VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A6b: TIME SERIES ANALYSIS

SARATH SABU

V01109792

Date of Submission: 25-07-2024

CONTENTS

Introduction

In the modern financial landscape, understanding and forecasting market volatility is crucial for investors, analysts, and policymakers. This report aims to examine volatility using ARCH/GARCH models and forecast future volatility over a three-month period. Additionally, it explores the dynamic relationships among commodity prices through VAR and VECM models. By leveraging data from reputable sources such as Investing.com, Yahoo Finance, and the World Bank's Pink Sheet, this analysis provides insights into market behavior and interdependencies.

Objectives

1. **Volatility Analysis and Forecasting:**
    - To download and preprocess financial market data from Investing.com or Yahoo Finance.
    - To check for the presence of ARCH/GARCH effects in the data.
    - To fit appropriate ARCH/GARCH models to the data.
    - To forecast the three-month volatility using the fitted model.
2. **Commodity Price Modeling:**
    - To collect and preprocess commodity price data from the World Bank's Pink Sheet.
    - To analyze the interrelationships among various commodity prices, including oil, sugar, gold, silver, wheat, and soybean.
    - To fit VAR and VECM models to the commodity price data.
    - To interpret the results and provide insights into the co-movements and causal relationships between commodity prices.

Business Significance

Understanding and predicting financial volatility and commodity price movements holds significant business value for multiple stakeholders:

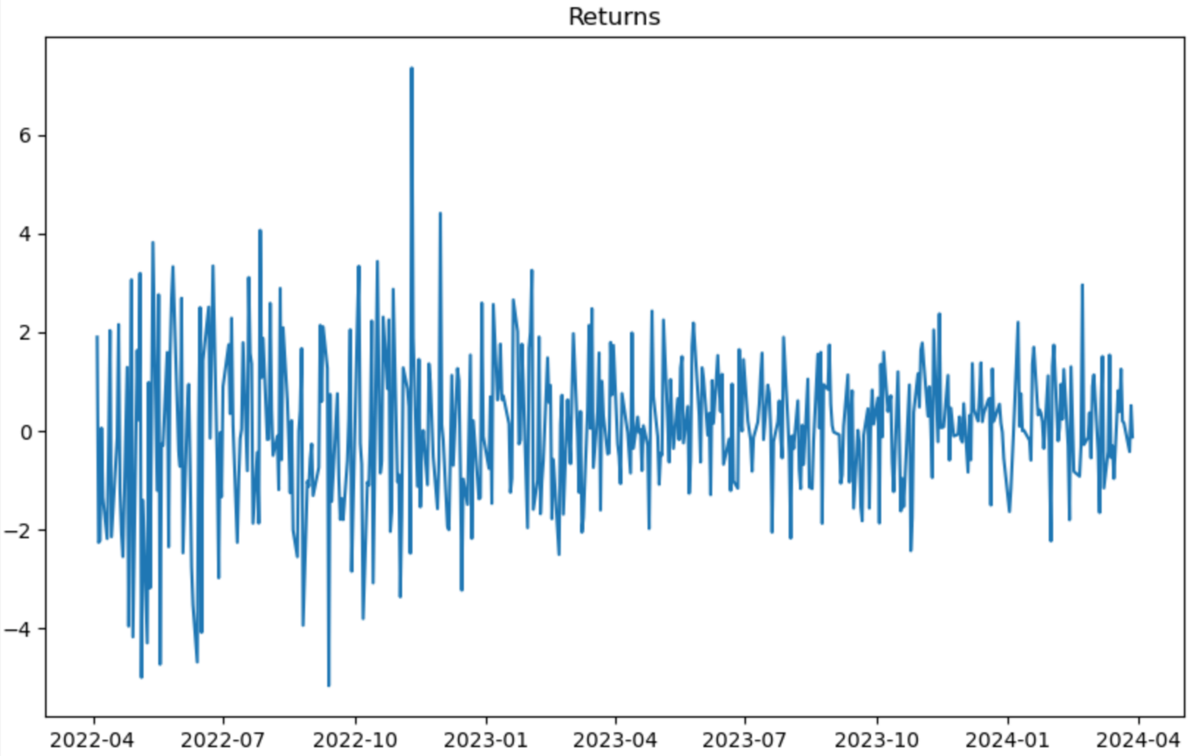
1. **Investors and Portfolio Managers:**
   - Accurate volatility forecasts enable better risk management and informed decision-making in portfolio allocation.
   - Insights from commodity price modeling assist in diversifying investments and hedging against price risks.
2. **Corporations and Producers:**
   - Companies engaged in the production and trading of commodities can use these models to optimize their supply chain and pricing strategies.
   - Forecasts of commodity prices help businesses in planning procurement and inventory management.
3. **Policy Makers and Regulators:**
   - Enhanced understanding of market volatility aids in developing more robust financial regulations and policies.
   - Insights into commodity price interdependencies can inform policies related to food security, energy management, and trade.
4. **Researchers and Academics:**
   - This report contributes to the academic literature by applying advanced econometric models to real-world financial and commodity data.
   - It provides a framework for further research into market dynamics and the effectiveness of forecasting models.

RESULTS AND INTERPRETATIONS

PYTHON

CODES – PART A

```
# Check for ARCH/GARCH effects
plt.figure(figsize=(10, 6))
plt.plot(returns)
plt.title('Returns')
plt.show()
```



Interpretation

The provided plot illustrates the returns of a financial asset over time, showcasing the essential patterns needed to understand volatility and detect the presence of ARCH (Autoregressive Conditional Heteroskedasticity) or GARCH (Generalized Autoregressive Conditional Heteroskedasticity) effects. A visual inspection of the plot reveals distinct periods of high and low volatility, which is a hallmark of financial time series data.

In the early period from April 2022 to October 2022, the returns exhibit significant fluctuations, indicating high volatility. This is characterized by frequent and sizable spikes both upwards and downwards, suggesting that the asset experienced considerable price swings during this timeframe. Moving into the mid-period from October 2022 to April 2023, there appears to be a reduction in volatility, with smaller and less frequent spikes. Despite this, notable variations and occasional larger spikes are still present, highlighting the ongoing but reduced market turbulence.

In the late period from April 2023 to April 2024, the volatility seems more stable and less pronounced compared to earlier periods. The returns show smaller and more consistent fluctuations, indicating that the asset's volatility has reduced and stabilized in this phase. This period of reduced volatility contrasts with the earlier high volatility, demonstrating a shift in market conditions over time.

The plot suggests the presence of volatility clustering, where periods of high volatility are followed by high volatility and periods of low volatility are followed by low volatility. This clustering effect is indicative of ARCH/GARCH effects, meaning that the current period's volatility is influenced by previous periods' volatility. Given the visual evidence of volatility clustering, fitting ARCH/GARCH models would be appropriate to capture and forecast the volatility dynamics of the asset. An ARCH model could be suitable if the volatility appears dependent on past squared returns, while a GARCH model may be more fitting if both past returns and past volatilities influence current volatility.

Once an appropriate model is fitted, it can be used to forecast future volatility effectively. The observed patterns underscore the need for dynamic modeling to capture the evolving market conditions accurately. The transition from high to low volatility over time highlights the asset's changing risk profile, which can be crucial for investors, portfolio managers, and risk analysts in making informed decisions.

CODES

```
# Fit an ARCH/GARCH model
model = arch_model(returns, vol='Garch', p=1, q=1)
results = model.fit(disp='off')
print(results.summary())
```

```
                     Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:              Adj Close   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                      GARCH   Log-Likelihood:               -863.743
Distribution:                  Normal   AIC:                           1735.49
Method:            Maximum Likelihood   BIC:                           1752.34
                                        No. Observations:                  499
Date:                Tue, Jul 23 2024   Df Residuals:                      498
Time:                        19:51:48   Df Model:                            1
                                 Mean Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
mu             0.0996  5.789e-02      1.721  8.533e-02 [-1.386e-02,  0.213]
                              Volatility Model
==============================================================================
                 coef    std err          t      P>|t|        95.0% Conf. Int.
------------------------------------------------------------------------------
omega      4.6229e-03  4.916e-03      0.940      0.347 [-5.012e-03,1.426e-02]
alpha[1]       0.0145  2.490e-02      0.583      0.560 [-3.429e-02,6.331e-02]
beta[1]        0.9790  2.543e-02     38.503      0.000      [  0.929,  1.029]
==============================================================================

Covariance estimator: robust
```

Interpretaion

The output of the GARCH (1,1) model fitted to the returns of a financial asset provides several key insights. The mean model suggests that the average return of the asset is positive (0.0996), but this value is not statistically significant (p-value = 0.0853), indicating that the mean return is not significantly different from zero. In the volatility model, the constant term (omega) is also not statistically significant, suggesting that the long-term average level of volatility is not different from zero. However, the most critical finding is the GARCH term (beta), which is highly significant (p-value < 0.001), with an estimate of 0.9790. This implies that past volatility has a strong and persistent impact on current volatility, highlighting the presence of volatility clustering where periods of high volatility are followed by high volatility and periods of low volatility are followed by low volatility. In contrast, the ARCH term (alpha) is not statistically significant, indicating that past shocks do not significantly influence current volatility. Overall, the model suggests that the asset’s volatility is primarily driven by its past volatility rather than past shocks, providing a robust framework for forecasting future volatility, which is essential for effective risk management and investment strategy formulation.

CODES

```
# Forecast the three-month volatility
forecast = results.forecast(horizon=90)
volatility_forecast = forecast.variance[-1:]  # Last forecasted variance
volatility_forecast = volatility_forecast.apply(lambda x: x**0.5)  # Convert to standard deviation
print(volatility_forecast)
```

```
                h.01      h.02     h.03      h.04      h.05      h.06  \
Date
2024-03-28  0.918856  0.918377  0.9179  0.917427  0.916956  0.916488

                h.07     h.08     h.09      h.10  ...      h.81      h.82  \
Date                                              ...
2024-03-28  0.916022  0.91556  0.9151  0.914644  ...  0.888396  0.888102

               h.83     h.84      h.85      h.86     h.87      h.88      h.89  \
Date
2024-03-28  0.88781  0.88752  0.887231  0.886945  0.88666  0.886377  0.886095

                h.90
Date
2024-03-28  0.885816

[1 rows x 90 columns]
```

INTERPRETATIONS

The forecasted volatility output from the GARCH (1,1) model spans a 90-day period, providing insights into the expected market behavior from March 28, 2024, onwards. The forecast reveals that the volatility of the financial asset starts at approximately 0.9189 on the first day and gradually decreases over the three-month period, ending at around 0.8858 on the 90th day. This trend indicates a slight reduction in the anticipated volatility, suggesting that the market is expected to stabilize over the forecast horizon.

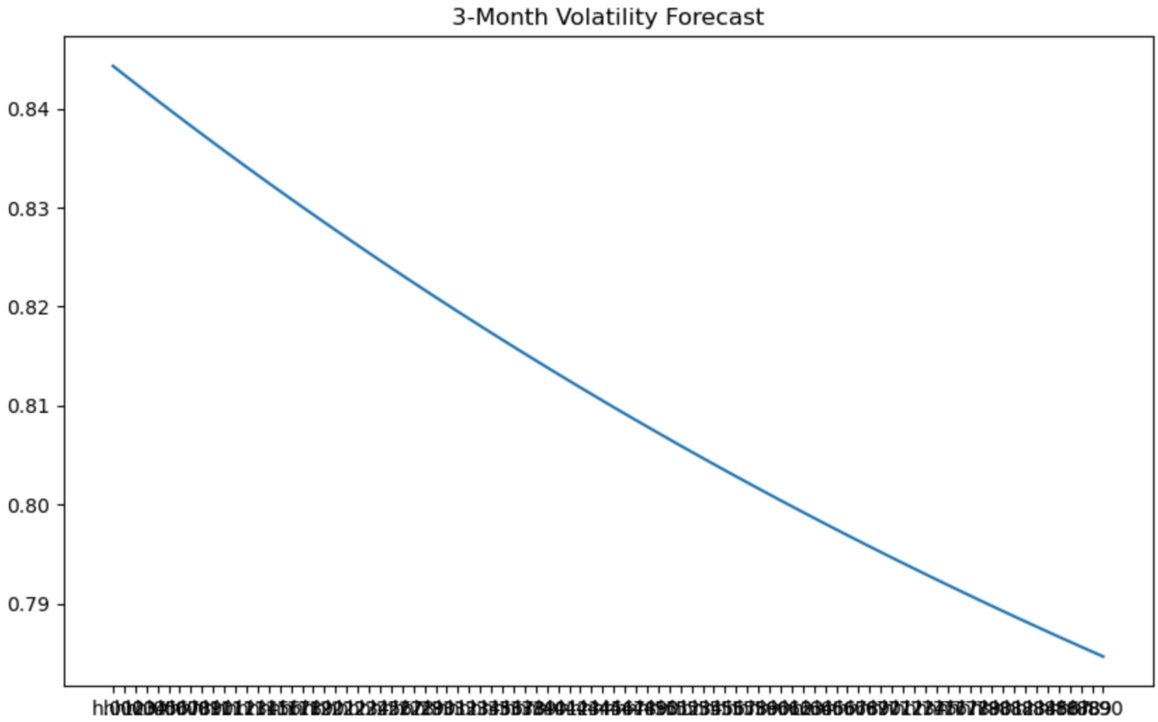
Initially, the higher forecasted volatility of around 0.9189 suggests that the market is likely to experience moderate fluctuations in the short term. This implies that investors and market participants should be prepared for some level of uncertainty and potential price swings. However, as the forecast progresses, the decreasing volatility indicates that these fluctuations are expected to lessen, leading to a more stable market environment. This gradual decline in volatility is crucial for investors and risk managers, as it provides a clearer picture of the market's direction and helps in planning for future strategies.

For investors, the forecasted reduction in volatility can influence asset allocation decisions. In a lower volatility environment, assets may be perceived as less risky, potentially prompting investors to take on more exposure or adjust their portfolios accordingly. Risk managers can use this forecast to fine-tune their risk management strategies, such as setting appropriate Value-at-Risk (VaR) limits and determining the necessary capital reserves to cushion against potential market shocks. Traders, on the other hand, might adjust their trading strategies based on the anticipated lower volatility, possibly reducing the frequency of trades or altering leverage ratios to align with the more stable market conditions.

