

# **B5W10: Change point analysis and statistical modelling of time series data**

Detect changes and associate causes on time series data

**An Interim Report**

Done by: Kokeb Belete

# Data Analysis Workflow

For this project, we will follow the following workflow to reach our data analysis goal.

## 1. Define the Problem

We are helping Birhan Energies, a consulting firm specializing in the energy sector, analyze data and pinpoint how events affect data. Specifically, we are trying to understand how political and economic events affect Brent oil prices. Knowing how events affect prices empowers investors to manage risk and make good decisions. It also helps energy companies to plan operations ahead and secure supply chains. To reach our objectives, we start by finding key events that have significantly impacted Brent oil prices over the past decade. We then measure how much these events affect price changes and providing clear, data-driven insights.

## 2. Data Collection

- We already have a CSV file containing the daily prices covering the range May 20, 1987, to September 30, 2022. This is the data that we will analyze to find the changes in price over the years using change point analysis.
- We will compile major geopolitical events, OPEC decisions, and economic shocks relevant to the oil market in the time frame that the data above has. We will use this data to explore and understand the changes that we saw happen in the oil price data. We can find this data from:
  - ✓ News archives for major events from renown news sites like CNN and BBC
  - ✓ OPEC website to see historical reports and event timelines.
  - ✓ Government websites
  - ✓ Industry reports from IMF and World Bank

### **3. Data Cleaning and Preparation**

After collecting the data we spend time cleaning and preparing it for further analysis. This includes:

- ✓ Checking for outliers and removing outliers that might affect data accuracy.
- ✓ Checking for missing values and removing the most detrimental ones or imputing those that can be imputed using mean or median
- ✓ Encoding categorical values so that they can be used in machine learning models.
- ✓ Scaling the data.
- ✓ Converting the data into the proper data format like date.

### **4. Exploratory Data Analysis**

After initial preparation of the data, we conduct exploratory data analysis to gain deeper insights to the data. It will help us form initial analysis. This includes:

- Summarizing key statistics: We can see values such as mean, median, standard deviation to gain insight to the data.
- Relationship distribution: plot histograms and bar charts to understand the data distribution better.
- Visualization: plot the time series to identify trends, seasonality, and periods of change.
- Check with events data to see if we can find and correlate the price changes with a real world event.
- Feature engineering to include important features that are useful our analysis.

### **5. Model Building**

The goal is to identify how Brent oil prices are affected by real world events. The model will help determine when changes occurred and how price dynamics shifted. We will apply Bayesian change point models to detect structural breaks in the time series using PyMC. Bayesian change

point model is a statistical model that detects when and how a time series changes — by treating those change points as unknown variables and using Bayesian inference to estimate them. We start with data containing Brent oil prices over time. At certain points — like during a war or an OPEC decision — the behavior of the prices changes. A Bayesian change point model helps us identify those shifts, while accounting for uncertainty in the data and the timing of the changes. Bayesian modeling treats unknowns (e.g., location of change point, mean before/after) as random variables, assigns prior distributions to them and uses observed data to compute posterior distributions over possible change point locations and parameters. This gives us probabilistic estimates and uncertainty quantification.

When using this model we are assuming:

- Changes in the statistical properties of the series reflect meaningful real-world events.
- External events may not cause but correlate with the identified change points.

To evaluate the model and validate it, we compare the detected change points in the data to our known event dataset and check alignment. We can also use posterior predictive checks.

## **6. Deployment**

We will build a dashboard application to visualize the result of analysis so that our work can reach stakeholders. We will create an intuitive and user-friendly interface to display analysis results using React. It will show how different events correlate with changes in oil prices.

### **Expected Outputs and Limitations of Change Point Analysis**

Change point analysis is used to identify significant shifts in the statistical properties of the Brent oil price time series over time. The primary outputs we expect from this analysis include:

- ✓ Estimated change point dates: These are specific time points where the model detects a structural shift in the behavior of the data, such as a sudden increase or decrease in prices.

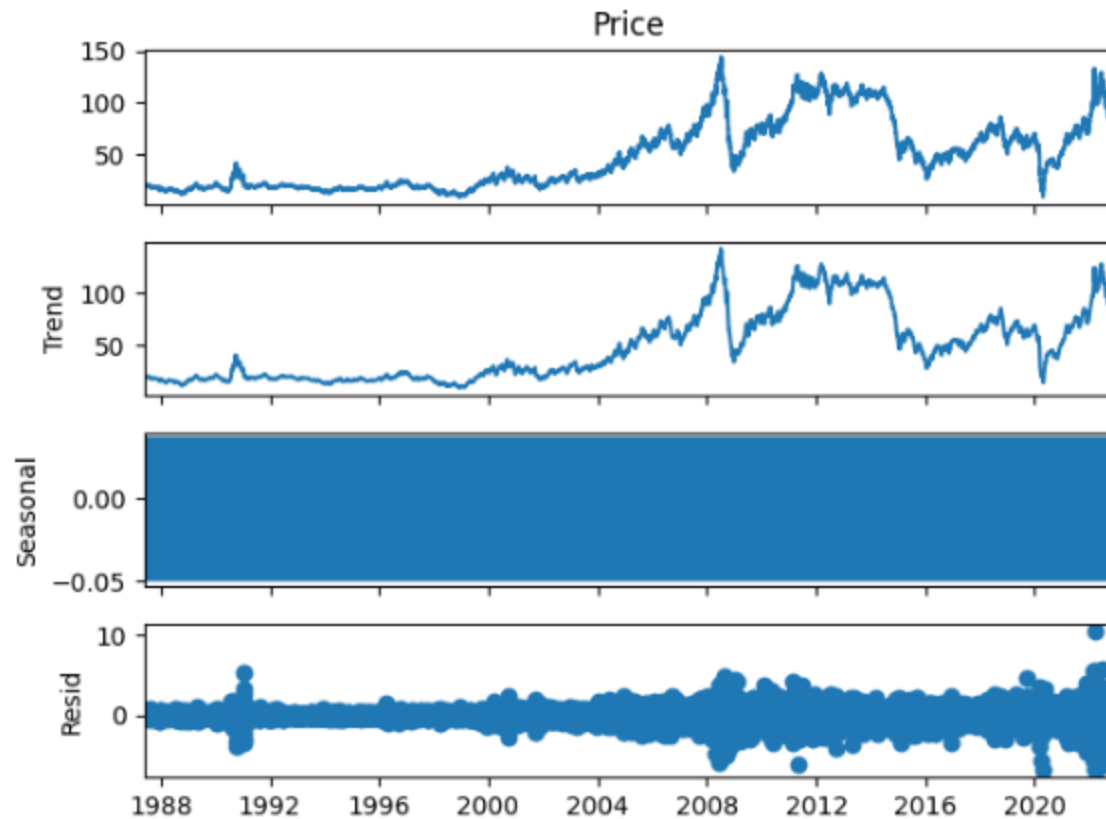
- ✓ Segmented parameter estimates: For each segment between change points, the model estimates new statistical properties — for example, a new mean level, trend slope, or variance.
- ✓ Uncertainty quantification: output will also include posterior distributions over the location of change points and over segment parameters, providing a measure of uncertainty rather than just single-point estimates.

#### The limitation in the models

- ✓ Lack of causal explanation: While change point models detect when changes happen, they do not explain why the changes occurred. Correlation doesn't mean causation. Just because oil prices changed around an event doesn't mean the event caused the change. We also are assuming the world events impact the prices immediately when in reality the effect might be delayed.
- ✓ Sensitivity to noise and modeling choices: The accuracy of change point detection can be affected by random fluctuations, outliers, and assumptions about the number or type of change points.
- ✓ Overfitting or underfitting risk: Choosing too many change points can lead the model to mistake noise for structure, while too few may cause it to miss important shifts.
- ✓ Using only a few subset of real-world events to conduct our analysis. This makes our analysis depend on these events when in reality oil prices depend on complex and multiple factors.

#### Time Series Analysis

The time series analysis on the oil prices gives us the following graph. We see here the decomposition of the time series analysis.



- ✓ We see the observed price at the top.
- ✓ We see the trend plot where we see the overall trend in the price
- ✓ We also see the seasonal behaviour which is flat which means there is no seasonal variation in the oil prices.
- ✓ And finally we see the residual plot showing us what the behaviour in the oil price is once the trend and the seasonal variations are removed. We see shocks and structural changes in the price. This is a great input for our change point analysis.