# Betasmartz Research – Confidential - Investment Cycle Labeling Algorithm

## Objective

The intention of this document is to introduce and outline all the steps that are involved in determination of the investment cycle labeling algorithm.

This algorithm creates the labels of the time; the algorithm is monthly based due to the nature of the data.

## Data Needed

There is no particular data needed if the algorithm wants to be run at random time series. It is recommended to input macro economical time series data, as the final output will be an economic cycle the cycling pattern determination should be made from an economical perspective.

## Step 1 – Input data and data cleansing

The economic data normally comes with noise so a proper data cleansing process should be in place. This involves:

* Removing the empty dates.
* Make sure the appropriate date indexing is in place: Date format.

## Step 2 – Data transformation if needed

There are certain transformations that due to statistical reasons (normality and homoscedasticity) could have the statistical properties that are more in correlation with the variable that wants to be explained.

## Step 3 – Decile ranking sub-Algorithm

The investment cycle labeling algorithm gets its base explanatory power in a decile ranking time series factoring. The factoring consists in a cumulative percentile ranking window

### Sub step 1 – Percentile ranking calculation

A percentile ranking calculation is defined as:

Where and x is the rank of the observation.

### Sub step 2 – Window definition

The intention of the percentile ranking calculation is to provide a distribution of stable percentile ranks through time, in order for a distribution to be stable over time (and a percentile distribution particularly) it has to have enough data to be considered robust.

A first burnout period of accumulation of data is defined as 7 years. This is period is defined due to the nature of the empirically research facts of the economic cycles (the normal average period of the economic cycle is 7 years).

### Sub step 3 – Accumulation ranking

The final step is to begin the rolling window of accumulation on the percentile ranking calculation to have a percentile ranking on each observation.

## Step 4 – Trend-extraction smoothing sub-algorithm

The decile ranking algorithm delivers a discrete time series with a lot of jumps, due to the nature of this time series, for the information to be trustworthy a filtering trending algorithm has to be implemented.

There is a theoretical result that is pretty useful; basically the Theorem states that every time series could be divided in a deterministic and a non-deterministic component.

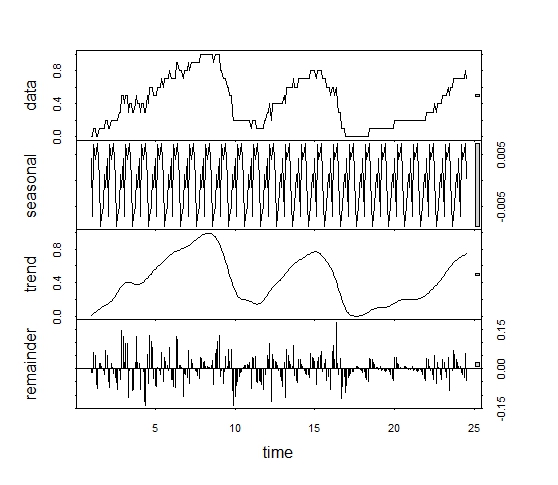
The theorem could be resumed in the following equation:

Where:

The algorithm being used is the one presented by Robert Cleveland, William Cleveland, Jean McRae and Irma Terpenning in the Journal of Official Statistics in 1990.

The trend component will be smooth and have been filtered the seasonality and white noise from the series.

An example of the decomposition is:



The trend component is the one that is going to be extract and taken as the cyclical time series to model. As it could be seen is smooth and de-noised.

## Step 5 – Critical Turning points – Sub-Algorithm

After getting the properly smoothed percentile rank time series in place, the next step is to define and identify the critical points of the series.

A critical point is defined by the help of its first derivative.

***Note:*** *This is the reason why a smoothed time series is needed, if the time series was not smoothed there might be no derivative in certain points, the derivative could head towards infinite or it could be 0 for a lot of consecutive periods and do not provide directional information.*

A critical point of a function given its domain is defined as the point where it first derivative is 0, within the domain. However, since we are dealing with stochastic data, that has noise and it has not defined a deterministic functional form an alternative approach should be taken.

The previous step provided a numerical approximation of the smoothed trend function of the time series, by definition; this smoothed function is continuous in its entire domain (the points where we have a numerical approximation of the series). Given this fact, we can assure that a derivative will exist in all the points, what is left is to calculate and use it for the critical point determination.

