frac{1}{14} \sum_{i=1}^{15} (r_{t+i} - \bar{r}_t)^2},$$

We then form two samples over our full dataset which during this test includes 104 957 one minutes intervals of  $P$  and Orderbook  $\Delta$ :

$$\mathcal{S}_{\text{out}} = \{\sigma_t^{(15)} : t \text{ is an outlier}\}, \quad \mathcal{S}_{\text{non}} = \{\sigma_t^{(15)} : t \text{ is not an outlier}\}.$$

Sample mean results:

$$\bar{\sigma}_{\text{out}}^{(15)} = 0.0006244, \quad \bar{\sigma}_{\text{non}}^{(15)} = 0.0005138,$$

This concludes an increase of  $r_t$  of roughly 21.5%

To check Statistical evidence

- a two-sample \*t\*-test (unequal variances), which yields

$$T = 24.72, \quad p = 4.79 \times 10^{-132},$$

- a Mann–Whitney \*U\*-test, which returns

$$p = 4.02 \times 10^{-157}.$$

### 3 Optimising for best Z-Score threshold for outliers

As stated inside of (1) we use a Z-Score threshold of 2 to detect outliers. I now want to see if by any chance there is a value where their realized volatility is higher than if we use 2 as a threshold.

To compare the outliers Volatility with the non outliers volatility I will use the following formular:

$$U(z) = \frac{\bar{\sigma}_{\text{out}}^{(15)}}{\bar{\sigma}_{\text{non}}^{(15)}} \quad (5)$$

First I run an optimization for the thresholds of the Z-Score to find the best threshold value for the outliers on a 45 days dataset. After that I compare the result with a 107880 minutes dataset. Where I also run an optimization for the thresholds of the Z-Score to find the best threshold value for the outliers.

Top 3  $z$ -values with largest volatility uplift 69811-minutes sample

$z$	$N_{\text{out}}$	$U(z)$ (%)	Mann-Whitney $p$
3.8	221	+58.36	$1.913 \times 10^{-51}$
3.9	166	+61.99	$4.227 \times 10^{-41}$
4.0	117	+58.36	$8.228 \times 10^{-28}$

Same  $z$ -values on extended dataset (Walk forward  $v1$ )

$z$	$N_{\text{out}}$	$U(z)$ (%)	Mann-Whitney $p$
3.8	644	+51.73	$2.549 \times 10^{-77}$
3.9	561	+53.28	$6.150 \times 10^{-88}$
4.0	479	+50.90	$5.517 \times 10^{-59}$

Top 3  $z$ -values with largest volatility uplift (Walk forward  $v2$ )

$z$	$N_{\text{out}}$	$U(z)$ (%)	Mann-Whitney $p$
4.8	214	+65.10	$6.475 \times 10^{-33}$
4.9	197	+66.81	$3.693 \times 10^{-29}$
5.0	181	+70.53	$6.599 \times 10^{-28}$

## 4 Measuring average return after price outlier detection

### 4.1 Formulas

Once a  $\Delta_t$  Outlier is detected we calculate the 15-min forward return of BTC/USD price

$$\text{Ret}_t^{(15)} = \frac{P_{t+15} - P_t}{P_t} \quad (4)$$

We then differentiate between a bullish and a bearish outlier. Which is already defined (1)

$$\overline{\text{Ret}}_{\text{bull}}^{(15)} = \frac{1}{|\mathcal{T}_{\text{bull}}|} \sum_{t \in \mathcal{T}_{\text{bull}}} \text{Ret}_t^{(15)} \quad (7)$$

$$\overline{\text{Ret}}_{\text{bear}}^{(15)} = \frac{1}{|\mathcal{T}_{\text{bear}}|} \sum_{t \in \mathcal{T}_{\text{bear}}} \text{Ret}_t^{(15)} \quad (8)$$

### 4.2 Dictionary of Terms

- Price at a certain time:  $P_t$
- 15-min forward return:  $\text{Ret}_t^{(15)}$

#compute 15-min forward return of BTC/USD price

## 5 Underlying strategy Bias

Every single parameter has to fight to be implemented into my strategy. To get some kind of filter since we are working with an asset which has clear trends so it isn't stationary we need to do some trend identification. Different trends are also called different regimes. I'll call it the underlying bias.

