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.

Table of Contents

Multicollinearity

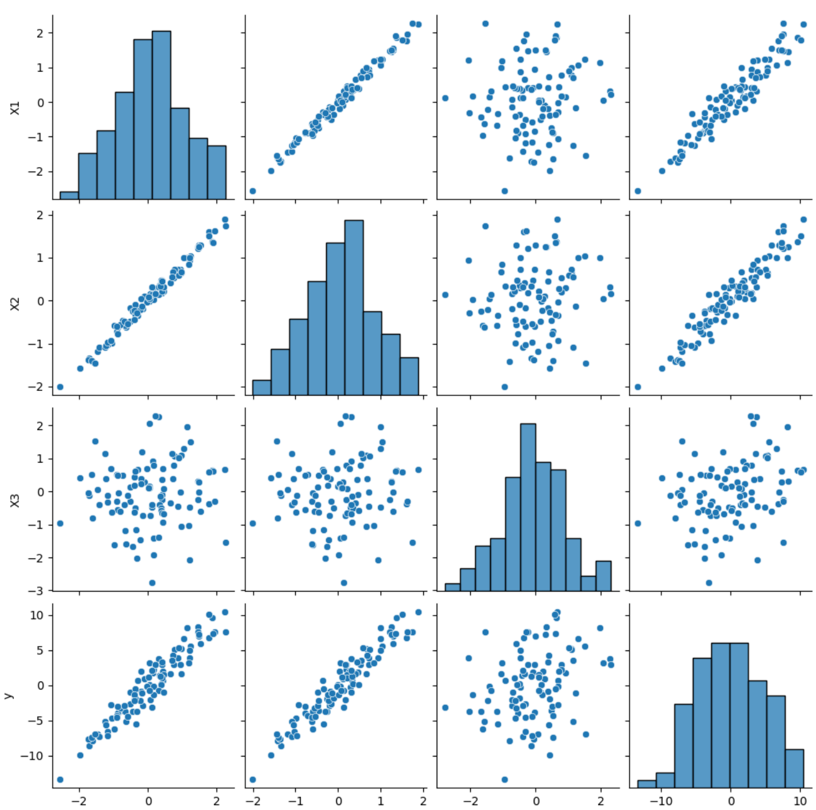
Definition
Multicollinearity occurs when two or more predictor variables in a regression model are highly correlated, meaning one can be linearly predicted from the others with a substantial degree of accuracy. Formally, if $X1$ and $X2$ are two predictor variables, multicollinearity is present if there exists a high correlation coefficient $\rho(X1, X2)$ close to ±1.

Description
Multicollinearity complicates the estimation of regression coefficients and makes it difficult to assess the individual effect of each predictor variable on the response variable.

Demonstration and Diagram
Using Python was generated a simulated dataset for 3 constants (X1, X2, and X3), after fitting a linear regression model was calculated the VIF for all of them.

	Variance Inflation Factor (VIF)
X1	64.366429
X2	64.421285
X3	1.008147

Was then plotted a pairplot to visualize correlations.



Diagnosis

There are 2 quick ways to recognize or test for multicollinearity in a dataset, the first would be by checking the correlation coefficients between predictor variables using a correlation matrix. High correlation (close to ±1) indicates multicollinearity. The second one would be by performing statistical tests by checking the VIF. A VIF above 10 is often taken as an indicator of multicollinearity.

In our dataset we can see that the VIF for X1 and X2 is way above 10 and the scatterplot shows a strong linear relationship, confirming high multicollinearity.

Damage

Multicollinearity can cause several issues in statistical modeling and data analysis:

- Unstable Estimates - coefficients can become very sensitive to changes in the model.
- Reduced Interpretability - difficulty in understanding the individual effect of correlated predictors.
- Inflated Standard Errors - making hypothesis tests for coefficients unreliable.

Directions

To address multicollinearity we can remove one of the correlated predictors, transform the correlated variables into a set of linearly uncorrelated components using Principal Components Analysis (PCA). We can also use Ridge Regression, this adds a penalty to the size of coefficients, which can reduce the impact of multicollinearity. And like PCA, we can use Partial Least Squares to transform predictors but it will also consider the response variable.

