Detecting Trend and Regime Changes

Authors: Mark Lee, Min Li, Dawn Lu, Qinyan Wu

Background

Market movements are particularly important in asset allocation. In the dynamic world of investing, understanding market movements is critical to effective asset allocation. As the market continues to evolve, investment strategies need to efficiently identify and adapt to these changes in order to profit in the market. Institutional changes represent critical shifts in economic and market conditions. These changes may be driven by a variety of factors, including changes in monetary policy, geopolitical events or significant global economic developments. Identifying these shifts is critical to adapting investment strategies to new market realities.

The main goal of our report is to develop methods for segmenting time series data to detect trends and identify important change points representing regime shifts. Use advanced data analysis and statistical methods to detect trends and institutional changes in financial markets. This includes developing and applying algorithms to identify key trends and change points from time series data, and assessing the impact of these changes on markets and investment strategies. potential impact. Through these analyses, our reports aim to provide investors with deeper market insights to help them make more informed decisions in the ever-changing market environment.

Specifically, in this report, we focus on exploring and detecting trends in the S&P 500 and its changes to reveal how economic indicators significantly impact stock market volatility. A detailed analysis of external economic factors such as gross domestic product (GDP), unemployment rate, consumer price index (CPI), benchmark interest rates, volatility index (VIX), and government bond yields. Their interaction and independent effects are crucial for predicting market movement patterns.

To accurately identify and quantify these trends and institutional changes, our research employs two main methodologies. The first method is the change point detection algorithm. This method can identify the exact location of structural changes in complex time series data, revealing the beginning, continuation, and end of trends. The second approach we use is Markov switching regression models, which are models that adaptively capture shifts in market conditions by estimating changes in model parameters to characterize transitions between different economic or market states. The combination of these methods provides a solid analytical foundation for investment strategies and can help investors make more informed decisions in the face of market uncertainty.

