Applications of the RSD Test in Stock Price Analysis

COMP4971C - Independent Study

Aaron Si-yuan Wang

Supervised by Dr. David Rossiter

Department of Computer Science and Engineering, HKUST

Abstract

Originally developed for detecting trend turnings in climate data, the running slope difference (RSD) test, proposed by Zuo *et al.* in [1], checks for statistical significance in the difference of the slope of two time series. This RSD test has many advantages over existing trend turning detection methods [1], making it favorable for generating trend lines in historical data. This report details an attempt at applying this RSD test into different areas of stock price analysis. In addition to utilizing this test to create an automatic stock price trend generator, this report also proposes a rudimentary trading algorithm based on the RSD test, which shows promising results when backtested against years of historical data.

Contents

1 Introduction

A time series is a collection of data points which represents the value of a variable at different times. Many things can be modelled as a time series, allowing us to study how these variables change over time, which can be beneficial. One typical example of a time series is that of stock prices, as knowing how the stock prices changes over time could allow the person to exploit the market to gain profit. When analyzing time series data, one common and useful technique used is to identify trends in the data. This can easily be done manually, however, there is not a clear way to effectively automate the process. This is because the same set of time series data can be interrelated in many different ways, showing different trends depending on the time scale.

There have been several attempts made to create an automatic trend line generator. One such attempt, made by Jitpakdee & Pravithana used image processing, namely Hough Transform, to generate trend lines [2]. Separate from applications in stock prices, trend line generation has been a very useful and well documented in the field of climatology, as the ability to detect multiple changes in trend is important for climate detection methods.

This report details an attempt to apply a recently developed statistical test to detect trend turnings in climate data called the running slope difference (RSD) t test to detect trend turnings in stock price data. The results show potential in using such test to detect trend turnings in historical stock price, allowing for more detailed and meaningful analysis of historical stock price trends. Apart from being used solely for detecting trend turnings, this report will also consider other applications of RSD in stock price analysis, namely with trading algorithms. To demonstrate the potential, one such RSD test based trading algorithm will be presented in this report and backtested against historical stock price data.

2 RSD Historical Trend Generator

2.1 RSD Test Overview

Trend turning occurs when two nearby periods of a time series show significantly different trends and is a powerful indicator commonly used when analyzing time series data. To detect these trend turning, the RSD test, developed by Zuo *et al.*, perform a t-distribution test on the difference of two linear time-series [1]. Let:

$$Y : \{y_i = \beta_Y i + \alpha_Y + \varepsilon_i \mid 1 \leq i \leq n\} \text{ and } Z : \{z_i = \beta_Z i + \alpha_Z + \varepsilon_i \mid 1 \leq i \leq m\}$$

be two sample series of length n and m respectively, with β_Y and β_Z be the slope of the series, α_Y and α_Z be the intercepts, and ε_i and ε_j being the error terms of Y and Z respectively. The error terms ε_i and ε_j are assumed to be normally distributed with zero mean and variance σ^2. If we let

$$\tilde{Y} : \{\tilde{y}_i = \tilde{\beta}_Y i + \tilde{\alpha}_Y \mid 1 \leq i \leq n\} \text{ and } \tilde{Z} : \{\tilde{z}_j = \tilde{\beta}_Z i + \tilde{\alpha}_Z \mid 1 \leq j \leq m\}$$

be the least-square linear regressions of Y and Z, then the general form of the slope difference t-distribution statistic, t_{slope}, between Y and Z is:

$$t_{\text{slope}} = \frac{\tilde{\beta}_Y - \tilde{\beta}_Z}{S_{\beta_Y,\beta_Z}},$$

where:

$$S^2_{\beta_Y,\beta_Z} = \frac{1}{C}\frac{1}{n+m-4}\left(\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 + \sum_{j=1}^{n}(z_j - \tilde{z}_j)^2\right)$$

is the variance of the regression errors, with:

$$C = \frac{NM}{N+M}, \text{ where } N = \sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^2 \text{ and } \sum_{j=1}^{m}\left(j - \frac{m+1}{2}\right)^2.$$

In this case, the null hypothesis is $\beta_Y = \beta_Z$, i.e. both sample series have the same slope. This null hypothesis is rejected at a significance level α if:

$$|t_{\text{slope}}| \geq t_{m+n-4,1-\frac{\alpha}{2}},$$

where $t_{m+n-4,1-\frac{\alpha}{2}}$ denotes the t distribution with $m+n-4$ degrees of freedom and with significance level α [1].