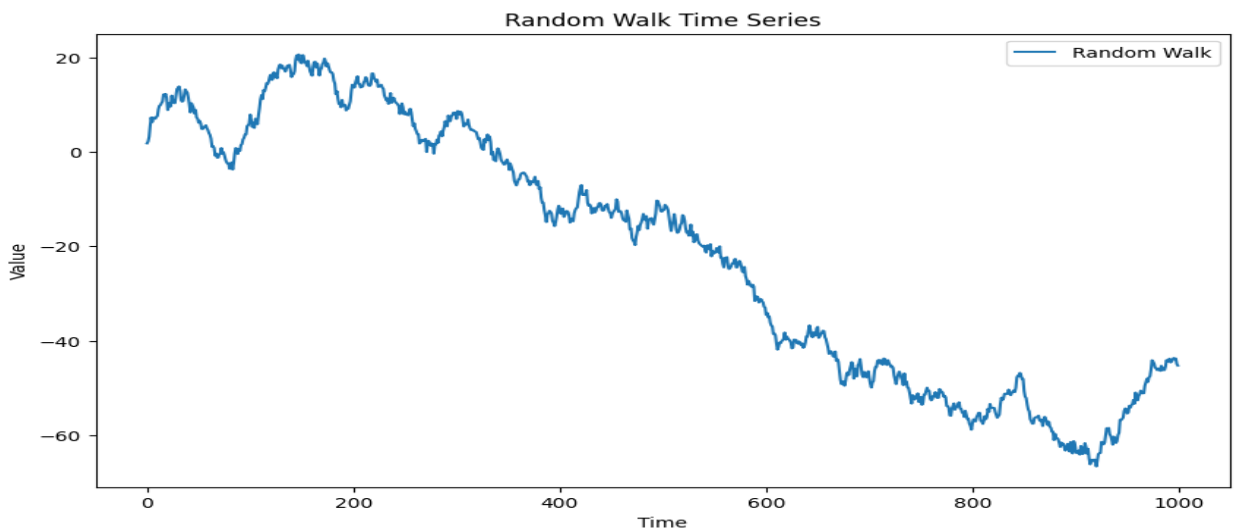
By implementing these strategies, we can mitigate the impact of multicollinearity on our model and data analysis, leading to more reliable and interpretable results.

Unit Root Testing

Definition
A **unit root** in a time series denotes that the series is non-stationary and possesses a stochastic trend. In simpler terms, it means that the time series has a tendency to wander over time without reverting to a long-term mean, making it unpredictable. Unit root testing involves statistical methods to determine whether a time series is stationary or contains a unit root (Dickey & Fuller, 1979).

Description
In time series analysis, a stationary process is one whose statistical properties, such as mean and variance, do not change over time. Conversely, a non-stationary process with a unit root implies that shocks to the system have permanent effects. Detecting unit roots is crucial because many statistical models, such as ARIMA, require the series to be stationary to make reliable inferences (Hamilton, 1994).
A common model for unit root testing is the Autoregressive (AR) model, particularly the AR(1) process.

Demonstration and Diagram



Diagnosis
To carry out a diagnosis of unit root testing using the Augmented Dickey-Fuller (ADF) test, the following analysis are done;

- **ADF Statistic**: A negative value indicates stronger evidence against the null hypothesis of a unit root.
- **p-value**: If the p-value is less than a chosen significance level (e.g., 0.05), we reject the null hypothesis, suggesting the series is stationary.
- **Critical Values**: Compare the test statistic to critical values at different significance levels (1%, 5%, 10%). If the statistic is less than the critical value, reject the null hypothesis.

Damage

The underlisted are the likely damages that might result from unit root testing;

- **Predictability**: Non-stationary series with unit roots are less predictable and can lead to unreliable forecasts.
- **Model Mis-specification**: Using inappropriate models for non-stationary data can lead to spurious results and misleading inferences (Granger & Newbold, 1974).
- **Volatility**: Non-stationary data can exhibit greater volatility and variability over time, complicating risk management and financial analysis.

