# Macro Data Clustering

AARSH SACHDEVA

# Table of Contents

# Overview

Since business cycles are natural, repetitive parts of the economy, economic data in one time period should resemble something we've encountered in history

- Eg: Data collected during recessions will most likely reveal low growth, low inflation, negative sentiment, high unemployment, etc.

An interesting investment strategy would be to compare current economic data to historical data and identify the investment strategies most likely to perform best given macroeconomic conditions

This project utilized approximately 66 different economic indicators to create 6 composite indicators that were then analyzed against historical data for clustering and strategy selection

I was inspired to pursue this project after coming across the idea in "Big Data and AI Strategies", a paper published by J.P. Morgan in May 2017

# Data Collection

# Economic Indices

Most of the data was collected on a Bloomberg Terminal. Others were collected from publicly available website (OECD).

66 raw features were used in this project with some data beginning in mid-1920.

Raw economic data however, needs to be manipulated in order to be useful as an indicator. This is why feature engineering was so essential to this project.

# Feature Engineering Process

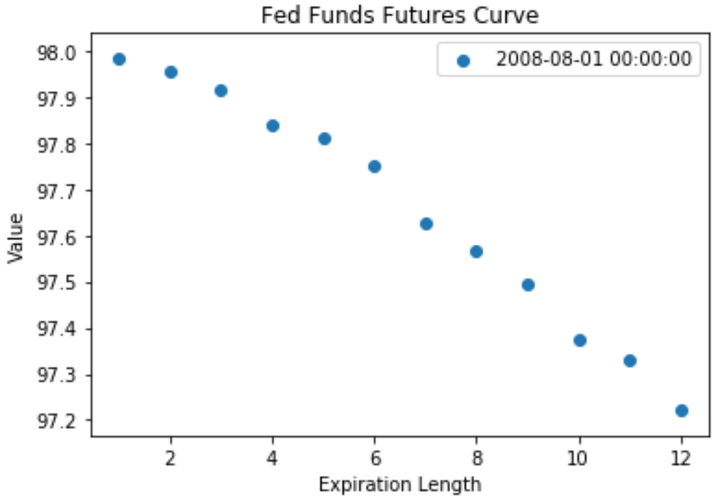
Prior to clustering data or creating composite indicators, it's important to make sure all the data is scaled.

Furthermore, when dealing with time series data, it's often important to make the data stationary or at least detrend the data, so it can be useful for future prediction or classification purposes.

Careful thought must be given to each feature and the targeted result in order to create high quality indicators.

# Feature Engineering Process

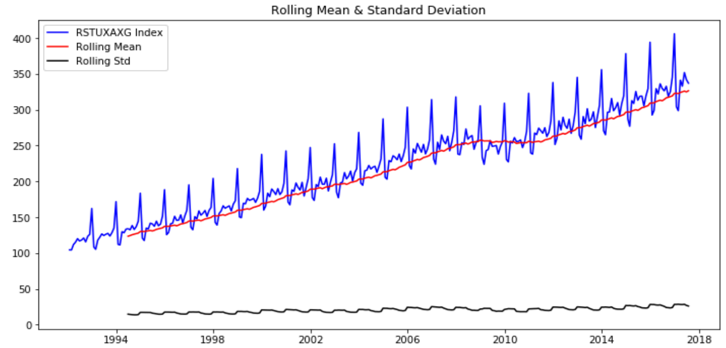
1. Understand what economic indicator is desired and what features must be used to produce that indicator
   - Often times, the indicator is a single feature itself. Other times, it must be calculated using multiple inputs.
   - Eg: An estimate of the term structure in Fed-Funds Futures requires regression analysis over different expiration dates (I used the coefficient attached to the quadratic term for every date).



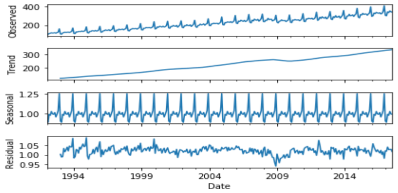
# Feature Engineering Process

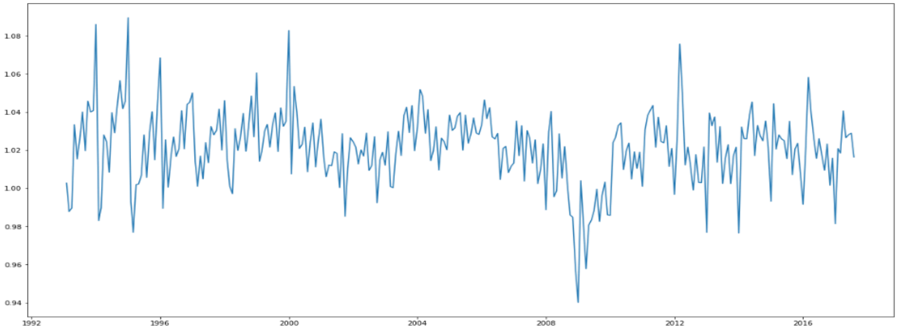
2. Analyze the time series of the indicator and determine whether it needs to be made stationary, and if so, what the best method to do so is (differencing vs decomposition).
   - Note: it's important to exercise careful judgement in this step, as it doesn't always make sense to manipulate certain data. Eg: You wouldn't want to detrend survey data, even if it isn't stationary
   - Also important to only use one-sided decomposition to avoid look-ahead bias

Retail Sales ex. Transportation Before and After Decomposition




```
Results of Dickey-Fuller Test:
Test Statistic                   0.176403
p-value                          0.970907
#Lags Used                      15.000000
Number of Observations Used    291.000000
Critical Value (1%)             -3.453023
Critical Value (5%)             -2.871524
Critical Value (10%)            -2.572090
dtype: float64
```



# Feature Engineering Process

3. Normalize all variables, this is necessary for clustering and composite indicator creation
   - Since the end goal of this project is to conduct a backtest, I split up my data with a 70/30 split. The training set ranged from approximately early-1997 to 2011 and the test set included everything from then on.
   - The training was used to calculate the "population" mean and standard deviation, which was then applied to both sets of data for normalization purposes. This way, look-ahead bias was avoided in my test set.
   - Note: I thought it would be best to remove all rows where at least 25% of my features were missing, which is why the training set began in 1997.

# Feature Engineering Process

4. Group features into 6 categories to create the composite indicators. This required analysis of the economic interpretation of the variable itself.

# Sentiment Features

ISM Manufacturing PMI

Michigan Consumer Sentiment

Barclays Credit Default Swaps 5Y and 10Y Spreads

10Y Bond Yield (detrended, multiplicative)

Fed Funds Futures Term Structure (estimated by quadratic regression coefficient on each date)

10Y/2Y Bond Yield Spread

Credit Managers Index

Loan Officers Survey (Average of collected surveys)

# Growth Features

Initial Jobless Claims ($1^{st}$ order differencing with a 52 week lag)

US Leading Indicator (decomposed, additive)

Global Leading Indicator (decomposed, additive)

Baltic Dry Index

Manufacturing New Orders ex. Transportation (decomposed, multiplicative)

New Housing Units Started

Dow Jones Transportation Average (12M Change)

Commercial Loan Growth

# Growth Features

Railroad Freight Index (12M Change)

Global GDP (12M Change)

Moody's BAA-AAA Spread

Barclay's High Yield Spread

M2 Money (12M Change)

House Price Index (12M Change)

ISM Prices Index (12M Change)

Leading Indicators-Lagging Indicators Spread

# Coincident Features

Citi US Economic Surprise Index

Citi Global Economic Surprise Index

ISM Non-Manufacturing NMI

Retail Sales ex. Transportation (decomposed, multiplicative)

# Quant-Macro Features

Margin Debt Level (12M Change)

US to World (detrended, additive)

VIX

Small-Cap to Large-Cap Outperformance (12M Change)

SPX to MZM 1Y Rolling Correlation

# Inflation Features

Purchasing Power Index (12M Change)

Import Price Inflation

Capacity Utilization

USD to World Trade Weighted Index

Unit Labor Costs (12M Change, detrended, additive)

Wage Trend Index (12M Change)

WTI Oil Price (decomposed, multiplicative)

Breakeven 10Y Inflation

Commodity Price Index (decomposed, additive)