I will differentiate between three different types of regimes:

- Uptrend
- Downtrend
- Ranging

Uptrend is defined as a period where the price is making higher highs and higher lows. Downtrend is defined as a period where the price is making lower highs and lower lows. Ranging is defined as a period where price is not making higher highs and higher lows or lower highs and lower lows. But for my approach I will just use this if we are getting mixed signals.

### 5.0.1 Trend identification version 1

To see in what kind of regime we currently are we first have to search for swing points. A swing point is a local extrema of a certain period. So a swing low is a local minimum and a swing high is a local maximum. Local is just that it is inside of our lookback period.

**Swing point** identification code:

```
if current_price > np.max(lookback_window):
    swing_high = current_price
if current_price < np.min(lookback_window):
    swing_low = current_price
```

This code snippet defines the swing low/high variables, it loops through the lookback window and checks if the current price is higher than the highest price in the lookback window or lower than the lowest price in the lookback window. If it is the case we update the swing low/high variable.

### Regime determination

```
window = 5
price_window = prices[i-window:i]
prev_window = prices[i-window-1:i-1]

#Then we calculate
price > local_high and local_high > prev_high
```

### Meaning:

Current price > local high (recent 5)

Recent 5-period high > previous 5-period high

This would mean we are in an **uptrend**.

For a **downtrend** analogous logic applies. If none of both applies we define the regime as **ranging**.

### 5.0.2 Trend identification version *v2*

Now after conducting research with chatgpt I found the Aroon indicator. This indicator is a technical indicator that measures the strength and direction of a trend. It is calculated by taking the difference between the current price and the highest price over the last  $n$  periods.

In my trend identification the Aroon indicator is used to determine how recently price has made a new high or low within our defenied lookback period.

The closer the Aroon up the more recent the high and the closer th Aroon Down the more recent the low.

```
aroon_indicator = aroonIndicator(period=window)
aroon_indicator.generate(df_resampled)
```

## 6 Finding an edge

### 6.1 Fees

So the first problem you run into if building a strategy is the fees. You have to pay fees for every trade you make. A fee is a fixed percentage amount of the trade size you take. On the exchange I choose, Hyperliquid the fee is about 0.003% of the trade size. Which is a normal amount inside the crypto space. Just to showcase how important fees are I will show you a simple example.

Inside of my Market simulation where I can test different kind of strategies I will set up a simple strategy. When ever a bullish outlier in the delta appears and the trend function sais that we are inside of an uptrend I will buy Bitcoin and wait for 240 Minutes.

```
#Define a mask for the bullish outliers inside of an uptrend
outlier_mask = (df_temp['outlier_context'] == 's') & (df_temp['trend'] == 'Uptrend')

#Define the entry and exit signals with the Vector BT library
pf = vbt.Portfolio.from_signals(
    close=price,
    entries=entry_signals,
    exits=exit_signals,
    init_cash=100, #Begining amount if cash
    freq='1T'
)
```

### 6.2 No Fees results

Table 1: No Fees	
Metric	Value
Holding Period	240 minutes
Total Return	2.599%
Mean sharp ratio	3.067088
Total Signals	187

### 6.3 With Fees results

Table 2: With Fees	
Metric	Value
Holding Period	240 minutes
Total Return	1.537%
Mean sharp	1.841555
Mean Z-Score	2.48

### 6.4 No clear path

Finding an edge is a very hard task. There is no clear path to success. You kinda have to try by trail and error every error could be a step further but also a potential path into a dead end.

There are three main steps to find an edge.

- Copying other strategies
- Finding an edge by yourself
- Build a systematic version of your discretionary trading.

While writting this I do not know yet if I was able to find an edge. Till now I searched for new things and combined different libraries from other people to develop a trend identification system. In the hyperlink below I linked the source of the trend identification system parts I copied.

