FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Eduardo Tadeu Santos Gomez Monteiro	United Kingdom	Eduardomonteiro128@gmail.com	
Ibrahim Adeshola Raji	Nigeria	rajiibrahim6@gmail.com	
Krishna Vamsi Regulavalasa	India	krishnavamsi8262@gmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).

Team member 1	Eduardo Tadeu Santos Gomez Monteiro
Team member 2	Ibrahim Adeshola Raji
Team member 3	Krishna Vamsi Regulavalasa

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.

Introduction

Time series data often exhibit non-stationarity, meaning their statistical properties change over time. To address this, we can use models like cointegration and error correction models (ECMs). This report explains how to model non-stationarity and find equilibrium using different datasets.

Our team chose to find Dataset that Illustrates Non Stationarity for the following three:

1) S&P 500 and Nasdaq
2) Amazon and Nasdaq
3) Bitcoin and Ethereum (USD)

Model Equations and Definitions

Cointegration and Vector Error Correction Model (VECM)

When two or more non-stationary time series are cointegrated, there exists a long-term equilibrium relationship between them. The VECM is an extension of the Vector Autoregression (VAR) model that incorporates cointegration.

Equations:

Johansen Cointegration Test:

$$\text{Let } \mathrm{Yt} = \begin{bmatrix} \mathrm{y}_{1,\mathrm{t}} \\ \mathrm{y}_{2,}\mathrm{t} \end{bmatrix}$$

The Johansen test for cointegration between $\mathrm{y}_{1,\mathrm{t}}$ and $\mathrm{y}_{2,}\mathrm{t}$ is based on the Vector Error Correction representation:

$$\Delta_{Yt} = \Pi\,{}^{Y}_{t-1} + \sum_{i=1}^{k-1} \Gamma\mathrm{i}\, \Delta_{t-i} + \epsilon\mathrm{t}$$

Where Π =αβ, α and β are matrices of parameters**, Γi** are short-term adjustment coefficients, and ϵt is a white noise error term.

Vector Error Correction Model (VECM):

- If cointegration is found, the VECM can be written as:

$$\Delta_{Yt} = u\ +\ \Pi\,{}^{Y}_{t-1} + \sum_{i=1}^{k-1} \Gamma\mathrm{i}\, \Delta_{t-i} + \epsilon\mathrm{t}$$

Where μ is a constant term, Π=αβ′, α represents the speed of adjustment to the equilibrium, and β represents the cointegrating vectors.

Importing, Structuring, and Modeling the Data

Python was used to import, structure and model the data for 3 pairs of datasets.

Calibrating the Parameters

Estimate Parameters: The parameters of the VECM include the cointegrating vectors (β), adjustment speeds (α), and short-term dynamics (Γ).

Interpretation:

α (Adjustment Speed): Indicates how quickly deviations from the equilibrium are corrected.

β (Cointegrating Vector): Represents the long-term relationship between the series.

Interpretation of Parameters

Adjustment Speed (α): If α is significant and negative, it indicates that deviations from the equilibrium are corrected in the subsequent periods, confirming the presence of a long-term equilibrium relationship.

Cointegrating Vector (β): The coefficients in β provide the long-term relationship between the variables. For instance, if β = [1, -0.5], it implies that in the long term, a 1 unit increase in NASDAQ is associated with a 0.5 unit decrease in Amazon's stock price.

Identifying reasons for choosing the Datasets.

S&P 500 and NASDAQ

Given the focus on modeling non-stationarity and finding equilibrium in time series data, a dataset that is rich in financial and economic information is preferred, as these often exhibit non-stationarity and potential long-term equilibrium relationships.

Why This Dataset Works Well:

- **Non-Stationarity:** Both indices are likely to exhibit non-stationarity, which can be confirmed using unit root tests.
- **Cointegration Analysis:** The relationship between S&P 500 and NASDAQ can be analyzed for cointegration, providing insights into their long-term equilibrium.
- **Practical Relevance:** Understanding the relationship between major stock indices is valuable for financial analysis and investment strategies.