Methodology

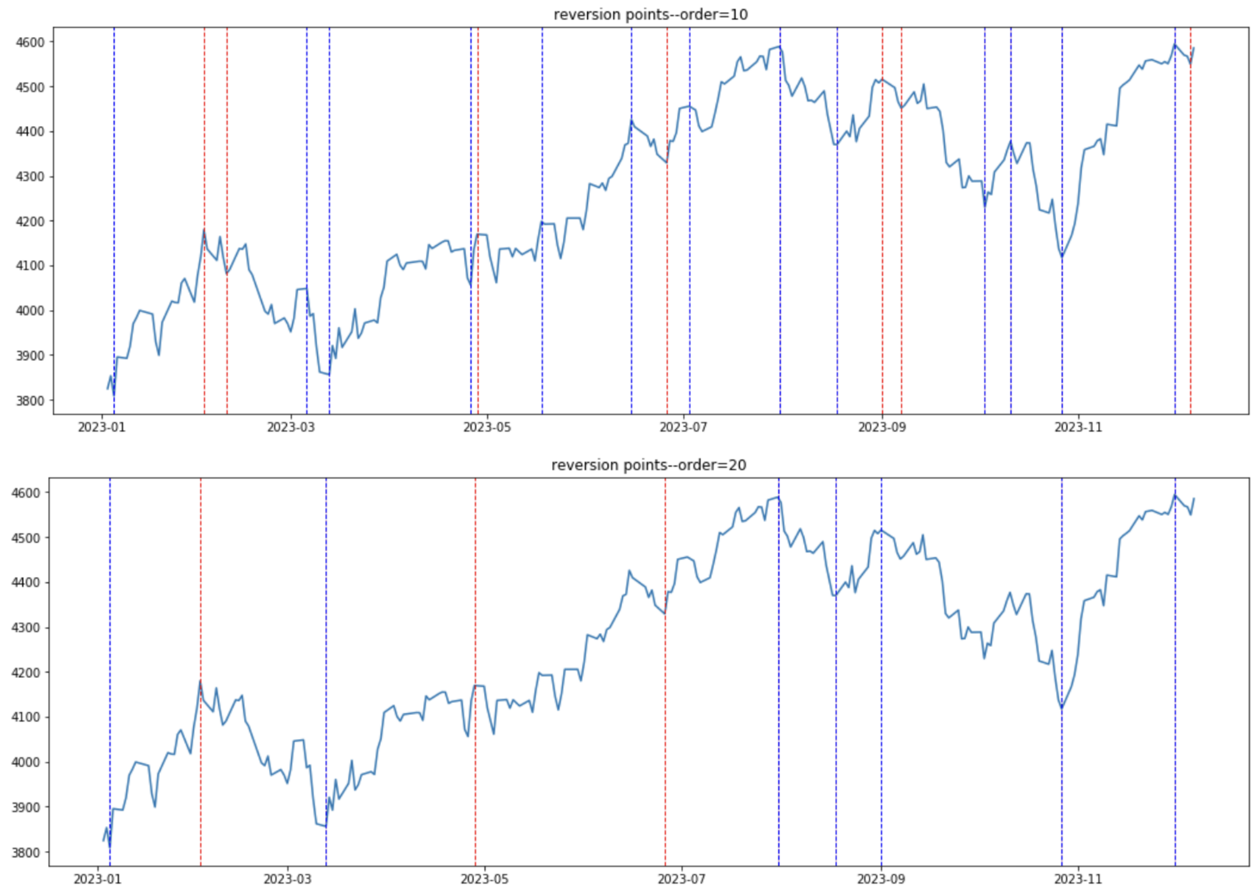
Data compilation and analysis

First we read the "S&P500.csv" file. The index columns were then converted to datetime format, the data set was sorted, and the data for 2023 and beyond were filtered out. After processing the data format, we visually displayed the data, showing the price trend in 2023. The chart shows continued price growth through 2023, laying the foundation for further statistical analysis and trend identification.

Argrelextrema Function

After processing the data, we processed the data through “argrelextrema” to find the local maximum or minimum points in a data - that is, in a given data sequence, these points are in their relative range of adjacent data points that are the largest or smallest. “argrelextrema” is a function in the Python scientific computing library SciPy and belongs to the “scipy.signal” module.

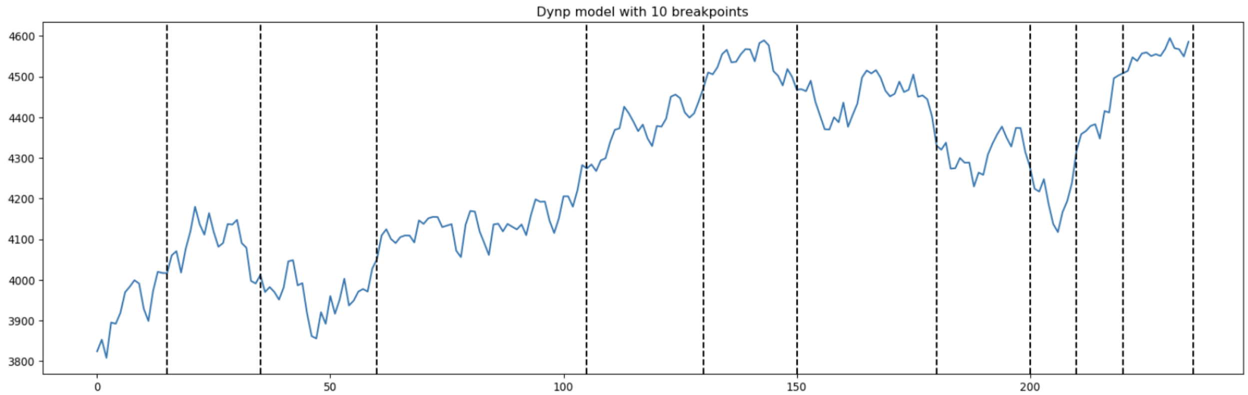
The dashed lines in the figure represent peaks and valleys, for identifying significant changes or turning points in time series data. We tuned the hyperparameter – “order”. Higher order values yield changes, lower values capture finer details. Setting a higher “order broader” value, the algorithm looks for changes in a wider range, effectively smoothing out short-term fluctuations and focusing on more significant trend reversals. This is useful for identifying long-term investment opportunities or major market movements. A lower “order” means the function will consider fewer points around each data point, allowing it to capture more short-term peaks and valleys. This is very valuable for traders who are interested in short-term market dynamics and want to take advantage of fast and frequent trades, which is more suitable for short-term traders.