2.2 Extracting Trend Turning Points

To locate the turning points in the data, $(t_i)_{\text{slope}}$ is calculated for each point t_i in the time series, with Y and Z being the time series data preceding and following t_i, each with a length of T data points. If $|(t_i)_{\text{slope}}| \geq t_{m+n-4,1-\frac{\alpha}{2}}$, the point is labeled as a potential turning point, as the slope of the time series data before and after t_i show significant differences. For each period of time $\{t_i, \ldots t_j\}$, where:

$$|(t_k)_{\text{slope}}| \geq t_{m+n-4,1-\frac{\alpha}{2}}, \quad \text{for all } k \in \{i, i+1, \ldots, j\},$$

we set t_ℓ to be the turning point, with:

$$(t_\ell)_{\text{slope}} = \underset{i \leq k \leq j}{\arg\max}\, |(t_k)_{\text{slope}}|.$$

In other words, t_ℓ is the point during the period $t \in \{t_i, \ldots, t_j\}$ for which t_{slope} reaches its maximum/minimum value. Once we have identified all of the turning points, we are able to do simple linear regression to get the trends. As such, by inputting different values of T, we will be able to generate trend lines of different time scales, since Y and Z will have more or fewer data points.

2.3 Results

The above trend-line generation algorithm has been implemented into *Stoxy*, a stock analysis program developed and provided by Dr. Rossiter. Figures 1 and 2 show the results of running the trend generator on CK Hutchison Holdings Ltd and Twitter Inc respectively. These different colored lines indicate the following:

- Black: Closing Price of Stock
- Blue: Trend Lines Generated with $T = 30$ days
- Orange: Start/End Dates of Trend Lines Generated with $T = 30$ days
- Red: Trend Lines Generated with $T = 100$ days
- Green: Start/End Dates of Trend Lines Generated with $T = 100$ days
- Thickness of Trend Lines: r-value of Trend (thick lines mean high r-value)

The quality of the trend lines generated is left up to the reader, as it is very difficult to quantify the generator's success objectively.

Figures 3 and 4 show a comparison between the values of t_{slope} when $T = 30$ and $T = 100$ days, for Twitter Inc from 11/7/2013 to 2/22/2019. The axes and lines represent the following:

- x-axis: days since start of trend generation period (11/7/2013)
- Black Line: value t_{slope} on the given day
- Red Lines: threshold value of t-test with $\alpha = 0.99$
- Blue Dashed Lines: days when $|t_{\text{slope}}|$ is maximized (start/end of trend line)

These figures show how the RSD test is able to capture trends of varying time frames, with the red trend lines in Figures 1 and 2 clearly capture long term trends, while the blue trend lines follow the stock price more closely on shorter time scales. This is further exemplified in Figures 3 and 4, as the blue dashed vertical lines appear more often in Figure 3, since t_{slope} passes the threshold more frequently for smaller values of T.

3 RSD Trading Algorithm

3.1 Motivation

Apart from generating trends in historical data to analyze them, another potential application of the RSD test is with developing trading algorithms, as it allows us to see whether there is a significant difference between trends of the same data. Creating a trading algorithm would allow us to take advantage of the rising or lowering prices of stocks for profit.

One key difference between historical trend generation and a practical trading algorithm is that for the trend generator, we are able to make use of prices after the given date to determine if it is a turning point. For trading algorithms, we are not able to make use of this information, as we are only able to use past information. As such, instead of being able to run the RSD test to compare the sample series before and after the date, we must compare two sample series before the given date.

