STA457 Time Series Analysis
Winter 2025
Final Project Guidline

1 Introduction

Modelling commodity prices is essential in various industries, including agriculture, finance, and trade. However, such modelling presents significant challenges, as commodity prices are influenced by multiple factors and often exhibit characteristics such as seasonality, volatility, and external shocks. These complexities make accurate forecasting a difficult but valuable task.

Cocoa, a key commodity in the global market, plays a crucial role in chocolate production and other food products. Its price is shaped by various factors, including weather conditions, geopolitical events, supply chain disruptions, and macroeconomic trends. Understanding and forecasting its price movements is of great interest to producers, traders, and policymakers.

In this project, you will apply time series analysis techniques to model and predict cocoa futures prices. You will explore various statistical methods to analyze historical price trends, assess model performance, and interpret the results. Through this project, you will gain hands-on experience working with real-world time series data and develop skills relevant to financial and economic forecasting.

You will have **considerable flexibility** in this project. There are **no restrictions** on the methods you choose—you are encouraged to explore and apply any techniques you find appropriate, even those beyond the scope of our course materials (This is an excellent opportunity for you to engage in self-study and further independent research while deepen your understanding). However, you must provide a clear justification for your chosen approach in your report. Additionally, aside from the data I provide, you may incorporate **any supplementary data** you deem relevant. This is a real-world, open-ended problem with no single correct answer. Your goal is to develop a well-reasoned and data-driven model for forecasting cocoa futures prices. The assessment criteria will be detailed in the rubric section.

Additionally, this is a problem I have personally encountered. About five years ago, while working in Singapore, I provided consulting services to a Singapore-based chocolate production company. One of their key concerns was understanding and predicting cocoa price fluctuations to optimize their procurement strategies. This experience highlighted the im-

portance and difficulty of modelling commodity prices effectively. I want you to step into my role at that time, tackling the same challenge I faced. This will give you a hands-on opportunity to apply time series analysis in a real-world business scenario.

2 Background Information

Cocoa is an essential agricultural commodity and the primary ingredient in chocolate production. It is derived from the seeds of the cacao tree (*Theobroma cacao*), which thrives in tropical climates with high humidity and consistent rainfall. The production of cocoa is highly sensitive to environmental and economic conditions, making price fluctuations a common challenge in the global market.

2.1 Cocoa Production and Growing Conditions

Cocoa trees require warm temperatures **(between 20°C and 30°C)**, regular rainfall, and well-drained, nutrient-rich soil. They typically take three to five years to reach maturity and begin producing cocoa pods. The trees bear fruit continuously throughout the year, but major harvests usually occur in **two main seasons**—one peak and one mid-crop season—depending on the region.

2.2 Major Cocoa-Producing Regions

Cocoa is primarily grown in the **equatorial belt** between **10 degrees north and south of the equator**, where climatic conditions are ideal. The **top cocoa-producing countries** include:

- **West Africa** – The region accounts for over 70% of global cocoa production, with **Ivory Coast** and **Ghana** being the largest producers, followed by **Nigeria** and **Cameroon**.
- **Latin America** – Countries such as **Ecuador**, **Brazil**, and **Peru** also contribute significantly to global cocoa production.
- **Southeast Asia** – **Indonesia** is the leading producer in this region, followed by Malaysia and Papua New Guinea.

2.3 Factors Affecting Cocoa Prices

Cocoa prices are influenced by a variety of factors, including:

1. **Weather Conditions** – Adverse weather, such as droughts or excessive rainfall, can impact cocoa yields. For example, El Niño events can lead to dry conditions that reduce production, while heavy rains can increase the spread of plant diseases.
2. **Pests and Diseases** – Cocoa trees are vulnerable to diseases such as black pod disease and pests like cocoa pod borers, which can significantly reduce output.

3. **Supply Chain and Political Stability** – Disruptions in key cocoa-producing countries, such as labor strikes, political instability, or export restrictions, can create supply shortages and drive up prices.
4. **Global Demand** – Rising demand from the chocolate industry, particularly in emerging markets like China and India, affects cocoa prices. Seasonal demand fluctuations, such as increased chocolate consumption during holidays, also contribute to price variability.
5. **Currency Exchange Rates** – Since cocoa is traded in U.S. dollars, fluctuations in currency values, particularly in producing countries, can impact global cocoa prices.
6. **Speculation and Market Dynamics** – Cocoa futures contracts are actively traded on commodity exchanges, and speculative trading activities by investors can influence short-term price movements.

Understanding these factors is crucial for modelling cocoa prices effectively, as they introduce various patterns, trends, and external shocks into time series data.

3 Data

In this project, you will work with two primary datasets: cocoa futures price data from the International Cocoa Organization (ICCO) and climate data (temperature and precipitation) from Ghana, sourced from the National Centers for Environmental Information (NCEI).

- Cocoa Futures Price Data:
  The file "Daily Prices_ICCO.csv" contains cocoa futures price obtained from the **International Cocoa Organization (ICCO)**. This dataset contains daily closing prices for cocoa futures contracts traded on major commodity exchanges. The time span of the data is from 1994.3.10 to 2025.2.27.
- Climate Data from Ghana:
  The file "Ghana_data.csv" contains climate data in Ghana, the largest cocoa-producing country in the world. The data is obtained from the National Centers for Environmental Information (NCEI) include:
  - STATION: The id of the observation station
  - NAME: The name of the observation station
  - DATE: The date of the observation
  - PRCP: The daily perception. If blank, then it means there is no perception on that day.
  - TAVG: The daily average temperature observed at 2 meters above ground
  - TMAX: The maximum daily temperature
  - TMIN: The minimum daily temperature

While these two datasets form the core of the analysis, you are encouraged to **incorporate any additional data sources** that you find relevant. For example, macroeconomic indicators, currency exchange rates, or other climate-related variables could further enhance your model. However, any additional data should be properly documented and justified in your report.