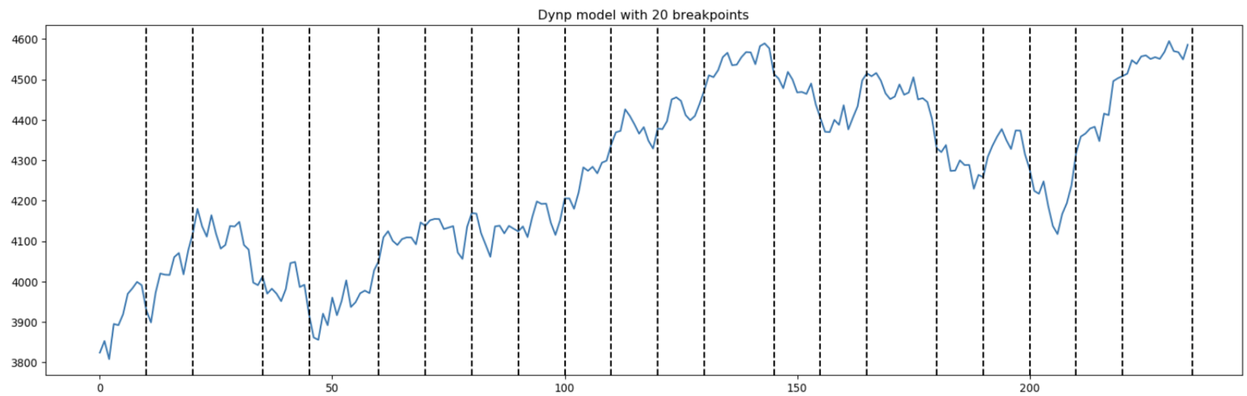


Dynamic Programming

We also use dynamic programming to visually analyze the data and detect change points. Dynamic Programming works by systematically reviewing all possible segments of a given signal to find the exact minimum of the sum of costs associated with each segment. The figure shows two different models, one model is preset with 10 change points, and the other is preset with 20 change points. These change points are represented by vertical dotted lines. Through this method, we can more intuitively identify important turning points in the data, thereby helping to analyze the structural changes in the data.

As can be seen from the graph, the number of change points has a significant impact on the sensitivity of the model in capturing changes in the data. Fewer change points (10) may cause the model to capture a wider range of trend changes, while more change points (20) allow the model to identify more details and short-term fluctuations. The choice between these two settings depends on the purpose of the analysis: looking for long-term trends may favor using fewer change points, while selecting more change points when conducting more refined data analysis. However, a key limitation of the DynP model is that the number of change points must be determined in advance, which requires users or analysts to have sufficient prior knowledge or alternative methods to estimate this parameter.