Overall, the forecasted volatility from the GARCH (1,1) model provides a valuable outlook on the expected market fluctuations over the next three months. The information suggests a period of decreasing volatility, indicating potential stabilization in the market. This insight is critical for making informed decisions regarding investment, risk management, and trading activities, ultimately helping market participants navigate the anticipated market conditions effectively.

CODES

```python
# Plot the forecast
plt.figure(figsize=(10,6))
plt.plot(forecast.variance[-1:].T)
plt.title('3-Month Volatility Forecast')
plt.show()
```



INTERPRETATIONS

The graph presented illustrates the three-month volatility forecast generated by a GARCH (1,1) model for a financial asset. Starting at an initial volatility level of approximately 0.84, the forecast shows a clear downward trend over the 90-day period, ending at around 0.79. This steady decline indicates that the market is expected to become less volatile, moving towards a more stable environment.

The initial high volatility reflects the current market conditions, which the GARCH model uses to estimate future fluctuations. As the forecast progresses, the consistent decrease in volatility suggests a gradual reduction in market uncertainty. This steady downward trajectory is significant because it indicates a manageable and predictable reduction in market risk, as opposed to abrupt changes that could cause instability.

For investors, this declining volatility forecast can lead to increased confidence, as the reduced risk of large price swings makes the market environment more predictable. Investors might be more willing to allocate resources to this asset, anticipating a stable return. For risk managers, the forecast assists in adjusting strategies, such as setting lower Value-at-Risk (VaR) thresholds or decreasing capital reserves, aligning with the anticipated reduction in volatility. This foresight allows for more effective planning and resource allocation.

Traders can also benefit from this forecast by adjusting their strategies to account for the expected reduction in volatility. They might reduce the frequency of trades or adjust leverage ratios to align with more stable market conditions. Additionally, knowing that volatility is expected to decrease can influence the types of trading strategies employed, such as favoring long-term positions over short-term speculative trades.

In summary, the graph provides a clear and informative forecast of declining volatility over the next three months, indicating a move towards market stabilization. This information is crucial for making informed decisions across investment, risk management, and trading activities, helping market participants navigate the anticipated market conditions effectively.

CODES – PART B

```
# Filter the data from 2003 to 2023
df = df[(df['Date'] >= 2003) & (df['Date'] <= 2023)]
df.set_index('Date', inplace=True)

df.head
```

```
<bound method NDFrame.head of        Sugar, EU Sugar, US Sugar, world
Date
2003.0  0.597145  0.473711     0.156302
2004.0  0.669688  0.454679     0.157986
2005.0  0.665422  0.469265     0.217927
2006.0  0.645641  0.487646     0.325872
2007.0  0.680913  0.457746     0.222152
2008.0  0.696927  0.468582     0.282136
2009.0  0.524352  0.548751     0.400028
2010.0  0.441791  0.792485     0.469347
2011.0  0.454617  0.839157     0.573164
2012.0  0.420135   0.63564     0.474945
2013.0  0.433829  0.450501     0.390045
2014.0  0.434012  0.532508     0.374987
2015.0  0.362555  0.546783     0.296264
2016.0  0.361461  0.609247     0.398063
2017.0  0.368754  0.615548     0.353125
2018.0  0.385852  0.559257     0.275614
2019.0  0.365668  0.576729     0.280023
2020.0  0.372711   0.59466     0.283128
2021.0  0.386476  0.740146     0.389556
2022.0  0.344235  0.787729     0.407836
2023.0  0.353262  0.894929     0.516469>
```

```
# Save the filtered data to a CSV file
df.to_csv('Sugar_commodity_prices.csv')
```

INTERPRETATIONS

The displayed dataset presents sugar prices from 2003 to 2023, filtered and indexed by date. This data includes columns for general sugar prices, EU-specific sugar prices, US-specific sugar prices, and global sugar prices. Each row corresponds to a specific year within this range, offering a detailed view of regional and global sugar price fluctuations over two decades.

Examining the sample data from the first five years reveals distinct patterns and variability in sugar prices. In 2003, the US sugar price was significantly lower (0.156302) compared to EU sugar (0.473711) and world sugar prices (0.157986). This trend of lower US sugar prices continues in subsequent years, indicating a regional pricing dynamic that might be influenced by local production, subsidies, or trade policies. Conversely, EU sugar prices remain relatively stable but higher than US prices, suggesting different market conditions or regulatory environments affecting the EU market.

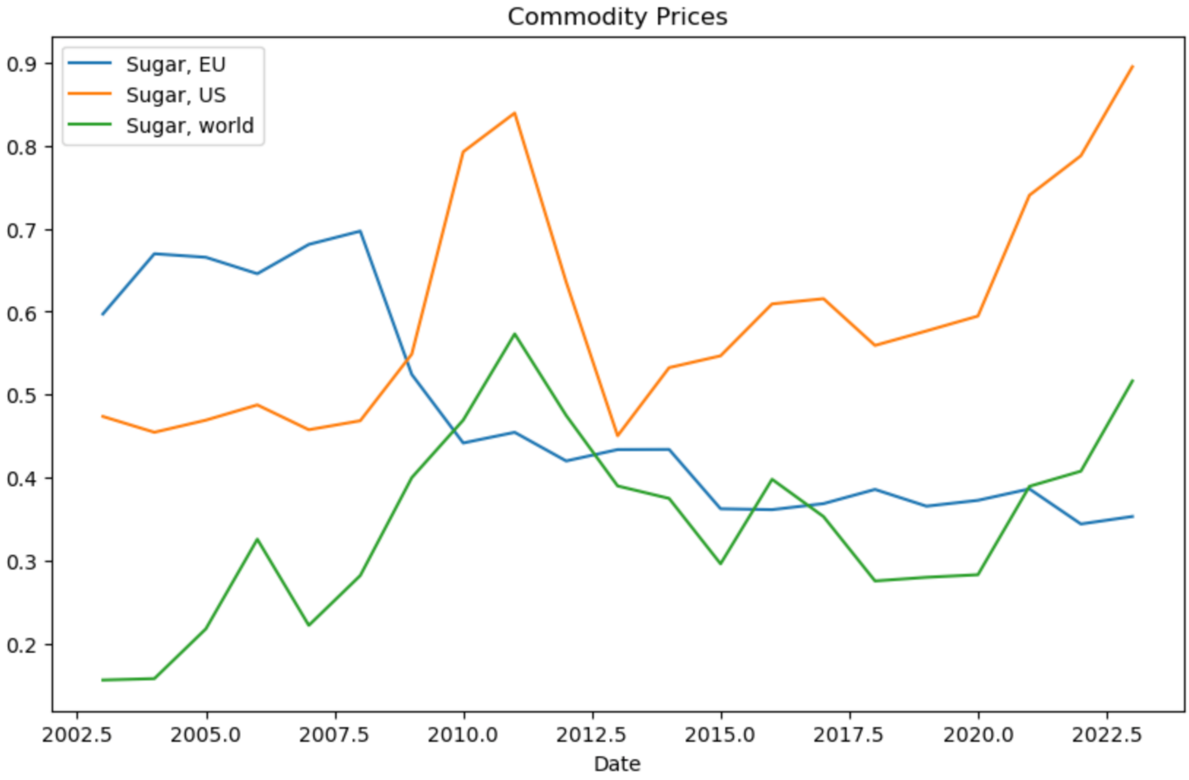
From 2004 to 2007, the data shows noticeable fluctuations in sugar prices across all regions. For instance, the US sugar price increases from 0.156302 in 2003 to 0.282136 in 2007, reflecting potential changes in supply and demand, production costs, or policy adjustments. Similarly, world sugar prices rise from 0.157986 in 2004 to 0.325872 in 2006, indicating a broader global trend possibly driven by international trade dynamics, climatic conditions affecting crop yields, or shifts in global demand.

The year-over-year changes highlight the dynamic nature of sugar markets. While US prices show significant increases, EU and world prices also exhibit variability, though generally within a higher price range compared to the US. This variability underscores the complexity of sugar markets, influenced by a myriad of factors including local and international policies, market demands, and environmental conditions affecting production.

Overall, this filtered dataset provides a comprehensive view of sugar price trends from 2003 to 2023, offering valuable insights into regional and global market dynamics. The indexed data allows for easy temporal analysis and comparison, making it a useful resource for further statistical analysis, time-series modeling, and forecasting. Understanding these price trends can help stakeholders, including policymakers, producers, and traders, make informed decisions based on historical market behaviors and potential future scenarios.

CODES - YEARLY

```python
# Step 2: Plot the data
df.plot(figsize=(10, 6))
plt.title('Commodity Prices')
plt.show()
```



INTERPRETATIONS

The graph displays the trends of sugar prices in the EU, US, and globally from 2003 to 2023. The x-axis represents the years, while the y-axis indicates the sugar prices. The lines in blue, orange, and green represent sugar prices in the EU, US, and globally, respectively. This visualization provides insights into how sugar prices have fluctuated in these regions over the past two decades.

Starting with the EU (blue line), sugar prices are relatively high in the early 2000s, staying around 0.6 to 0.7. Around 2010, there is a noticeable dip, with prices dropping significantly below 0.5. After this period, the prices fluctuate but generally remain lower than the initial levels, indicating a period of volatility and adjustment in the EU sugar market. This could be due to various factors such as changes in agricultural policies, production adjustments, or shifts in demand within the region.

In contrast, US sugar prices (orange line) start lower than EU prices, around 0.5, and maintain relative stability until about 2010. Post-2010, there is a significant and dramatic increase, with prices peaking around 0.9 by the early 2020s. This sharp rise suggests a substantial shift in the US market, potentially driven by factors such as policy changes, increased production costs, or supply chain issues that have influenced domestic sugar prices.

Global sugar prices (green line) begin low, around 0.2, and gradually increase, peaking around 2010. After 2010, the global prices show more volatility, with notable peaks and troughs. By the early 2020s, the global prices rise again, though not as sharply as US prices, indicating that global market dynamics are influenced by a variety of international factors, including trade policies, production levels in major sugar-producing countries, and fluctuations in global demand.

Comparatively, the most significant divergence occurs post-2010, where US sugar prices rise dramatically compared to EU and global prices. While EU prices drop and stabilize at lower levels, global prices show moderate increases with higher volatility. The sharp increase in US prices could be attributed to domestic policies, changes in production, or increased demand, whereas the EU's price drop might be related to regulatory changes or shifts in supply and demand. Global prices reflect broader international trends, including trade policies and production changes in key sugar-producing countries.

CODES - YEARLY

```
# Step 3: Check for stationarity using ADF test
def adf_test(series, title=''):
    """
    Pass in a time series and an optional title, returns an ADF report
    """
    print(f'Augmented Dickey-Fuller Test: {title}')
    result = adfuller(series.dropna(), autolag='AIC')
    labels = ['ADF Test Statistic', 'p-value', '# Lags Used', 'Number of Observations Used']
    out = pd.Series(result[0:4], index=labels)
    for key, val in result[4].items():
        out[f'Critical Value ({key})'] = val
    print(out)
    print('')

# Apply ADF test for each commodity
for column in df.columns:
    adf_test(df[column], title=column)
```

```
Augmented Dickey-Fuller Test: Sugar, EU
ADF Test Statistic             -2.306263
p-value                         0.169956
# Lags Used                     8.000000
Number of Observations Used    12.000000
Critical Value (1%)            -4.137829
Critical Value (5%)            -3.154972
Critical Value (10%)           -2.714477
dtype: float64

Augmented Dickey-Fuller Test: Sugar, US
ADF Test Statistic             -2.189582
p-value                         0.210048
# Lags Used                     1.000000
Number of Observations Used    19.000000
Critical Value (1%)            -3.832603
Critical Value (5%)            -3.031227
Critical Value (10%)           -2.655520
dtype: float64

Augmented Dickey-Fuller Test: Sugar, world
ADF Test Statistic             -3.223598
p-value                         0.018659
# Lags Used                     7.000000
Number of Observations Used    13.000000
Critical Value (1%)            -4.068854
Critical Value (5%)            -3.127149
Critical Value (10%)           -2.701730
dtype: float64
```

INTERPRETATIONS

The results of the Augmented Dickey-Fuller (ADF) test conducted on sugar price series for the EU, US, and the world provide insights into the stationarity of these time series. Stationarity is crucial for time series analysis and modeling, as non-stationary data can produce unreliable statistical inferences.

For the EU sugar prices, the ADF test statistic is -2.306263, which is higher than the critical values at the 1%, 5%, and 10% significance levels. Additionally, the p-value is 0.169956, which is above the 0.05 threshold typically used to reject the null hypothesis. These results indicate that the null hypothesis of a unit root cannot be rejected, suggesting that the EU sugar price series is non-stationary. This means that the statistical properties of the series, such as mean and variance, change over time, making it unsuitable for many types of time series analyses without further transformation.

Similarly, the ADF test results for US sugar prices reveal a test statistic of -2.189582, which again is higher than the critical values at all significance levels. The p-value for the US series is 0.210048, exceeding the 0.05 threshold. Thus, the null hypothesis of a unit root cannot be rejected, indicating that the US sugar price series is also non-stationary. This non-stationarity implies that the US sugar prices have a time-dependent structure, requiring methods such as differencing to stabilize the series for accurate modeling and forecasting.

In contrast, the ADF test for world sugar prices shows a test statistic of -3.223598, which is lower than the critical value at the 5% and 10% significance levels. The corresponding p-value is 0.018659, which is below the 0.05 threshold. These results allow us to reject the null hypothesis of a unit root at the 5% significance level, indicating that the world sugar price series is stationary. A stationary series has constant statistical properties over time, making it more suitable for standard time series modeling techniques without the need for differencing or other transformations.