# Liquidity Features

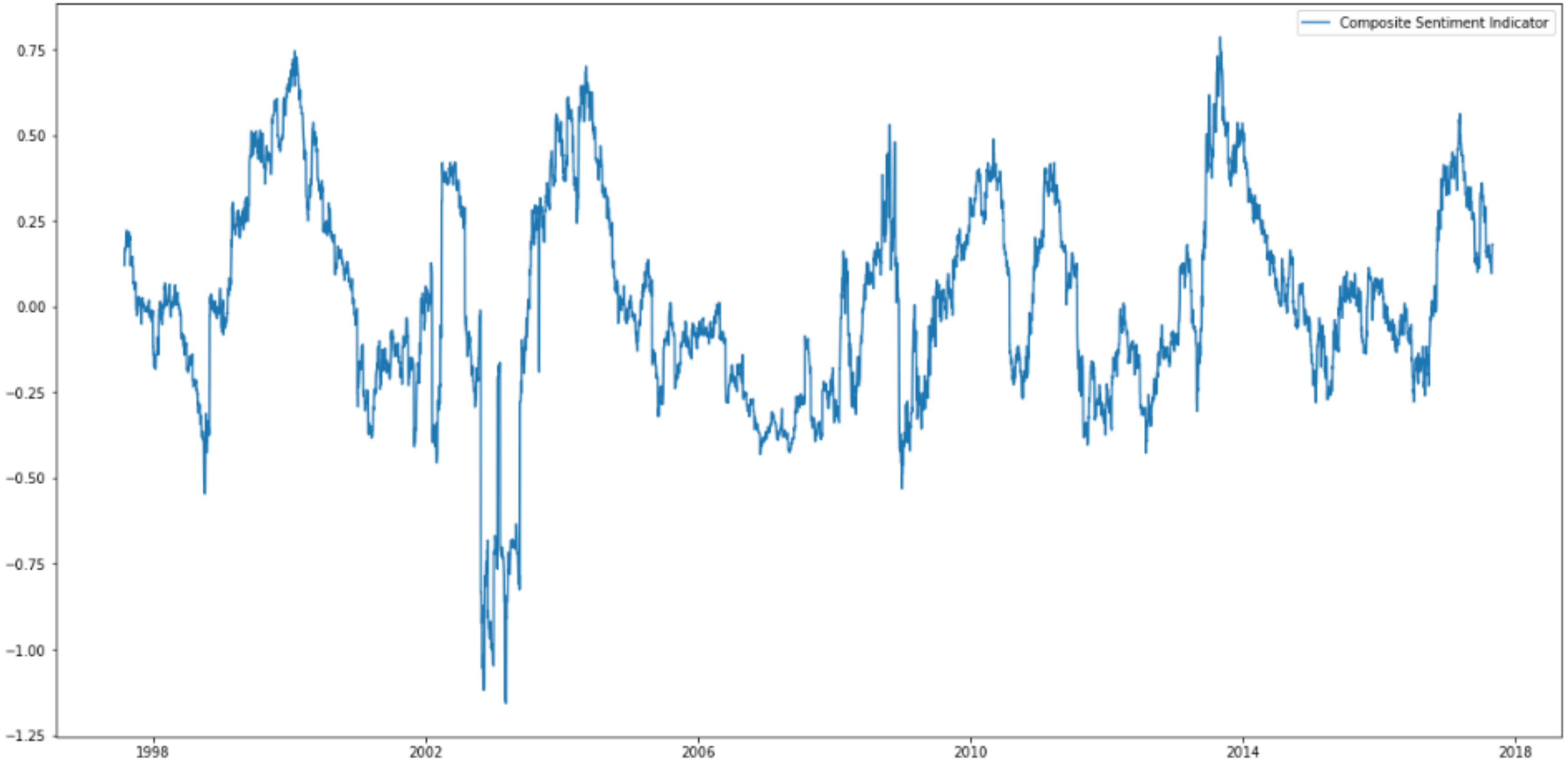
Average Treasury Volume to Market Debt (detrended, multiplicative)

Citi US Market Liquidity

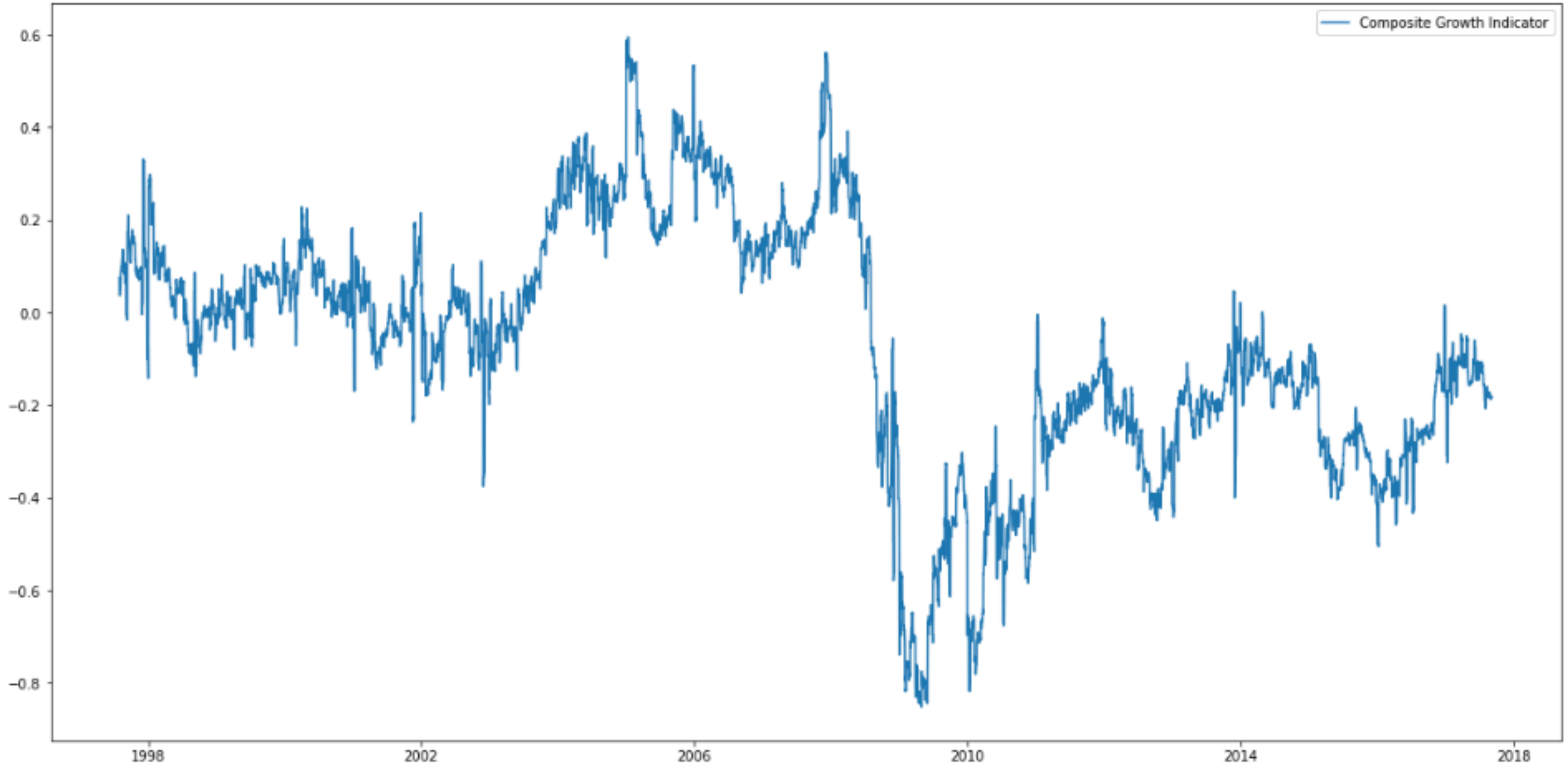
# Feature Engineering Process

5. Once the normalized features had been grouped for the composite indicators, a simple average was taken for each date to create the composite indicator's value

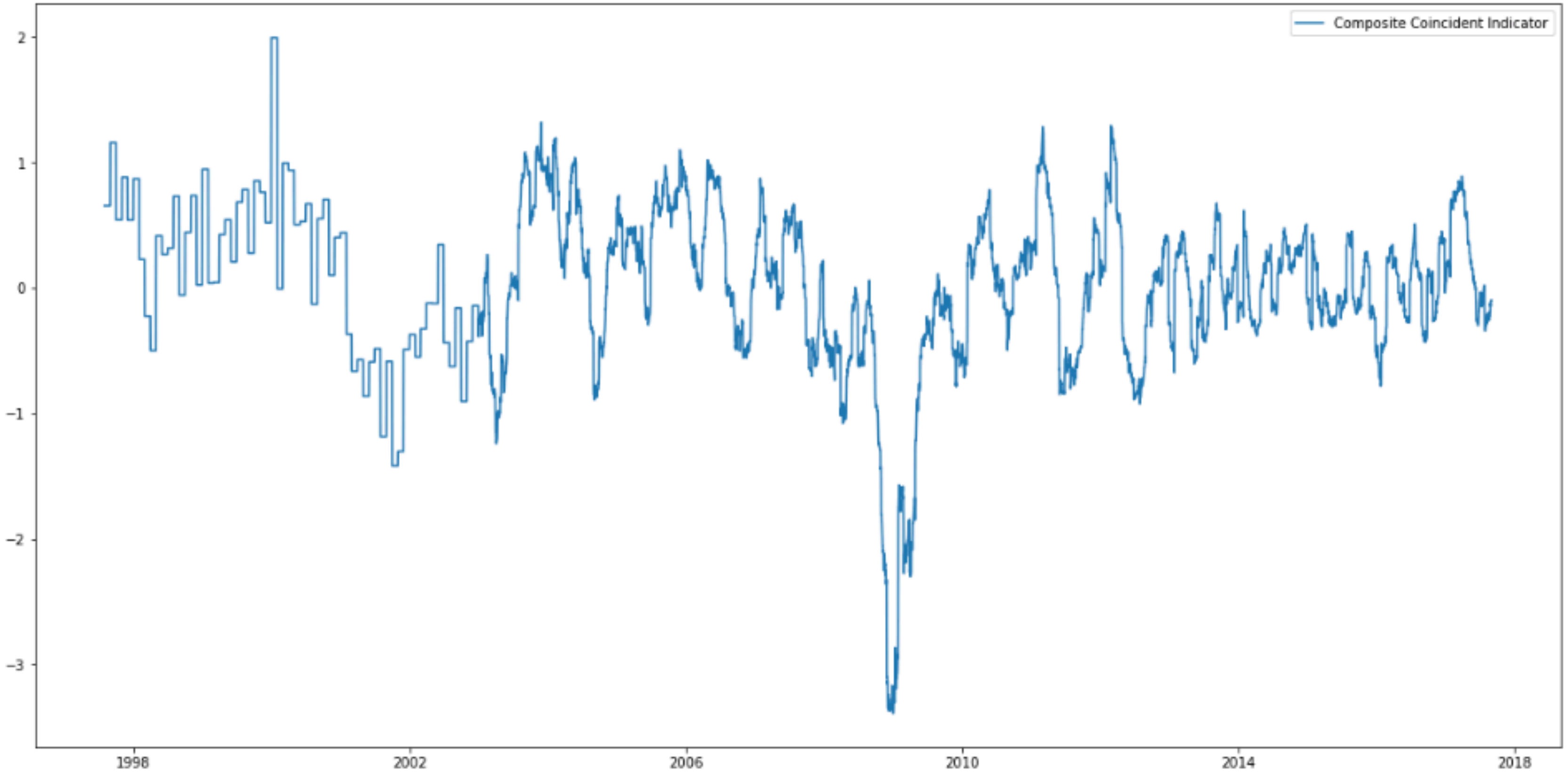
# Sentiment Indicator



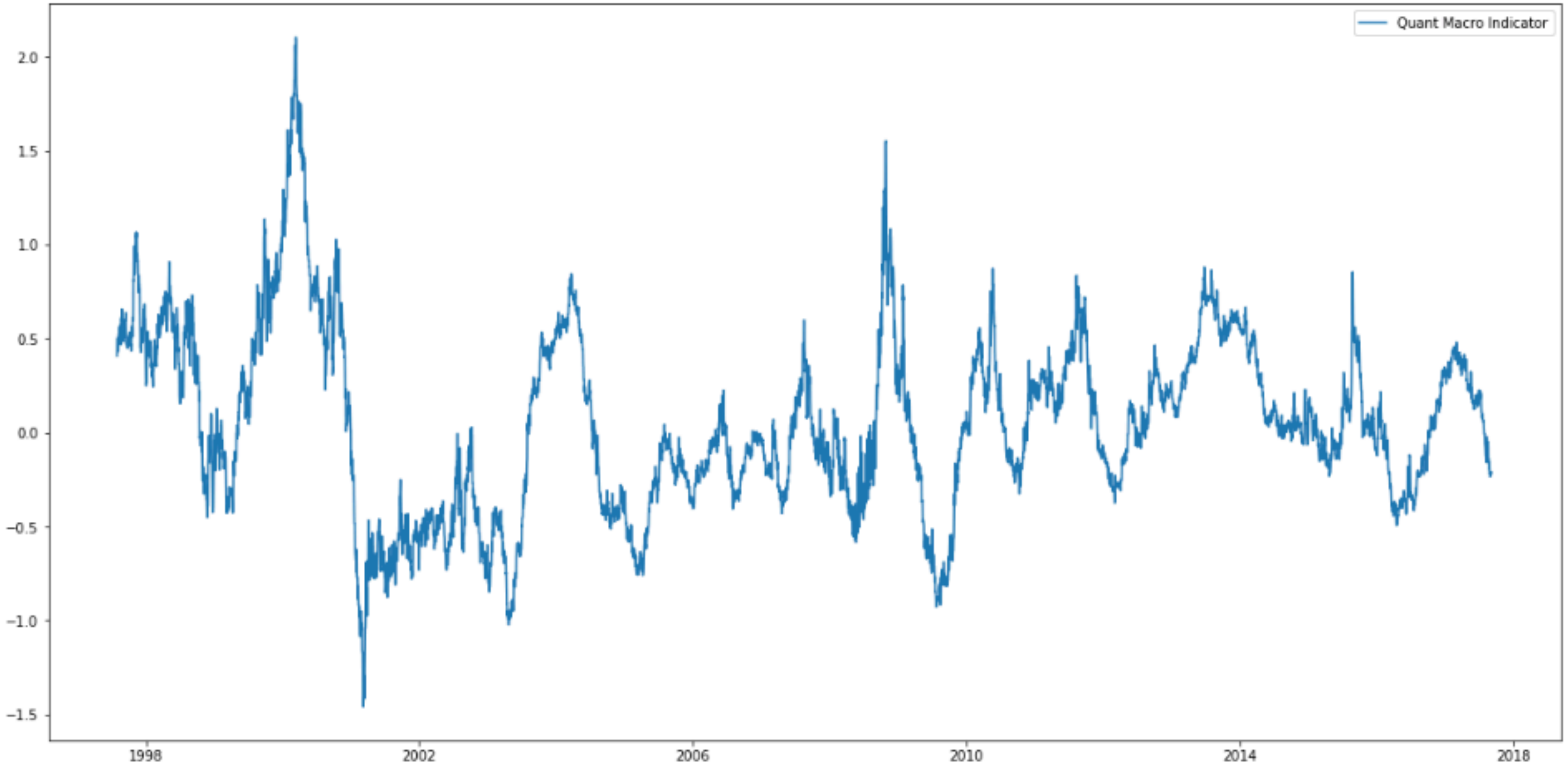
# Growth Indicator



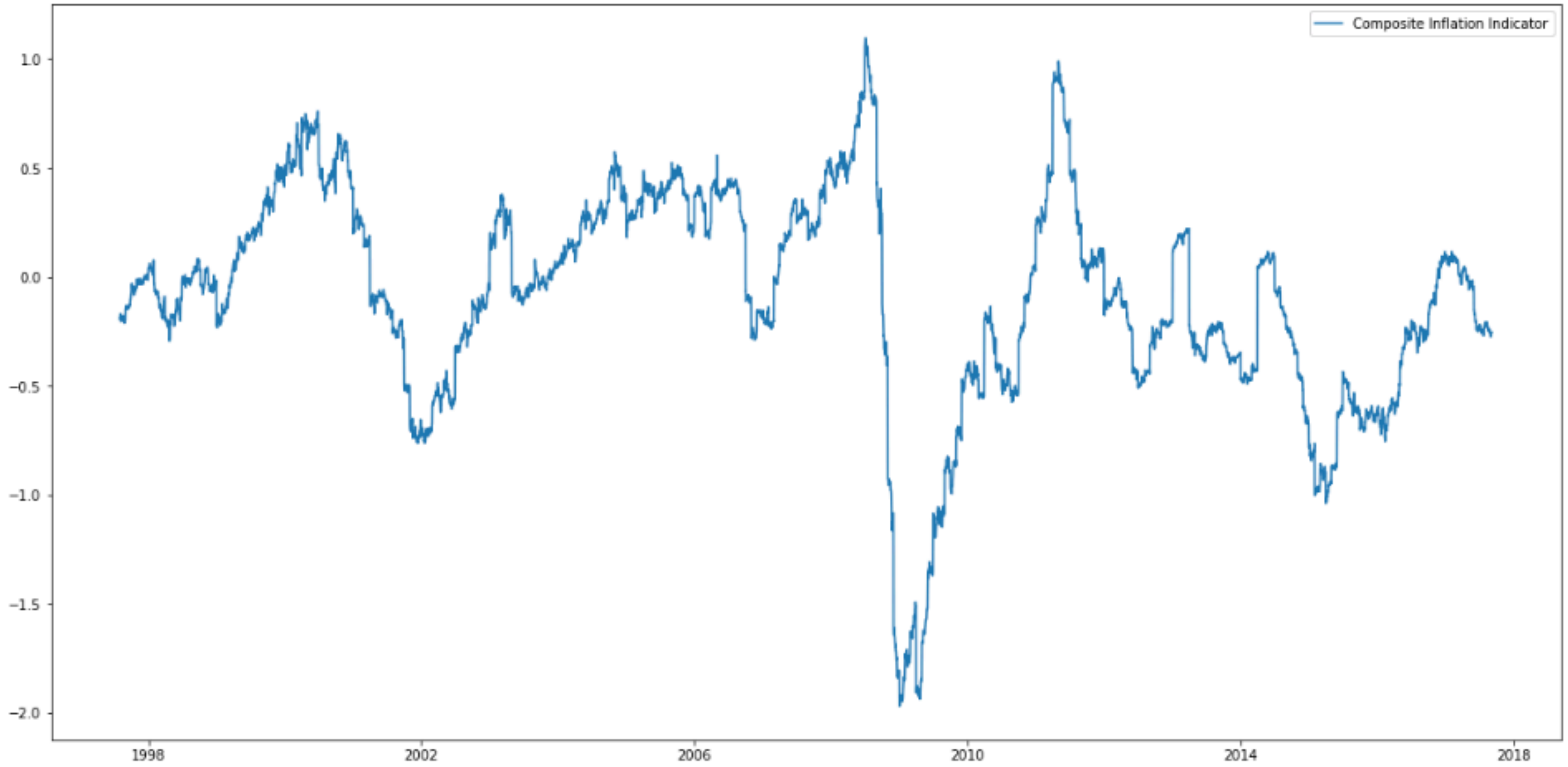
# Coincident Indicator



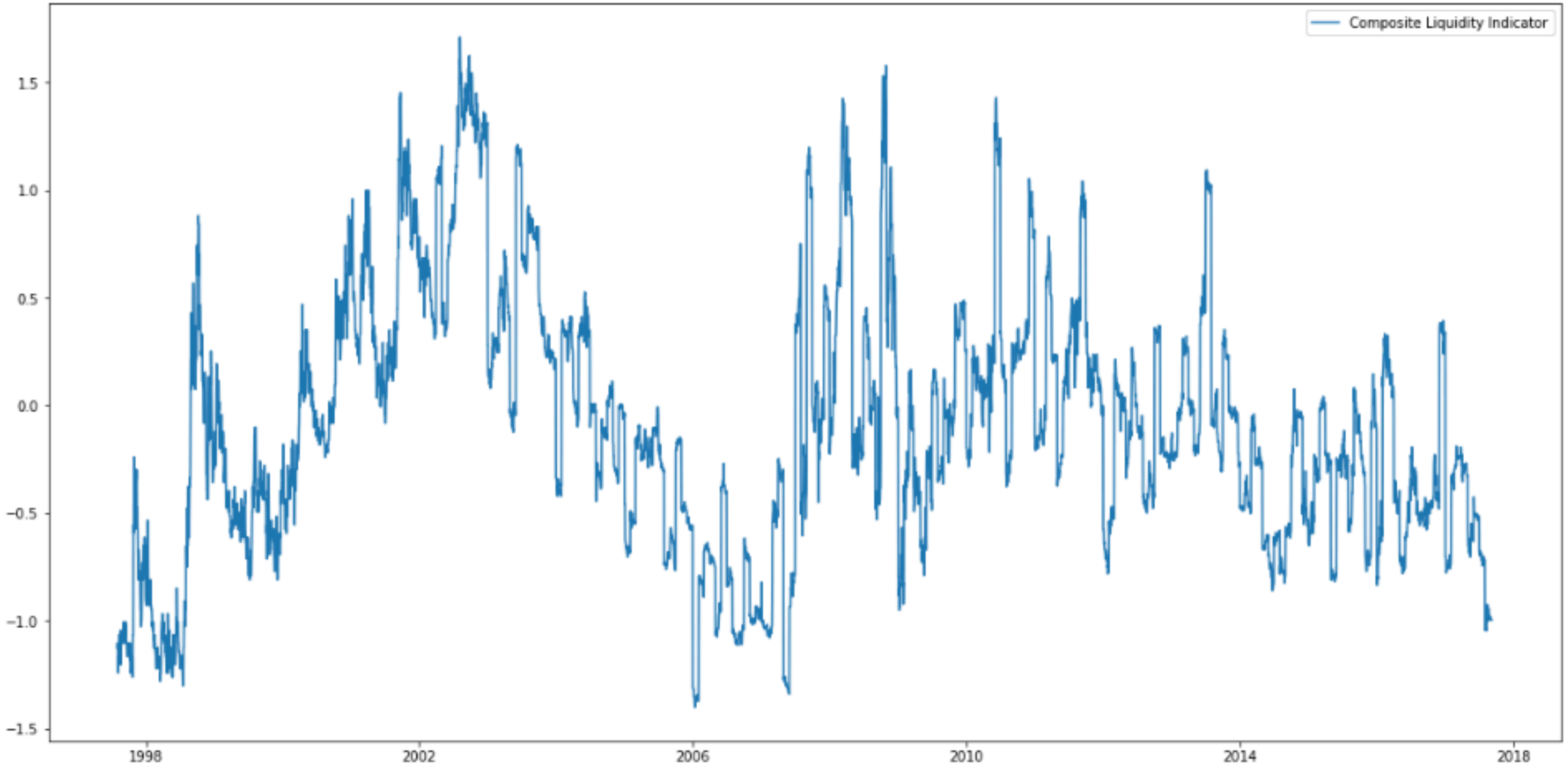