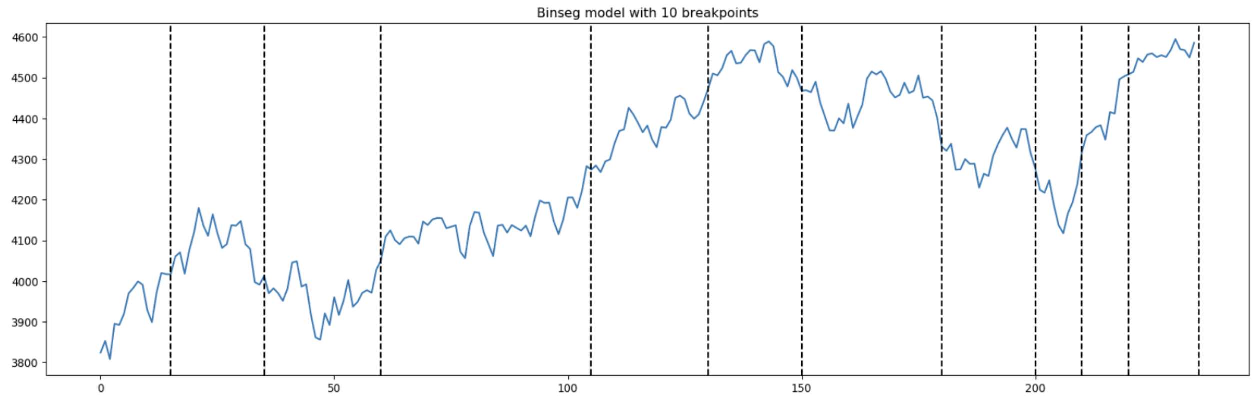


Binary Segmentation

Binary segmentation is an algorithm for finding a single change point. It first searches for a change point in the entire signal. After finding it, it divides the signal into two parts and repeats the same process for the two parts. This process continues until no more change points are found or a certain stopping criterion is reached. Decisions in this approach are based on the best option available at the moment, without considering the overall impact on the final outcome.

Two different models, one with 10 change points and the other with 20 change points. These change points are marked with vertical dashed lines. It can be seen that when the number of change points increases, the model is able to capture more details, reflecting smaller structural changes in the time series. However, this can also lead to the model being too responsive to random fluctuations in the data and ignoring long-term trends. Conversely, fewer change points may miss some important short-term changes, but they may better capture and emphasize the main trends in the data.

Choosing an appropriate number of change points requires comprehensive consideration of the characteristics of the data, the purpose of the analysis, and the possible risk of overfitting.



Markov Switching Regression Model

The second approach we used to detect and predict regime changes in the S&P 500 is the Markov switching regression model, which is often used in econometrics and time series analysis to capture the dynamic and often unpredictable nature of the economy and financial markets. Unlike traditional regression models that tend to assume a constant relationship between variables over time, Markov switching models allow for regime shifts or changes in the underlying structure of the data.

At its core, a Markov switching regression model integrates the concept of a Markov chain, which is a stochastic process where the likelihood of transitioning from one state to another depends solely on the current state. The transition between these states is governed by a state transition probability matrix, which defines the probabilities of moving from one state to another. In the context of regression analysis and our project, the states represent different economic regimes and market conditions.

The model assumes that the data undergoes discrete changes between different states, and each state is associated with a distinct set of regression coefficients. These coefficients capture the relationship between the dependent variable and the independent variables in that particular state. The regression coefficients are estimated by maximizing the likelihood function, which measures how well the model explains the observed data. The likelihood function is constructed based on the joint probability of observing the entire sequence of states and data. This involves multiplying the probabilities of transitioning between states represented by the state transition probability matrix and the likelihood of observing the data given the current state.

The key advantage of using a Markov switching regression model lies in its ability to account for periods of stability and volatility within the data. Therefore, the model can potentially capture shifts between high and low volatility states, which is essentially the goal of our project. By allowing the parameters of the regression to vary across different states, the model provides a more realistic representation of the underlying dynamics, making it particularly useful for forecasting and risk management in environments characterized by changing economic conditions.

We first employed a simple Markov regime regression model using only historical S&P 500 data as a model parameter. We first gathered S&P 500 price data from 2000 and 2023 prior to training and testing the model. The model is then trained to predict whether the index is in a state of high volatility or low volatility and tested for performance.

