Page-Hinkley Notes and Psudocode

February 8, 2026

1 Overview

The Page-Hinkley test is used for change detection in a time series. It tries to capture sudden shifts in the mean of the data by separating the time series into two regimes, each of which assumes its own probability distribution modeling the behaviour of the data. It involves continually computing a sum of small drifts in the recent data points and alerting a regime shift when the sum passes a threshold value.

2 How it Works

let x_n reprisent the rolling standard deviation in the return of the stock price. We'll define the quantity

$$X_n = \log_e(f_1(x_n)) - \log_e(f_0(x_n))$$

Here, f_1 is the assumed probability distribution of the of the rolling standard deviation in the high volatility regime, and f_0 is the assumed probability distribution of the rolling standard deviation in the low volatility regime. The idea is to first assume a probability density function which governs the volatility data. Then at each time step we take the current volatility and evaluate each of the densities at that points as a way to evaluate how probable it is that the data point came from either of the two distributions. Now define:

$$S_0 = 0, S_n = \sum_{k=1}^{n} X_k$$

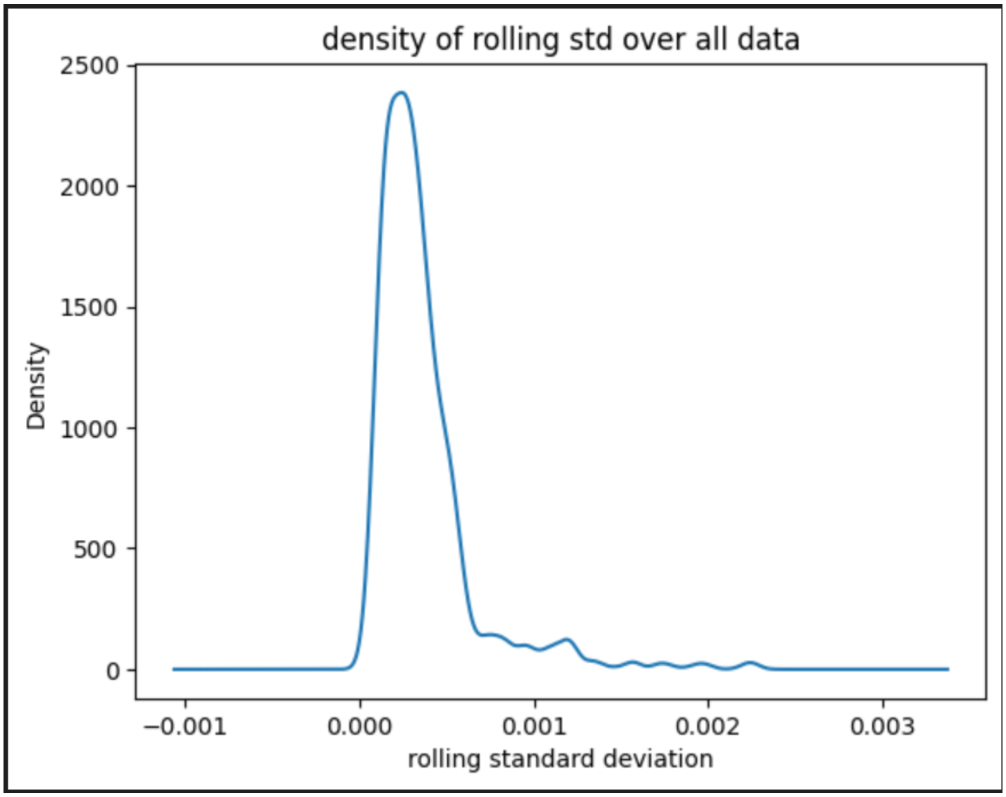
and the Page-Hinkley Detector is defined as:

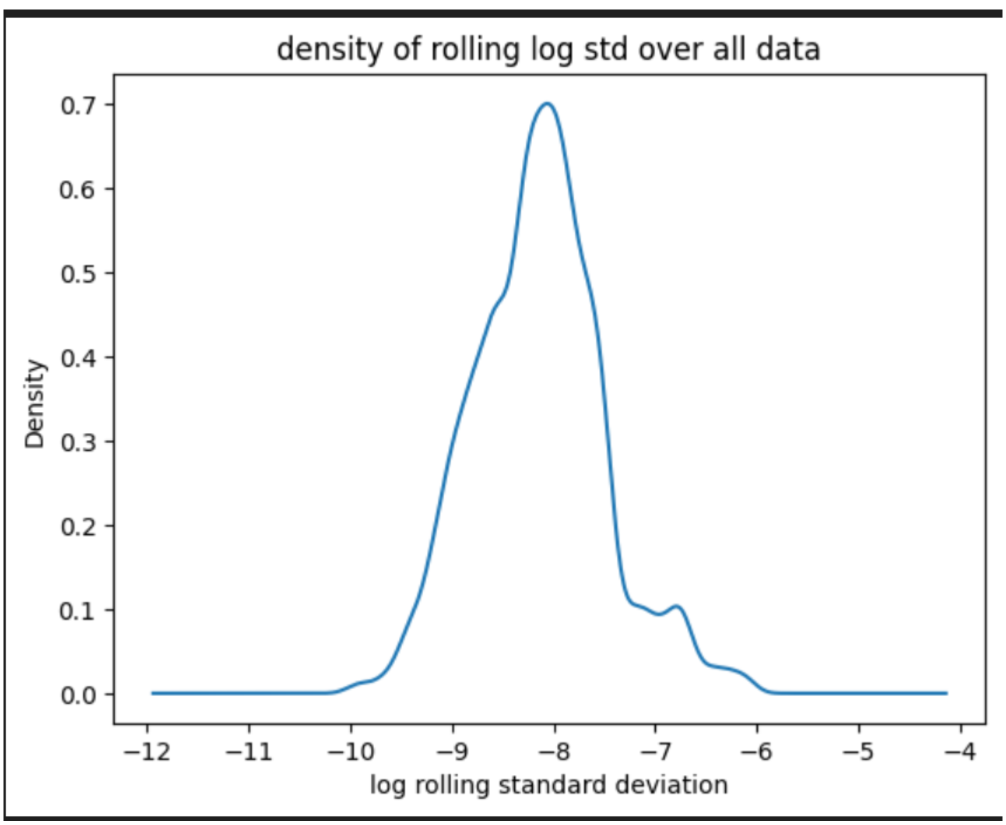
$$g_n = S_n - \min_{0 \leq k \leq n} S_k$$

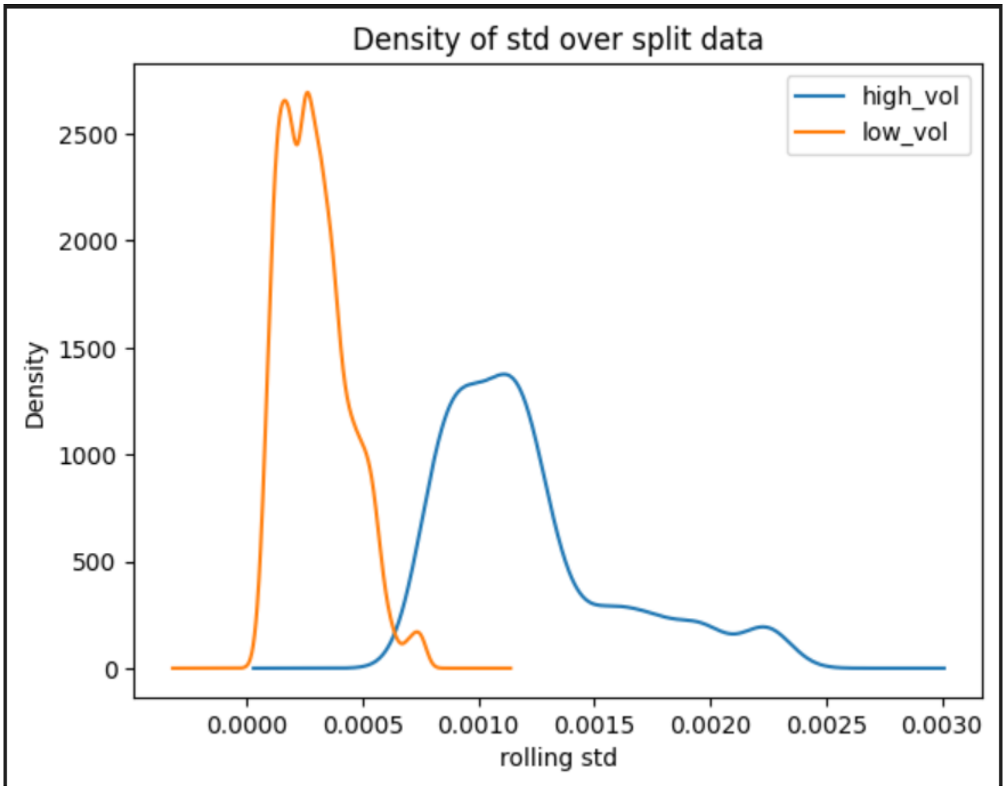
S_n tracks the cumulative sum of the X_n's. If X_n is consistently positive for a period of time this indicates a higher probability the data is being sampled from the high vol probability distribution and which causes S_n to increase. If X_n is negative it indicates the low vol regime and S_n decreases. g_n measures how much higher the current S_n is than the lowest recorded value. When g_n is higher than a threshold δ, an alarm triggers that the data is in a high vol regime.