4 Output Requirements

Your final output for this project should be a well-structured report that clearly presents your approach to modelling cocoa futures prices. The report should be written in a formal style and include appropriate explanations, justifications, and interpretations of results. Below is a suggested structure for your report:

1. **Introduction** The introduction should provide an overview of the project, including a brief description of the study, the motivation of the study and its real-world significance. You may also briefly outline the objectives of your analysis and the key challenges involved.
2. **Literature Review** This section should summarize relevant research on time series forecasting and commodity price modelling. Highlight how your approach builds upon or differs from existing methods.
3. **Methodology** Clearly describe the forecasting methods you choose to apply. As mentioned earlier, this may include classical time series models covered in this course, or more sophisticated, e.g., machine learning approaches. Justify your choice of models based on theoretical considerations, past research and/or the pattern of the data. Explain in detail any steps, including preprocessing steps such as handling missing data, stationarity checks, transformation or feature engineering.
4. **Data** Provide a detailed description of the datasets used in the analysis. For any external data, please indicate the source of the data. You may also include Summary statistics and visualizations that help illustrate the characteristics of the data.
5. **Forecasting and Results** Present the results of your forecasting models, including:
    - Model training and validation process.
    - Performance evaluation using appropriate metrics.
    - Forecasted values and any observed patterns.
    - Graphical representations of predictions versus actual prices.
6. **Discussion and Conclusion** Interpret your results and discuss their implications. You may also consider including the limitations of your approach, the challenges you face and how can your analysis be improved or extended.
7. **Appendix**

The appendix should include any supplementary materials that support your analysis but are not suitable for the main body of the report. The full code used for the project **must** be provided in the Appendix. The appendix may also include: Additional elaborations, derivations, or mathematical justifications and extended statistical analyses, tables, or figures that complement the main discussion.

8. **References** Include a properly formatted list of references for any external sources cited in your report. Follow a standard citation style (e.g., APA, IEEE, or any format specified in the submission guidelines). All literature, datasets, and tools used in your analysis should be appropriately credited.

Your report should be well-organized and written in a clear and concise manner. The use of formulas, tables, figures, and references is encouraged to support your analysis and they should be well labelled. Ensure that your submission adheres to proper citation standards.

5 Marking Rubric

As mentioned in the course syllabus, the assessment of the group project is divided into two components: 20 marks for the overall quality of the project and 10 marks for individual contributions. For the first 20 marks, all group members will receive the same score, and a detailed rubric will be provided as follows. The remaining 10 marks will be awarded based on peer evaluations, reflecting each individual's contribution to the project.

The first 20 marks for the overall quality of the project will be assessed based on three key criteria: Writing and Organization, Method Correctness, and Creativity. Below is a breakdown of how each section will be evaluated.

1. **Writing and Organization (5 marks, 25%)** - Clear and concise writing that effectively communicates ideas. - Well-organized structure with appropriate use of headings, subheadings, and sections. - Logical flow of content with smooth transitions between sections. - Proper spelling, grammar, and formatting throughout the report. - Appropriate use of tables, figures, and visualizations to support the analysis.

2. **Method Correctness (5 marks, 25%)** - Appropriateness of the chosen methods and models for the problem with accurate application. - Justification of model choices, with clear explanations of why certain methods were used. - Proper handling of data, including necessary preprocessing steps. - Clear presentation of the models.

3. **Prediction Performance (5 marks, 25%)** - Accuracy of the forecasts based on relevant performance metrics (e.g., RMSE, MAE, MAPE). - Interpretation of forecast results and discussion of their reliability.

4. **Creativity (5 marks, 25%)** - Originality and creativity in the approach to the problem. - Exploration of techniques beyond the course material, including any self-driven research. - Innovative solutions or novel insights derived from the data and analysis. - Effective use of external data or additional sources to enhance the model or analysis. - Thoughtfulness in addressing the limitations of the chosen methods and

suggesting improvements or future work.

6 Individual Contribution

The individual contribution mark is calculated as follows:

First, each group member will assign a contribution score between 0 and 10 to every other member in their group, reflecting their perception of each member's effort and involvement in the project.

Then, each student's final individual contribution mark will be computed using the following formula:

$$\begin{aligned}&\text{Individual Contribution Mark}\\ =&10 \times \text{number of students in the group} \times \left(\frac{\text{Total Marks Received by the individual}}{\text{Total Marks Assigned by Group Members}}\right)\end{aligned}$$

For example, consider a group of four students. The following table represents their peer evaluation results, where each row corresponds to the marks assigned by an individual, and each column represents the marks received by each student.

		Receiver				Total
		A	B	C	D	
Grader	A		n_{AB}	n_{AC}	n_{AD}	
	B	n_{BA}		n_{BC}	n_{BD}	
	C	n_{CA}	n_{CB}		n_{CD}	
	D	n_{DA}	n_{DB}	n_{DC}		
Total		n_A	n_B	n_C	n_D	n_{total}

Table 1: A table elaborating the calculation of individual contribution mark

Then the grade received by the ith student will be

$$10 \times 4 \times \frac{n_i}{n_{total}}.$$

For example, marks received by Student A is $40 \times \frac{n_A}{n_{total}}$. Under this marking scheme, a student may receive an individual contribution mark greater than 10. Any excess marks will be treated as bonus points and included in the final evaluation for this course.

This adjustment ensures that students who contribute more significantly to the project receive appropriate recognition, while those who contribute less will have a lower individual contribution mark.

7 Deadline

The deadline of the final project is **Friday April 4th, 2025 23:59 PM**.