Github account

## 7 Combining Indicators

Here I visualised the swing points, the EMA spread and the 100 outliers with the highest Z-Score in the same plot.



Figure 1: combined indicators png

1

<sup>1</sup>Chart made with Matplotlib and Seaborn



## 8 From outliers to strategy

### 8.1 Progress description

I had a pretty hard time going from developing a logic for the outliers of the Delta. Finding that volatility increases after outliers was a good first find which didn't take long and the process was pretty straight forward. But finding some definition condition is meet directional bias was very difficult. By directional bias I mean a price direction which is not random after some kind of condition. It was pretty clear from the begging on that the outliers by there self won't offer any kind of directional bias

An outlier is defined as:

- (1).  $\Delta < \mu(\Delta) - 2\sigma(\Delta)$  or  $\Delta > \mu(\Delta) + 2\sigma(\Delta)$

Then since Bitcoin clearly is not a stationary asset I needed to find some kind of trend identification. I found that the swing points are a good indicator for this. We determined a trend by using looking at swingpoints. A swing point is a local extrema of a certain period.

### 8.2 Where did I start my research

```
#compute swing points
swing_points = swing_points(df['price'], period=n)
```

Parameters of the swing\_points function:

- $n$  is the lookback period
- $df['price']$  is the price column of the dataframe (Which is a timeseries)
- $P_t$  is a value of  $df['price']$  at time  $t$

We basically got through the time series dataset and look back  $n$   $P_t$  values. The highest and lowest points inside of that specific lookback period are the swing points. After that we determine if price is making higher highs or lower lows. We check this by storing the last swing points and wait till we are either making a higher high or a lower low.

## 8.3 Finding an edge



Figure 2: Visualization of  $\Delta_t$  outliers and regime identification



Figure 3: Visualization of  $\Delta_t$  outliers and regime identification

Here you can see two visualisations of outliers and regime identification. On two different timeseries datasets. This was my trend identification system *v1* which is only based on swing point mapping.

After plotting different kind of parts of Timeseries Data sets I created I was pretty sure that somewhere I would be finding an edge since the outliers often where at good entry points for a strategy. Only problem was I wasn't able to get any statistical proof of this. After finding a Github Repo which used Vectorbt to test 1000 strategies at the same time I tried a similar approach and tested different conditions to see if there was some kind of edge.

## 8.4 Monte Carlo Testing

First of I have to explain how I search for a strategy and how it nearly killed my computer. We start by defining a different kind of sample sizes of all the  $\Delta_t$  outliers time periods after the outlier detection.

We want to know the return after our condition meets and then check in after  $z$  minutes after that

```
sampling_percentages = [0.1, 0.2, 0.6, 0.8, 1.0]
holding_periods = [60, 120, 240, 360]
```

The sampling percentage defines how many of the existing outliers we use and test the returns on. So let's say we have  $x$  amount of  $\Delta_t$  Outliers. We then randomly select  $z$  % amount of outliers and test the returns on them. We do this 1000 times and then take the average return of all the 1000 tests. The second step is to test each sampling percentage on the different holding periods. So let's say we have a sampling percentage of  $z$  % and a holding period of  $y$  minutes. We then take the  $z$  % amount of outliers and test the returns on them after  $y$  minutes. This test is basically copied from the Vectorbt Github Respository. I just changed the code to fit my needs and added some extra information returns. My code gives back the average return, the Sharpe, expectancy and mean Z-Score.

- **Sharpe Ratio:** The Sharpe ratio is a measure of the risk-adjusted return of a strategy. It is calculated by dividing the average return of the strategy by the standard deviation of the returns.
- Calculated by  $\frac{\bar{r}}{\sigma}$  Where  $\bar{r}$  is the average return and  $\sigma$  is the standard deviation of the returns.
- **Expectancy:** The expectancy is a measure of the profitability of a strategy. It is calculated by dividing the average return of the strategy by the average loss of the strategy.
- Calculated by  $\frac{\bar{r}}{\bar{l}}$  Where  $\bar{r}$  is the average return and  $\bar{l}$  is the average loss.
- **Mean Z-Score:** The mean Z-Score is a measure of the average Z-Score of the strategy. It is calculated by taking the average of the Z-Scores of the strategy.

