Change Point Detection via Mixed Integer Optimisation

with an application in European carbon emission

Tomoya Ozawa
Supervisor: Dr Jessica Leung

17th April 2023

ETF5550 First Presentation

What are we trying to do?

- We aim to detect when changes in relationship between Y and X happened by finding the changes in coefficients ($\beta_1 \to \beta_2$).

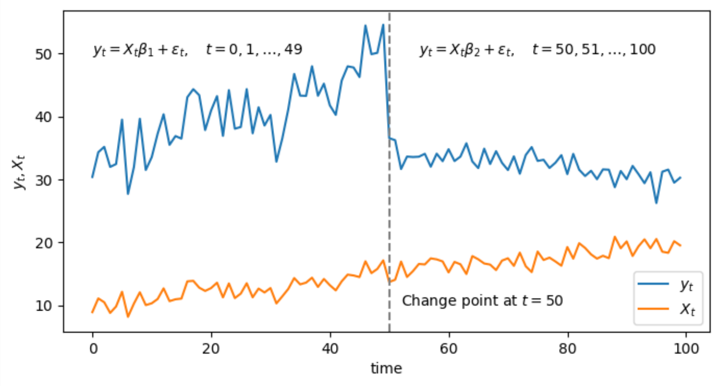

Figure 1: An example

What is our research?

Our research topic is to propose a new method to Change Point Detection using Mixed Integer Optimisation. Our research questions are:

RQ1 Can our proposed method outperform the existing methods?

RQ2 When does our method perform better than benchmarks?

This study consists of 2 parts.

I. Simulation study

- To answer our research questions

II. Empirical study

- To demonstrate how our proposed method performs in the real-world data set, the European CO2 emissions.

What is Change Point Detection?

- Change Point Detection (CPD) is a method to detect abrupt changes in time series data.
- Methods have been developed by the literature in engineering, bioinformatics, and econometrics.
- These methods are widely adopted for real-world applications such as signal processing, medical condition monitoring and climate change detection.
- CPD has many categories in terms of objectives (Online / Offline) and formulations (Supervised / Unsupervised).
- Our proposed method is classified as offline supervised CPD.

Previous studies in offline supervised CPD

[Chow, 1960]

- First study. Hypothesis testing whether the coefficients of the regression model vary before and after the structural break.

[Bai, 1994]

- This model used dynamic programming to detect a single change point. It is extended to multiple change points in [Bai and Perron, 1998, Bai and Perron, 2003]

[Wang and Emerson, 2015]

- A bayesian approach. It generalized the Product Partition Model proposed in [Barry and Hartigan, 1992].

What is Mixed Integer Optimization?

- Mixed Integer Optimization (MIO) is one of the classes of optimization problems.
- MIO is an optimization problem with some decision variables (x) that are restricted to be integers.
- Recently, the application of MIO has flourished due to the advancement of algorithms, solvers and computational resources.

$$
\begin{aligned}
\text{minimise} \quad & 3x_1 + 2x_2 + x_3 - 10 \\
\text{subject to} \quad & x_1 - 2x_2 + x_3 = 2.5 \\
& 2x_1 + x_2 = 1.5 \\
& x_j \geq 0 (j = 1, 2, 3) \\
& x_2, x_3 \text{ are integer}
\end{aligned}
$$

Motivation: Why do we do this research?

Three key points are identified through literature review:

1. CPD is a problem where MIO can be utilised. [RQ1]
    - MIO can provide optimal solutions for some variables that are restricted to be integers (the number of change points).
2. Almost all methods assume distribution assumption. [RQ1]
    - It is significant to re-visit CPD from a Data-driven perspective.
3. Few studies discuss when a proposed method outperforms existing methods. [RQ2]
    - Clarifying when our method works better than others is meaningful from an empirical perspective.

Our proposed method

Consider the following regression model:

$$Y_t = X_t^T \beta_t + u_t \quad t = 1, 2, \ldots, T \tag{1}$$

- Y_t is the data point of the objective variable and X_t is a p-dimensional column vector of predictors.
- β_t is a p-dimensional column vector of coefficients, depending on time t.
- u_t is a residual with no distribution assumption.
- *Toeplitz matrix* and *Vandermonde matrix* can be applied to the formulation. Lag and polynomial terms of the original variables in X can be expressed.

Our proposed method

$$\text{minimise} \quad \sum_{t=1}^{T}(Y_t - X_t^T\beta_t)^2 + \lambda \sum_{t=1}^{T-1} z_t \tag{2a}$$

$$\text{subject to} \quad |\beta_{t+1,j} - \beta_{t,j}| \leq Mz_t \quad \text{for } t = 1, 2, \ldots, T-1, \text{ and } j = 1, \ldots, p \tag{2b}$$

$$z_t \in \{0, 1\} \quad \text{for } t = 1, 2, \ldots, T-1 \tag{2c}$$

- λ is a hyperparameter. M is an arbitrarily large scalar, called Big M.
- z_t is a binary variable that takes 1 if t is detected to be a change point and 0 otherwise.

Simulation study

Objective:

- To evaluate the performance of our proposed method. We can answer our research questions.

How it works

- We simulate the data with change points based on regression models, implement the proposed method and benchmarks to the simulated data and compare how they perform.
- We do multiple simulation settings with different values for the two:
  - Signal to Noise Ratio (SNR), which indicates how much Y is explained by X relative to an error term, and correlation between predictors, X.

Simulation study

What we have done

- Implemented our proposed method and one of the benchmarks from scratch in Python.
- We are doing simulations of our proposed method.

What we are going to do: week 7 - 9

- Complete simulations for all methods
- Compare the results and answer our research questions

Empirical study

Objective:

- To demonstrate how our proposed method works in the real-world data set, we implement our method to the CO_2 emission data (caused by transportation) in the EU.

How it works

- We will pick up some countries and formulate a regression form for each country.
- Our method tells us that when changes in the relationships between CO_2 emissions, Y, and possible factors, X, such as *GDP*, *population* and *CO_2 emission of the neighbouring countries* occur.

Empirical study

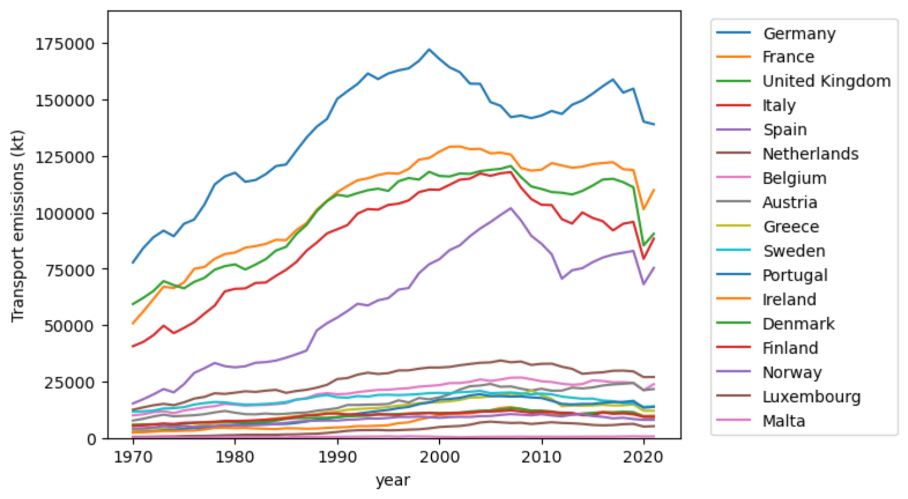

Figure 2: CO2 emissions for each country

Empirical study

What we have done

- Collected the latest dataset for CO2 emissions (from JRC EDGAR, [Crippa et al., 2022]), GDP and Population (from World Development Indicators, [The World Bank, 2021a, The World Bank, 2021b]), and merged them into one dataset by country and year.
- Conducted explanatory data analysis.

What we are going to do: week 9 - 10

- Decide model specification.
- Implement our purposed method to the data set and discuss results from the EU's climate policy perspective.

Conclusion

- Our research topic is to propose a new method to Change Point Detection using Mixed Integer Optimisation.
- To compare our proposed method to benchmark methods, we do Simulation Study.
- To demonstrate how our proposed method works in real-world data, we do Empirical study, implementing our method in the CO2 emissions dataset in the EU.
- In the next presentation and final thesis, we will report:
    - the results of Simulation study and Empirical study and discuss the obtained result.
    - the limitations and potential extensions of this study.