Directions

To address unit roots in a time series, one can apply differencing by transforming the series yty_tyt to Δyt=yt−yt−1\Delta y_t = y_t - y_{t-1}Δyt=yt−yt−1 to achieve stationarity. Additionally, transformations such as logarithms or detrending can be used. It is crucial to select models that can handle non-stationary data, like ARIMA models with integration parameters. For robustness, further tests such as the Phillips-Perron test or the KPSS test can be utilized (Phillips & Perron, 1988).

Random Change Models

Definition

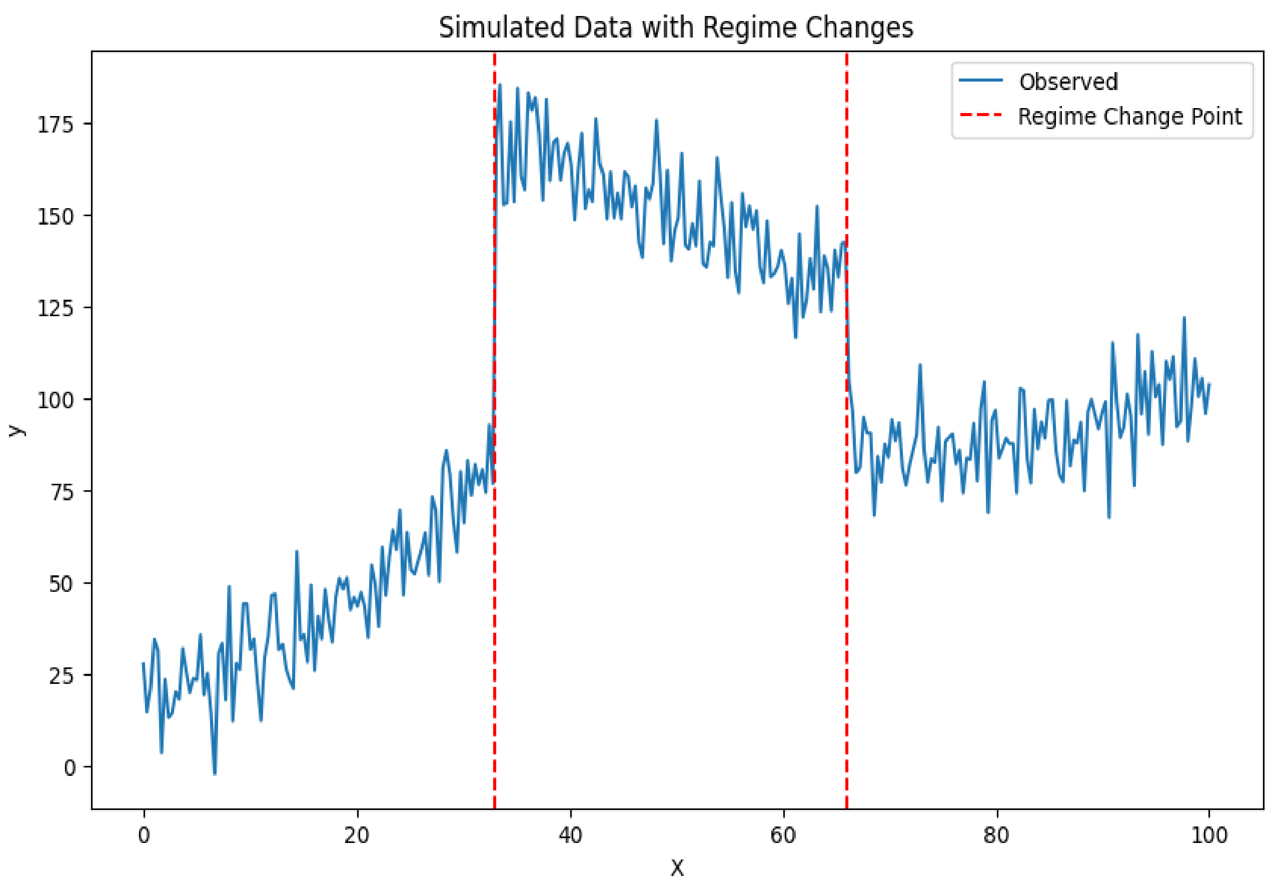
Regime change models, also known as regime-switching models, are statistical models used in quantitative finance to capture changes in the behavior or structure of financial time series. These models assume that financial markets can be in different "regimes" or states, each characterized by its own set of statistical properties such as mean, variance, and correlations. The regime can switch based on certain probabilistic rules, often modeled using hidden Markov models (HMMs) or other state-transition frameworks (Hamilton, 1989).