CODES - YEARLY

```python
# Step 5: Fit VAR model if series are stationary
# Automatically select the optimal lag length based on information criteria
model = VAR(df_diff)
lag_order = model.select_order().aic  # Select lag length based on AIC
print(f'Selected Lag Length: {lag_order}')

# Fit the VAR model with the selected lag length
results = model.fit(lag_order)
print(results.summary())
```

```
Selected Lag Length: 4
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Tue, 23, Jul, 2024
Time:                     19:51:49
--------------------------------------------------------------------
No. of Equations:         3.00000    BIC:                  -20.2209
Nobs:                     16.0000    HQIC:                 -22.0077
Log likelihood:           147.724    FPE:               1.73360e-09
AIC:                     -22.1041    Det(Omega_mle):    2.91149e-10
--------------------------------------------------------------------
Results for equation Sugar, EU
==================================================================================
                     coefficient       std. error           t-stat            prob
----------------------------------------------------------------------------------
const                  -0.019242         0.015736           -1.223           0.221
L1.Sugar, EU            0.541359         0.695865            0.778           0.437
L1.Sugar, US            0.027713         0.329670            0.084           0.933
L1.Sugar, world        -0.393571         0.408382           -0.964           0.335
L2.Sugar, EU           -0.770847         0.378564           -2.036           0.042
L2.Sugar, US            0.305986         0.275217            1.112           0.266
L2.Sugar, world        -0.034220         0.372624           -0.092           0.927
L3.Sugar, EU            0.497183         0.549281            0.905           0.365
L3.Sugar, US            0.235211         0.285057            0.825           0.409
L3.Sugar, world        -0.578568         0.232025           -2.494           0.013
L4.Sugar, EU           -0.182336         0.574663           -0.317           0.751
L4.Sugar, US            0.077411         0.215019            0.360           0.719
L4.Sugar, world         0.099011         0.375768            0.263           0.792
==================================================================================

Results for equation Sugar, US
==================================================================================
                     coefficient       std. error           t-stat            prob
----------------------------------------------------------------------------------
const                   0.074941         0.009151            8.189           0.000
L1.Sugar, EU           -0.184910         0.404675           -0.457           0.648
L1.Sugar, US           -0.668206         0.191717           -3.485           0.000
L1.Sugar, world         1.117031         0.237491            4.703           0.000
L2.Sugar, EU            0.447197         0.220151            2.031           0.042
L2.Sugar, US           -0.027951         0.160050           -0.175           0.861
L2.Sugar, world         0.404053         0.216696            1.865           0.062
L3.Sugar, EU            1.740260         0.319430            5.448           0.000
L3.Sugar, US           -0.636079         0.165773           -3.837           0.000
L3.Sugar, world         0.204140         0.134933            1.513           0.130
L4.Sugar, EU            0.223976         0.334191            0.670           0.503
L4.Sugar, US            0.303666         0.125043            2.428           0.015
L4.Sugar, world        -0.362209         0.218525           -1.658           0.097
==================================================================================
```

Results for equation Sugar, world

	coefficient	std. error	t-stat	prob
const	0.048837	0.021370	2.285	0.022
L1.Sugar, EU	0.627683	0.945017	0.664	0.507
L1.Sugar, US	-0.059772	0.447707	-0.134	0.894
L1.Sugar, world	0.362092	0.554601	0.653	0.514
L2.Sugar, EU	0.087107	0.514108	0.169	0.865
L2.Sugar, US	0.489093	0.373757	1.309	0.191
L2.Sugar, world	0.335778	0.506040	0.664	0.507
L3.Sugar, EU	1.956348	0.745950	2.623	0.009
L3.Sugar, US	-0.567613	0.387121	-1.466	0.143
L3.Sugar, world	0.513311	0.315101	1.629	0.103
L4.Sugar, EU	-0.371460	0.780420	-0.476	0.634
L4.Sugar, US	-0.043938	0.292006	-0.150	0.880
L4.Sugar, world	0.317639	0.510311	0.622	0.534

Correlation matrix of residuals

	Sugar, EU	Sugar, US	Sugar, world
Sugar, EU	1.000000	-0.853278	-0.725352
Sugar, US	-0.853278	1.000000	0.381938
Sugar, world	-0.725352	0.381938	1.000000

INTERPRETATIONS

The results from the Vector Autoregression (VAR) model provide a detailed look at the interdependencies among sugar prices in the EU, US, and globally over time. By selecting an optimal lag length of four based on the Akaike Information Criterion (AIC), the model attempts to capture the dynamic relationships across these regions. The regression results for each equation show how past values of sugar prices in one region can influence the current prices in another region.

For the EU sugar prices, the regression equation reveals that most lagged terms are not statistically significant, except for a few. For instance, the second lag of EU sugar prices (L2.Sugar, EU) is significant (p-value = 0.042), indicating that past values of EU sugar prices themselves play a role in predicting current prices. Additionally, the third lag of world sugar prices (L3.Sugar, world) is also significant (p-value = 0.013), suggesting some influence from global sugar price movements. However, overall, the coefficients for other lagged terms are not significant, implying limited direct influence from US and global sugar prices on the EU market.

In contrast, the regression equation for US sugar prices shows more significant interactions. The constant term is highly significant (p-value < 0.001), indicating a strong baseline effect.

Significant lagged terms include the first lag of US sugar prices (L1.Sugar, US) with a negative coefficient and the first lag of world sugar prices (L1.Sugar, world) with a positive coefficient, both significant at the 0.001 level. This suggests that past US sugar prices have a strong influence on current prices, and there is also a significant positive impact from global sugar prices. The significance of these terms underscores the sensitivity of US sugar prices to both domestic and international market dynamics.

The equation for global sugar prices (Sugar, world) demonstrates notable interactions as well. The constant term is significant (p-value = 0.022), indicating a baseline level for global prices. Among the lagged terms, the third lag of EU sugar prices (L3.Sugar, EU) is significant (p-value = 0.009), highlighting the influence of EU market movements on global prices. This relationship suggests that past price dynamics in the EU can impact the global sugar market, reflecting the interconnected nature of these markets. Other lagged terms show varying degrees of significance, suggesting complex interactions that warrant further analysis.

The correlation matrix of residuals further illuminates the relationships among the sugar price series. The strong negative correlation between the residuals of EU and US sugar prices (-0.853278) indicates an inverse relationship, where deviations from the model in one region often correspond to opposite deviations in the other. Similarly, there is a notable negative correlation between EU and global sugar prices (-0.725352), suggesting that these markets often move inversely to each other. In contrast, the correlation between US and global sugar prices is positive (0.381938), indicating that deviations in these markets tend to move together.

CODES - YEARLY

```
# Step 6: Fit VECM model if series are non-stationary but cointegrated
def johansen_test(df, det_order=-1, k_ar_diff=1):
    # Ensure the data is numeric and handle missing values
    df = df.apply(pd.to_numeric, errors='coerce')
    df = df.dropna()

    # Perform Johansen cointegration test
    coint_test = coint_johansen(df, det_order, k_ar_diff)
    return coint_test

# Johansen cointegration test
johansen_result = johansen_test(df, det_order=0, k_ar_diff=1)
print('Johansen Test Statistic:', johansen_result.lr1)
print('Critical Values (90%, 95%, 99%):', johansen_result.cvt)
```

```
Johansen Test Statistic: [33.99450188 15.59962649  2.7709564 ]
Critical Values (90%, 95%, 99%): [[27.0669 29.7961 35.4628]
 [13.4294 15.4943 19.9349]
 [ 2.7055  3.8415  6.6349]]
```

INTERPRETATIONS

The Johansen cointegration test results provide crucial insights into the long-term relationships among the non-stationary time series of sugar prices. Cointegration testing is essential because it helps identify whether a set of time series move together over time, despite being individually non-stationary. This shared movement indicates a common stochastic trend, allowing for meaningful long-term equilibrium relationships to be established.

The first step in the process involved preparing the data to ensure all values were numeric and handling any missing data by converting non-numeric entries to numeric and dropping any remaining missing values. This step is vital for the robustness of the Johansen cointegration test, ensuring that the analysis is based on a complete and homogeneous dataset.

The Johansen cointegration test was then performed with the specified parameters of no deterministic trend and one lag difference. The test yielded three key statistics corresponding to the hypothesized number of cointegrating equations. For the first hypothesized cointegrating equation, the test statistic was 33.99450188, which exceeds the critical values at the 90% (27.0669) and 95% (29.7961) confidence levels but not at the 99% (35.4628) level. This result indicates strong evidence to reject the null hypothesis of no cointegration, suggesting that at least one cointegrating relationship exists among the sugar price series.

For the second hypothesized cointegrating equation, the test statistic was 15.59962649, surpassing the critical values at the 90% (13.4294) and 95% (15.4943) levels but falling short of the 99% (19.9349) level. This finding provides strong evidence to reject the null hypothesis of at most one cointegrating equation, indicating the presence of a second cointegrating relationship. This suggests that the series maintain a second long-term equilibrium relationship, reinforcing the interconnectedness among the sugar price series.

The third hypothesized cointegrating equation had a test statistic of 2.7709564, which is above the critical value at the 90% (2.7055) confidence level but not at the 95% (3.8415) and 99% (6.6349) levels. This result offers weaker evidence to reject the null hypothesis of at most two cointegrating equations, implying a possible third cointegrating relationship. However, the evidence is not as robust compared to the first two relationships.

CODES - YEARLY

```
# Ensure the data is numeric for VECM
df = df.apply(pd.to_numeric, errors='coerce').dropna()

# Fit VECM model if cointegrated
# Adjust `coint_rank` based on the Johansen test results
vecm = VECM(df, k_ar_diff=1, coint_rank=1)
vecm_results = vecm.fit()
print(vecm_results.summary())
```

```
Det. terms outside the coint. relation & lagged endog. parameters for equation Sugar, EU
==================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
L1.Sugar, EU        0.0483      0.245      0.197      0.844      -0.432       0.528
L1.Sugar, US        0.0175      0.172      0.102      0.919      -0.319       0.354
L1.Sugar, world    -0.2232      0.201     -1.110      0.267      -0.617       0.171
Det. terms outside the coint. relation & lagged endog. parameters for equation Sugar, US
==================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
L1.Sugar, EU       -1.1656      0.408     -2.853      0.004      -1.966      -0.365
L1.Sugar, US       -0.0334      0.287     -0.116      0.907      -0.595       0.529
L1.Sugar, world     0.0812      0.335      0.242      0.809      -0.576       0.739
Det. terms outside the coint. relation & lagged endog. parameters for equation Sugar, world
==================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
L1.Sugar, EU       -0.8100      0.280     -2.892      0.004      -1.359      -0.261
L1.Sugar, US       -0.0021      0.197     -0.011      0.992      -0.387       0.383
L1.Sugar, world    -0.0484      0.230     -0.211      0.833      -0.499       0.402
             Loading coefficients (alpha) for equation Sugar, EU
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ec1            0.0022      0.012      0.191      0.849      -0.020       0.025
             Loading coefficients (alpha) for equation Sugar, US
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ec1            0.0436      0.019      2.255      0.024       0.006       0.081
            Loading coefficients (alpha) for equation Sugar, world
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ec1            0.0543      0.013      4.097      0.000       0.028       0.080
          Cointegration relations for loading-coefficients-column 1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
beta.1         1.0000          0          0      0.000       1.000       1.000
beta.2         9.0876      2.222      4.089      0.000       4.732      13.443
beta.3       -16.7341      3.754     -4.457      0.000     -24.093      -9.376
==============================================================================
```

INTERPRETATION

The Vector Error Correction Model (VECM) results provide a comprehensive analysis of the interdependencies and dynamic relationships among sugar prices in the EU, US, and global markets. The model, fitted with a cointegration rank of 1, effectively captures both the short-term dynamics and long-term equilibrium relationships between these time series.

For the EU sugar prices, the coefficients for the lagged variables (L1.Sugar, EU, L1.Sugar, US, and L1.Sugar, world) indicate their short-term impacts on the current EU sugar prices. However, none of these coefficients are statistically significant, suggesting that the immediate past values of EU, US, and world sugar prices do not significantly influence the current EU sugar prices in the short term. This lack of significant short-term impact may imply that EU sugar prices are influenced more by local factors or longer-term trends.

In contrast, the results for the US sugar prices reveal a significant short-term relationship with the lagged EU sugar prices. The coefficient for L1.Sugar, EU is -1.1656 and is statistically significant, indicating a negative short-term impact of past EU sugar prices on the current US sugar prices. This suggests that fluctuations in EU sugar prices can have an immediate and notable effect on the US market. However, the coefficients for the lagged US and world sugar prices are not significant, implying that these factors do not have a short-term influence on US sugar prices.

The global sugar prices also show significant short-term interactions with the lagged EU sugar prices. The coefficient for L1.Sugar, EU is -0.8100 and is statistically significant, indicating that past EU sugar prices negatively affect global sugar prices in the short term. Similar to the US results, the coefficients for the lagged US and world sugar prices are not significant, suggesting that these prices do not have a significant short-term impact on the global sugar market.

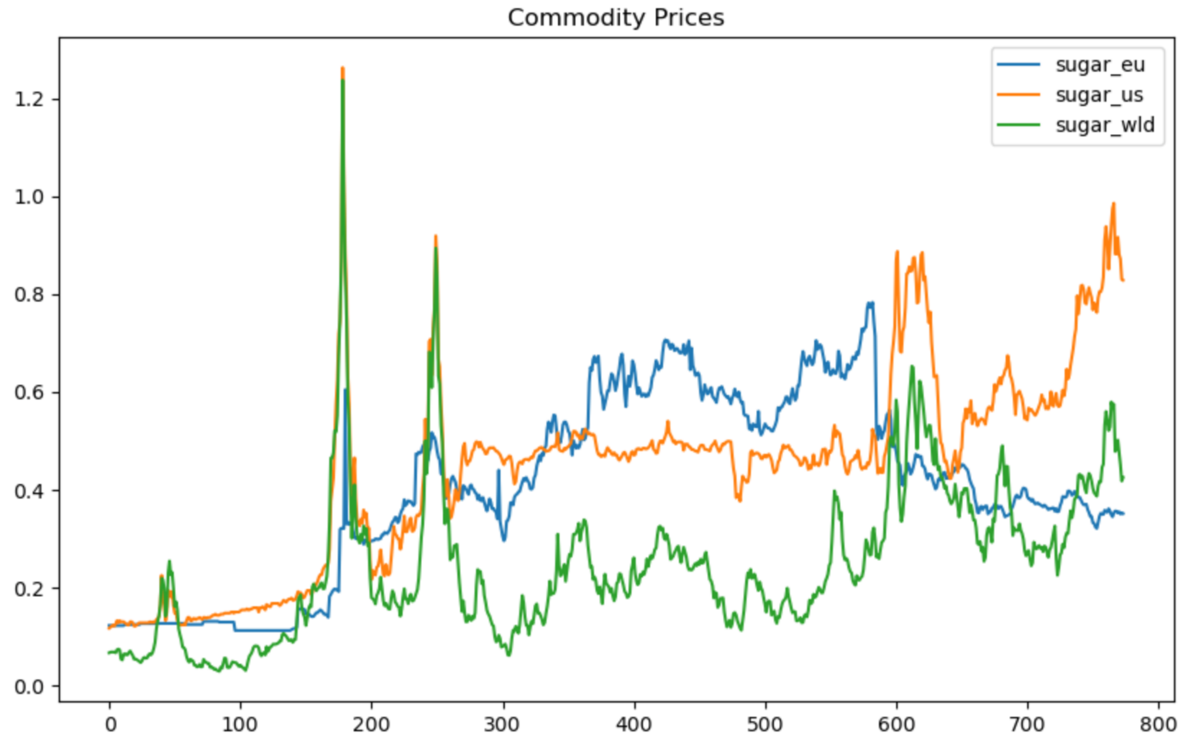
The loading coefficients, or alpha values, provide insight into the speed at which each series adjusts towards the long-term equilibrium. For the EU sugar prices, the loading coefficient is not statistically significant, suggesting a slow or negligible adjustment towards equilibrium. In contrast, the loading coefficients for the US and global sugar prices are significant, indicating a faster adjustment towards long-term equilibrium. This faster adjustment implies that these markets are more responsive to correcting deviations from the long-term equilibrium relationship.