Besides the simple univariate Markov regime regression model, we also incorporated exogenous factors to examine if they could improve the performance of the model. The external factors include GDP, unemployment rate, CPI, interest rate, VIX, and treasury yield. The model with extra factors besides the S&P 500 is also trained and tested to analyze how well it predicts whether the index is in a state of high or low volatility. The performances of the two models are also compared with one another to see if the external factors increase the predictive power of the model.

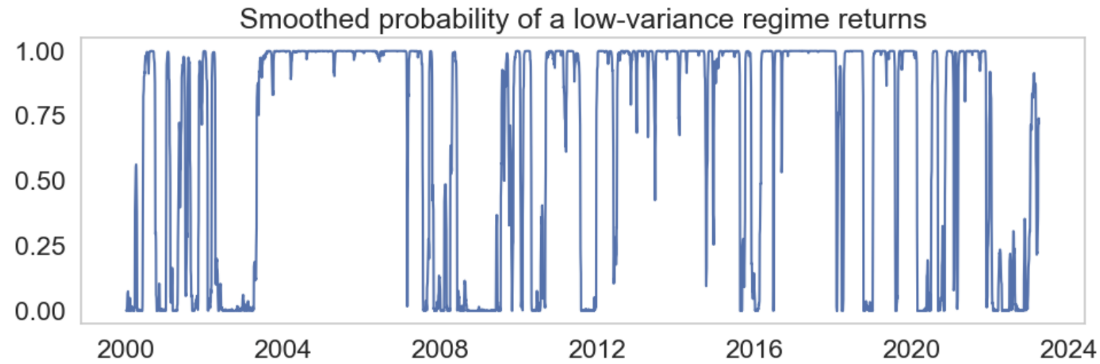
Results

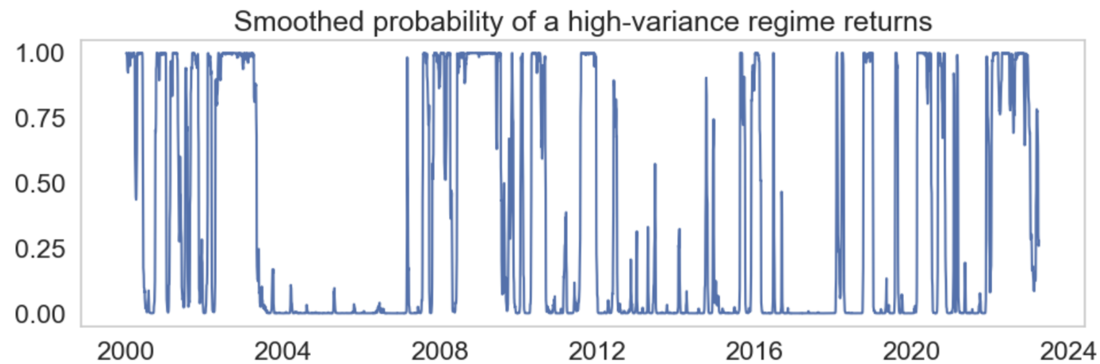
Trends and Change Point Detection

Dynamic programming, binary partitioning and the “argrelextrema” function are three different time series analysis methods. Dynamic programming finds cost-minimizing solutions by evaluating all possible segmentations, but it requires the analyst to know the number of change points in advance, which can be a limitation in practice. The binary segmentation rule adopts a step-by-step method to iteratively find and segment the change points. Although this method is more flexible and computationally efficient, it may miss the global optimal solution. Compared with these two methods, the “argrelextrema” function directly identifies local extreme points in the data, which allows it to capture significant changes without knowing the number of change points in advance. It is relatively sensitive to noise and sensitive to parameter selection. Overall, these three methods each have their own advantages and limitations. Which one to choose should be determined based on specific data characteristics and analysis needs. Combining multiple methods may lead to more accurate and robust results.