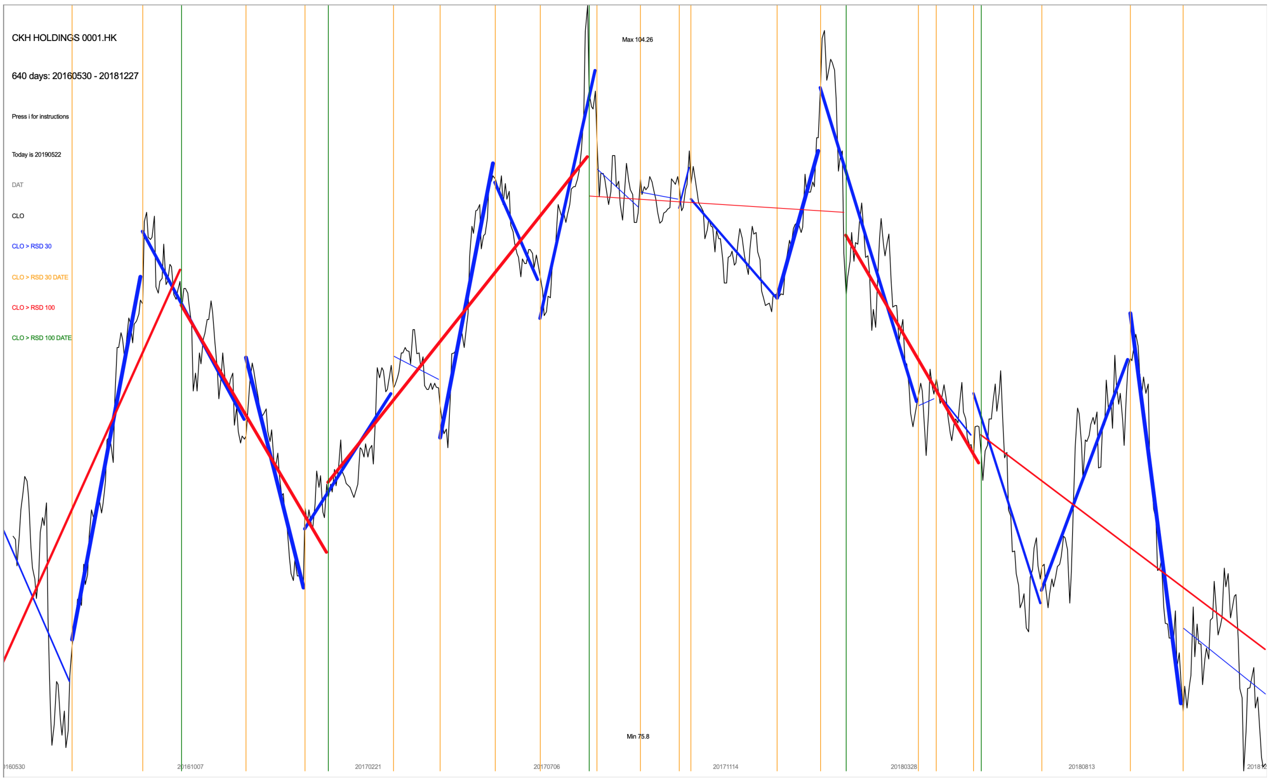

Figure 1: CK Hutchison Holdings Ltd (HKG:0001) From 5/30/2016 to 12/27/2018

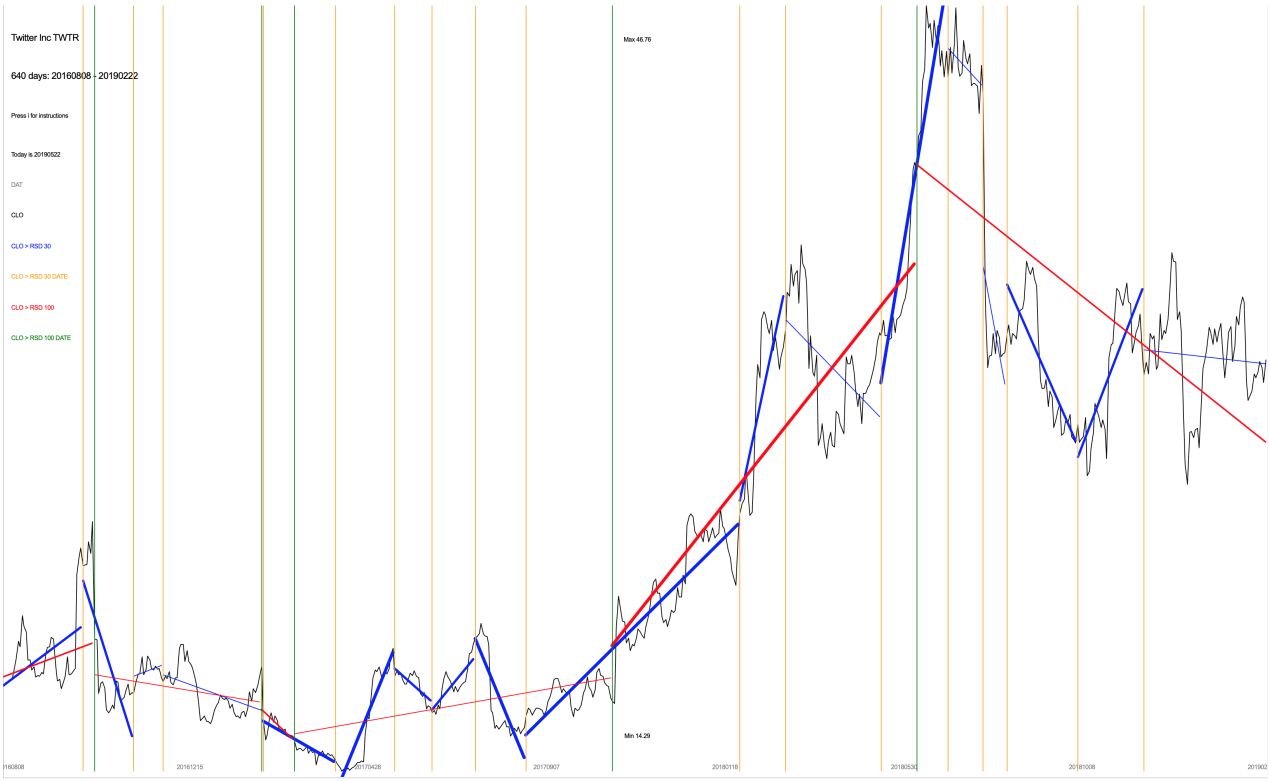

Figure 2: Twitter Inc (NYSE: TWTR) From 8/8/2016 to 2/22/2019

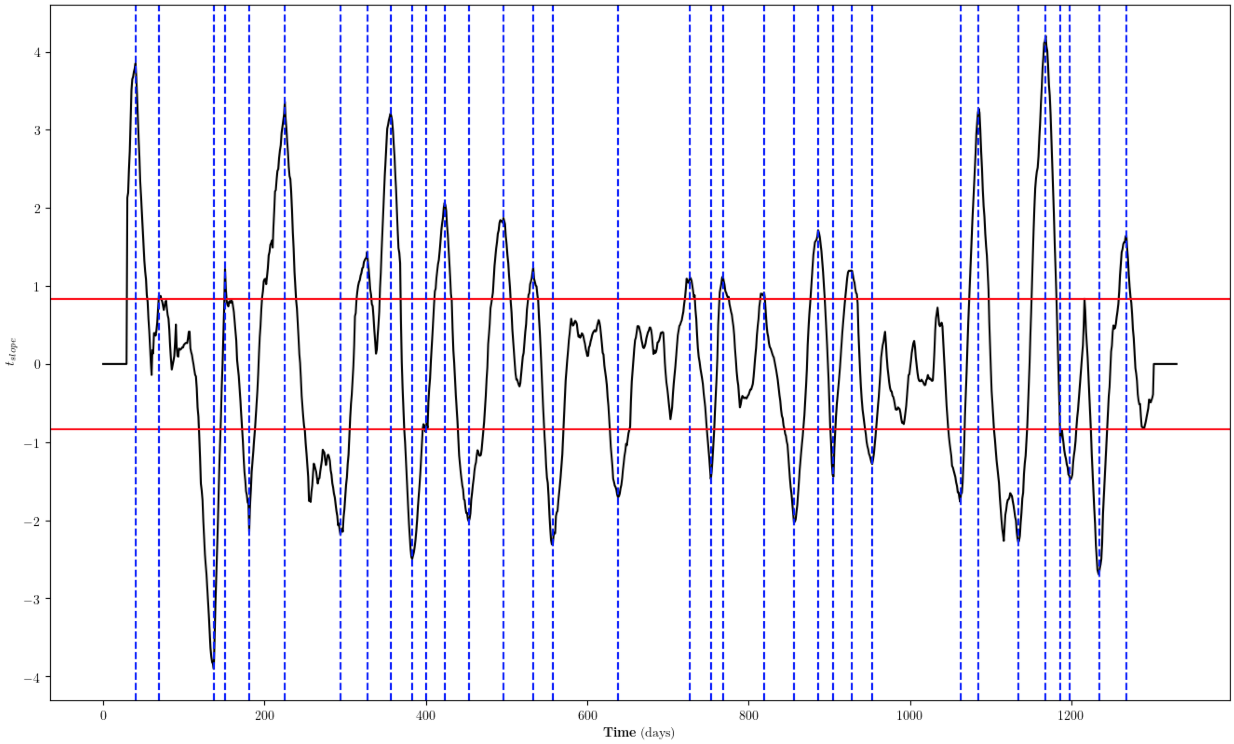

Figure 3: t_{slope} of Twitter Inc for $T = 30$ days

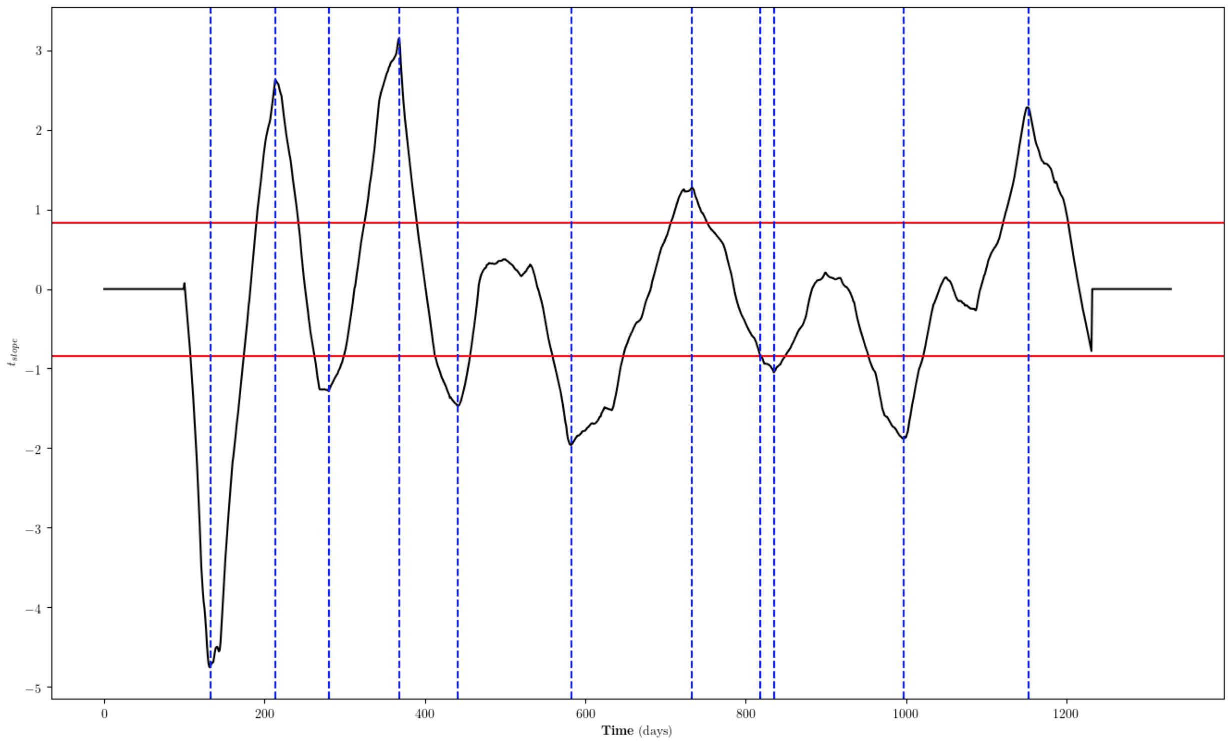

Figure 4: t_{slope} of Twitter Inc for $T = 100$ days

3.2 RSD Trading Algorithm Overview

Because the RSD test can be done with time series of different lengths, one natural idea would be to use the test to see if there is a significant difference between long and short term, to see if there will be an inflection point in the price. This trading algorithm will work as follows:

- Each day, we will calculate $t_{\text{slope}}(S, L)$ by comparing the slope of the most recent S and L days, with $S < L$.
- If $|t_{\text{slope}}| \leq t_{S+L-4,1-\frac{\alpha}{2}}$, then there is no significant change in the trend, and as such, we do not buy or sell the stock.