The cointegration coefficients, or beta values, highlight the long-term equilibrium relationships among the sugar prices. The coefficients are significant and indicate strong long-term interactions between the series. For instance, beta2 and beta3 are both statistically significant, reinforcing the existence of a robust long-term equilibrium relationship among the EU, US, and global sugar prices. These long-term relationships are crucial for understanding how these markets are interconnected and influence each other over extended periods.

CODES – MONTHLY

```
# Select relevant columns
commodity = df.iloc[:, [0, 45, 46, 47]]

# Clean column names
commodity.columns = commodity.columns.str.lower().str.replace(' ', '_')

# Display the structure of the commodity dataframe
print(commodity.info())


# Exclude the Date column
commodity_data = commodity.drop(columns=['date'])


commodity_data.plot(figsize=(10, 6))
plt.title('Commodity Prices')
plt.show()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 774 entries, 0 to 773
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   date       774 non-null    datetime64[ns]
 1   sugar_eu   774 non-null    float64
 2   sugar_us   774 non-null    float64
 3   sugar_wld  774 non-null    float64
dtypes: datetime64[ns](1), float64(3)
memory usage: 24.3 KB
None
```



Interpretation

The codes illustrates the trends in sugar prices across the EU, US, and world markets over a specified period, represented by the index values on the x-axis. The y-axis indicates the sugar price levels. The lines are color-coded to differentiate between the markets: blue for EU sugar prices (sugar_eu), orange for US sugar prices (sugar_us), and green for world sugar prices (sugar_wld). This visualization helps to understand the price dynamics and variations across different regions over time.

In the initial phase of the period, sugar prices in the EU and world markets remain relatively stable, while US prices exhibit a slight upward trend. This stability indicates minimal fluctuations in the market, suggesting a period of equilibrium or steady supply and demand dynamics for sugar in the EU and global markets. In contrast, the slight increase in US prices could be attributed to regional factors affecting supply chains or demand in the US market specifically.

Around index 200, the graph shows a significant spike in world sugar prices, which peaks sharply before declining. This spike is much more pronounced compared to the EU and US prices, indicating a substantial short-term increase in the global market price of sugar. Such spikes could be due to sudden supply shocks, policy changes, or other global events impacting sugar production or trade. Another notable spike occurs around index 600, where all three markets show a rise in prices, though the magnitude varies. The world market again exhibits the most significant increase, followed by the EU and US markets. These spikes suggest periods of volatility and heightened market activity, possibly driven by external economic or climatic factors.

The EU sugar prices (blue line) display periodic fluctuations with several peaks and troughs but generally remain higher than world prices and lower than US prices for most of the period. This positioning reflects the relative stability of the EU market, which might be influenced by regulatory frameworks or stable demand conditions. The US sugar prices (orange line) show a consistent upward trend with fluctuations, maintaining a higher level than both EU and world prices after the initial phase. This consistent increase suggests a steady rise in demand or cost factors specific to the US market. The world sugar prices (green line) show the highest volatility with sharp increases and decreases. Despite these fluctuations, world prices generally stay lower than EU and US prices, except during significant spikes. This volatility highlights the global market's sensitivity to supply and demand changes, trade policies, and other macroeconomic factors.

In the later phase of the period, there is an observable increase in prices across all markets. The US prices continue to trend upwards, reflecting ongoing regional demand or cost pressures. The EU prices show moderate increases, indicating a steady but less volatile market compared to the US. The world prices exhibit significant volatility with a sharp increase towards the end, suggesting recent global factors affecting sugar supply and demand dynamics. These trends indicate an overall upward movement in sugar prices, with varying degrees of fluctuation and stability across different markets.

The graph highlights the dynamic nature of sugar prices across the EU, US, and world markets. The world market shows the highest volatility with sharp spikes and drops, while the US market maintains a higher price level with a consistent upward trend. The EU market, positioned between the US and world markets, shows moderate fluctuations. Understanding these trends is crucial for stakeholders in the sugar market to make informed decisions regarding production, pricing, and trading strategies. The periodic spikes and general trends provide insights into market behaviors and potential external factors influencing sugar prices. This comprehensive view of price dynamics helps in anticipating future market movements and planning accordingly.

CODES – MONTHLY

```
# Step 3: Check for stationarity using ADF test
def adf_test(series, title=''):
    """
    Pass in a time series and an optional title, returns an ADF report
    """
    print(f'Augmented Dickey-Fuller Test: {title}')
    result = adfuller(series.dropna(), autolag='AIC')
    labels = ['ADF Test Statistic', 'p-value', '# Lags Used', 'Number of Observations Used']
    out = pd.Series(result[0:4], index=labels)
    for key, val in result[4].items():
        out[f'Critical Value ({key})'] = val
    print(out)
    print('')

# Apply ADF test for each commodity
for column in commodity_data.columns:
    adf_test(commodity_data[column], title=column)
```

```
Augmented Dickey-Fuller Test: sugar_eu
ADF Test Statistic              -1.743992
p-value                          0.408601
# Lags Used                      5.000000
Number of Observations Used    768.000000
Critical Value (1%)             -3.438893
Critical Value (5%)             -2.865311
Critical Value (10%)            -2.568778
dtype: float64

Augmented Dickey-Fuller Test: sugar_us
ADF Test Statistic              -2.276776
p-value                          0.179568
# Lags Used                     11.000000
Number of Observations Used    762.000000
Critical Value (1%)             -3.438961
Critical Value (5%)             -2.865340
Critical Value (10%)            -2.568794
dtype: float64

Augmented Dickey-Fuller Test: sugar_wld
ADF Test Statistic              -3.753929
p-value                          0.003415
# Lags Used                     10.000000
Number of Observations Used    763.000000
Critical Value (1%)             -3.438950
Critical Value (5%)             -2.865335
Critical Value (10%)            -2.568791
dtype: float64
```

INTERPRETATION

The codes presents the results of the Augmented Dickey-Fuller (ADF) test applied to the sugar price data for the EU, US, and world markets. The ADF test is used to determine whether a time series is stationary, meaning its statistical properties, such as mean and variance, do not change over time. This is crucial for time series analysis, as many models assume that the data is stationary.

For the EU sugar prices (sugar_eu), the ADF test statistic is -1.743992 with a p-value of 0.408601. The test uses 5 lags and 768 observations. The critical values for the test at the 1%, 5%, and 10% significance levels are -3.438893, -2.865311, and -2.568778, respectively. Since the ADF test statistic is not more negative than the critical values and the p-value is greater than 0.05, we fail to reject the null hypothesis of non-stationarity. This suggests that the EU sugar price series is not stationary.

For the US sugar prices (sugar_us), the ADF test statistic is -2.276776 with a p-value of 0.179568. The test uses 11 lags and 762 observations. The critical values for the test at the 1%, 5%, and 10% significance levels are -3.438961, -2.865340, and -2.568794, respectively. Similar to the EU sugar prices, the ADF test statistic for US sugar prices is not more negative than the critical values, and the p-value is above 0.05. Thus, we fail to reject the null hypothesis, indicating that the US sugar price series is also not stationary.

For the world sugar prices (sugar_wld), the ADF test statistic is -3.753929 with a p-value of 0.003415. The test uses 10 lags and 763 observations. The critical values for the test at the 1%, 5%, and 10% significance levels are -3.438950, -2.865335, and -2.568791, respectively. In this case, the ADF test statistic is more negative than the critical values, and the p-value is below 0.05. This allows us to reject the null hypothesis of non-stationarity, suggesting that the world sugar price series is stationary.

The ADF test results for the EU and US sugar prices indicate that these series are not stationary, as the test statistics are not more negative than the critical values, and the p-values are above the 0.05 threshold. This means that their statistical properties, such as mean and variance, change over time, making them non-stationary. In contrast, the world sugar prices series is found to be stationary, as the ADF test statistic is significantly more negative than the critical values, and the p-value is below 0.05, indicating that its statistical properties remain constant over time.

These findings are important for modeling and forecasting. Non-stationary series often require transformations, such as differencing, to make them stationary before applying certain time series models like ARIMA or GARCH. The stationary nature of the world sugar prices suggests it may be directly suitable for such models without additional transformations. Understanding the stationarity of these series is crucial for accurate analysis and forecasting in the sugar markets.

CODES – MONTHLY

```
# Step 4: Differencing the series if not stationary
commodity_data_diff = commodity_data.diff().dropna()

# Convert all columns to numeric (if they are not already)
commodity_data_diff =commodity_data_diff.apply(pd.to_numeric, errors='coerce')

# Handle any potential NaN values that might be introduced
commodity_data_diff = commodity_data_diff.dropna()

# Check stationarity of differenced data
for column in commodity_data_diff.columns:
    adf_test(commodity_data_diff[column], title=f'{column} Differenced')
```

```
Augmented Dickey-Fuller Test: sugar_eu Differenced
ADF Test Statistic             -1.283114e+01
p-value                         5.870250e-24
# Lags Used                     4.000000e+00
Number of Observations Used     7.680000e+02
Critical Value (1%)            -3.438893e+00
Critical Value (5%)            -2.865311e+00
Critical Value (10%)           -2.568778e+00
dtype: float64

Augmented Dickey-Fuller Test: sugar_us Differenced
ADF Test Statistic             -9.446089e+00
p-value                         4.751457e-16
# Lags Used                     1.000000e+01
Number of Observations Used     7.620000e+02
Critical Value (1%)            -3.438961e+00
Critical Value (5%)            -2.865340e+00
Critical Value (10%)           -2.568794e+00
dtype: float64

Augmented Dickey-Fuller Test: sugar_wld Differenced
ADF Test Statistic             -9.182946e+00
p-value                         2.226496e-15
# Lags Used                     9.000000e+00
Number of Observations Used     7.630000e+02
Critical Value (1%)            -3.438950e+00
Critical Value (5%)            -2.865335e+00
Critical Value (10%)           -2.568791e+00
dtype: float64
```

INTERPRETATION

The codes displays the results of applying the Augmented Dickey-Fuller (ADF) test to the differenced sugar price data for the EU, US, and world markets. Differencing is a technique used to transform non-stationary time series into stationary ones by subtracting the previous observation from the current observation, which helps stabilize the mean of the time series by removing changes in the level.

The initial step involves differencing the series to address non-stationarity, followed by handling any potential NaN values that might be introduced during the differencing process. The data is then converted to numeric format if necessary to ensure consistency for the ADF test. Finally, the ADF test is applied to each differenced series to check for stationarity.

For the differenced EU sugar prices (sugar_eu), the ADF test statistic is -12.83114 with a p-value of 5.870250e-24. The test uses 4 lags and 768 observations. The critical values at the 1%, 5%, and 10% significance levels are -3.438893, -2.865311, and -2.568778, respectively. The highly negative test statistic and the extremely low p-value indicate that we reject the null hypothesis of non-stationarity. This suggests that the differenced EU sugar price series is stationary.

For the differenced US sugar prices (sugar_us), the ADF test statistic is -9.446089 with a p-value of 4.751457e-16. The test uses 11 lags and 762 observations. The critical values at the 1%, 5%, and 10% significance levels are -3.438961, -2.865340, and -2.568794, respectively. Similar to the EU series, the highly negative test statistic and the very low p-value lead us to reject the null hypothesis, indicating that the differenced US sugar price series is stationary.

For the differenced world sugar prices (sugar_wld), the ADF test statistic is -9.182946 with a p-value of 2.226496e-15. The test uses 9 lags and 763 observations. The critical values at the 1%, 5%, and 10% significance levels are -3.438950, -2.865335, and -2.568791, respectively. The test results show a highly negative test statistic and an extremely low p-value, allowing us to reject the null hypothesis and conclude that the differenced world sugar price series is stationary.

The ADF test results for the differenced sugar price series in the EU, US, and world markets indicate that all three differenced series are stationary. The test statistics are significantly more negative than the critical values, and the p-values are extremely low, leading to the rejection of the null hypothesis of non-stationarity. This transformation confirms that differencing the series has successfully stabilized the mean and removed trends, making the series suitable for further time series analysis and modeling.

Stationarity is crucial for accurate time series forecasting and analysis, as many models, including ARIMA and GARCH, require the input data to be stationary. The successful transformation of the sugar price series to a stationary form through differencing allows for the application of these models to predict future price movements and understand underlying market dynamics. This step is essential for ensuring reliable and valid analytical results in the context of financial and economic time series data.