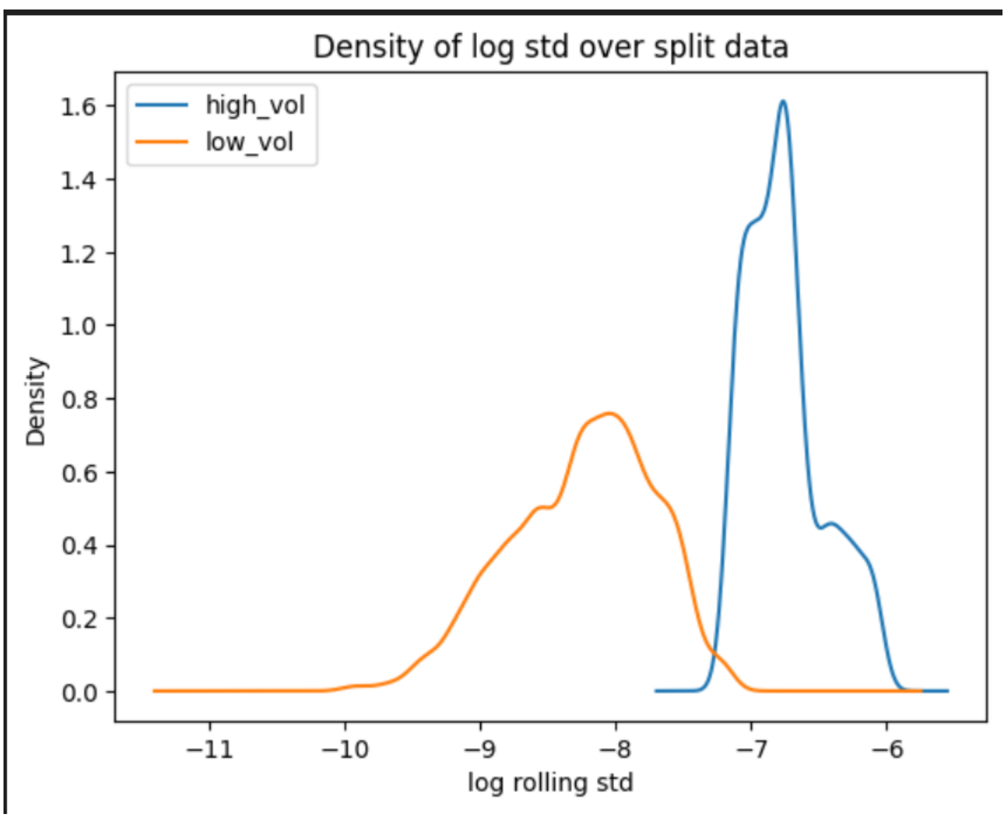
3 Estimating probability densities

The definition of X_n requires an assumed density for each of the two regimes. A common assumption to make is that the volatility of a stocks returns follows a log normal distribution. In order to test the idea, I used the same data the CUSUM model was tested with and used the CUSUM alarms to classify the data into high and low volatility regimes. I then plotted the kernel density estimation of the data in each regime.


density of rolling std over all data
Density
2500
2000
1500
1000
500
0
−0.001
0.000
0.001
0.002
0.003
rolling standard deviation


density of rolling log std over all data
Density
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−12
−11
−10
−9
−8
−7
−6
−5
−4
log rolling standard deviation





Given the data above, it seems reasonable to model the stock volatility as a log normal distribution in each of the regimes. Also note the data set used was limited and there are almost 20 times as many data points on the low vol period than the high vol given that it is more rare.

4 Pseudocode

```
#x is an array of rolling standard deviation
def Page_Hinkley_Test(x, threshold):
    S = []
    S[0] = 0
    for i = 1 to len(x):
        X[i] = log(f1(x[i])) - log(f0(x[i]))
        S[i] = S[i-1] +  X[i]
        if i == 1: minimum = S[i]
        minimum = min(S[i], minimum)

        g = S[i] - minimum
        if g > threshold:
            print("high volatility period")
            return i
```

5 Main Source

https://www.sciencedirect.com/science/article/abs/pii/S0167691111000727