- Calculated by  $\frac{1}{n} \sum_{i=1}^n z_i$  Where  $z_i$  is the Z-Score of the strategy at time  $i$  and  $n$  is the number of Z-Scores.
- **Average Return:** The average return is a measure of the average return of the strategy. It is calculated by taking the average of the returns of the strategy.
- Calculated by  $\frac{1}{n} \sum_{i=1}^n r_i$  Where  $r_i$  is the return of the strategy at time  $i$  and  $n$  is the number of returns.

Now it is time to test some strategies. The first strategy I want to test is pretty intuitive. Measure the average return of bullish outliers inside of an uptrend.

spacing

label A outlier is bullish if the the orderbook  $\Delta_t$  is two standard deviations above the mean of the last 1440  $\Delta_t$  values and bearish if it is two standard deviations below the mean of the last 1440  $\Delta_t$  values.

Now when we backtest a strategy we have to have a few things in mind. First of all we are backtesting on on historical data and if we just use different kind of entry conditions we might just change the entry conditions till we have a good end results. This would be overfitting and not work on future data. In order to prevent overfitting we have to use a holdout sample. I have one dataset on which we test our entry conditions to see if they even have potential.

If these conditions are met for a condition I test on some out of sample data. (Out of sample data just means that we test on new data) Most of the time things end up not even passing the first out of sample test and lose their potential instantly. If not I have another out of sample test and if that one is passed we are ready to do some more monte carlo testing and look how potential equity curves look like. And to make things even worse we can assume that the sharpen ratio will decrease by atleast 20% compared to the backtest simulation.

## 8.5 Backtesting framework

My backtesting framework is built in Python using the VectorBT library and consists of several key components:

### Data Pipeline

Connection to Postgres database hosted on railway.app with global access. Live data mining progress

- Historical price data from Coinbase (BTC/USD)
- Orderbook delta data at different depths ( $\Delta_{1\%}$ ,  $\Delta_{2.5\%}$ ,  $\Delta_{5\%}$ )
- 1-minute timeframe for base calculations
- Direct ema calculation which a lenght of 50 on the 1h timeframe

### Testing Methodology

The framework implements a three-stage testing process:

#### 1. Initial Sample Testing

- Test strategy on initial dataset
- Multiple sampling percentages: [0.1, 0.2, 0.6, 0.8, 1.0]
- Various holding periods: [60, 120, 240, 360] minutes

#### 2. Out-of-Sample Validation

- Test promising strategies on separate dataset (Two different datasets)
- Require consistent performance across datasets

#### 3. Monte Carlo Simulation

- Random sampling of trade opportunities
- 1000 iterations per test configuration
- Analysis of distribution of outcomes

### Performance Metrics

Each strategy is evaluated using:

- **Sharpe Ratio:**  $\frac{\bar{r}}{\sigma}$
- **Expectancy:**  $\frac{\bar{r}}{l}$
- **Average Return:**  $\frac{1}{n} \sum_{i=1}^n r_i$

## Risk Management

The framework incorporates:

- Trading fees (0.003% on Hyperliquid)
- Slippage of (0.0001% on Hyperliquid)
- Maximum drawdown limits
- Stop-loss and take-profit levels

## Implementation

```
# Example of strategy implementation
def backtest_strategy(data, params):
    # Define entry/exit signals
    outlier_mask = (data['outlier_context'] == 's') &
                    (data['trend'] == 'Uptrend')

    # Create portfolio simulation
    pf = vbt.Portfolio.from_signals(
        close=data['price'],
        entries=outlier_mask,
        exits=exits_after_n_bars(outlier_mask, n=params['holding_period']),
        init_cash=100,
        freq='1T'
    )

    return calculate_metrics(pf)
```

This framework allows for rapid testing of multiple strategy variations while maintaining strict validation criteria to prevent overfitting.