- If $t_{\text{slope}} > t_{S+L-4,1-\frac{\alpha}{2}}$, then the slope of the short term trend is significantly greater than that of the long term trend, suggesting that the price will start rising. As such, we will buy at this point.
- If $t_{\text{slope}} < -t_{S+L-4,1-\frac{\alpha}{2}}$, then the slope short term trend is significantly less than the long term trend. As such, we should sell the stock, as the price is likely to continue to drop.

Note that the amount of buying/selling can both be fixed or variable. To demonstrate this trading algorithm in action, the algorithm has been backtested against historical data of various stocks. The backtesting procedure is as follows:

- At the very start of the testing, we start with 1,000,000 HKD in cash.
- Each day, depending on t_{slope}, we do one of the following:
  - If $t_{\text{slope}} > t_{S+L-4,1-\frac{\alpha}{2}}$, then we will spend 25% of the cash on hand to purchase stocks.
  - If $t_{\text{slope}} < -t_{S+L-4,1-\frac{\alpha}{2}}$, then we will sell all of the stock on hand for cash.
- Continue until we reach the last closing price available.

3.3 Results

Tables 1, 2, 3, and 4 show the annualized return of the trading algorithm given above with different values of S and L for different stocks, ranging from 3 days to 90 days. This is compared to the "Buy and Hold" strategy, where the 1,000,000 HKD is spent on the first day, and held until the last. The rows in red indicate when the trading algorithm has performed better than the "Buy and Hold" strategy.

The backtested results show that this type of algorithm does show potential, as it can sometimes outperform the "Buy and Hold" strategy. However, the return on investment does vary based on S and L, with some combinations showing losses in the backtest. This suggests that S and L must be chosen carefully or should be changed dynamically. Besides deciding on the values of S and L, changing the buying/selling amount will also affect