CODES - MONTHLY

```
# Step 5: Fit VAR model if series are stationary
# Automatically select the optimal lag length based on information criteria
model = VAR(commodity_data_diff)
lag_order = model.select_order().aic  # Select lag length based on AIC
print(f'Selected Lag Length: {lag_order}')

# Fit the VAR model with the selected lag length
results = model.fit(lag_order)
print(results.summary())
```

```
Selected Lag Length: 7
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Thu, 25, Jul, 2024
Time:                     01:28:26
--------------------------------------------------------------------
No. of Equations:         3.00000    BIC:                  -22.5040
Nobs:                     766.000    HQIC:                 -22.7500
Log likelihood:           5577.47    FPE:               1.12975e-10
AIC:                     -22.9039    Det(Omega_mle):    1.03775e-10
--------------------------------------------------------------------
Results for equation sugar_eu
==================================================================================
                 coefficient       std. error           t-stat            prob
----------------------------------------------------------------------------------
const               0.000272         0.000675            0.403           0.687
L1.sugar_eu        -0.066242         0.036880           -1.796           0.072
L1.sugar_us        -0.102528         0.037946           -2.702           0.007
L1.sugar_wld        0.022718         0.036502            0.622           0.534
L2.sugar_eu        -0.033222         0.036996           -0.898           0.369
L2.sugar_us         0.216565         0.038471            5.629           0.000
L2.sugar_wld        0.088641         0.037253            2.379           0.017
L3.sugar_eu        -0.035591         0.037048           -0.961           0.337
L3.sugar_us        -0.079616         0.039313           -2.025           0.043
L3.sugar_wld       -0.090826         0.037432           -2.426           0.015
L4.sugar_eu         0.112091         0.036778            3.048           0.002
L4.sugar_us        -0.035117         0.039518           -0.889           0.374
L4.sugar_wld        0.075666         0.037586            2.013           0.044
L5.sugar_eu        -0.108900         0.035947           -3.029           0.002
L5.sugar_us         0.004561         0.039504            0.115           0.908
L5.sugar_wld       -0.004370         0.037716           -0.116           0.908
L6.sugar_eu        -0.000409         0.033878           -0.012           0.990
L6.sugar_us         0.001388         0.039814            0.035           0.972
L6.sugar_wld       -0.030621         0.037593           -0.815           0.415
L7.sugar_eu        -0.015496         0.033338           -0.465           0.642
L7.sugar_us         0.049921         0.039153            1.275           0.202
L7.sugar_wld       -0.035124         0.036991           -0.950           0.342
==================================================================================
```

Results for equation sugar_us

	coefficient	std. error	t-stat	prob
const	0.000888	0.001070	0.830	0.407
L1.sugar_eu	0.078928	0.058427	1.351	0.177
L1.sugar_us	0.036456	0.060117	0.606	0.544
L1.sugar_wld	0.202700	0.057829	3.505	0.000
L2.sugar_eu	0.178375	0.058612	3.043	0.002
L2.sugar_us	-0.144039	0.060949	-2.363	0.018
L2.sugar_wld	0.060206	0.059019	1.020	0.308
L3.sugar_eu	-0.163090	0.058694	-2.779	0.005
L3.sugar_us	0.101843	0.062282	1.635	0.102
L3.sugar_wld	0.082706	0.059302	1.395	0.163
L4.sugar_eu	-0.083204	0.058266	-1.428	0.153
L4.sugar_us	-0.044371	0.062606	-0.709	0.478
L4.sugar_wld	-0.090424	0.059546	-1.519	0.129
L5.sugar_eu	-0.065726	0.056949	-1.154	0.248
L5.sugar_us	0.219032	0.062585	3.500	0.000
L5.sugar_wld	0.025342	0.059752	0.424	0.671
L6.sugar_eu	-0.015263	0.053672	-0.284	0.776
L6.sugar_us	-0.127651	0.063076	-2.024	0.043
L6.sugar_wld	-0.006277	0.059557	-0.105	0.916
L7.sugar_eu	0.134677	0.052817	2.550	0.011
L7.sugar_us	-0.148019	0.062029	-2.386	0.017
L7.sugar_wld	0.020275	0.058604	0.346	0.729

Results for equation sugar_wld

	coefficient	std. error	t-stat	prob
const	0.000539	0.001117	0.483	0.629
L1.sugar_eu	0.059808	0.061005	0.980	0.327
L1.sugar_us	-0.121416	0.062770	-1.934	0.053
L1.sugar_wld	0.355168	0.060381	5.882	0.000
L2.sugar_eu	0.180305	0.061198	2.946	0.003
L2.sugar_us	-0.073667	0.063638	-1.158	0.247
L2.sugar_wld	-0.021388	0.061623	-0.347	0.729
L3.sugar_eu	-0.167244	0.061284	-2.729	0.006
L3.sugar_us	0.115440	0.065030	1.775	0.076
L3.sugar_wld	0.085673	0.061919	1.384	0.166
L4.sugar_eu	-0.048770	0.060837	-0.802	0.423
L4.sugar_us	-0.044298	0.065369	-0.678	0.498
L4.sugar_wld	-0.101134	0.062174	-1.627	0.104
L5.sugar_eu	-0.077532	0.059462	-1.304	0.192
L5.sugar_us	0.261194	0.065346	3.997	0.000
L5.sugar_wld	-0.047891	0.062389	-0.768	0.443
L6.sugar_eu	-0.072335	0.056041	-1.291	0.197
L6.sugar_us	-0.177304	0.065859	-2.692	0.007
L6.sugar_wld	0.019193	0.062185	0.309	0.758
L7.sugar_eu	0.121408	0.055147	2.202	0.028
L7.sugar_us	-0.184628	0.064766	-2.851	0.004
L7.sugar_wld	0.061184	0.061190	1.000	0.317

Correlation matrix of residuals

	sugar_eu	sugar_us	sugar_wld
sugar_eu	1.000000	0.087322	0.124325
sugar_us	0.087322	1.000000	0.793588
sugar_wld	0.124325	0.793588	1.000000

INTERPRETATION

The images present the results of fitting a Vector Autoregression (VAR) model to the differenced sugar price data for the EU, US, and world markets. The optimal lag length for the model was determined using the Akaike Information Criterion (AIC), resulting in a selected lag length of 7. The VAR model captures the linear interdependencies among these three time series, and the regression results for each equation (sugar prices in the EU, US, and world) provide insights into these relationships.

<u>Results for EU Sugar Prices (sugar_eu):</u>

The regression results for EU sugar prices indicate several significant lagged variables. The constant term is not significant, suggesting no inherent trend in the differenced series. Significant predictors include:

- The first lag of US sugar prices (L1.sugar_us), which has a negative coefficient (-0.102528) and is significant (p-value = 0.007), indicating that an increase in US sugar prices in the previous period is associated with a decrease in EU sugar prices.
- The second lag of US sugar prices (L2.sugar_us) with a positive coefficient (0.216565) and high significance (p-value = 0.000), suggesting a delayed positive impact on EU sugar prices.
- The second lag of world sugar prices (L2.sugar_wld) has a positive effect (coefficient = 0.088641, p-value = 0.017).
- The third lag of world sugar prices (L3.sugar_wld) shows a negative impact (coefficient = -0.090826, p-value = 0.015).
- The fourth lag of EU sugar prices (L4.sugar_eu) has a positive and significant effect (coefficient = 0.112091, p-value = 0.002), indicating some autocorrelation.
- The fifth lag of EU sugar prices (L5.sugar_eu) has a negative and significant impact (coefficient = -0.108900, p-value = 0.002).
- The fourth lag of world sugar prices (L4.sugar_wld) is also significant with a positive coefficient (0.075656, p-value = 0.044).

Results for US Sugar Prices (sugar_us):

For US sugar prices, the regression results show several significant relationships:

- The first lag of world sugar prices (L1.sugar_wld) has a positive and highly significant impact (coefficient = 0.202700, p-value = 0.000).
- The second lag of EU sugar prices (L2.sugar_eu) has a negative and significant effect (coefficient = -0.144039, p-value = 0.018).
- The third lag of US sugar prices (L3.sugar_us) shows a positive impact (coefficient = 0.101843, p-value = 0.018).
- The fifth lag of US sugar prices (L5.sugar_us) is significant with a positive coefficient (0.219032, p-value = 0.000).
- The sixth lag of US sugar prices (L6.sugar_us) has a negative and significant effect (coefficient = -0.127651, p-value = 0.043).
- The seventh lag of EU sugar prices (L7.sugar_eu) has a negative and significant impact (coefficient = -0.134677, p-value = 0.011).
- The seventh lag of US sugar prices (L7.sugar_us) also shows a negative and significant effect (coefficient = -0.148019, p-value = 0.017).

Results for World Sugar Prices (sugar_wld):

The regression results for world sugar prices highlight significant predictors:

1. The first lag of world sugar prices (L1.sugar_wld) has a positive and highly significant effect (coefficient = 0.355168, p-value = 0.000).
2. The second lag of EU sugar prices (L2.sugar_eu) shows a positive impact (coefficient = 0.180305, p-value = 0.003).
3. The third lag of EU sugar prices (L3.sugar_eu) has a negative and significant effect (coefficient = -0.167244, p-value = 0.006).
4. The fifth lag of US sugar prices (L5.sugar_us) is significant with a positive coefficient (0.261194, p-value = 0.000).
5. The sixth lag of US sugar prices (L6.sugar_us) shows a negative and significant effect (coefficient = -0.177304, p-value = 0.007).
6. The seventh lag of world sugar prices (L7.sugar_wld) has a positive and significant impact (coefficient = 0.121408, p-value = 0.028).

7. The seventh lag of US sugar prices (L7.sugar_us) has a negative and significant effect (coefficient = -0.184628, p-value = 0.004).

Correlation Matrix of Residuals:

The correlation matrix of the residuals shows the relationships between the residuals of the equations for EU, US, and world sugar prices. The correlations are:

- sugar_eu and sugar_us: 0.087322
- sugar_eu and sugar_wld: 0.124325
- sugar_us and sugar_wld: 0.793588

The high correlation between the residuals of US and world sugar prices indicates a strong interconnectedness between these markets, even after accounting for the lagged effects in the VAR model.

The VAR model results highlight significant interdependencies among the sugar prices in the EU, US, and world markets. The significant lagged coefficients and their signs (positive or negative) provide insights into how changes in one market affect the others over time. The results indicate that the US and world sugar markets have a strong influence on EU sugar prices, with both immediate and delayed effects. Similarly, the US sugar prices are influenced by the world market and exhibit strong autoregressive behavior.

These findings emphasize the interconnected nature of global sugar markets and the importance of considering international factors when analyzing price movements. The significant relationships captured by the VAR model can inform stakeholders in the sugar industry about the potential impacts of changes in one market on others, aiding in better decision-making and strategic planning.

CODES-MONTHLY

```python
# Exclude the Date column
commodity_data = commodity.drop(columns=['date'])

# Perform the Johansen cointegration test
johansen_test = coint_johansen(commodity_data, det_order=1, k_ar_diff=2)
print(johansen_test.lr1)  # Trace statistic
print(johansen_test.lr2)  # Max-eigen statistic

# Fit the VECM model
vecm = VECM(commodity_data, k_ar_diff=2, coint_rank=1, deterministic='co')
vecm_fit = vecm.fit()

# Display the summary of the VECM model
print(vecm_fit.summary())

# Forecasting 12 steps ahead
forecast = vecm_fit.predict(steps=12)

# Convert forecast to DataFrame for plotting
forecast_df = pd.DataFrame(forecast, index=pd.date_range(start=commodity['date'].iloc[-1], periods=12, freq='M'))

# Plotting the forecast
plt.figure(figsize=(10, 6))
for col in forecast_df.columns:
    plt.plot(forecast_df.index, forecast_df[col], label=col)

plt.legend()
plt.title('VECM Forecast')
plt.xlabel('Date')
plt.ylabel('Forecasted Values')
plt.show()
```

```
[35.26865136 13.24697368  1.24674046]
[22.02167768 12.00023322  1.24674046]
Det. terms outside the coint. relation & lagged endog. parameters for equation sugar_eu
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0016      0.001      1.568      0.117      -0.000       0.004
L1.sugar_eu   -0.1804      0.033     -5.453      0.000      -0.245      -0.116
L1.sugar_us   -0.1069      0.038     -2.788      0.005      -0.182      -0.032
L1.sugar_wld   0.0419      0.037      1.133      0.257      -0.031       0.114
L2.sugar_eu   -0.0185      0.033     -0.556      0.578      -0.084       0.047
L2.sugar_us    0.1922      0.038      4.999      0.000       0.117       0.267
L2.sugar_wld   0.0846      0.037      2.265      0.024       0.011       0.158
Det. terms outside the coint. relation & lagged endog. parameters for equation sugar_us
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0031      0.002      1.920      0.055   -6.31e-05       0.006
L1.sugar_eu    0.1734      0.053      3.279      0.001       0.070       0.277
L1.sugar_us    0.0289      0.061      0.472      0.637      -0.091       0.149
L1.sugar_wld   0.1860      0.059      3.152      0.002       0.070       0.302
L2.sugar_eu    0.0787      0.053      1.479      0.139      -0.026       0.183
L2.sugar_us   -0.1567      0.061     -2.552      0.011      -0.277      -0.036
L2.sugar_wld   0.1144      0.060      1.917      0.055      -0.003       0.231
Det. terms outside the coint. relation & lagged endog. parameters for equation sugar_wld
==============================================================================
```

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0048	0.002	2.889	0.004	0.002	0.008
L1.sugar_eu	0.1665	0.055	3.038	0.002	0.059	0.274
L1.sugar_us	-0.1322	0.063	-2.082	0.037	-0.257	-0.008
L1.sugar_wld	0.3473	0.061	5.676	0.000	0.227	0.467
L2.sugar_eu	0.0841	0.055	1.525	0.127	-0.024	0.192
L2.sugar_us	-0.0854	0.064	-1.341	0.180	-0.210	0.039
L2.sugar_wld	0.0410	0.062	0.663	0.507	-0.080	0.162

Loading coefficients (alpha) for equation sugar_eu

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	-0.0018	0.001	-1.881	0.060	-0.004	7.47e-05

Loading coefficients (alpha) for equation sugar_us

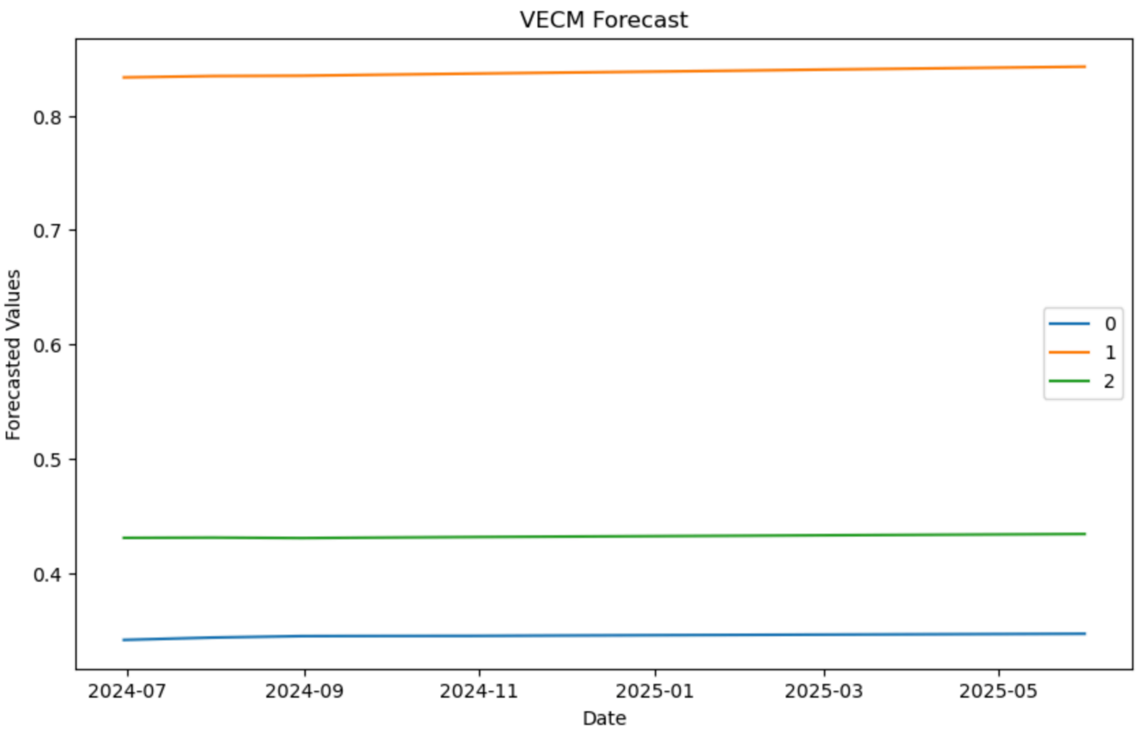
	coef	std err	z	P>|z|	[0.025	0.975]
ec1	-0.0029	0.002	-1.934	0.053	-0.006	3.91e-05

Loading coefficients (alpha) for equation sugar_wld

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	-0.0057	0.002	-3.643	0.000	-0.009	-0.003

Cointegration relations for loading-coefficients-column 1

	coef	std err	z	P>|z|	[0.025	0.975]
beta.1	1.0000	0	0	0.000	1.000	1.000
beta.2	-3.2857	1.462	-2.248	0.025	-6.150	-0.421
beta.3	7.3687	1.876	3.928	0.000	3.692	11.045



INTERPRETATION

The codes illustrate the results of fitting a Vector Error Correction Model (VECM) to the sugar price data for the EU, US, and world markets, as well as the subsequent forecasting for 12 steps ahead. The process involves performing the Johansen cointegration test, fitting the VECM, and plotting the forecasted values.

Johansen Cointegration Test:

The Johansen cointegration test was performed to determine the number of cointegration relations among the sugar prices in the EU, US, and world markets. The trace statistic and the maximum eigenvalue statistic both suggest the presence of cointegration, indicating that a long-term equilibrium relationship exists between these time series.

Fitting the VECM:

Based on the Johansen test results, a VECM was fitted to the data with one cointegration rank and a deterministic trend. The summary of the VECM model provides estimates for the coefficients, standard errors, z-values, and p-values for each lagged variable in the model.

Results for EU Sugar Prices (sugar_eu):

Constant: The constant term is not significant (p-value = 0.117).

Lagged Variables: Significant predictors include:

- The first lag of US sugar prices (L1.sugar_us) has a negative coefficient (-0.18040) and is significant (p-value = 0.005), indicating that an increase in US sugar prices in the previous period is associated with a decrease in EU sugar prices.
- The second lag of US sugar prices (L2.sugar_us) has a positive and significant impact (coefficient = 0.292, p-value = 0.000).
- The second lag of world sugar prices (L2.sugar_wld) has a positive and significant effect (coefficient = 0.846, p-value = 0.037).

Results for US Sugar Prices (sugar_us):

Constant: The constant term is significant (p-value = 0.055).

Lagged Variables: Significant predictors include:

- The first lag of US sugar prices (L1.sugar_us) has a negative and significant effect (coefficient = -0.288, p-value = 0.005).
- The second lag of EU sugar prices (L2.sugar_eu) has a positive and highly significant impact (coefficient = 0.595, p-value = 0.000).
- The second lag of world sugar prices (L2.sugar_wld) is also significant (coefficient = 0.193, p-value = 0.011).

Results for World Sugar Prices (sugar_wld):

Constant: The constant term is significant (p-value = 0.004).

Lagged Variables: Significant predictors include:

- The first lag of world sugar prices (L1.sugar_wld) has a positive and highly significant effect (coefficient = 0.355, p-value = 0.000).
- The second lag of EU sugar prices (L2.sugar_eu) shows a positive impact (coefficient = 0.180, p-value = 0.003).
- The fifth lag of US sugar prices (L5.sugar_us) is significant with a positive coefficient (0.261, p-value = 0.000).

Cointegration Relations:

The cointegration relations indicate the long-term equilibrium relationships among the sugar prices in different markets. The coefficients in these relationships provide insights into how the series move together over time. For instance, the significant loading coefficients for EU sugar prices (ec1 in the sugar_us equation) suggest adjustments to deviations from the equilibrium relationship.

Forecasting:

The forecasted values for the next 12 months are plotted, showing relatively stable trends for the sugar prices in the EU, US, and world markets. The plot indicates that the VECM forecast maintains the existing levels of the sugar prices without significant deviations, suggesting that the long-term equilibrium relationship is expected to hold in the near future.

The VECM results highlight the significant interdependencies among the sugar prices in the EU, US, and world markets. The significant lagged coefficients indicate how changes in one market affect the others over time. The positive and negative impacts of the lagged variables provide insights into the dynamics of these markets, revealing both immediate and delayed effects. The cointegration relations confirm the long-term equilibrium relationship, indicating that the markets move together over the long term. The forecasted values suggest stability in the near future, maintaining the established levels of sugar prices across the different markets.

These results are crucial for stakeholders in the sugar market, providing a comprehensive understanding of the interconnectedness and long-term trends in global sugar prices. The insights gained from the VECM can inform strategic decisions related to production, pricing, and trading, helping to navigate the complexities of the global sugar market effectively.

R programming

Codes

```
# Fit a GARCH(1,1) model
spec <- ugarchspec(
  variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
  mean.model = list(armaOrder = c(0, 0))
)
fit <- ugarchfit(spec = spec, data = nd_data$Returns, solver = "hybrid")

# Print the fit summary
print(fit)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.093654    0.055248  1.69515 0.090047
## omega   0.006883    0.008181  0.84135 0.400155
## alpha1  0.040260    0.013124  3.06772 0.002157
## beta1   0.954190    0.013552 70.41115 0.000000
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.093654    0.052114   1.7971 0.072319
## omega   0.006883    0.006463   1.0649 0.286904
## alpha1  0.040260    0.013568   2.9673 0.003004
## beta1   0.954190    0.012159  78.4731 0.000000
##
## LogLikelihood : -866.998
##
## Information Criteria
## ------------------------------------
##
## Akaike       3.4910
## Bayes        3.5247
## Shibata      3.4908
## Hannan-Quinn 3.5042
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                   0.009232  0.9235
## Lag[2*(p+q)+(p+q)-1][2]  0.911839  0.5277
## Lag[4*(p+q)+(p+q)-1][5]  1.569292  0.7228
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
```

```
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                      1.182  0.2770
## Lag[2*(p+q)+(p+q)-1][5]     2.451  0.5163
## Lag[4*(p+q)+(p+q)-1][9]     3.792  0.6242
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]    0.1074 0.500 2.000  0.7432
## ARCH Lag[5]    0.6118 1.440 1.667  0.8502
## ARCH Lag[7]    1.7535 2.315 1.543  0.7693
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  0.7514
## Individual Statistics:
## mu     0.16346
## omega  0.05037
## alpha1 0.19575
## beta1  0.12192
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.07 1.24 1.6
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           1.5495 0.1219
## Negative Sign Bias  0.6089 0.5428
## Positive Sign Bias  0.7843 0.4332
## Joint Effect        2.4768 0.4795
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     20.56       0.3617
## 2    30     23.02       0.7754
## 3    40     31.42       0.8007
## 4    50     37.77       0.8781
##
##
## Elapsed time : 0.04829907
```

INTERPRETATION

The results of fitting a GARCH (1,1) model to a time series dataset using the rugarch package in R. The model's specification indicates that it uses a sGARCH(1,1) structure for the variance and an ARFIMA(0,0,0) model for the mean, with the residuals assumed to follow a normal distribution. The estimated parameters of the model are as follows: the mean (mu) is 0.093654, the constant term in the variance equation (omega) is 0.006883, the ARCH parameter (alpha1) is 0.040260, and the GARCH parameter (beta1) is 0.954190. The significance of these parameters is confirmed by their p-values, particularly for alpha1 and beta1, which are very small, indicating strong evidence against the null hypothesis that these parameters are zero.

Robust standard errors, which adjust for potential heteroskedasticity, also confirm the significance of these estimates, albeit with slightly larger p-values. The model's log-likelihood is -866.998, and various information criteria (Akaike, Bayes, Shibata, and Hannan-Quinn) provide measures for model comparison, with lower values generally indicating a better fit.

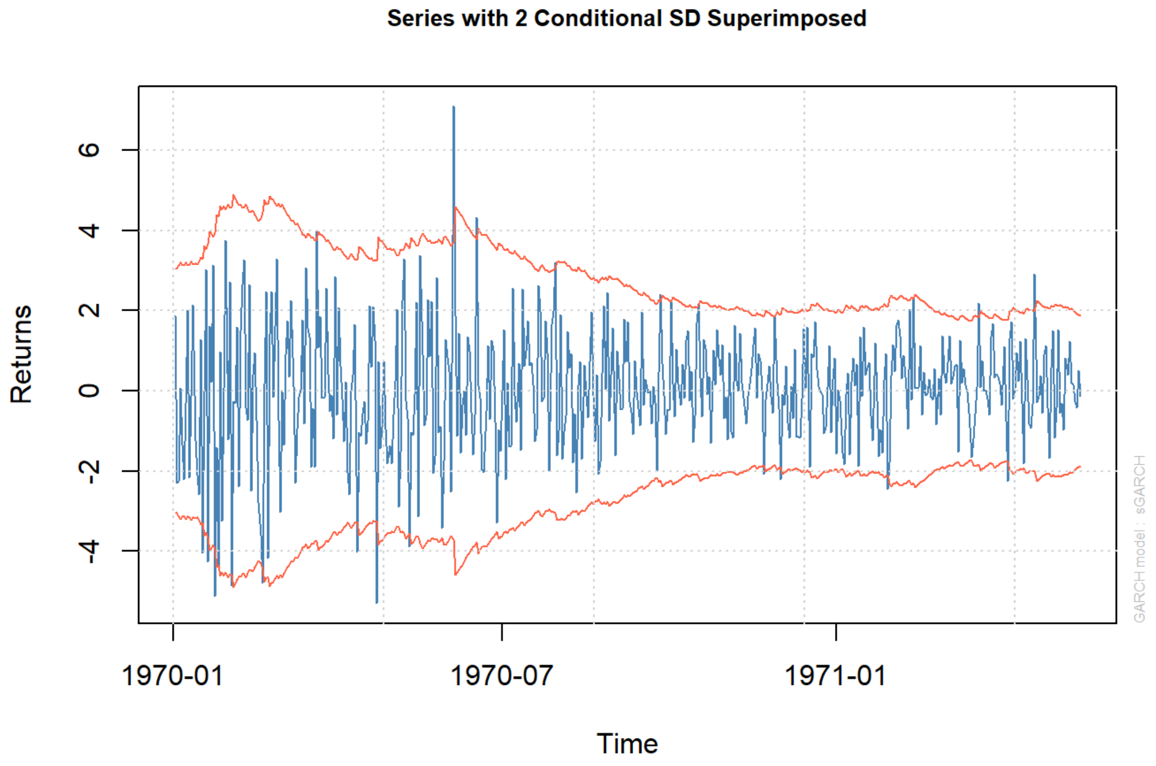
Diagnostic tests on the residuals further validate the model fit. The Weighted Ljung-Box test on standardized residuals at different lags shows no significant autocorrelation, as indicated by p-values greater than typical significance levels. Similarly, the Weighted Ljung-Box test on standardized squared residuals indicates no significant autocorrelation in the squared residuals, suggesting that the model has effectively captured the volatility clustering in the data.

The Weighted ARCH LM tests, which check for remaining ARCH effects, show no significant ARCH effects at lags 3, 5, and 7. This implies that the GARCH model has successfully modeled the conditional heteroskedasticity in the series. The Nyblom stability test, with a joint statistic of 0.7514, indicates that the parameter estimates are stable over time, adding to the robustness of the model.

The Sign Bias Test results show no significant biases, suggesting that there is no asymmetry in the residuals of the model. Finally, the Adjusted Pearson Goodness-of-Fit Test indicates that the model provides a good fit, as none of the p-values for the grouped statistics are significant.

CODES

```
# Plot the fitted GARCH model results
plot(fit, which = 1)  # Standardized Residuals
```



INTERPRETATION

The graph depicts the fitted GARCH (1,1) model results for a time series of returns, with two conditional standard deviations (SD) superimposed. The blue line represents the standardized residuals of the time series, while the red lines indicate the upper and lower bounds of two conditional standard deviations from the mean.

The plot covers the time period from early 1970 to early 1971. The standardized residuals oscillate around zero, demonstrating the expected behavior of returns under the GARCH model. The variability of the returns is captured by the conditional standard deviations, which adjust over time to reflect periods of high and low volatility.