## 9 Developing a strategy

### 9.1 Initial thoughts

I will only able to show a few strategies backtests since I tested on about 150 different logics and only 1 of them passed all the out of sample tests. I will show the best ideas I had and the results of the backtests.

### 9.2 Reverse strategy

So logical thinking we could assume that long strategy can be reversed into a short strategy and vice versa. The problem is that we have fees and slippage. So we need a certain amount of profit to cover the fees and slippage. And since my strategy has a rather high frequency we need to have a lot of trades to cover the fees and slippage.

### 9.3 First strategy test

So the first strategy I test was before I developed the backtesting framework. I was testing on a simple logic. If the price on the 1h timeframe closed above the 50 period EMA and the  $\Delta_5$  of the orderbook was positive we would enter a buy position if price was in a range of 0.05% of the  $EMA_{50}$ .

In addition to that I was using a trailing stop. The trailing stop was set to 1% of the price. A trailing stop is a stop loss that is adjusted to the price of the asset. So if the price goes up the stop loss goes up with it. The idea behing this is that we can catch a bigger upwards in contrast to a fixed stop profit.

I only tested this on a short sample of data because I just had started my data mining process and didn't have any more data. But I can assure that more data didn't make it more profitable just worse.

#### 9.3.1 Code

```
# Define entry/exit signals
range_pct = 0.0005
trailing_stop_pct = 0.01

entries = (
    (df['bias'] > df['ema']) &
    (df['price'] >= df['ema'] * (1 - range_pct)) &
    (df['price'] <= df['ema'] * (1 + range_pct))
)
```

### 9.3.2 Strategy 1 results

Table 3: Detailed backtest results for the first strategy test.

Metric	Value
<b>General</b>	
Start	2025-03-12 00:00:01
End	2025-03-17 07:24:01
Period (minutes)	7,636
<b>Performance</b>	
Start Value	100.00
End Value	97.13
Total Return [%]	-2.87
Benchmark Return [%]	0.38
Total Fees Paid	2.21
Open Trade PnL	0.23
<b>Trades</b>	
Total Trades	23
Closed Trades	22
Win Rate [%]	31.82
Profit Factor	0.75
Expectancy	-0.14
Best Trade [%]	3.15
Worst Trade [%]	-1.24
Avg Winning Trade [%]	1.35
Avg Losing Trade [%]	-0.83

## 9.4 Learning from the first strategy

So from march till july I didn't backtest any strategies. I was focusing on developing signals and indicators. This backtest clearly showed me that I have to do research first and have some kind of background idea of what I am doing.



## 9.5 Explaining figure

So on the  $Y$  axis we have the average return of the strategy. And on the  $X$  axis we have the sampling percentage. A sampling percentage of 10% means that we are using 10 random percentage of the  $\Delta_t$  outliers my system identified over the backtesting timeseries dataset and then testing the returns on them. We conduct this 1000 times and then take the average return of all the 1000 tests. In addition to that we are testing the returns on different holding periods. So for example if we are testing the returns on a holding period of 60 minutes we are testing the returns on the 10 random percentage of the outliers after 60 minutes. We conduct this 1000 times and then take the average return of all the 1000 tests.

