**Part 1: Data**

Daily Tick data of SPY/SH/SSO/SDS/UPRO/SPXU.

SPY: SPDR S&P500 seeks daily investment results, before fees and expenses, that correspond to (1x) the price and yield performance of the S&P 500.

SH: ProShares Short S&P500 seeks daily investment results, before fees and expenses, that correspond to the inverse (-1x) of the daily performance of the S&P 500.

SSO: ProShares Ultra S&P500 seeks daily investment results, before fees and expenses, that correspond to two times (2x) the daily performance of the S&P 500.

SDS: ProShares UltraShort S&P500 seeks daily investment results, before fees and expenses, that correspond to two times the inverse (-2x) of the daily performance of the S&P 500.

UPRO: ProShares UltraPro S&P500 seeks daily investment results, before fees and expenses, that correspond to three times (3x) the daily performance of the S&P 500.

SPXU: ProShares UltraPro Short S&P500 seeks daily investment results, before fees and expenses, that correspond to three times the inverse (-3x) of the daily performance of the S&P 500.

Sample of the data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time | Bid | Ask | Bid Size | Ask Size |
| 4 | 181.59 | 181.6 | 118 | 6 |
| 11 | 181.59 | 181.6 | 118 | 6 |
| 12 | 181.59 | 181.6 | 118 | 6 |
| 62 | 181.58 | 181.6 | 118 | 6 |
| 75 | 181.58 | 181.6 | 228 | 6 |
| 78 | 181.58 | 181.59 | 228 | 6 |

Time is measured in millisecond and the market opening time was set as 0.

**Part 2: Transaction Cost vs. Order Imbalance**

Example: SPY

|  |  |
| --- | --- |
| Time | |
|  | Due to the internet connection, we have missed some data for a very short of time period and we see a discontinuous line in terms of the time. Otherwise, we should expect a line without a significant jump. |

|  |  |
| --- | --- |
| Bid and ask | |
|  | Those straight line segments are due the disconnection of data recording. |

|  |  |
| --- | --- |
| Order Imbalance | |
|  | The formula of order imbalance is |

|  |  |
| --- | --- |
| The change of mid-price | |
|  | With a larger time lag (from 1ms to 10000ms), the change of mid-price is more likely to have larger values. |

|  |  |  |
| --- | --- | --- |
| Transaction Cost vs. Order Imbalance | | |
| Histogram of change of mid-price in terms of order imbalance | | |
| 0 ≤ Order Imbalance < 0.1 | 0.5 ≤ Order Imbalance < 0.6 | 0.9 ≤ Order Imbalance < 1 |
| 1ms | | |
|  |  |  |
| 10ms | | |
|  |  |  |
| 100ms | | |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| 1000ms | | |
|  |  |  |
| 10000ms | | |
|  |  |  |
| Though it is not significant, for larger time lag, the distribution of the change of mid-price tends to skew to right when order imbalance is small and tends to skew to left then order imbalanec is large. | | |

|  |
| --- |
| Transaction cost (change of mid-price) vs. order imbalabce in bucket average |
|  |
|  |
|  |
|  |
|  |
|  |
| Generally speaking, the lag of 1000ms is a proper value for measuring the relationship between transaction cost and order imbalance. |

**Part 3: Expected profit vs. Residual value**

Example: SPY-SH

|  |  |
| --- | --- |
| Scatter plot: mid-price of SPY vs. mid-price of SH | |
|  | Since the SH is the inverse of the S&P 500 index performance, we are expecting to see a strong negative correlation between the mid-price and the slope is -1. |

|  |  |
| --- | --- |
| Intercept B and residual value | |
|  | The histogram of B has removed the zero values, which are due to beginning of the data and the maximum values, which are due to disconnection of the data.  The residual value is shown from -0.002 to 0.002 after removing the beginning of the data. For where the date get disconnected, the residual value will be zero and hence negligible. |
|  | The scatter plot between B and residual values shows a negative correlation. |

|  |  |
| --- | --- |
| Expected Profit | |
|  | Similar the change of mid-price, with a larger time lag (from 100ms to 100s), the PnL is more likely to have larger values.  Formula:  (in the case of SPY-SH, we will long both the SPY and SH, since the slope is -1) |

|  |  |  |
| --- | --- | --- |
| Expected profit vs. Residual value | | |
| Histogram of expected profit in terms of residual value | | |
| -1≤ Residual\*103 < -0.8 | 0≤ Residual\*103 < 0.2 | 0.8≤ Residual\*103 < 1 |
| 100ms | | |
|  |  |  |
| 1s | | |
|  |  |  |
| 10s | | |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| 100s | | |
|  |  |  |
| Though it is not significant, for larger time lag, the distribution of the change of mid-price tends to skew to left when order imbalance is small and tends to skew to right then order imbalanec is large. | | |

|  |  |
| --- | --- |
| Expected Profit vs. Residual Value in bucket average | |
| Histogram of residual values | Expected profit vs. residual value |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |