MACROSYNERGY

Classifying market regimes

By Editor December 25, 2021 21,972

Market regimes are clusters of persistent market conditions. They affect the relevance of investment factors and the success of trading strategies. The practical challenge is to detect market regime changes quickly and to backtest methods that may do the job. Machine learning offers a range of approaches to that end. Recent proposals include [1] supervised ensemble learning with random forests, which relate the market state to values of regime-relevant time series, [2] unsupervised learning with Gaussian mixture models, which fit various distinct Gaussian distributions to capture states of the data, [3] unsupervised learning with hidden Markov models, which relate observable market data, such as volatility, to latent state vectors, and [4] unsupervised learning with Wasserstein k-means clustering, which classifies market regimes based on the distance of observed points in a metric space.

Sources are linked next to the below quotes. Headings, cursive text, and text in brackets has been added.

This post ties in with this site's summary on quantitative methods for macro information efficiency, particularly the section on unsupervised learning.

Why market regimes matter

"Financial markets have the tendency to change their behaviour over time, which can create regimes or periods of fairly persistent market conditions…Modelling various market regimes…can enable macroeconomically aware investment decision-making and better management of tail risks." [Two Sigma]

"It is well understood that return series are non-stationary in the strong sense, and exhibit volatility clustering…An observed sequence of asset returns exhibits periods of similar behaviour, followed by potentially distinct periods that indicate a significantly different underlying distribution. Such periods are often referred to as market regimes…Within the arena of…deep learning-based methods…detection of significant shifts in market behaviour is a central tool for their model governance since it serves as an indicator for the need to retrain the machine learning model". [Horvath, Issa and Muguruza]

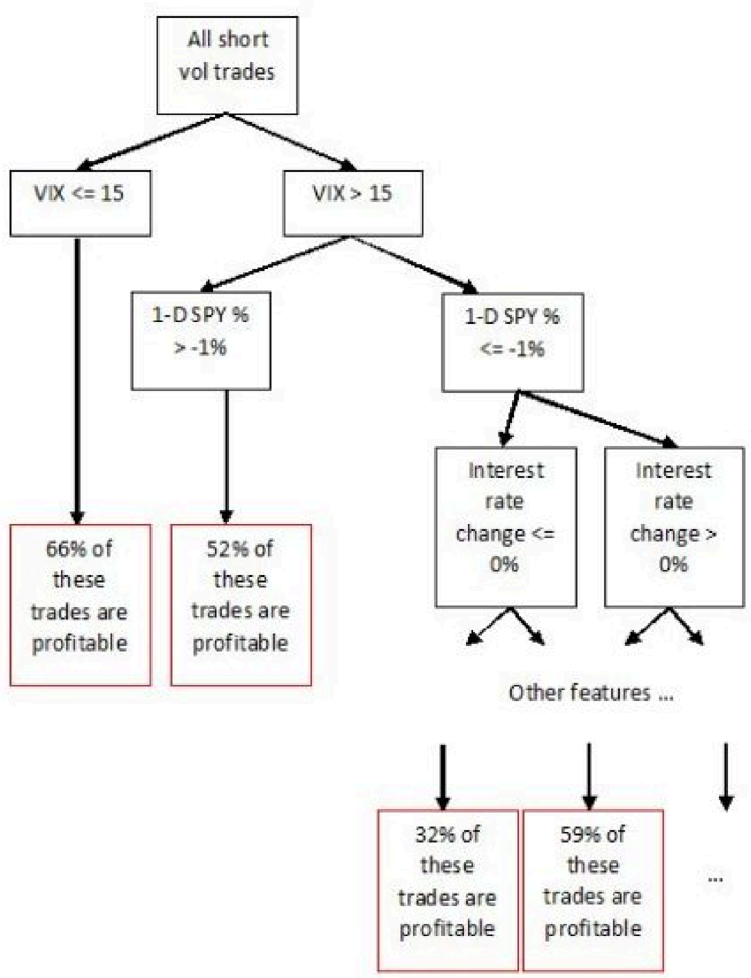
"[For example, if] you are trading a short volatility strategy…during a very calm market, it is likely that your conditional probability of profit would be quite high. If you are trading during a financial crisis, it could be very low. The conditions that can determine the probability [of positive strategy returns] may even be quantifiable." [Chan]

Approaches to classifying market regimes

Random forest

"The way to compute this conditional probability [of a trading strategy yielding positive returns] is machine learning…Intuition tells you that there are some variables that you didn't take into account in your original, simple, trading strategy. There are just too many of these variables, and you don't know how to incorporate them to improve your trading strategy…But that's not a problem for machine learning…The machine learning algorithm will get rid of the useless features via…feature selection." [Chan]

"Let's say we only care about whether [trades] are profitable or not [and] ignore the magnitude of returns.[We] label those trades that are profitable 1, otherwise 0. These are called '**metalabels**' by Marcos Lopez de Prado, who pioneered this financial machine learning technique…The metalabels [indicate] whether those base predictions are correct or not…[For example] a random forest algorithm may discover the hypothetical relationship between VIX, 1-day SPY return and whether your short vol trade will be profitable as illustrated in [the below] schematic diagram." [Chan]

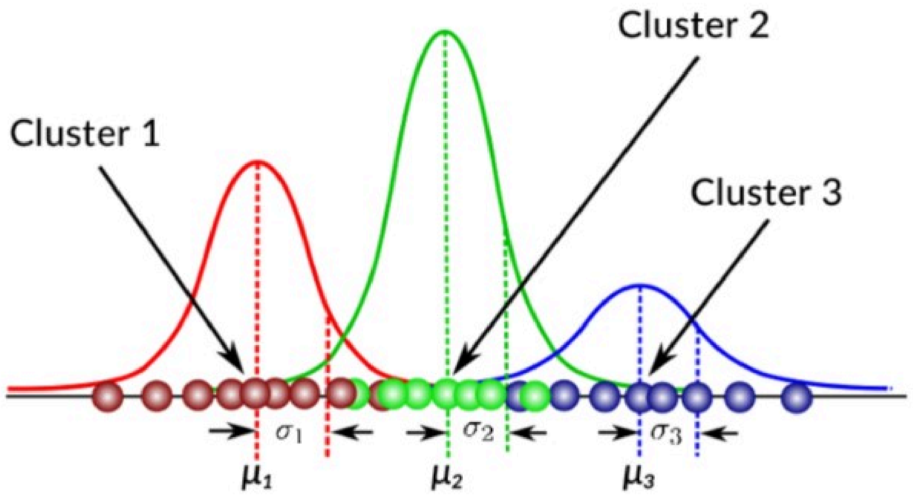

All short vol trades
VIX <= 15
VIX > 15
1-D SPY % > -1%
1-D SPY % <= -1%
Interest rate change <= 0%
Interest rate change > 0%
66% of these trades are profitable
52% of these trades are profitable
Other features ...
32% of these trades are profitable
59% of these trades are profitable
...

"The sklearn.ensemble module includes averaging algorithms based on randomized decision trees [such as] the RandomForest algorithm…A diverse set of classifiers is created by introducing randomness in the classifier construction. The prediction of the ensemble is given as the averaged prediction of the individual classifiers…In random forests each tree in the ensemble is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. Furthermore, when splitting each node during the construction of a tree, the best split is found either from all input features or a random subset…The purpose of these two sources of randomness is to decrease the variance of the forest estimator. Indeed, individual decision trees typically exhibit high variance and tend to overfit. The injected randomness in forests yield decision trees with somewhat decoupled prediction errors. By taking an average of those predictions, some errors can cancel out." [scikit-learn user guide]

Gaussian Mixture Model

"A Gaussian Mixture Model (GMM)…is a type of unsupervised learning method [that] uses various Gaussian distributions to model different parts of the data. As a simple example, imagine we had a single time series of an asset's returns. As we know, returns of financial assets do not always follow a normal distribution. So a GMM would fit various Gaussian distributions to capture different parts of the asset's return distribution, and each of those distributions would have its own properties, like means and volatilities. In [the exhibit] the green Cluster 2 captures the centre part of the asset's return data, while the red and blue Clusters 1 and 3 capture the tails." [Two Sigma]

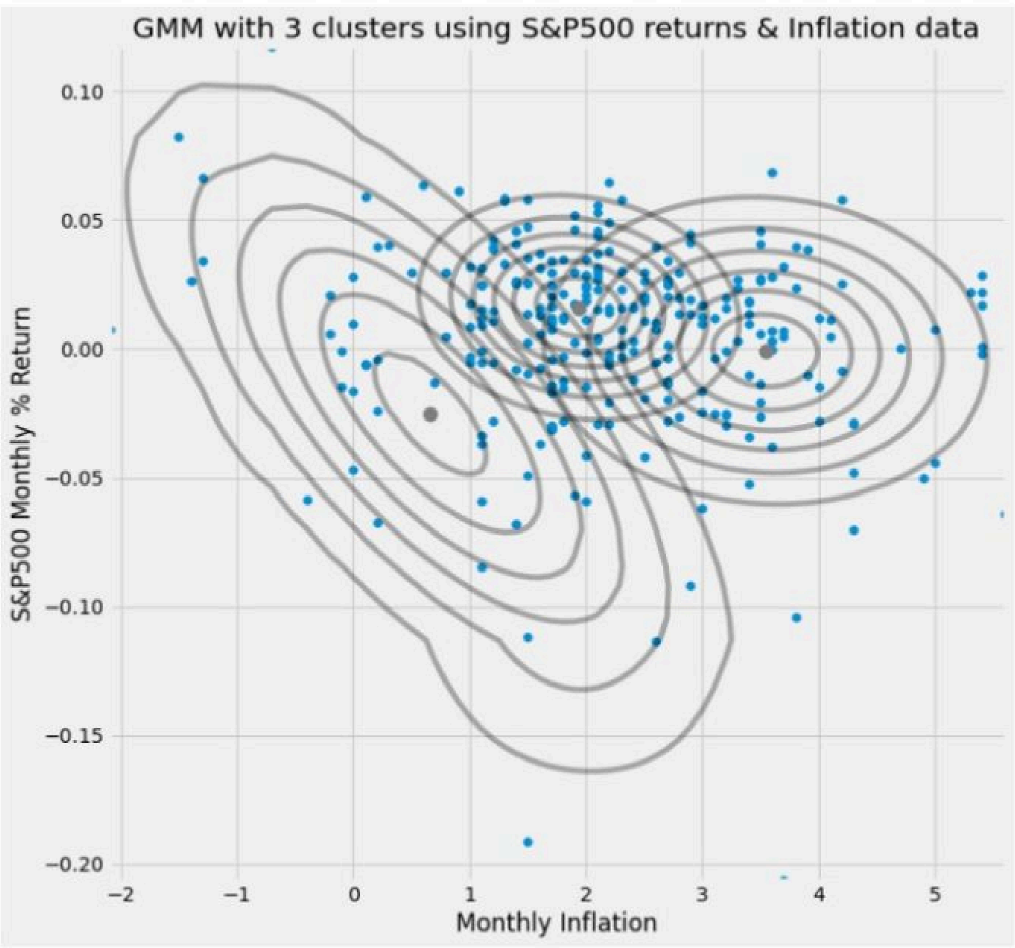
Exhibit 1: Illustrative Example of Three Gaussians that Describe a Single Time Series



"The gaussian mixture model is the overlapping of multi-normal distributions in p-dimensional space. The dimension of the space is generated by the number of variables. For example, if we had one variable (S&P 500 returns), the GMM would be fit based one dimensional data. The GMM can be used to model the state of the stock market along with other financial applications…An advantage of the GMM approach is that it is entirely data-driven. The data given to the model will form the clusters…The python implementation for the GMM on one dimensional data is quite simple [by using scikit-learn]" [Johnson-Skinner]

"sklearn.mixture is a package which enables one to learn Gaussian Mixture Models (diagonal, spherical, tied and full covariance matrices supported), sample them, and estimate them from data. Facilities to help determine the appropriate number of components are also provided…A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians." [**scikit-learn user guide**]

"One [key] benefit of using the GMM for unsupervised clustering is the space encompassing each cluster can take on a ellipse shape. Gaussian mixture models take not only means into account but also co-variance to form a cluster…The graph demonstrates [that] the normal distributions can produce the ellipse shape and this property comes from the covariance matrix." [**Johnson-Skinner**]



"The result of the GMM using…factor data was four different clusters, or what we think may correspond to four different types of market conditions…The factors are constructed to be lowly correlated with one another, especially over long periods…Each market condition from our GMM is characterized by a 17-dimensional Gaussian distribution.

- The most appropriate label for Market Condition 1 would be **Crisis**…In this market condition, we see that several of the core and secondary macro factors exhibited extremely poor performance on average…The Interest Rates factor, representing global sovereign bonds, exhibited a positive mean return…Market Condition 1 exhibited the highest average absolute correlation between the factors.
- Market Condition 2 [can be labelled as] **Steady State** [and]…seems to cover the most normal and healthy market periods, as there are no obviously large drawdowns for any factor…Equity, Credit, and nearly every style factor performed well on average.

- Market Condition 3 [can be labelled as] **Inflation**…The U.S.-specific Local Inflation factor exhibited a double-digit mean return, the highest mean return for that factor across the four market conditions…We find that the global Equity and Interest Rates factors have small positive mean returns, underperforming most, if not all, of the other four market conditions.
- Market Condition 4…looks like this market condition potentially captures risk-on market periods where bubbles might exist or be forming. We label it **Walking on Ice**…Global equity markets (as proxied by the Equity factor) do well here, but with a higher volatility than their long-term average." [Two Sigma]

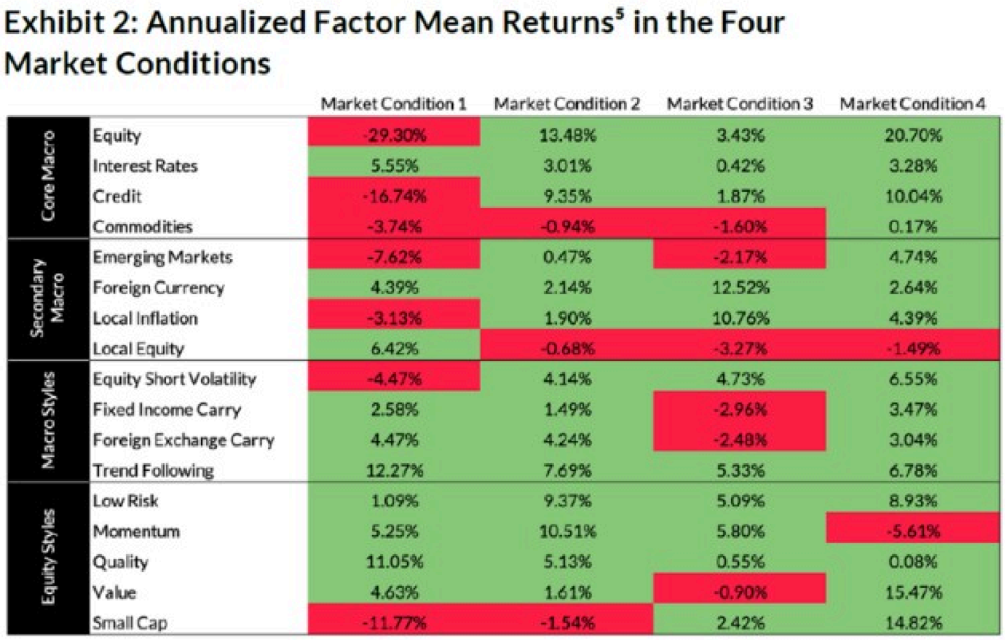


Exhibit 2: Annualized Factor Mean Returns[5] in the Four Market Conditions

		Market Condition 1	Market Condition 2	Market Condition 3	Market Condition 4
Core Macro	Equity	-29.30%	13.48%	3.43%	20.70%
	Interest Rates	5.55%	3.01%	0.42%	3.28%
	Credit	-16.74%	9.35%	1.87%	10.04%
	Commodities	-3.74%	-0.94%	-1.60%	0.17%
Secondary Macro	Emerging Markets	-7.62%	0.47%	-2.17%	4.74%
	Foreign Currency	4.39%	2.14%	12.52%	2.64%
	Local Inflation	-3.13%	1.90%	10.76%	4.39%
	Local Equity	6.42%	-0.68%	-3.27%	-1.49%
Macro Styles	Equity Short Volatility	-4.47%	4.14%	4.73%	6.55%
	Fixed Income Carry	2.58%	1.49%	-2.96%	3.47%
	Foreign Exchange Carry	4.47%	4.24%	-2.48%	3.04%
	Trend Following	12.27%	7.69%	5.33%	6.78%
Equity Styles	Low Risk	1.09%	9.37%	5.09%	8.93%
	Momentum	5.25%	10.51%	5.80%	-5.61%
	Quality	11.05%	5.13%	0.55%	0.08%
	Value	4.63%	1.61%	-0.90%	15.47%
	Small Cap	-11.77%	-1.54%	2.42%	14.82%

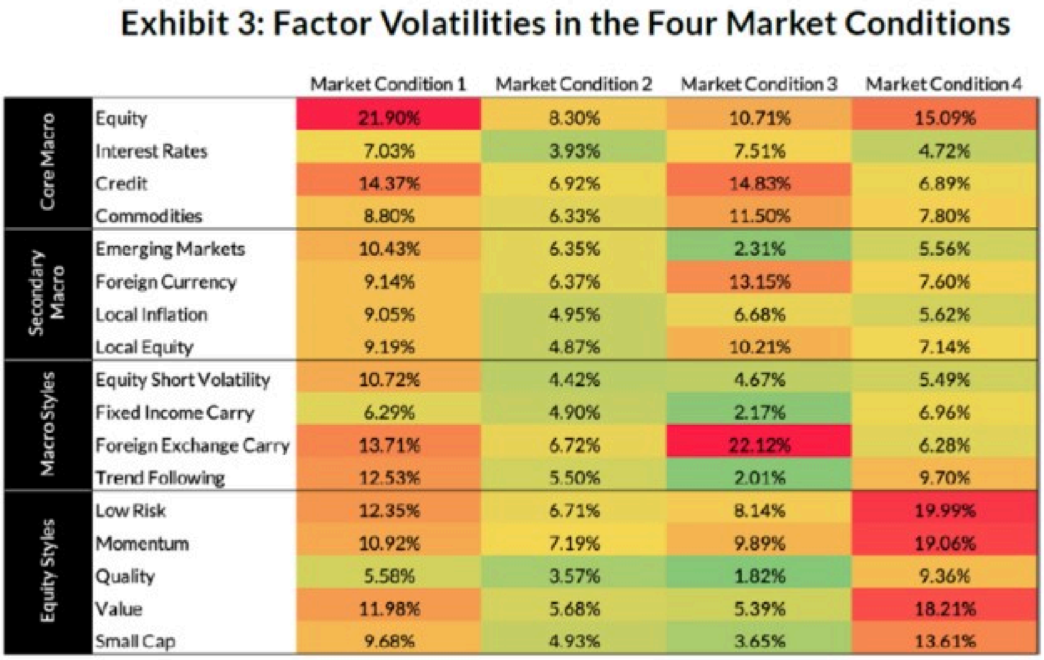
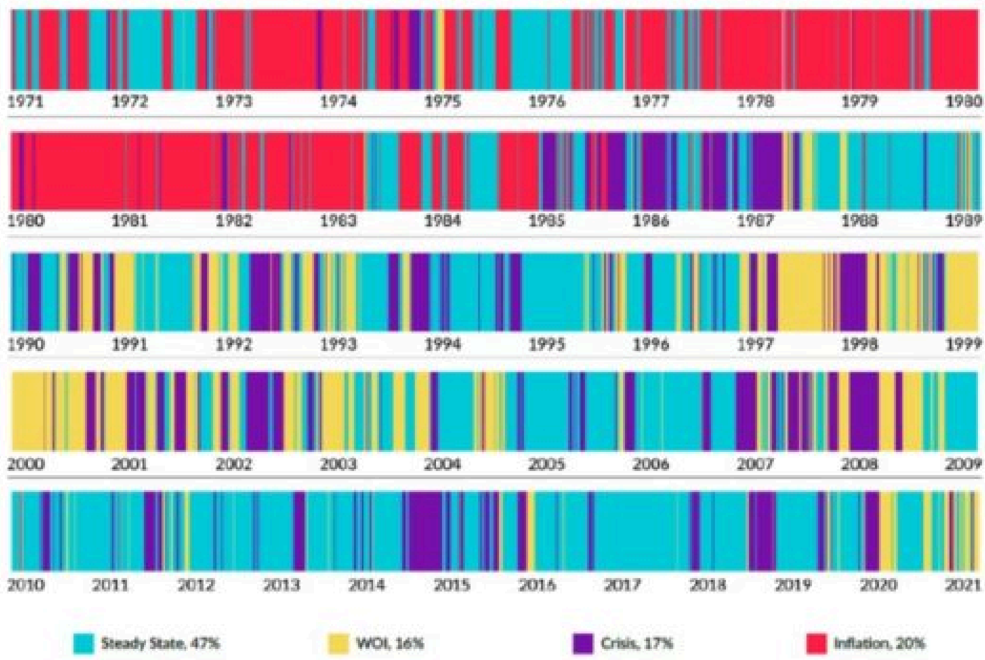


Exhibit 3: Factor Volatilities in the Four Market Conditions

		Market Condition 1	Market Condition 2	Market Condition 3	Market Condition 4
Core Macro	Equity	21.90%	8.30%	10.71%	15.09%
	Interest Rates	7.03%	3.93%	7.51%	4.72%
	Credit	14.37%	6.92%	14.83%	6.89%
	Commodities	8.80%	6.33%	11.50%	7.80%
Secondary Macro	Emerging Markets	10.43%	6.35%	2.31%	5.56%
	Foreign Currency	9.14%	6.37%	13.15%	7.60%
	Local Inflation	9.05%	4.95%	6.68%	5.62%
	Local Equity	9.19%	4.87%	10.21%	7.14%
Macro Styles	Equity Short Volatility	10.72%	4.42%	4.67%	5.49%
	Fixed Income Carry	6.29%	4.90%	2.17%	6.96%
	Foreign Exchange Carry	13.71%	6.72%	22.12%	6.28%
	Trend Following	12.53%	5.50%	2.01%	9.70%
Equity Styles	Low Risk	12.35%	6.71%	8.14%	19.99%
	Momentum	10.92%	7.19%	9.89%	19.06%
	Quality	5.58%	3.57%	1.82%	9.36%
	Value	11.98%	5.68%	5.39%	18.21%
	Small Cap	9.68%	4.93%	3.65%	13.61%

Exhibit 4: Highest Probability Market Conditions Throughout History

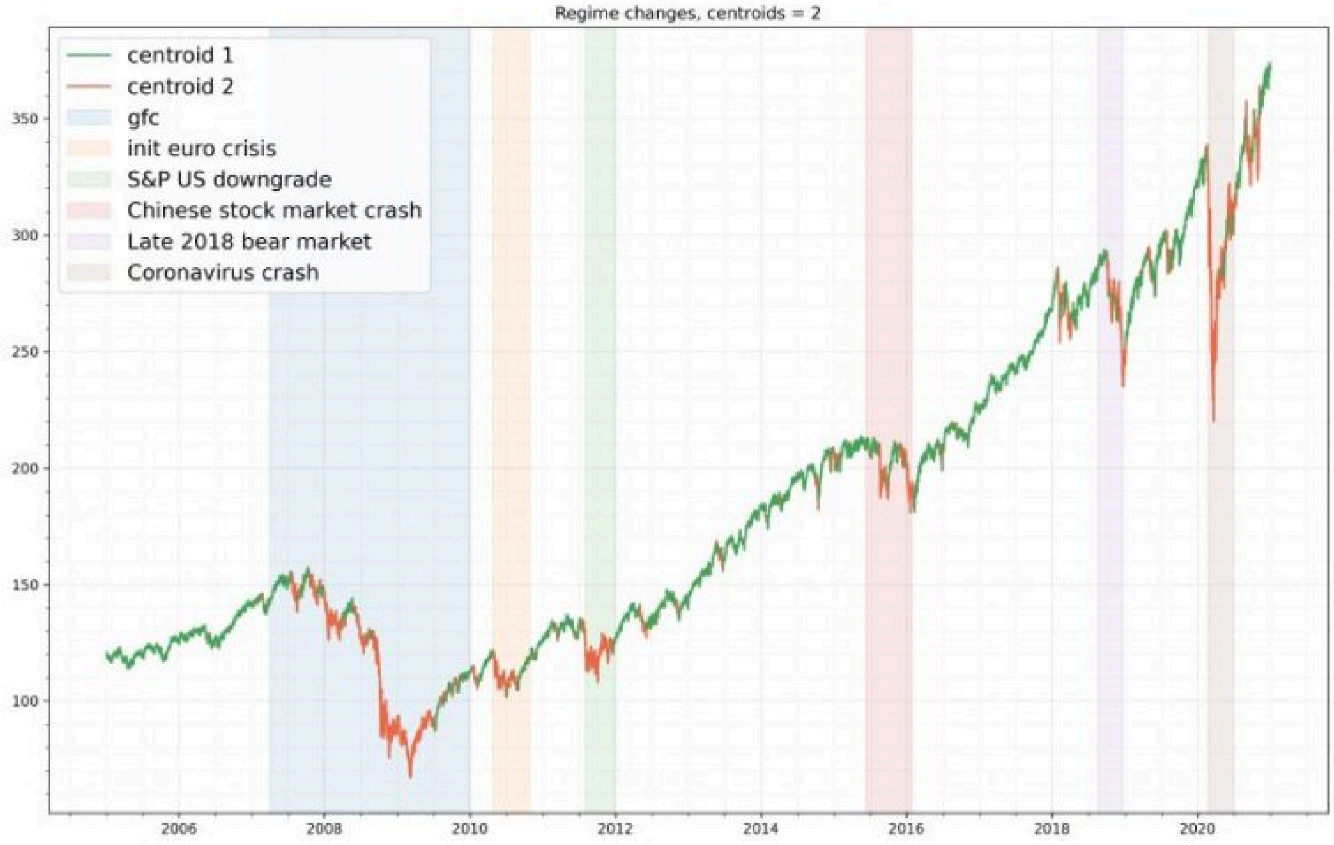


Hidden Markov model

“The question…is how to classify the [market] into states. To answer this we need to first define what is a state, I will define this mathematically as a state space. A state-space model is a probabilistic model that describes a system as a set of input, output and state variables. A probabilistic relationship exists between a latent state and observations characterized by the state and observation equations. In simpler words a state-space can be used to model a time series where each state at a given time is probabilistic depending on the previous state and current information. This can be extended where each state is said to be hidden or not observable, and is inferred from observable data.” [Johnson-Skinner]

“The Hidden Markov model is a stochastic process with an underlying stochastic process that is non-observable. The Hidden Markov Model is from the family of Markov models and inherits the properties from a Markov process, where future states depend only on the current state…Hidden Markov Models are fitted with [an] expectation maximization algorithm…an iterative method to find local maximum posterior estimates of parameters in statistical models.” [Johnson-Skinner]

“I [use] the volatility of the S&P500 [return] to [estimate[hidden states…Since the Hidden Markov Model state is based on the volatility of the S&P500, I will simply produce the one day ahead volatility prediction using the Data-Driven Exponential Weighted Moving Average and feed that into the Hidden Markov Model…The primary idea will be to classify the S&P500 into three segments based on the modelled volatility and using each segment to modify the algorithmic trading strategy.” [Johnson-Skinner]

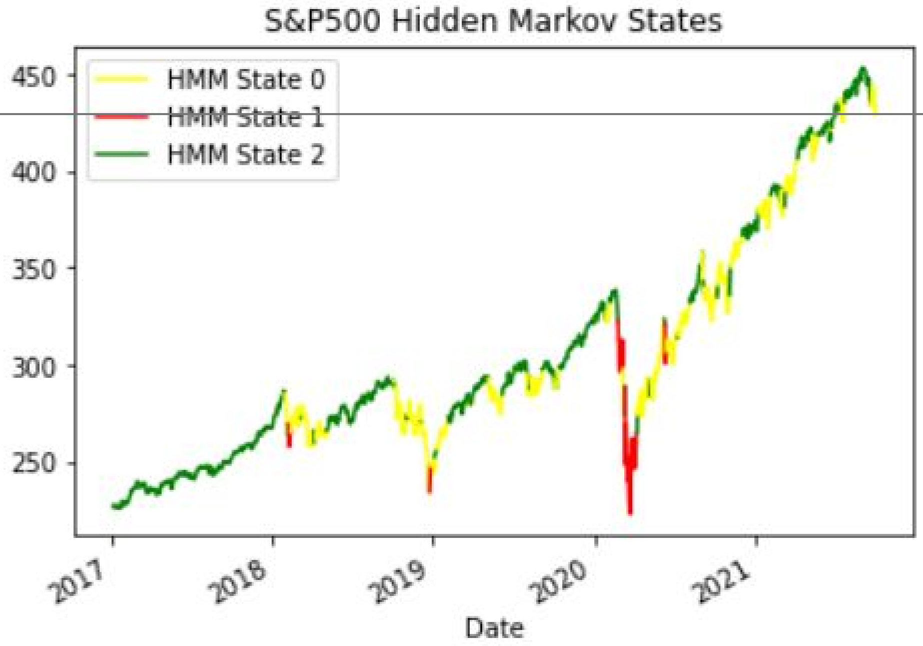


Wasserstein k-means clustering

“We outline an unsupervised learning algorithm for clustering financial time-series into a suitable number of…market regimes…We develop a robust algorithm that automates the process of classifying market regimes…the Wasserstein k-means algorithm…The method is robust in the sense that it does not depend on modelling assumptions of the underlying time series as our experiments with real datasets show. [It is] a modified, versatile version of the classical k-means clustering…The way to modify the classical algorithm is twofold: Firstly, by a shift of perspective, we consider the clustering problem as one on the space of distributions with finite pth moment, as opposed to one on Euclidean space. Secondly, our choice of metric on this space is the pth Wasserstein distance, and we aggregate nearest neighbours using the associated Wasserstein barycenter [center of mass of two or more bodies]…The Wasserstein distance is a natural choice for comparing distributions of points on a metric space.” [Horvath, Issa and Muguruza]

“[The] Wasserstein distance…is also known as the earth mover’s distance, since it can be seen as the minimum amount of ‘work’ required to transform one set of values into another, where ‘work’ is measured as the amount of distribution weight that must be moved, multiplied by the distance it has to be moved.” [scipy API references]

“We compare our Wasserstein k-means approach with a more traditional clustering algorithms by studying the so-called maximum mean discrepancy scores between, and within clusters. In both cases it is shown that the Wasserstein k-means means algorithm vastly outperforms all considered competitor approaches.” [Horvath, Issa and Muguruza]


S&P500 Hidden Markov States
HMM State 0
HMM State 1
HMM State 2
450
400
350
300
250
2017
2018
2019
2020
2021
Date