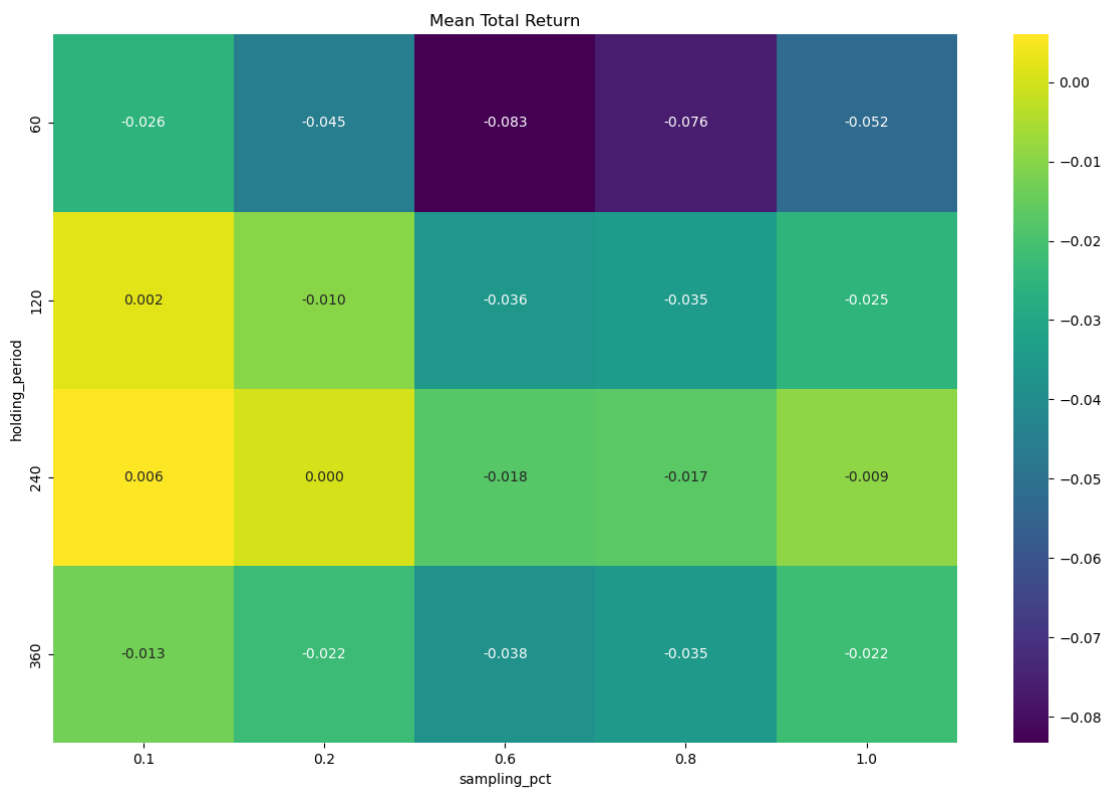


Figure 4: Strategy 2 results

## 9.6 Strategy 2

So this was the second strategy I tested and the first time I used my new backtesting framework. Results from this backtest: 4

- A  $\Delta_{5,t}$  value is identified as a bullish outlier (see Section 1).
- Trend is identified as uptrend

## 10 Algorithms Implementation

### 10.1 Outlier detection version 2

After not finding any edge with the normal z-score outlier detection I decided to implement a new outlier <sup>2</sup> detection logic.

The IQR method divides data into four equal groups.

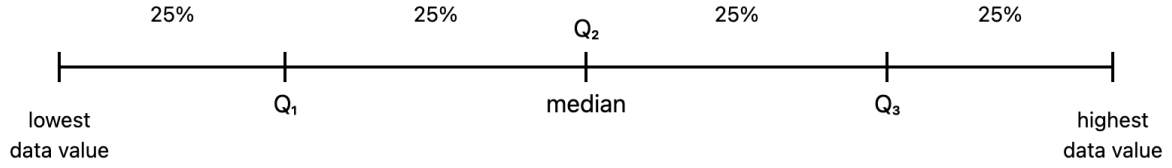


Figure 5: IQR method

- $Q_2$  (median): The middle value that divides the entire dataset into two equal halves
- $Q_1$  (first quartile): The median of the lower half of the data (values below  $Q_2$ )
- $Q_3$  (third quartile): The median of the upper half of the data (values above  $Q_2$ )

Each quartile represents 25% of the data, making it easy to identify the middle 50% of values (between  $Q_1$  and  $Q_3$ ).

An outlier is now defined as a value which falls outside of the interquartile range multiplied by a factor  $z$  in my case  $z = 2$ . IQR is defined as  $IQR = Q_3 - Q_1$ .