By using a dataset of daily closing prices for major stock indices like the S&P 500 and NASDAQ, we can effectively demonstrate the concepts of non-stationarity and cointegration, making our analysis both robust and relevant.

Amazon Stock Price and NASDAQ

The choice of selecting Amazon stock price and NASDAQ for non-stationarity modeling is influence major by 3 factors:

1. **High Volatility and Trend Components**: Amazon's stock price and the NASDAQ index exhibit significant volatility and clear trends over time, which are essential for identifying and analyzing non-stationary behavior in financial time series.

2. **Real world relevance:** Considering the need to make use of real data, the Amazon stock price and NASDAQ provides a true representation of market influence, and a clear understanding of market dynamics, portfolio management and economic forecasting.

3. **Data Availability and Quality**: Both Amazon and NASDAQ provide extensive, high-quality historical data, enabling rigorous statistical analysis and model validation.

Cryptocurrencies (Ethereum & Bitcoin - USD)

The dataset was retrieved from Yahoo Finance and imported into Notebook using API - Application programming interface. These cryptocurrencies are in USD ($) and selected for the following reasons:

1) High frequency data availability: Cryptocurrencies most often give high frequency data, enabling more accurate short term dynamics modeling. While capturing cointegration interactions, this can be useful.

2) Lack of Capital controls & Decentralization: since the selected cryptocurrencies are decentralized, they are not under capital restrictions that can impede the flow of money.

3) Market Dynamics & New asset class: They represent new asset classes with distinct market dynamics. The dynamics such as rapid price movements & high volatility, can present interesting opportunities for cointegration modeling and analysis.

Team Selects Cryptocurrency Dataset:

Our Team chose to work on Cryptocurrencies for their unique benefits and challenges in modeling stationarity and cointegration. Here are the main reasons:

A. Volatility: Cryptocurrencies are highly volatile, which makes them ideal for studying non-stationarity.
B. Global Significance: Cryptocurrencies are used worldwide, providing larger sample sizes. This helps find correlations across different markets & regions.
C. Market Efficiency: Crypto Markets quickly incorporate new information, which allows for the exploration of cointegration relationship and market efficiency hypothesis.
D. Technology and data availability: The Cryptocurrencies uses blockchain technology, offering transparent and detailed data transactions. This high frequency helps in analyzing non stationary and cointegration relationships more accurately.
   The accessibility of such information may enable more precise modeling and examination of economic ideas.

Modeling non-stationary & Finding an Equilibrium in Bitcoin & Ethereum Cryptocurrency

Definition:
In time series analysis, non-stationary data is to be transformed to become stationary for accurate analysis and forecasting. To find long term equilibrium relationships between variables, a concept we know as "cointegration", We utilize Augmented dickey fuller method to check for non-stationarity.

Then we have used the Engle granger method to test cointegration by running a regression on cryptocurrencies. If the residuals are stationary, it indicates a long-term relationship.

To address models lacking long run solutions, where it cannot determine equilibrium relation between cryptocurrencies, we introduce a class of models (error correction model) that might overcome this problem, providing a solution to maintain equilibrium relation.

Description:
Non-stationary modeling involves dealing with a situation where the properties in prices of cryptocurrencies change overtime, leading to trends, fluctuation irregularities that complicates the accurate analysis and forecasting.
In contrast, finding an equilibrium refers to the concept that if cryptocurrencies have a long-term equilibrium relation, any deviations will be corrected in the short term from this equilibrium.

Demonstration:
To show non-stationarity and finding an equilibrium, we have used Bitcoin and Ethereum cryptocurrency.
First, we started checking for stationarity with the ADF method. And the hypothesis is:

HO: variable has a unit root.
H1: variable doesn't have a unit root.
We reject the null hypothesis if the p-value of H0 is <0.05. If not, we don't reject it. For bitcoin and Ethereum level form, the ADF probability values are 0.821 and 0.720 respectively. Since both are >0.05, we don't reject H0. We can conclude that Ethereum and bitcoin both have a unit root.