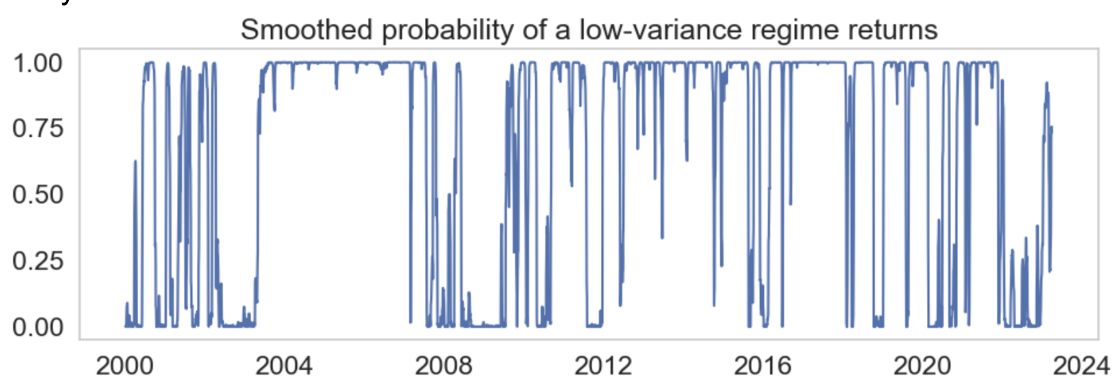
Markov Switching Regression Model

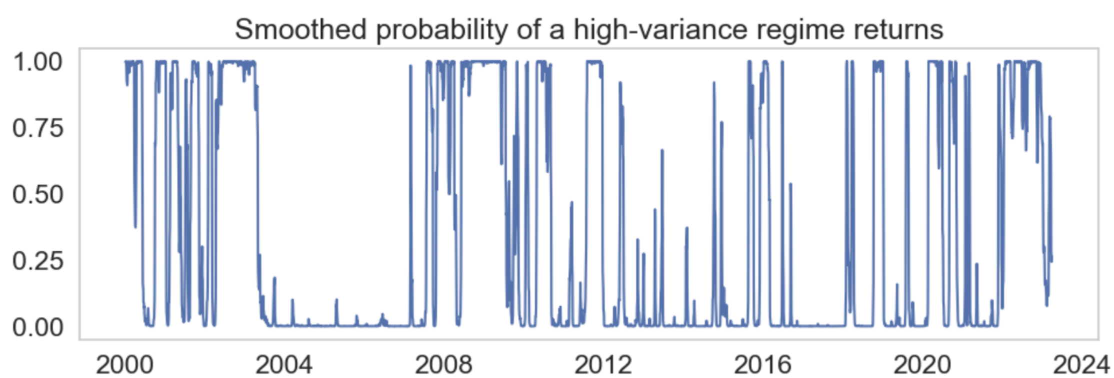
The Markov switching regression model predicts whether the S&P 500 is in a high volatility state or a low volatility state. The graphs below exhibit the probability of the index being in each respective state over time given the predictions of the simple Markov switching regression model. A high probability indicates that the model is highly confident in its prediction of the volatility of the S&P 500 and a low probability indicates that the index is most likely in the opposite volatility state.





The graphs below display the same information but for the Markov switching regression model with extra external factors. Overall, the two models don't seem to perform much differently from each other.





The graph below gives a better visual representation of the model prediction results, with each dot representing the daily closing price of the S&P 500. The green dots indicate that the model believes that the index is currently in a state of low volatility, and conversely, the red dots indicate that the model believes that the index is currently in a state of high volatility. The top graph shows the prediction results of the simple Markov switching regression model and the bottom graph shows the prediction results of the model with exogenous factors. Once again, the results of the two prediction models don't seem to be so different.

We also seek to compare the performance of the two models by looking at the mean-squared error of each. The simple Markov Switching model yields an MSE score of 0.0001554 and the model with exogenous variables yields an MSE score of 0.0001551, showing that the economic indicators only marginally improve the performance of the model.