Outlier definition:

$$\Delta_t < Q1 - z * (IQR) \quad \text{or} \quad \Delta_T > Q3 + z * (IQR)$$

---

<sup>2</sup>an extrem high or low value compared to other values inside of the dataset

## IQR method python implementation

```
delta5_15 = df15['delta5']
roll = delta5_15.rolling(window, min_periods=window // 4)
#window is a parameter inside of function
#calculating rolling quartiles
q1 = roll.quantile(0.25).shift(1) #first quartile
q3 = roll.quantile(0.75).shift(1) #third quartile
iqr = q3 - q1 # interquartile range

#calculating bounds
lower_bound = q1 - z_value * iqr
upper_bound = q3 + z_value * iqr

#identifying outliers
outliers = (delta5_15 < lower_bound) | (delta5_15 > upper_bound)
```

For my usecase the IQR method is better fitting than the z-score method.

## Problems with z-score threshold

If the data is skewed or has fat tails the z-score method will be biased. The IQR method is more robust to outliers and is not sensitive to the distribution of the data. In simple words if the market is volatile the z-score method will be biased. The standard deviation be higher than it should and the mean will be affected too by strong volatility.

## 11 Algorithmic infrastructure implementation

Now since my strategy allows multiple positions open at the same time the infrastructure turns out to be a bit more complex. If we open two positions on the *BTC/USD* pair with a size of 1000\$ we end up having one positions with a size of 2000\$.

### Vectorbt implementation

```
entries = (df_temp['outlier_context'] == 's') & (df_temp['trend'] == 'Uptrend')
exits = exits_after_n_bars(entries, n=120) #120 minutes after entry

pf = vbt.Portfolio.from_signals(
    close=price,
    entries=entry_signals,
    exits=exit_signals,
    accumulate=True, #allows multiple positions open at the same time
    init_cash=100, #Beginning amount if cash
    freq='1T'
)
```

As you can see in the figure 6 there are times where multiple positions are open at the same time.



Figure 6: Trades

## 11.1 Handling trades asynchronously

I decided to handle the issues of multiple positions with different exit times by using an asynchronous approach. Every positions is handled in a seperate python thread. The thread just sells the exact amount of one single position which is 10% of the portfolio. The maximum of open positions is set to 20.

## 11.2 Strategy 3

So this was the third strategy I tested. I was testing on a simple logic. If the price on the 1h timeframe closed above the 50 period EMA and the  $\Delta_5$  of the orderbook was positive we would enter a buy position if price was in a range of 0.05% of the  $EMA_{50}$ .

## **12 sources**

### **12.1 How good or random is your trading**

Twitter article about random walk theory and how to test if a strategy is random or not.

### **12.2 Tweets**

A lot of my ideas come from thiy guys Tweets, to research into this. He is a disgresionary trade what means he trades based on his own decisions and has a mental framework to trade. I learned what an orderbook delta is from him.

### **12.3 TRDR This platform allow you to use different kind of metrics on different Timeseries datasets (BTC/USD Price, orderbookdelta and Open Interest**

his platform allow you to use different kind of metrics on different Timeseries datasets like: BTC/USD Price orderbookdelta and Open Interest

### **12.4 Trend line automation**

Used this for the *v3* version of my trend identification system

### **12.5 Vectorbt github documentation respository.**

I implemented some ideas from this repository for my backtesting framework and used vectorbt for the fees and slippage implementations

### **12.6 Deep research by ChatGPT**

The OpenAI subscription enables you use deepsearch function from chatGPT I would say this tool is probably useful for research but I didn't find any value from this besides finding out about new indicators. The only thing finding I made from these two pdfs was the Aroon indicator which I implemented in the second verions of my trend indentification system *v2*

### **12.7 Markets in profile – James Dalton**

Altho this book is for a disgresionary trading approach it taught me a lot about the Efficient Market Hypothesis and trading psychology.

### **12.8 Do retail traders stande a chance?**

Good read about if retail traders even have a chance to beat the market.