Figure 1 - Daily Prices of Bitcoin and Ethereum between 2020 and 2024

Figure 2 - First Difference of Daily Prices of Bitcoin and Ethereum between 2020 and 2024

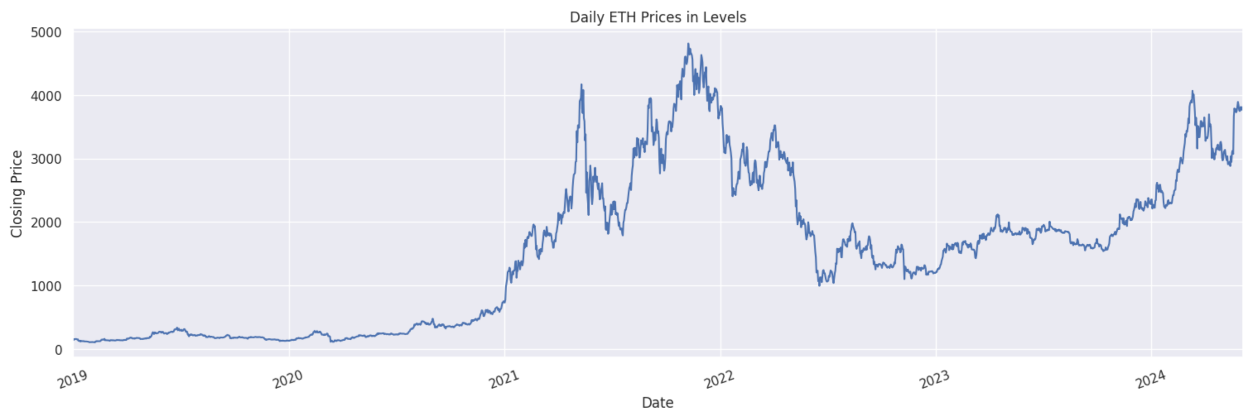

Figure 3 - The plot for the Daily ETH Prices in Levels:

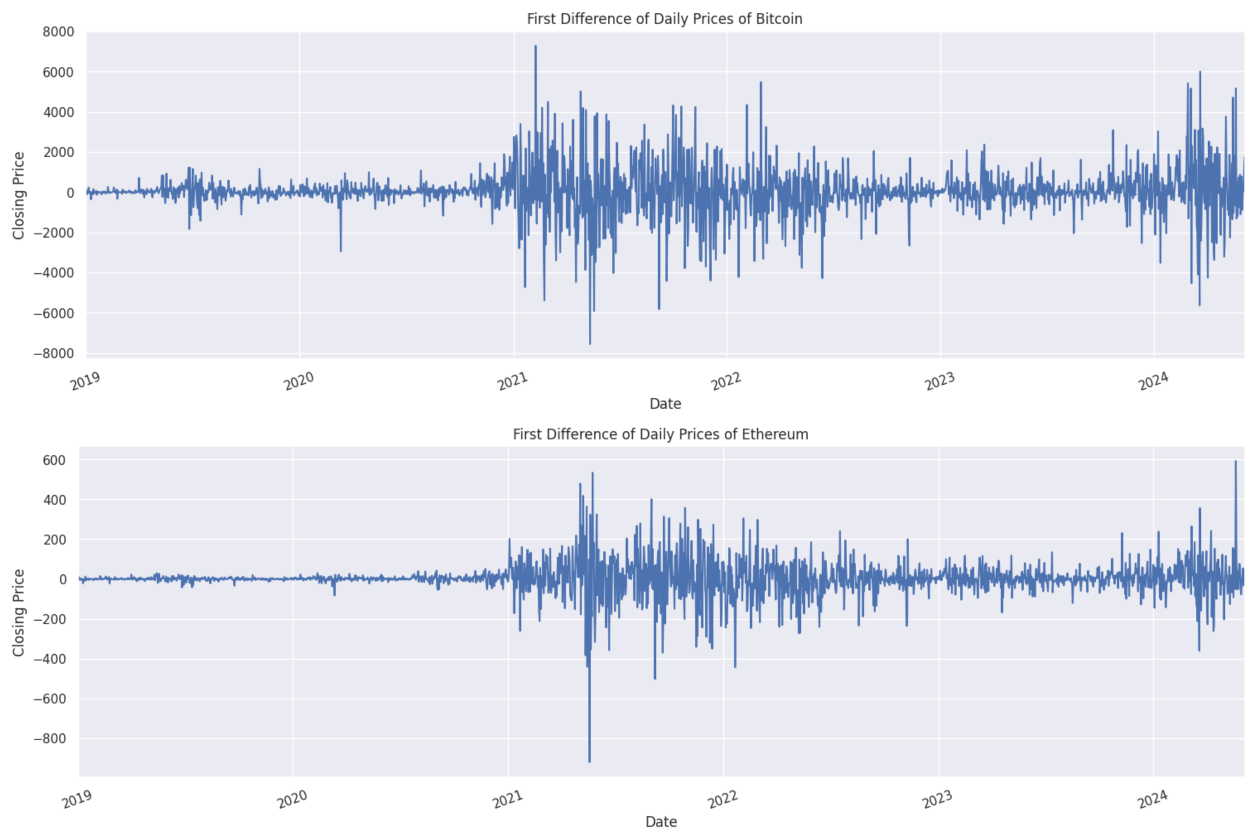

Figure 4 - Plot for the First Difference of Bitcoin prices and Ethereum Prices

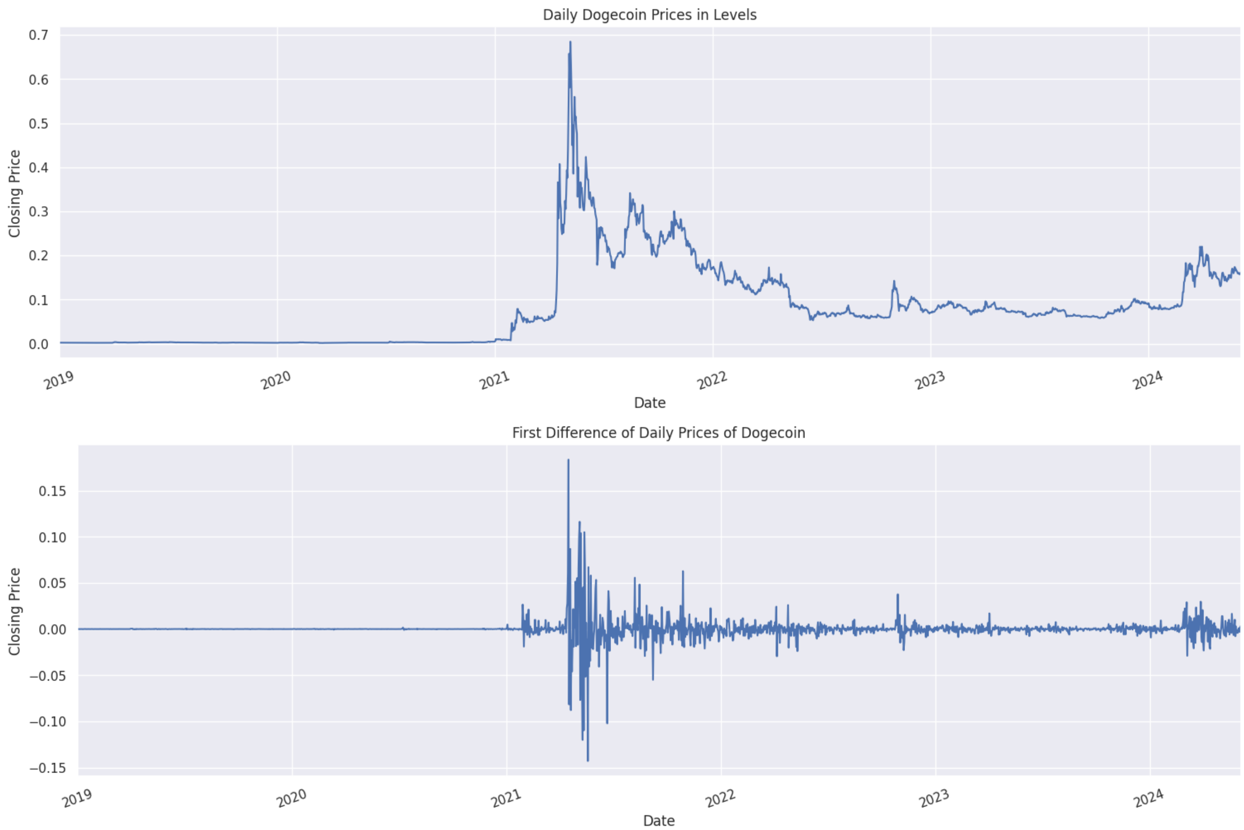

Figure 5 - plot for the First Difference of Dogecoin prices

Diagnosis:

To show that the selected currencies are useful to this project, we included plots of their daily prices in levels and after taking the first difference, as shown in the Diagram section.

1) The availability of daily prices offers a unique opportunity for high frequency data (good for time series analysis). Because it provides a larger sample size.
2) The fluctuations during the covid-19 period, are externally determined and free from capital restrictions.

Visualizing the diagrams, the initial plots of ETH and BTC do not show constant mean, variance, indicating non-stationarity. After taking the first difference, the plots are mean reverting and exhibit stable variance and autocovariances.

We will demonstrate the stationarity and non-stationarity of the time series using the Augmented dicker fuller test. We also test for cointegration using the Engle-granger method and the Johansen cointegration test.

Based on results we shall use VAR and VEC methods. Finally, we apply Granger causality tests to check whether lagged values of one variable conditionally help predict another variable.

Damage:

The Engle-Granger method isn't good for cases having more than two variables. This is because it only tests cointegration between two variables, so it can't handle multivariable situations.
The Johansen test can analyze cointegration among multiple variables, Economic relations in real life usually involve more than two variables.

The Johansen procedure can also recognize multiple cointegrating vectors if they exist.
This makes the Johansen test more robust and efficient.

Directions:

To reflect a real-world scenario, we introduced Dogecoin as a 3rd cryptocurrency (DOGE-USD). By doing so, we turned to the Johansen cointegration procedure to establish a long-term relationship between cryptocurrencies.

1. Lag selection procedure: Choosing the right lag length is crucial for the Johansen Cointegration Procedure's accuracy.
    a) Akaike Information Criterion: This helps determine the lag length. Especially good for small sample sizes.
    b) High order VAR Drawbacks: While adding more lags can lead to a loss of degrees of freedom and biased estimates, it is often acceptable to ensure error terms are white noise.
    c) Wald test statistics: this method can also be used to check if coefficients at each lag are zero.
2. Trace/Maximum Eigenvalue Statistic for Johansen Cointegration Test:
   Once we have obtained a VAR model with the suitable number of lags and white noise residuals, we can move on to determine the number of cointegrating vectors, denoted by r0, using the Trace and Maximum Eigenvalue statistics.
    d) **VAR Model and Cointegration**: After selecting the appropriate lag length and ensuring white noise residuals, we determine the number of cointegrating vectors using the Trace and Maximum Eigenvalue tests
    e) **Granger Causality Tests**: Conducting Granger causality tests with the VAR model to see if lagged values of one variable predict another
    f) **Model Specification**: Ensuring stationarity at the first difference, with variables detrended and allowing for an intercept.
    g) **Causality Criteria**: Use chi-square statistics for hypothesis testing.

Deployment:

The big change in our model is that we introduced another cryptocurrency known as dogecoin (DOGE-USD). For completeness, we confirm that the prices of Ethereum are stationary at first difference.
From python script results, the optimal lag length was found to be 3 with a minimum AIC of 11.296 approximately. Based on maximum eigenvalue statistics for JC test analysis in python script, we don't reject the null hypothesis and that means there is no cointegrating rank of VEC for the cryptocurrencies.
A VAR model will be used to estimate the granger causality tests for cryptocurrencies.

References

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427-431.

Hamilton, J. D. (1994). Time series analysis. Princeton University Press.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrics, 59(6), 1551-1580.

Johansen, S. (1995). Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press.

Yahoo Finance. “Yahoo Finance - Business Finance, Stock Market, Quotes, News.” YahooFinance, 2024, finance.yahoo.com.