Now that the existence of the derivative has been assured, the first derivative is calculated as the difference between contiguous points, this definition will provide an approximate numerical value of the derivative for each point (It could be interpreted as the slope of the curve in that point). However, the 0 value of the derivative has 0 probability of appearance (Since this is a numerical approximation of the real function), so a critical point could not be defined in the classical mathematical sense but in a numerical Newton-Raphson sense:

***“A critical point is defined as the observation where the first derivative changes of sign”***

This definition of the derivative is heavily used in all kind of root search algorithms, and is particularly used in the famous Newton-Raphson one. The result comes from a corollary of the medium value theorem.

## Step 6 – Cyclical relevant turning points – Sub-Algorithm

Having defined all the critical points that the smoothed time series could have, there is a need for filtering the points that define a non-cyclical turning point as it could be seen below:

There are critical points that only represent a minor decrease in the cycle (a short term swing of the economy), this could provide a false economic period if taken into account.

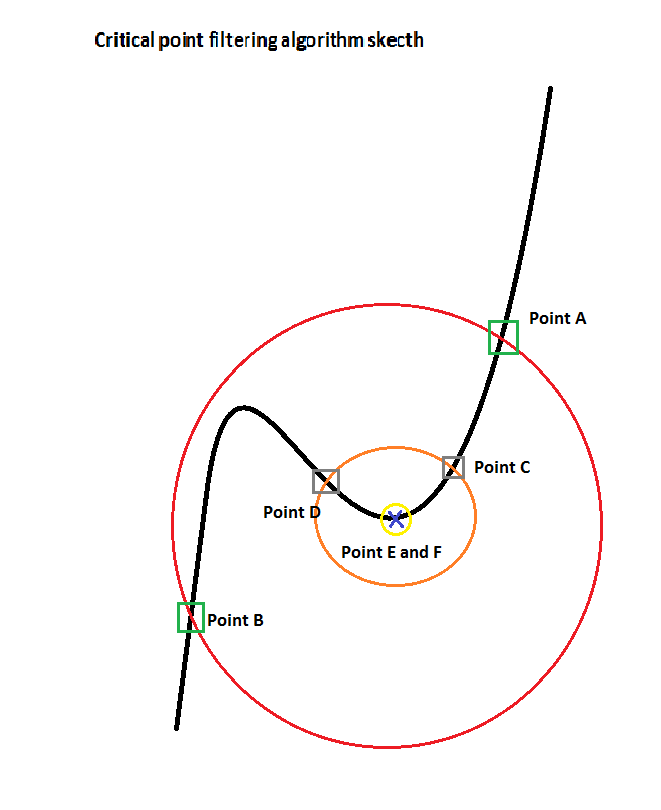
The aforementioned problem raises the need for a critical point filtering algorithm. The critical point filtering algorithm should take into account the relevance of the critical point within its neighbor and determine if the evaluated critical point will be considered a relevant cyclical turning point.

The filtering algorithm is conceived as:

***“A critical turning point is considered a cyclical turning point if the critical point presents a change in its derivative for at least X periods”***

This is better explained in the following charts.

The following chart will be the base for the critical point filtering explanation:



A typical critical point (as defined by the previous algorithm) is defined in the yellow neighbor as the point where a change of sign occurs between the **Points E and F.** However in the case of this particular critical point it will provide a false positive for a cyclical turning point, the filtering algorithm is designed to filter this type of points. The algorithm consists in filtering out the point that does not comply with the criteria of slope change in a bigger neighbor; it will be outline as follows:

* The Point is selected as critical due to the change in the first derivative taking as reference the derivatives in **Points E and F**.
* The Point will be **NOT** filtered if for a given neighbor the change of the first derivative is **STILL** in place. This case is represented as **the Point C and Point D** situation, it could be seen that the slope of the two points differ in sign.
* The Point will be filtered if for a given neighbor the first derivative does not change sing. The point will not be selected as a cyclical point. This is the case of **the Point A and Point B** where it could be seen that the derivative takes the same sign in both points.

***Important note:*** *From this algorithm one hyper parameter to calibrate arises, the deep duration (neighbor radius), the optimal solution for this parameter could be answered with the following question:*

*“How deep should a cycle change need to be considered a cycle change?”*

## Step 7 – Business Cycle definition