Description

Regime change models address the non-linear dynamics of financial markets by allowing for shifts between different market conditions. Traditional financial models, like the Black-Scholes model, assume a single set of parameters governing the entire dataset. However, financial markets are influenced by various economic factors, policy changes, and investor behaviors, leading to different periods or regimes (Ang & Timmermann, 2012).

Demonstration and Diagram

Using Python a simulated dataset with regime changes was generated and then plotted.



Diagnosis

Diagnosing regime changes involves identifying the current state of the market and predicting the probability of transitioning to another state. This can be achieved through:

- Using algorithms like the Baum-Welch algorithm to estimate the probabilities of being in each regime at each point in time (Baum et al., 1970).
- Comparing the model's predictions with actual market data to evaluate its performance, often using metrics like log-likelihood, AIC, or BIC.
- Evaluating the model under various hypothetical scenarios to understand its robustness and reliability (Ang & Timmermann, 2012).

Damage

Improper handling of regime changes can lead to significant damage, including:

- Incorrectly estimating asset prices and derivatives can lead to substantial financial losses.
- Inadequate risk models can result in underestimating the risk during regime shifts, potentially leading to significant financial distress (Ang & Timmermann, 2012).
- Strategies based on static assumptions might fail to perform well during regime changes, affecting portfolio returns and increasing volatility.

Direction

The direction of regime change modeling is towards more sophisticated and integrated approaches, including:

- Machine Learning: Leveraging machine learning techniques to improve regime detection and transition prediction (Nystrup et al., 2018).
- High-Frequency Data: Incorporating high-frequency data to capture more granular regime changes and improve the accuracy of models.
- Multivariate Models: Developing models that can handle multiple assets and their interactions, providing a comprehensive view of market regimes.

Joining Time Series with Different Frequencies

Definition: Joining time series with different frequencies involves combining two or more time series datasets that have observations recorded at different intervals. The process involves matching up the times from different datasets and combining them into one big dataset. This is a big topic in time series analysis and working with data. (McKinney, 2017)

Description: Combining time series with different frequencies usually requires resampling one or both series to a common frequency to ensure compatibility for analysis. This process is important for insights. This is often needed when looking at data from different places or when the data is in different sizes or details. (Brockwell & Davis, 2016, pp).

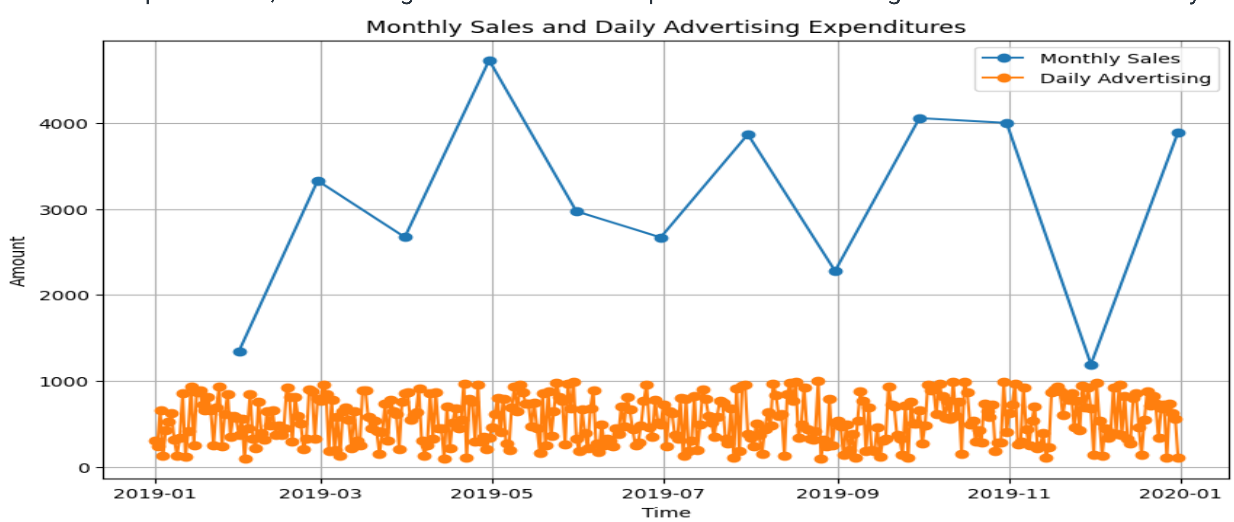
Demonstration: let's consider a scenario. Imagine we have the following Datasets:

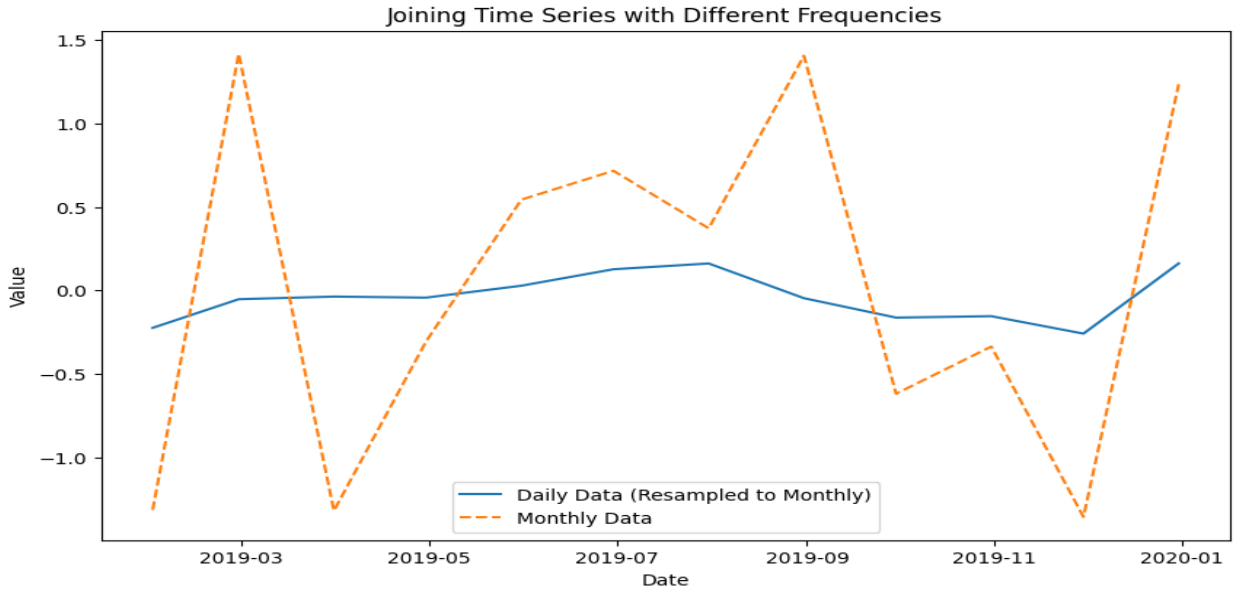
1. Monthly Sales Data: Total sales for each month.
2. Daily Advertising Expenditure Data: Advertising expenses for each day.

Steps:

Simulate Data: We will create a sample data for both monthly sales and daily advertising expenses.
Merge Data: Aggregate daily advertising expenses to monthly totals.
Analyse Relationship: We Examine the correlation between monthly advertising spending and sales.
By merging and analysing these datasets, we are aiming to understand how monthly sales are impacted by daily advertising.

Diagram: The plot represents the resampled daily data alongside the monthly sales data and daily advertise expenditure , illustrating how different frequencies can be aligned for combined analysis.





Diagnosis:
To identify the problem of different frequencies, we can inspect the time intervals of the datasets and observe any discrepancies in the frequencies. If the intervals do not align (e.g., one is daily and the other is monthly), the data frequencies are mismatched. (Hyndman & Athanasopoulos, 2018, pp. 76-92).