# Quant-Macro Indicator



# Inflation Indicator



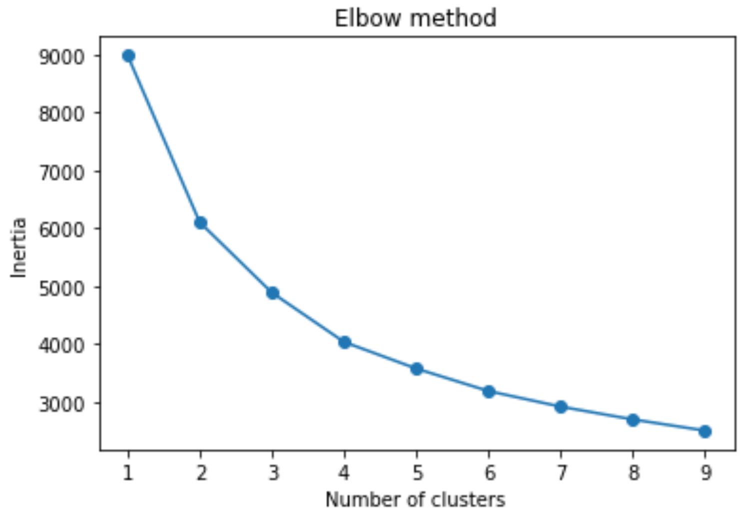
# Liquidity Indicator

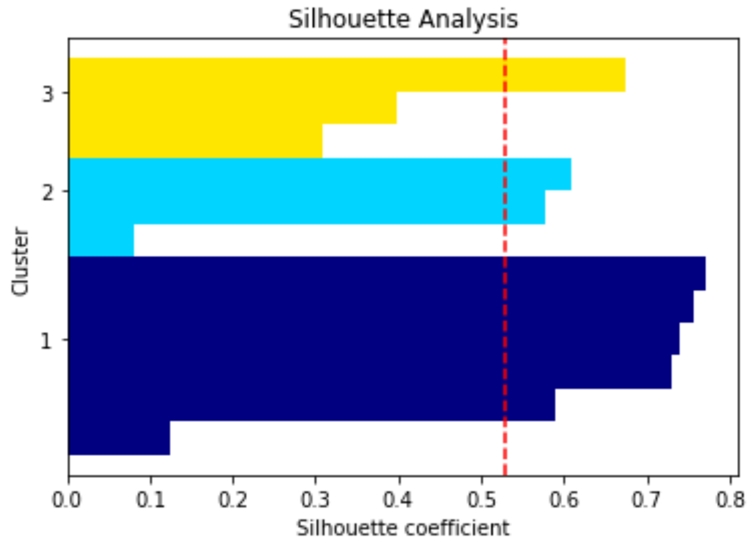


# Clustering

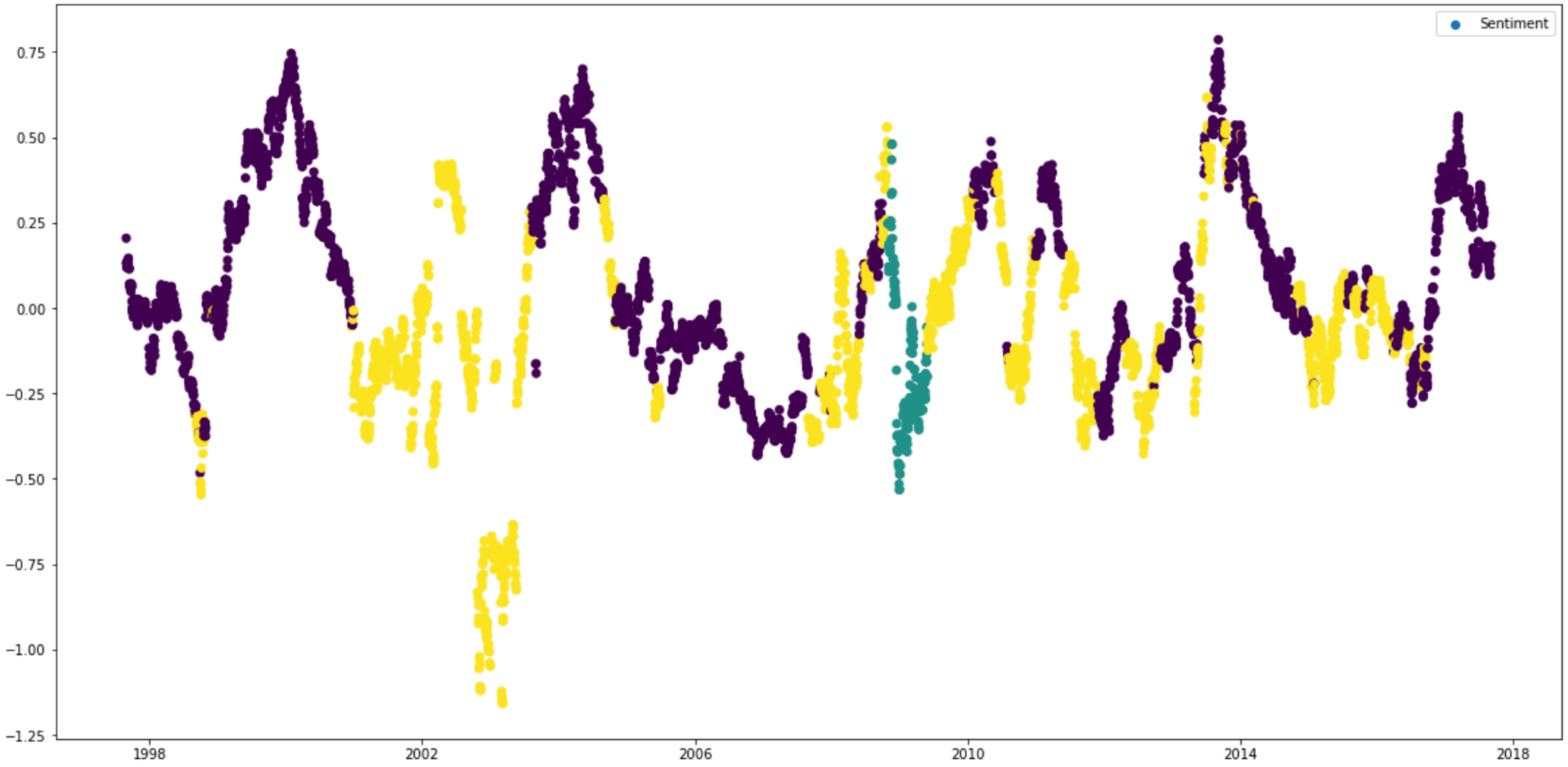
# Number of Clusters

I decided to go with 3 clusters

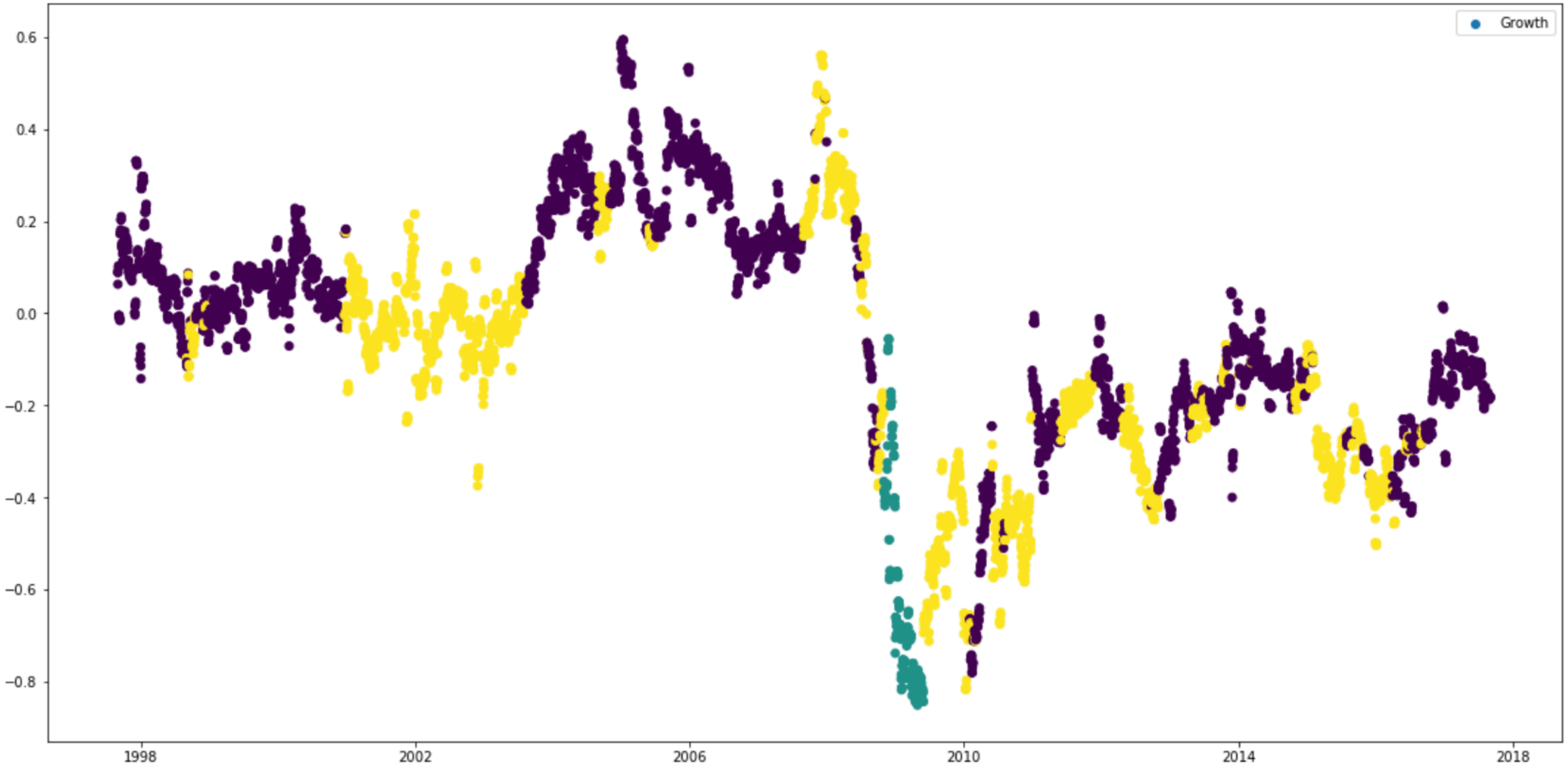