the results and performance of the trading algorithm. As such, there is still a lot of work to be done to refine this algorithm and ensure its stability.

S	L	Annualized Return
3 days	7 days	3.07 %
3 days	14 days	−9.84 %
3 days	30 days	−15.06 %
3 days	60 days	−9.75 %
3 days	90 days	−15.18 %
7 days	14 days	−1.9 %
7 days	30 days	−7.12 %
7 days	60 days	−4.1 %
7 days	90 days	−16.58 %
14 days	30 days	−10.79 %
14 days	60 days	−14.71 %
14 days	90 days	−18.08 %
30 days	60 days	8.3 %
30 days	90 days	−11.23 %
60 days	90 days	−10.46 %
Buy and Hold		−9.1 %

Table 1: Twitter Inc (NYSE: TWTR)

From 11/7/2013 to 2/22/2019

S	L	Annualized Return
3 days	7 days	−1.1 %
3 days	14 days	−1.25 %
3 days	30 days	1.3 %
3 days	60 days	4.83 %
3 days	90 days	8.04 %
7 days	14 days	−0.43 %
7 days	30 days	−0.07 %
7 days	60 days	3.19 %
7 days	90 days	3.98 %
14 days	30 days	8.91 %
14 days	60 days	4.61 %
14 days	90 days	3.98 %
30 days	60 days	9.77 %
30 days	90 days	0.01 %
60 days	90 days	−2.45 %
Buy and Hold		6.09 %