Damage: If not addressed, joining time series with different frequencies can give rise to several issues:

- Misalignment of data points
- Incorrect analysis
- Misleading results
- Impact the accuracy of analysis
  For example, monthly trends might not align with daily variations, causing inaccuracies in any combined analysis or forecasting.

Directions:
- Resampling: Adjusting the frequency of one time series to match the other (e.g., resample daily data to monthly by taking the mean).
- Interpolation: We can estimate values for the lower frequency series to match the higher.
- Time Series Models: Use of models that can handle mixed frequencies, like "state-space models"

How two Challenges can relate:

1. Multicollinearity
2. Unit Root Testing

While multicollinearity and unit root testing are distinct statistical issues, they share several similarities, and they can influence each other in various contexts. Understanding their relationship is necessary for effective time series analysis, model accuracy.

Small Sample Size
Both multicollinearity and unit root testing can be significantly impacted by small sample sizes. A limited sample size might lead to an unreliable statistical results and reduced power to detect a unit root. Similarly, small sample sizes can cause unstable coefficient estimates and inflated confidence intervals in regression models, exacerbating multicollinearity.

Misspecification of Alternative Hypothesis
The reliability of unit root tests and multicollinearity diagnostics can be affected by the misspecification of the alternative hypothesis. Unit root tests, such as the Augmented Dickey-Fuller test, depends on the right specification of the alternative hypothesis.
If the true data-generating process should differ from the assumed alternative hypothesis, the test results may become unreliable. Similarly, in multicollinearity testing, if the alternative hypothesis is incorrectly specified, the results can be misleading.

Structural Breaks
Both unit root testing and multicollinearity can be influenced by structural breaks in the data. Unit root tests assumes a continuous data generation process. The presence of structural breaks can lead to wrong conclusions about the presence of a unit root.
In multicollinearity, structural breaks can alter the correlation patterns between variables, potentially increasing or decreasing multicollinearity.
We can address this common challenge by incorporating structural breaks into the unit root analysis as well as in solving multicollinearity problems.

Nonlinear Dependencies
Unit root tests are typically going to assume linear dependencies in the data. If the underlying relationships are nonlinear, traditional unit root tests may give false positives. Similarly, multicollinearity tests assume linear relationships between variables.

Data Transformations
Data transformations can help addressing both unit root and multicollinearity issues. Differencing the time series data can help eliminate unit roots, making the series stationary. For multicollinearity, we can reduce the correlation between predictors, mitigating multicollinearity by transforming variables (e.g., taking logs)

Conducting many tests on the same data results in false positives, this applies to both unit root and multicollinearity tests. To avoid these errors, we should adjust significance thresholds or use correction techniques to ensure results are reliable.

Conclusion
Although unit root testing and multicollinearity are separate challenges, they share common factors such as the

1. Impact of sample size
2. Structural breaks
3. Nonlinear dependencies.

Addressing these issues requires careful model specification, appropriate data transformations, and consideration of multiple testing. By understanding how these problems work together, analysts can make their time series models more accurate and trustworthy.

References

1. 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(366), 427-431.
2. Granger, C. W. J., & Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, 2(2), 111-120.
3. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
4. HAYES, ADAM. “Multicollinearity.” Investopedia, 4 Jan. 2022, www.investopedia.com/terms/m/multicollinearity.asp#:~:text=Multicollinearity%20is%20a%20statistical%20concept.
5. Phillips, P. C. B., & Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, 75(2), 335-346.
6. Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer. Available at SpringerLink. (pp. 123-145)
7. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts. Available online. (pp. 76-92)
8. Peng, R. D. (2008). "Simpleboot: Simple Bootstrap Routines". R package version 1.1-3. CRAN. (pp. 1226-1227)
9. Ang, A., & Timmermann, A. (2012). Regime Changes and Financial Markets. Annual Review of Financial Economics, 4(1), 313-337.
10. Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. The Annals of Mathematical Statistics, 41(1), 164-171.
11. Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica, 57(2), 357-384.
12. Nystrup, P., Lindström, E., Madsen, H., & Pinson, P. (2018). Detecting Change Points in Time Series: A Review of Bayesian and Frequentist Methods. Energies, 11(5), 1206.