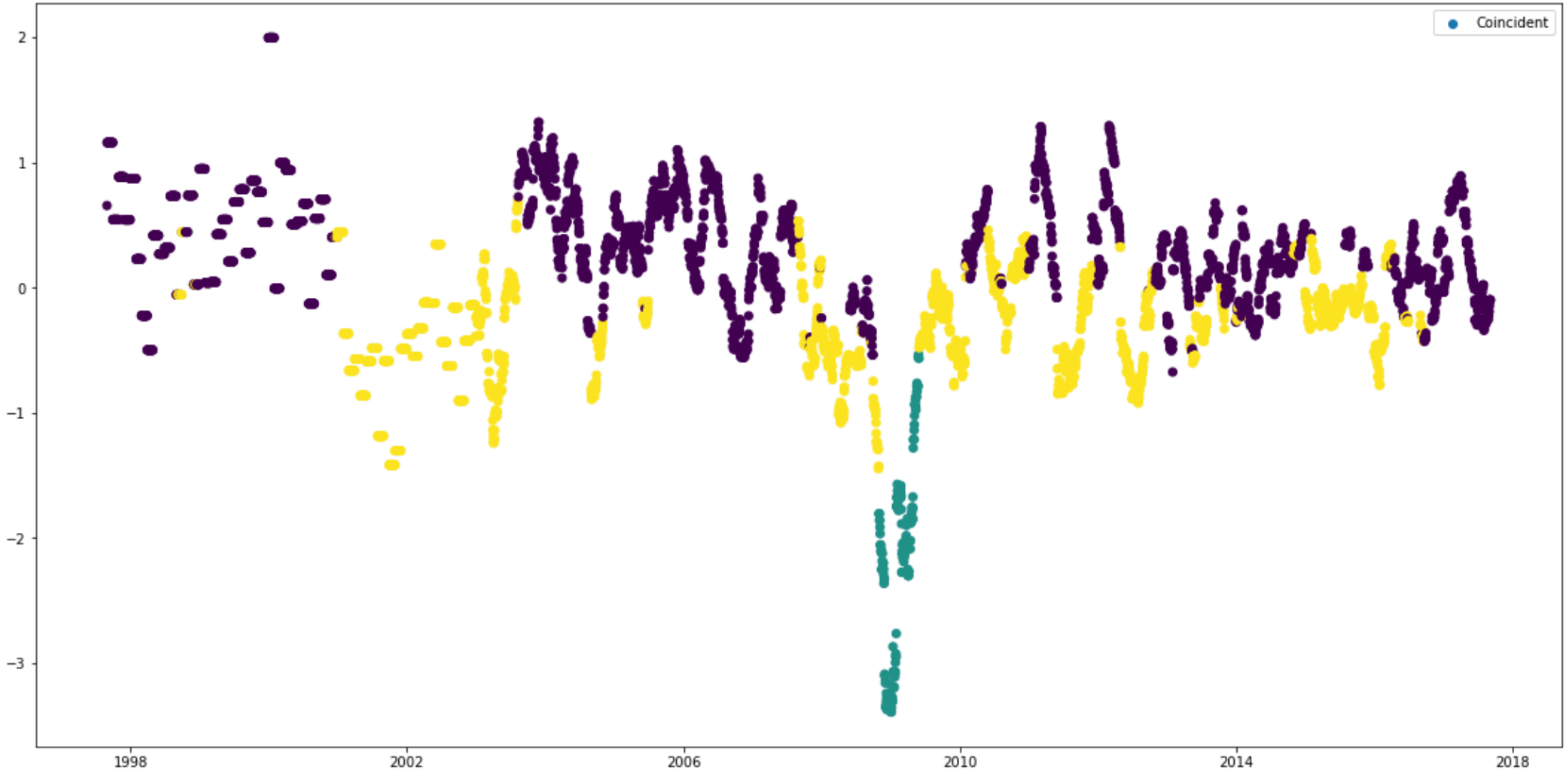
# Sentiment Indicator Clustered



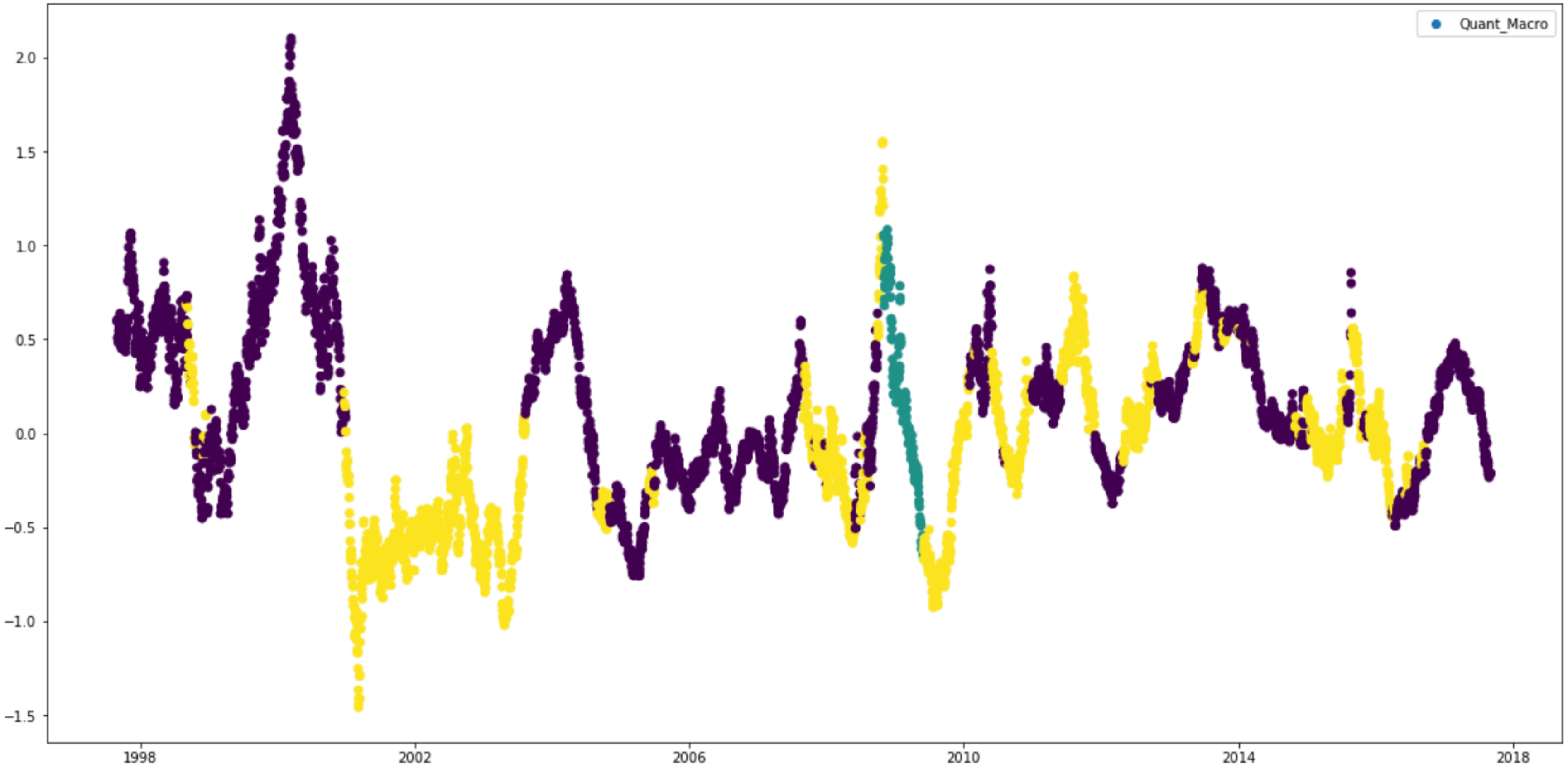
# Growth Indicator Clustered



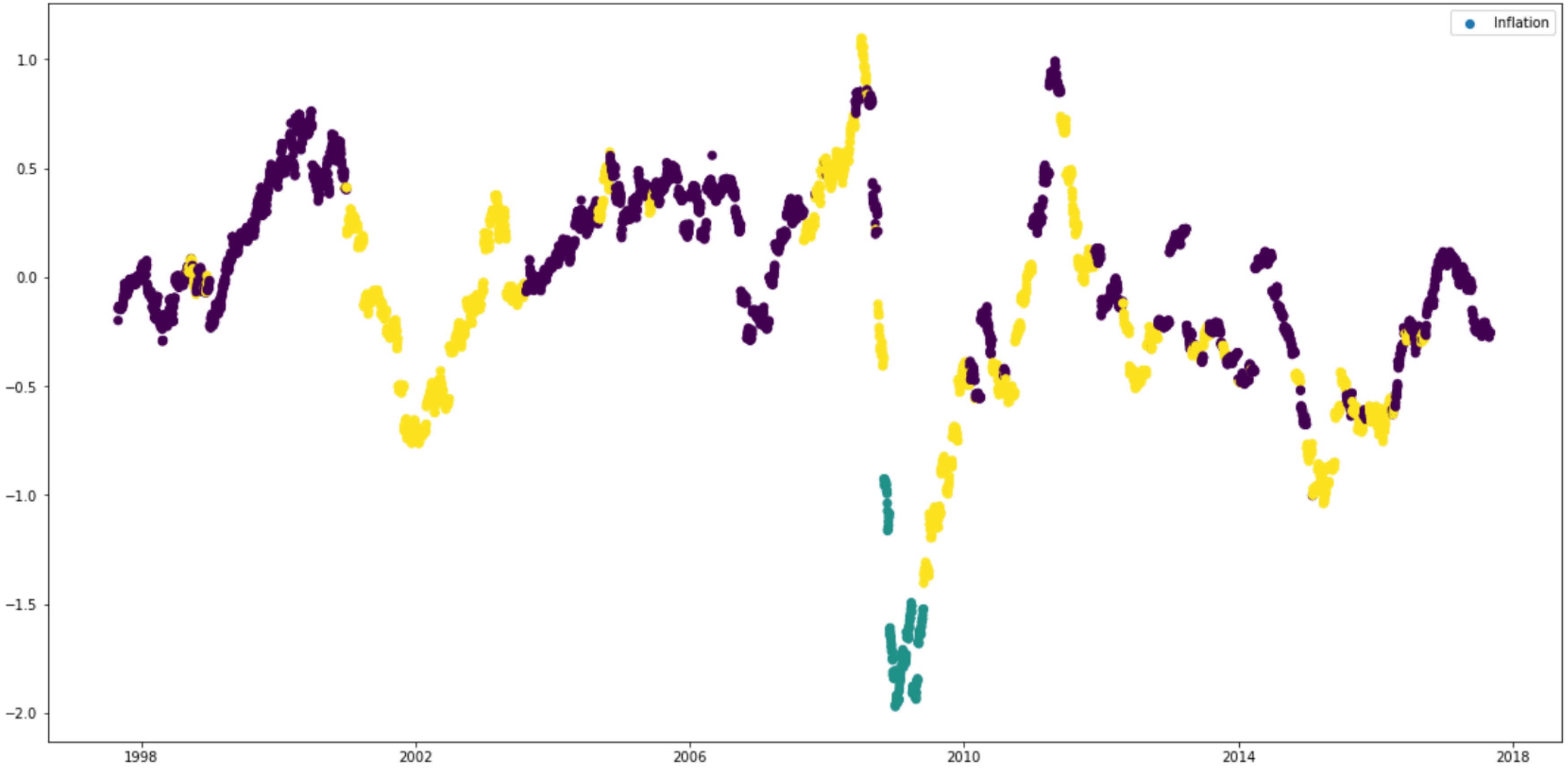
# Coincident Indicator Clustered



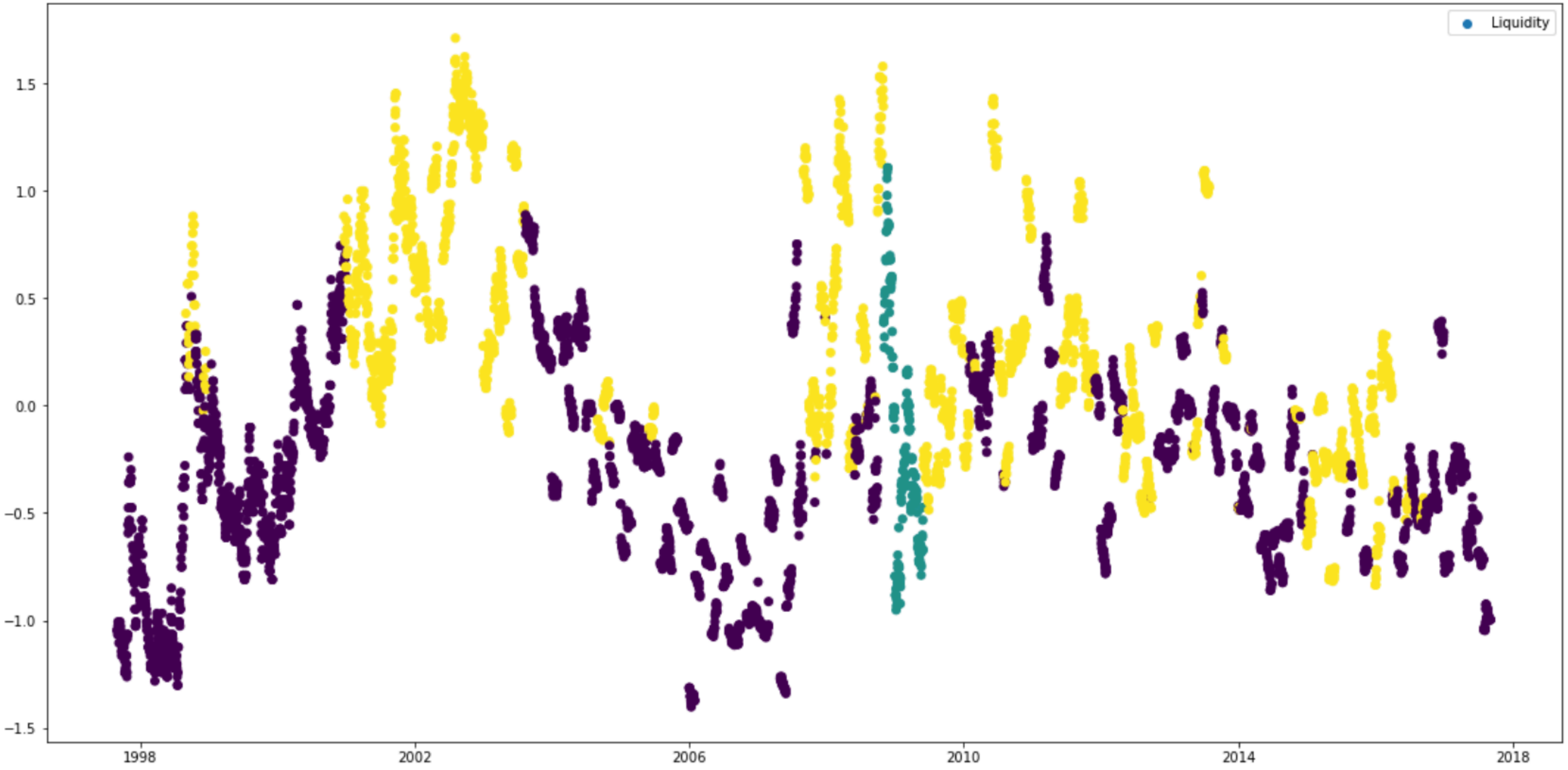