Reference i

Bai, J. (1994).
Least squares estimation of a shift in linear processes.
Journal of Time Series Analysis, 15(5):453–472.

Bai, J. and Perron, P. (1998).
Estimating and testing linear models with multiple structural changes.
Econometrica, pages 47–78.

Reference ii

- Bai, J. and Perron, P. (2003).
  Computation and analysis of multiple structural change models.
  Journal of applied econometrics, 18(1):1–22.
- Barry, D. and Hartigan, J. A. (1992).
  Product partition models for change point problems.
  The Annals of Statistics, pages 260–279.
- Chow, G. C. (1960).
  Tests of equality between sets of coefficients in two linear regressions.
  Econometrica: Journal of the Econometric Society, pages 591–605.

Reference iii

- Crippa, M., Guizzardi, D., Banja, M., Solazzo, E., Muntean, M., Schaaf, E., Pagani, F., Monforti-Ferrario, F., Olivier, J., Quadrelli, R., et al. (2022). **Co2 emissions of all world countries.** *Luxembourg: Publications Office of the European Union. doi*, 10:730164.
- The World Bank (2021a). **World development indicators.** GDP (constant 2015 US$), Retrieved from https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.

Reference iv

The World Bank (2021b).
World development indicators.
Population, total, Retrieved from
https://data.worldbank.org/indicator/SP.POP.TOTL.

Wang, X. and Emerson, J. W. (2015).
Bayesian change point analysis of linear models on graphs.
arXiv preprint arXiv:1509.00817.

Appendix: Toeplitz matrix

Toeplitz matrix is a matrix in that any diagonal elements are the same.

When Toeplitz matrix for $\boldsymbol{X} = [x_1, x_2, \cdots, x_T]$ is defined as:

$$T(\boldsymbol{X}) = \begin{bmatrix} x_2 & x_1 \\ x_3 & x_2 \\ \vdots & \vdots \\ x_T & x_{T-1} \end{bmatrix}$$

A regression model with the quadratic terms can be expressed as follows:

$$T(\boldsymbol{X})\boldsymbol{\beta} = \begin{bmatrix} x_2 & x_1 \\ x_3 & x_2 \\ \vdots & \vdots \\ x_T & x_{T-1} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Appendix: Vandermonde matrix

Vandermonde matrix for $\boldsymbol{X} = [x_1, x_2, \cdots, x_T]$ with $n-1$ degree of freedom is defined as:

$$V_{T\times n}(\boldsymbol{X}) = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \cdots & x_1^{n-1} \\ 1 & x_T & x_T^2 & \cdots & x_T^{n-1} \end{bmatrix}$$

Appendix: Vandermonde matrix

Multiplying the Vandermonde matrix for $\boldsymbol{X}$ by $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_n]$, the polynomial regression form can be obtained.

$$V_{T\times n}(\boldsymbol{X})\boldsymbol{\beta} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \cdots & x_1^{n-1} \\ 1 & x_T & x_T^2 & \cdots & x_T^{n-1} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}$$

Appendix: Signal To Noise (SNR)

- In the context of a regression model $Y = X\beta + \epsilon$, SNR is defined as the follows:

$$SNR = \frac{Var[X\beta]}{Var[\epsilon]}$$

- A higher SNR indicates that the predictors, X, have a greater impact in generating y compared to error terms ϵ.

Appendix: Simulation setting

Data-generating process:

$$\begin{aligned} t &= 1, 2, \ldots, 500 \\ Y_t &= X_t^T \beta_1 + \epsilon_t \\ \beta_1 &= [1, 1, 1, 0, 1] \\ \epsilon_t &\sim N\left(0, \frac{Var[X_{t=1:500}\beta_1]}{SNR}\right) \end{aligned}$$

$$\begin{aligned} t &= 501, 502, \ldots, 1000 \\ Y_t &= X_t^T \beta_2 + \epsilon_t \\ \beta_2 &= [-1, -1, -1, 0, -1] \\ \epsilon_t &\sim N\left(0, \frac{Var[X_{t=501:1000}\beta_2]}{SNR}\right) \end{aligned}$$

$$X_t \sim MVN(0, \Sigma) \qquad \text{,where } \Sigma_{i,j} = \rho^{|i-j|}$$