Table 2: CKH Holdings Ltd (HKG:0001)

From 1/4/2000 to 12/27/2018

S	L	Annualized Return
3 days	7 days	0.97 %
3 days	14 days	−2.39 %
3 days	30 days	−1.41 %
3 days	60 days	0.7 %
3 days	90 days	0.94 %
7 days	14 days	−1.57 %
7 days	30 days	−0.13 %
7 days	60 days	2.18 %
7 days	90 days	0.63 %
14 days	30 days	6.89 %
14 days	60 days	14.39 %
14 days	90 days	13.03 %
30 days	60 days	10.93 %
30 days	90 days	9.76 %
60 days	90 days	−1.18 %
Buy and Hold		12.3 %

Table 3: Hang Lung Group Ltd (HKG:0010)

From 1/4/2000 to 9/12/2018

S	L	Annualized Return
3 days	7 days	0.39 %
3 days	14 days	−0.01 %
3 days	30 days	1.23 %
3 days	60 days	0.69 %
3 days	90 days	3.3 %
7 days	14 days	1.94 %
7 days	30 days	1.56 %
7 days	60 days	3.52 %
7 days	90 days	9.55 %
14 days	30 days	6.55 %
14 days	60 days	−5.75 %
14 days	90 days	4.35 %
30 days	60 days	0.42 %
30 days	90 days	12.03 %
60 days	90 days	0.0 %
Buy and Hold		2.05 %

Table 4: Tracker Fund HK (HKG:2800)

From. 12/31/2007 to 1/11/2019

4 Further Explorations

4.1 Further RSD Trading Algorithm Development

The trading algorithm proposed in this report is very rudimentary, and as such, there are many other possible trading algorithms that can be made based off of this RSD test. In addition, the algorithm utilizes a very basic buying/selling strategy, spending 25% of cash available when buying, and selling all stocks when selling. In the future, a more sophisticated strategy could be developed to allow the user to profit more.

4.2 Forecasting Algorithm

Besides historical trend generation and trading algorithms, one important concept in stock price analysis which has yet to been addressed is with forecasting. As such, one further exploration possible would be to design a forecasting algorithm based on the RSD test. One difficulty in doing so is effectively converting t_{slope} into some model to project into the future. However the suitability and practicality of such forecasting algorithm has yet to be determined.

4.3 Refining Assumptions Made

In the derivation of the RSD test distribution, one assumption made is that the true signal is linear with Gaussian noise. However, this assumption is not necessarily true, as stock prices tend to show exponential trends. As such, one improvement would be to change this assumption in the original derivation of the RSD test, and see how it would affect the trend generation and trading algorithm. However, even without doing so, the trend line generator and trading algorithm are still able to provide useful information, as they are able to identify long and short term trends in the market.

5 Conclusion

As demonstrated by both the trend line generator and the performance of the trading algorithm, the RSD test is shown to have many applications in stock price analysis. The trend line generator created has shown great potential in automatically generating trend lines of stock prices, supporting both long and short term trend detection. One change that could be made to the RSD test is to change the assumptions made, as it assumes that the true signal is linear with Gaussian noise. Through backtesting with historical data, the RSD based trading algorithm appears to have potential, as it is able to outperform the "Buy and Hold" strategy on occasions, despite being very rudimentary. As such, there is further potential to develop the trend generator and trading algorithm further or to try and apply the RSD to other stock price analysis aspects, such as with forecasting.

6 References

1. Zuo, B., Li, J., Sun, C. & Zhou, X. A new statistical method for detecting trend turning. *Theoretical and Applied Climatology* (Feb. 2019).
2. Jitpakdee, P. & Pravithana, N. A Novel Method of Trend Lines Generation Using Hough Transform Method. *International Journal of Computing Academic Research* **vol. 6,** pp. 125–135 (Aug. 2017).