# Quant-Macro Indicator Clustered



# Inflation Indicator Clustered

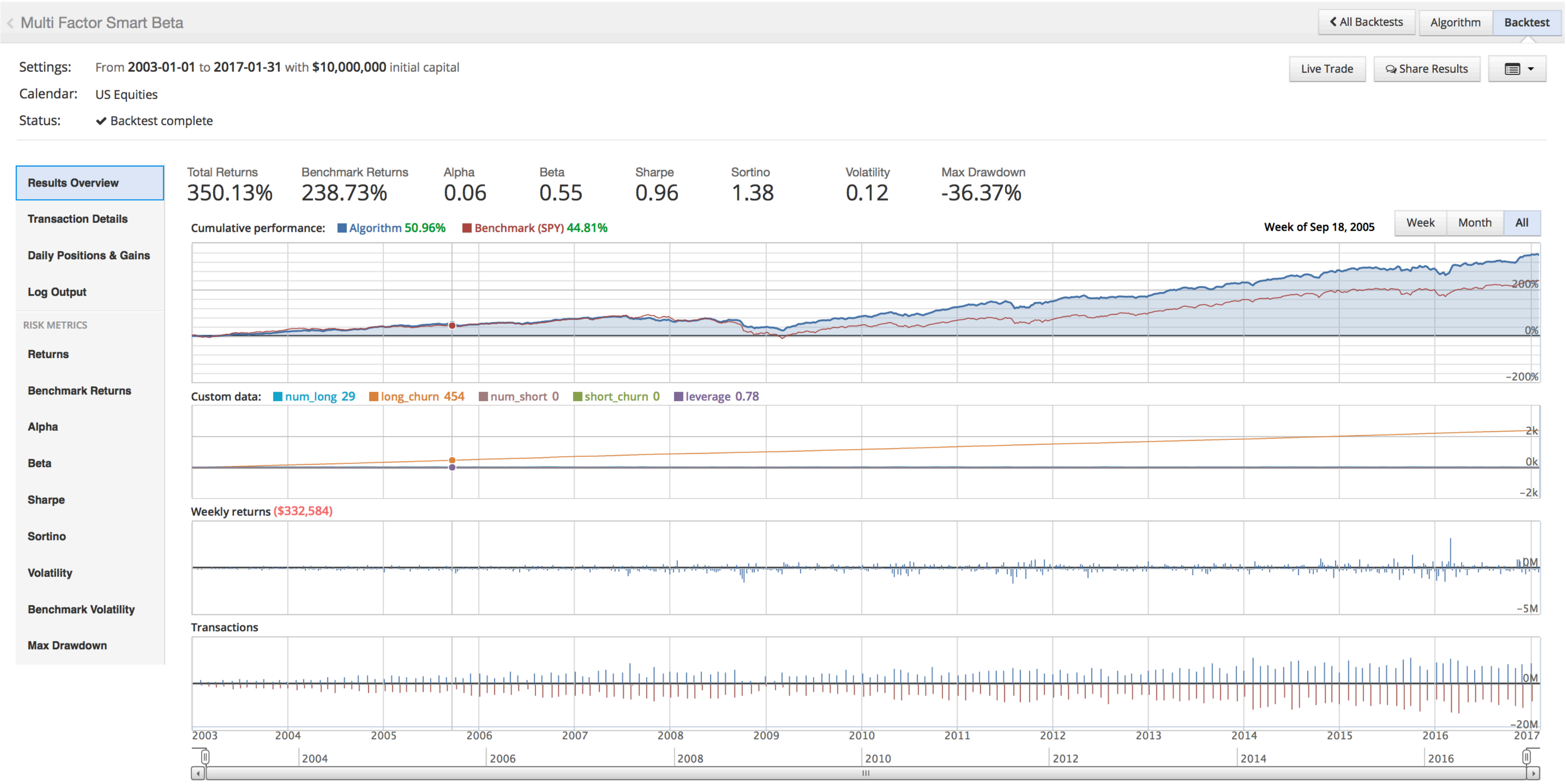


# Liquidity Indicator Clustered

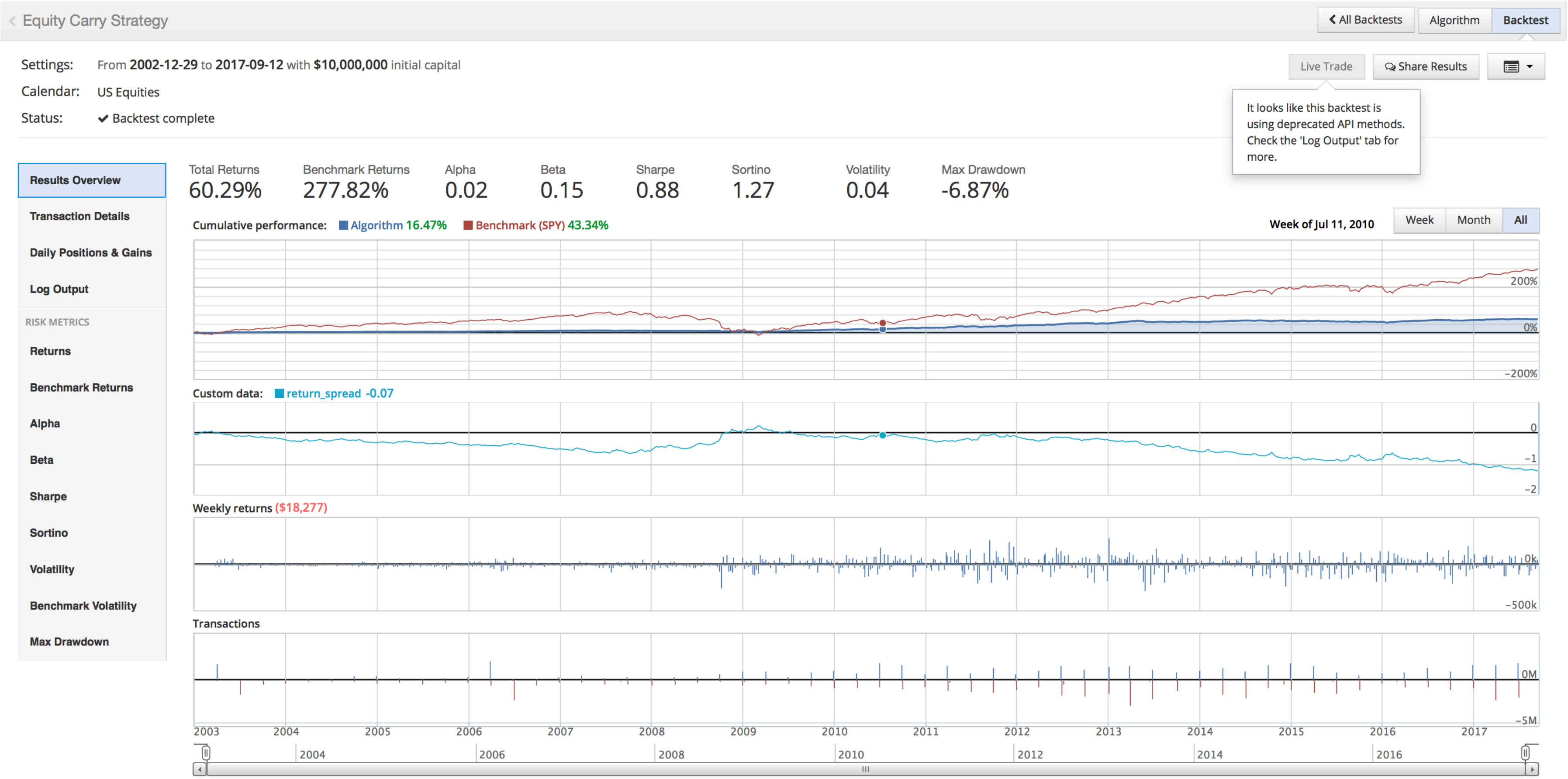


# Quantopian Strategies

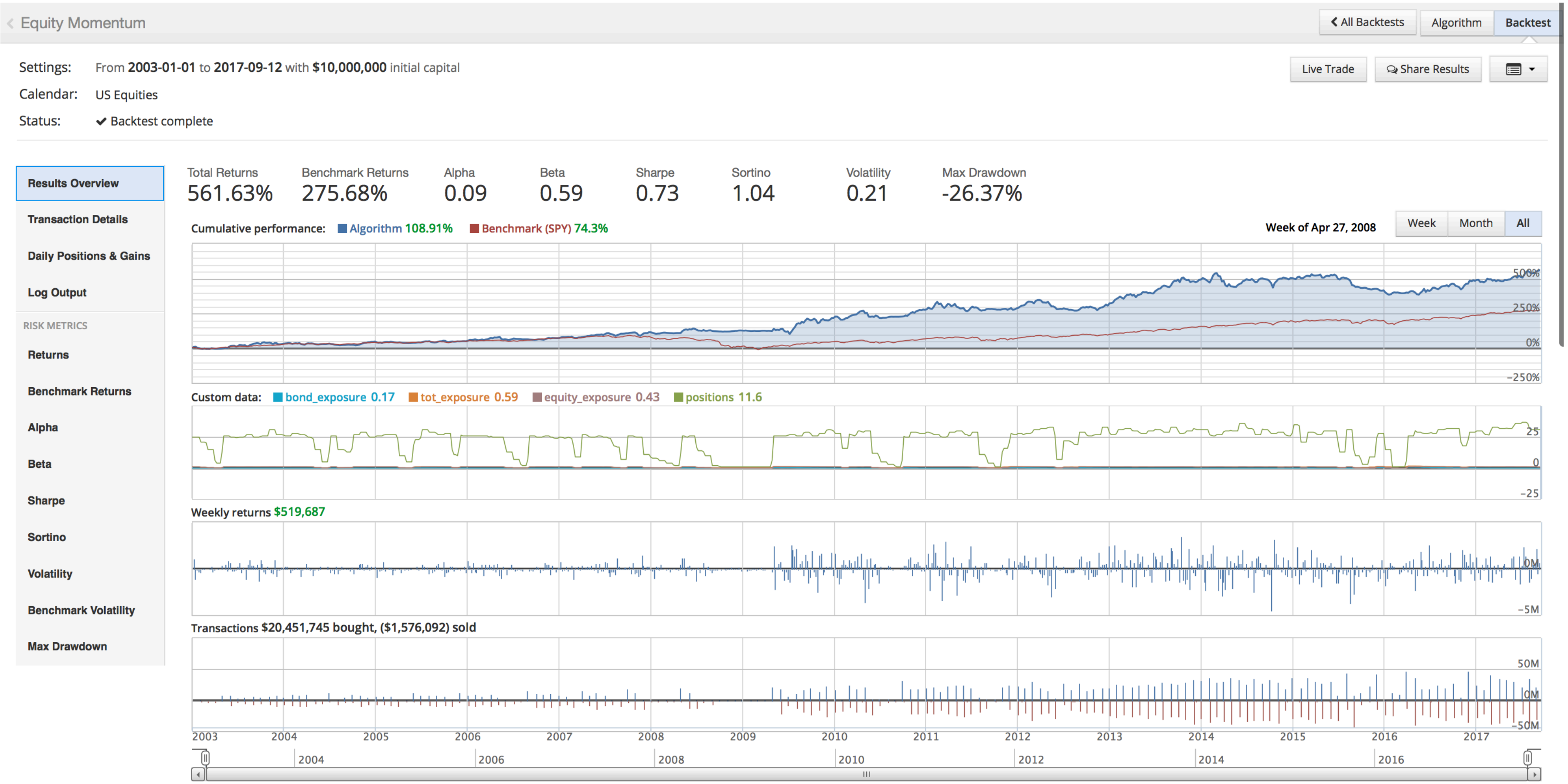
# Multi Factor Smart Beta



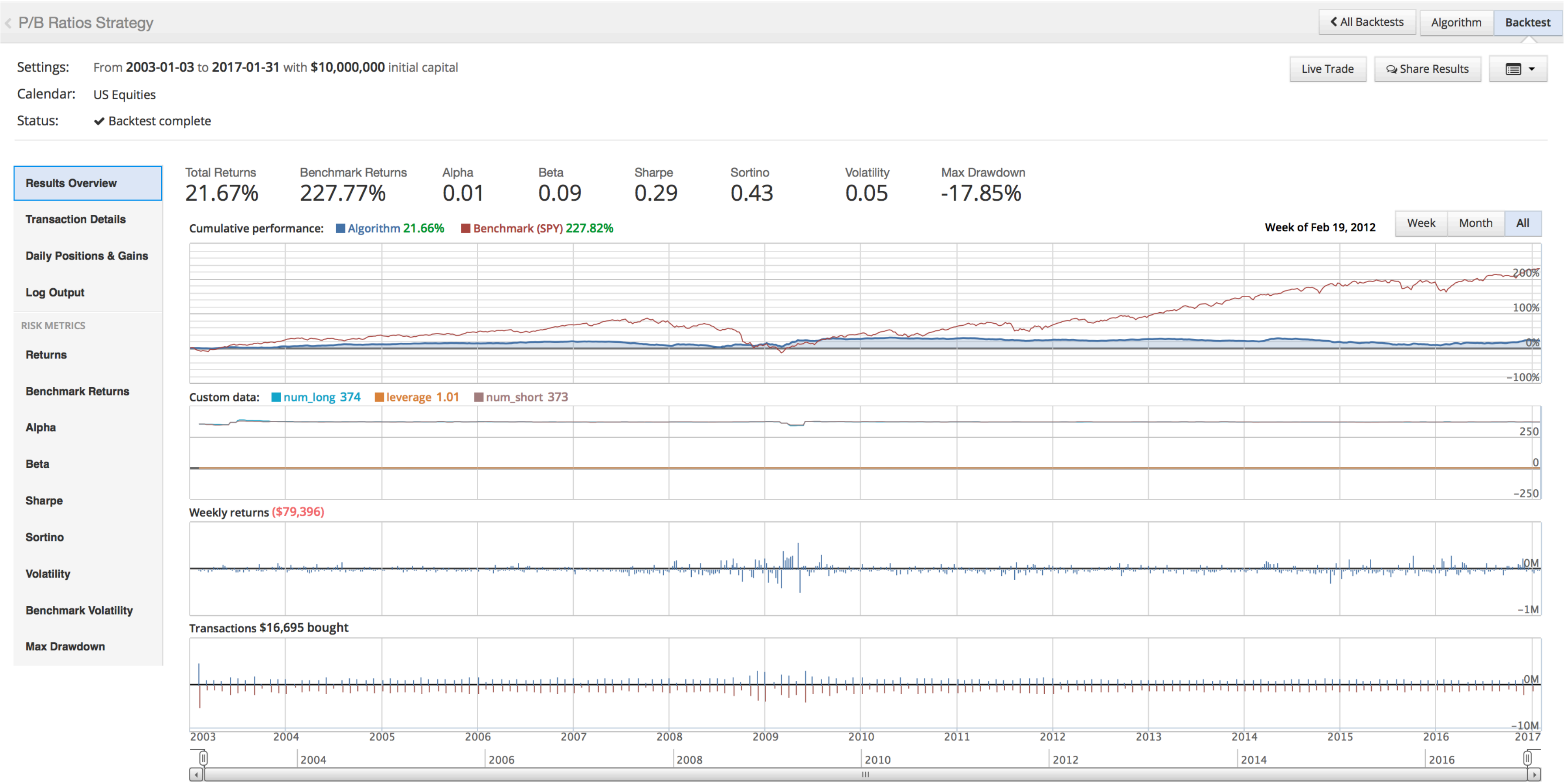
# Equity Carry



# Equity Momentum



# Value Strategy



# Next Steps

# Backtesting

Get my normalized composite indicators uploaded to Quantopian (I've been having trouble with that)

For each month, calculate returns for all 4 strategies

Start a test where the strategy uses K-Nearest Neighbors to find the closest macro points in the past 10 years

Evaluate what strategies worked best during those historical time points

Invest in chosen strategies at the moment and see how returns play out based on this positioning