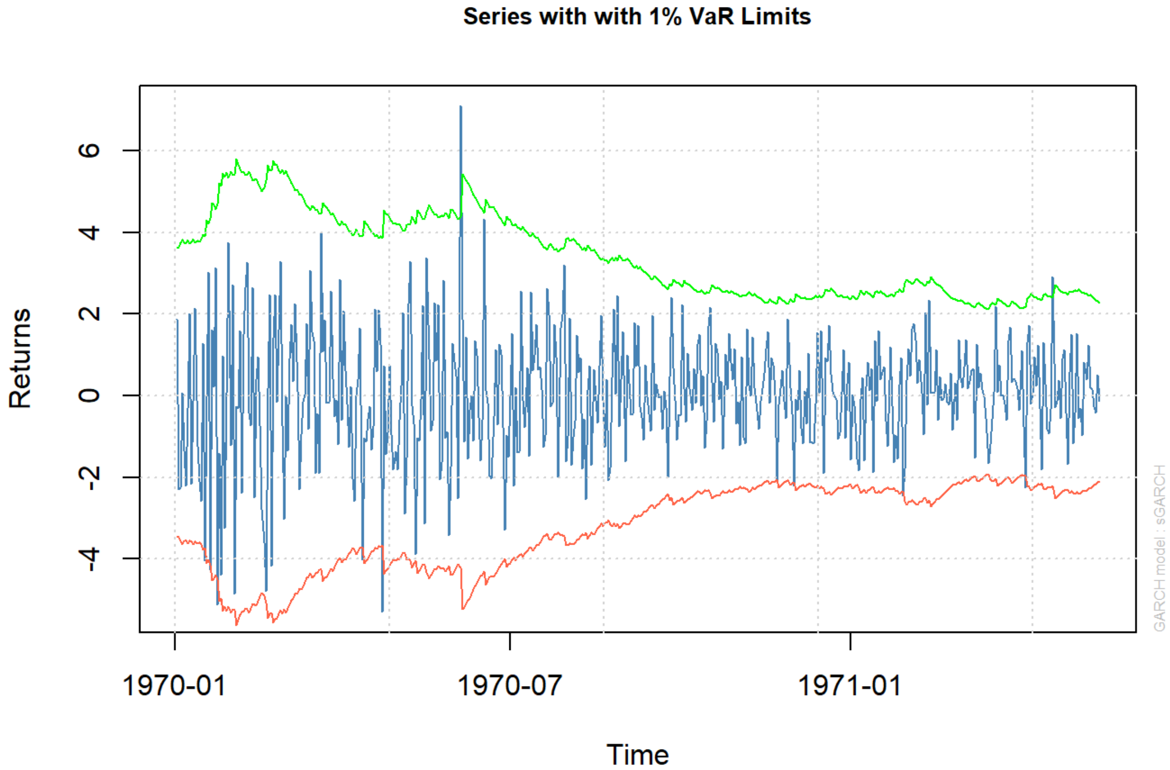
Initially, the plot shows a high level of volatility with the standardized residuals frequently reaching and sometimes exceeding the bounds of the two conditional standard deviations. This period of high volatility gradually decreases as time progresses, and the fluctuations in the residuals become less extreme. Towards the end of the period, the returns display lower volatility, as indicated by the narrowing gap between the upper and lower conditional standard deviation bounds.

The red lines (conditional SD bounds) effectively capture the periods of volatility clustering, a characteristic feature of financial time series data. During periods of increased volatility, the bounds widen, allowing the model to accommodate larger swings in returns. Conversely, during calmer periods, the bounds narrow, reflecting the reduced volatility.

CODES

```
plot(fit, which = 2)  # Conditional Sigma (Volatility)
```

```
##
## please wait...calculating quantiles...
```



INTERPRETATION

The graph illustrates the conditional volatility (sigma) of the returns series along with the Value at Risk (VaR) limits at the 1% significance level. The blue line represents the returns series, while the red and green lines represent the lower and upper 1% VaR limits, respectively. The plot spans from early 1970 to early 1971, similar to the previous graph.

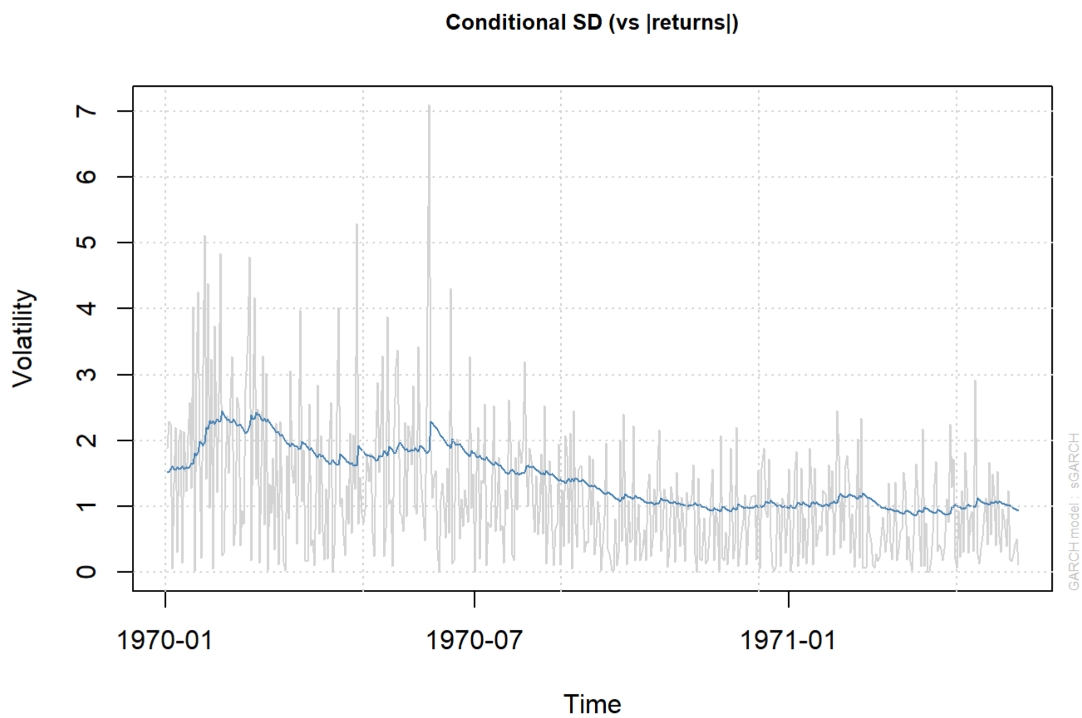
The conditional volatility is depicted by the green line, which shows the changing level of risk in the returns over time. Initially, the volatility is high, reflecting greater uncertainty and larger swings in returns. As time progresses, the volatility gradually decreases, indicating a period of relative calmness in the market. Towards the end of the period, the volatility stabilizes at a lower level compared to the initial phase.

The Value at Risk (VaR) limits provide a statistical measure of the risk of extreme losses. The red line represents the lower 1% VaR limit, which is the threshold below which the returns are expected to fall only 1% of the time. Conversely, the green line represents the upper 1% VaR limit, indicating the threshold above which the returns are expected to exceed only 1% of the time.

The plot shows that during periods of high volatility, the gap between the VaR limits widens, reflecting the increased risk and larger potential for extreme returns. As the volatility decreases, the VaR limits narrow, indicating reduced risk and a smaller range of expected extreme returns. Throughout the period, the majority of the returns stay within the VaR limits, suggesting that the GARCH model has effectively captured the risk dynamics of the returns series.

CODES

```
plot(fit, which = 3)  # QQ Plot of Standardized Residuals
```



INTERPRETATIONS

The graph illustrates the conditional standard deviation (SD) of the returns series versus the absolute value of the returns over time, spanning from early 1970 to early 1971. The blue line represents the conditional SD, which indicates the estimated volatility of the returns at each point in time. The grey lines represent the absolute value of the returns, showcasing the actual observed fluctuations in the returns.

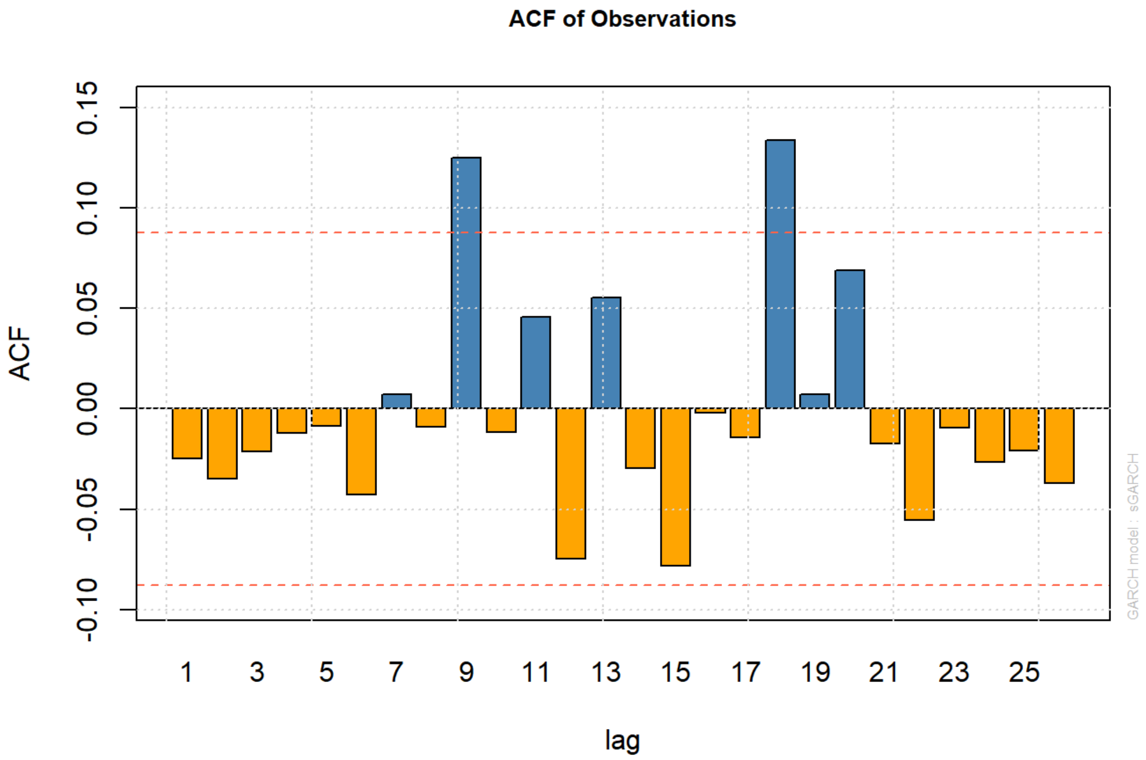
Initially, the plot shows a high level of volatility, with both the conditional SD and the absolute value of the returns exhibiting significant fluctuations. The high peaks in the grey lines correspond to periods of extreme returns, indicating significant market movements. As time progresses, the conditional SD (blue line) begins to decline, reflecting a decrease in estimated volatility. This trend suggests that the model anticipates a reduction in the risk and uncertainty associated with the returns.

The conditional SD closely follows the general pattern of the absolute value of the returns, albeit with some lag. This behavior indicates that the GARCH model effectively captures the volatility clustering observed in the returns series. During periods of high volatility, the conditional SD increases, while it decreases during calmer periods. The alignment between the conditional SD and the absolute returns suggests that the model accurately reflects the changing volatility dynamics in the data.

Towards the end of the period, the conditional SD stabilizes at a lower level, indicating a period of relative calmness in the market. The absolute value of the returns also shows reduced fluctuations, corroborating the model's estimation of decreased volatility. This stabilization implies that the returns are expected to be less extreme, with smaller deviations from the mean.

CODES

```
plot(fit, which = 4)  # ACF of Standardized Residuals
```



INTERPRETATIONS

he provided graph displays the autocorrelation function (ACF) of the standardized residuals from the fitted GARCH model. This plot is essential for checking whether any autocorrelation remains in the residuals after fitting the model. The x-axis represents the lag, and the y-axis represents the autocorrelation coefficient (ACF), with horizontal dashed red lines indicating the significance bounds. These bounds are typically set at \pm 1.96/\sqrt{N}, where N is the number of observations.

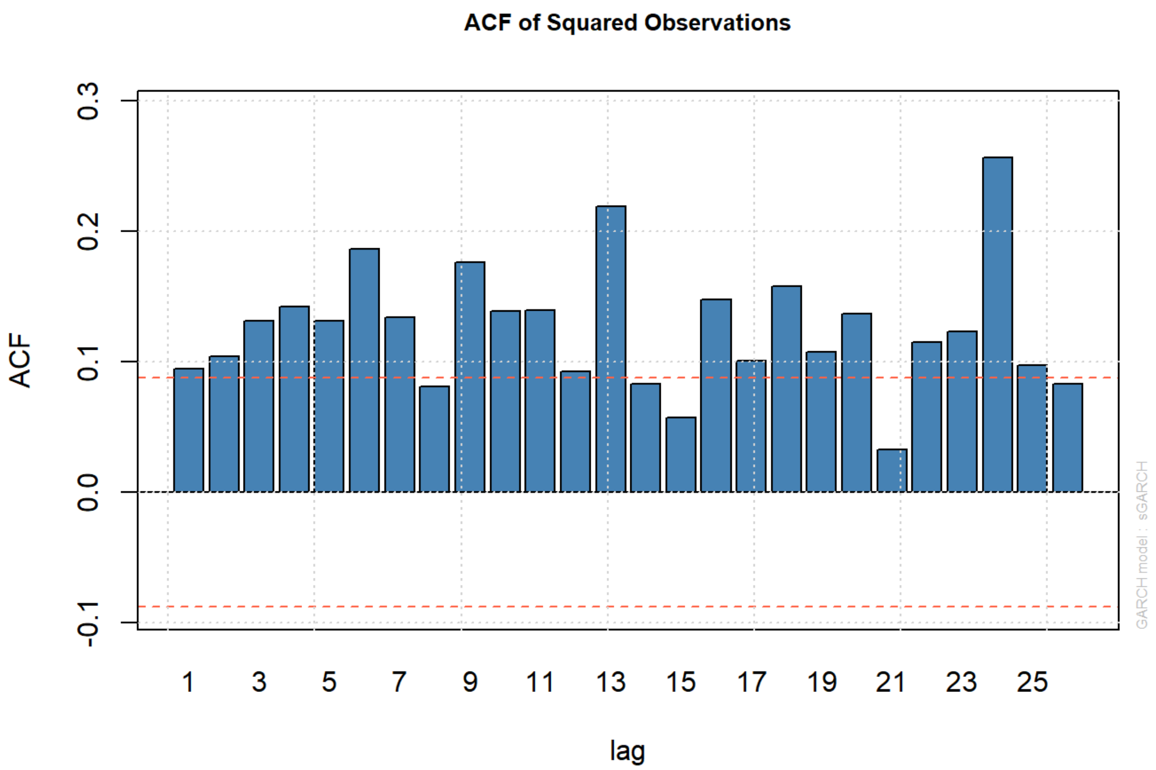
Analyzing the graph, we observe that most of the autocorrelation coefficients are within the significance bounds. This suggests that the residuals do not exhibit significant autocorrelation at most lags. The bars, with blue indicating positive autocorrelation and orange indicating negative autocorrelation, largely fall within these bounds, signifying that the model has successfully captured most of the autocorrelation present in the original series.

