

# How the Scalar function works

The main problem of the data generated by the model is the scale of the data itself. As we can see in the following picture, the serie seems realistic if we just look at the trend. However, the magnitude of the data is totally out of scale. The algorithm below is designed to force the series to converge to the real data range.

## Introduction

The scaler aims at scaling the time serie without affecting the trend. The idea is that when we have a peak of returns in the time serie, the significance of the peak depends on its distance from the second highest peak in the series. For exemple let's take the following series from the dataset of stock values:

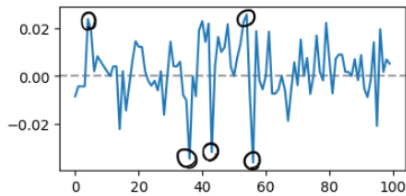


Figure 1

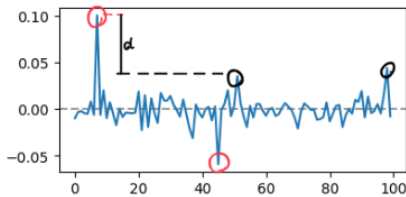
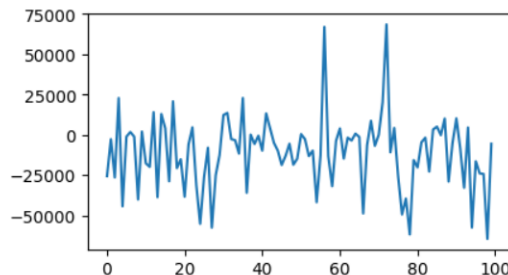


Figure 2

As we can see in (Figure 1) the 2 highest peaks are not very far from each other and that makes the order of magnitude lower compared to the peak we can see in the (Figure 2). Let's analyze two different meanings a peak can refer to:

### Exemple:



- a strong movement in price
- a shock, due to news or particular events.

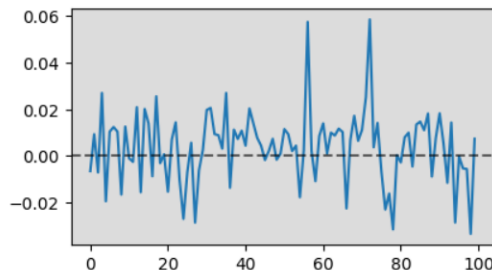
Taking into consideration the nature of the sample taken, it is very unlikely to have two big shocks in the same window. Following this logic, when the 1<sup>st</sup> and 2<sup>nd</sup> peaks in a serie are close, we can conclude they are probably just strong movements in the price. Again, when the 1<sup>st</sup> peak in terms of magnitude is way bigger than the 2<sup>nd</sup>, it is unlikely to have price movements with this elevated magnitude, in very small ranges. We can note that since the usual price movement of these kind of stocks falls between a range of -0.01% and 0.01%, It seems reasonable to apply the scalar function to get an ultimate generated serie that can be used for data analysis purposes while reflecting the true nature of the financial asset.

## Scalar Logic

Baset on the above reasonment the Scalar aims assign to "lonley" peaks in the generated series the value that usually shocks have in the real data and assign to close peaks value that are typical of strong market movements.

## Scalar Implementation

To do It the scalar compute the quantiles of the max values distribution and same thing for the distribution of the difference between the 1<sup>st</sup> and 2<sup>nd</sup> peaks. Then the algorithm take the sample and see where the distance between the 1<sup>st</sup> and 2<sup>nd</sup> peaks are among the quantiles of the differences for real data. Then pick a random value from a uniform distribution where the lower and upper bonudaries are the corresponding quantiles in the max distribution. Same thing for the negative peaks.



In the above exemple we can see on the left the series generated by our model and on the righth the series after the we apply the Scaler. As we can see the scaler preserves the pattern generated by our model and just rescale the data so we have realistic data.