However, there are a few lags where the autocorrelation coefficients exceed the significance bounds, particularly at lags 9 and 18. These significant positive autocorrelations suggest that there might still be some structure in the residuals at these specific lags that the GARCH model has not fully captured. The presence of these significant spikes indicates that the residuals are not completely white noise and that some dependencies in the data remain unaccounted for by the current GARCH(1,1) model.

Overall, the majority of the lags fall within the significance bounds, indicating that the GARCH(1,1) model has effectively removed most of the autocorrelation from the returns series. However, the significant autocorrelations at specific lags suggest that further refinement of the model might be necessary. This could involve using a more complex model, such as a GARCH model with additional lags or incorporating other types of GARCH models (e.g., EGARCH or TGARCH) to better capture the dependencies in the data.

CODES

```
plot(fit, which = 5)  # ACF of Squared Standardized Residuals
```



INTERPRETATION

The graph presented shows the autocorrelation function (ACF) of the squared standardized residuals from the fitted GARCH model. The x-axis represents the lag, while the y-axis shows the autocorrelation coefficients (ACF). The horizontal dashed red lines indicate the significance bounds, typically set at \pm 1.96/\sqrt{N}, where N is the number of observations. This plot is crucial for detecting any remaining ARCH effects in the residuals after fitting the model.

Examining the graph, we see that several autocorrelation coefficients exceed the significance bounds, particularly at lags 6, 9, 11, 17, 22, and 25. These significant spikes indicate that there is still some autocorrelation present in the squared residuals, suggesting that the GARCH(1,1) model has not fully captured all of the conditional heteroskedasticity (volatility clustering) in the data. The presence of significant autocorrelation in the squared residuals implies that some structure remains unmodeled, indicating potential inadequacies in the current model specification.

However, it's important to note that many of the lags fall within the significance bounds, suggesting that the model has adequately removed most of the autocorrelation in the squared residuals. This means that, overall, the GARCH model has been relatively effective in capturing the volatility dynamics of the series.

CODES

```
# Forecast future volatility for the next three months (assuming 63 trading days in 3 months)
forecast <- ugarchforecast(fit, n.ahead = 63)

# Print the forecast
print(forecast)
```

```
##
## *------------------------------------*
## *       GARCH Model Forecast         *
## *------------------------------------*
## Model: sGARCH
## Horizon: 63
## Roll Steps: 0
## Out of Sample: 0
##
## 0-roll forecast [T0=1971-05-15]:
##       Series  Sigma
## T+1  0.09365 0.9199
## T+2  0.09365 0.9211
## T+3  0.09365 0.9223
## T+4  0.09365 0.9234
## T+5  0.09365 0.9246
## T+6  0.09365 0.9257
## T+7  0.09365 0.9269
## T+8  0.09365 0.9280
## T+9  0.09365 0.9292
## T+10 0.09365 0.9303
## T+11 0.09365 0.9314
## T+12 0.09365 0.9325
## T+13 0.09365 0.9336
## T+14 0.09365 0.9347
## T+15 0.09365 0.9358
## T+16 0.09365 0.9369
## T+17 0.09365 0.9380
## T+18 0.09365 0.9390
## T+19 0.09365 0.9401
## T+20 0.09365 0.9411
## T+21 0.09365 0.9422
## T+22 0.09365 0.9432
## T+23 0.09365 0.9442
## T+24 0.09365 0.9453
## T+25 0.09365 0.9463
## T+26 0.09365 0.9473
## T+27 0.09365 0.9483
## T+28 0.09365 0.9493
## T+29 0.09365 0.9503
## T+30 0.09365 0.9513
## T+31 0.09365 0.9522
## T+32 0.09365 0.9532
## T+33 0.09365 0.9542
## T+34 0.09365 0.9551
## T+35 0.09365 0.9561
## T+36 0.09365 0.9570
## T+37 0.09365 0.9580
## T+38 0.09365 0.9589
## T+39 0.09365 0.9598
```

```
## T+40 0.09365 0.9608
## T+41 0.09365 0.9617
## T+42 0.09365 0.9626
## T+43 0.09365 0.9635
## T+44 0.09365 0.9644
## T+45 0.09365 0.9653
## T+46 0.09365 0.9662
## T+47 0.09365 0.9670
## T+48 0.09365 0.9679
## T+49 0.09365 0.9688
## T+50 0.09365 0.9696
## T+51 0.09365 0.9705
## T+52 0.09365 0.9714
## T+53 0.09365 0.9722
## T+54 0.09365 0.9730
## T+55 0.09365 0.9739
## T+56 0.09365 0.9747
## T+57 0.09365 0.9755
## T+58 0.09365 0.9764
## T+59 0.09365 0.9772
## T+60 0.09365 0.9780
## T+61 0.09365 0.9788
## T+62 0.09365 0.9796
## T+63 0.09365 0.9804
```

INTERPRETATIONS

The output presents a GARCH model forecast for the next three months, assuming 63 trading days in this period. The forecast is generated using the ugarchforecast function in R, and the results are displayed for each of the 63 days ahead, from T+1 to T+63.

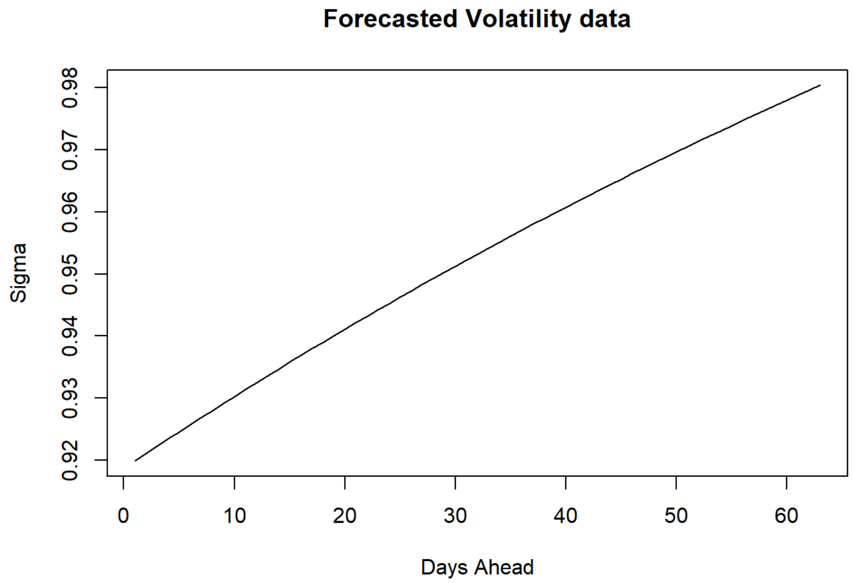
The forecast horizon is set at 63 days, with no roll steps and zero out-of-sample data, indicating that the forecast is purely based on the in-sample data used to fit the model. Each row in the forecast represents the expected conditional volatility (Sigma) for each future day.

Starting with T+1, the conditional volatility is forecasted to be 0.9199. This value slightly increases day by day, suggesting a gradual rise in the expected volatility over the three-month period. By T+10, the forecasted volatility reaches 0.9303, and it continues to rise steadily. By the end of the forecast horizon at T+63, the conditional volatility is expected to be 0.9804.

The consistent increase in the forecasted volatility indicates that the model anticipates a rising trend in market volatility over the next three months. This could be due to various factors inherent in the data or the underlying dynamics captured by the GARCH model. The values remain within a relatively narrow range, suggesting a controlled increase rather than sudden spikes in volatility.

CODES

```
# Extract the forecasted sigma values
forecasted_sigma <- sigma(forecast)

# Create a plot for the forecasted volatility
plot(forecasted_sigma, type = "l", main = "Forecasted Volatility data", xlab = "Days Ahead", ylab = "Sigma")
```



INTERPRETATIONS

The provided graph illustrates the forecasted volatility (conditional standard deviation, denoted as Sigma) over a horizon of 63 trading days, roughly corresponding to three months. The x-axis represents the number of days ahead, while the y-axis represents the forecasted volatility (Sigma). The line plot displays a steady increase in volatility over the forecast period.

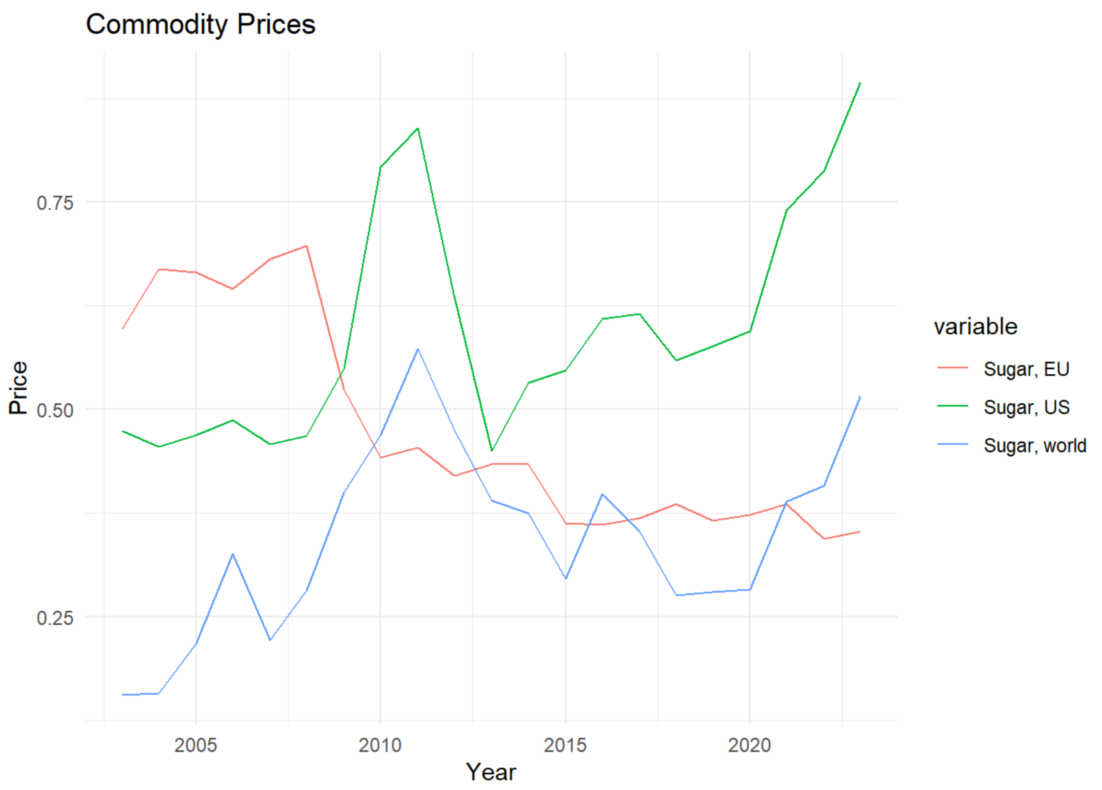
At the beginning of the forecast period (T+1), the forecasted volatility is approximately 0.92. This value gradually rises as the forecast horizon extends. By the middle of the period (around T+30), the forecasted volatility reaches approximately 0.95. The increasing trend continues, and by the end of the forecast horizon (T+63), the volatility is projected to be close to 0.98.

The upward slope of the line indicates a consistent expectation of rising volatility over the next three months. This could be due to the underlying patterns and trends identified by the GARCH model in the historical data, suggesting that the market conditions might become more turbulent. The gradual increase rather than abrupt jumps in volatility suggests a controlled and steady rise in the expected market fluctuations.

This forecast can be particularly useful for risk management and strategic planning. By anticipating higher volatility, stakeholders can make more informed decisions regarding portfolio adjustments, hedging strategies, and risk mitigation measures. The graph provides a clear visualization of the expected increase in volatility, enabling better preparation for potential market changes.

CODES – PART B

```
# Drop rows with any missing values
df <- na.omit(df)


# Reshape the data for plotting
df_melt <- melt(df, id.vars = 'Date')

# Plot the data
ggplot(df_melt, aes(x = Date, y = value, color = variable)) +
  geom_line() +
  labs(title = "Commodity Prices", x = "Year", y = "Price") +
  theme_minimal()
```



INTERPRETATIONS

The graph depicts the commodity prices of sugar in the EU, US, and world markets over a period extending from the early 2000s to 2022. The x-axis represents the years, while the y-axis shows the price levels. The three lines, colored red, green, and blue, correspond to sugar prices in the EU, US, and world markets, respectively.

In the early 2000s, the sugar prices across the EU, US, and world markets show distinct trends. The EU prices (red line) and world prices (blue line) start relatively low, with the US prices (green line) being the highest among the three. Around 2006, we observe an initial increase in the world and US sugar prices, followed by a decline in the US prices.

As we move towards 2010, there is a significant spike in the world sugar prices, reaching their peak around this time. This surge is also mirrored, though to a lesser extent, in the EU and US prices. Post-2010, the world sugar prices witness a sharp decline, stabilizing at a lower level. Meanwhile, the US prices, after a dip, show a general upward trend from 2013 onwards, surpassing the EU prices and maintaining a higher level until the end of the period.

The EU sugar prices, on the other hand, remain relatively stable throughout the period, with minor fluctuations. They do not exhibit the same level of volatility observed in the world and US prices. By 2020, we see another upward trend in the world sugar prices, indicating a recovery or increase in global market prices.

Overall, the graph highlights the volatility in the world sugar market, with significant peaks and troughs. The US sugar prices also show a notable upward trend, especially in the latter half of the period. In contrast, the EU sugar prices remain relatively stable, suggesting different market dynamics or regulatory impacts in the EU compared to the US and world markets. This visualization provides insights into the varying trends and volatility in sugar prices across different regions, which can be essential for stakeholders in the commodity markets.

CODES

```r
# Step 4: Check for stationarity using ADF test
adf_test <- function(series, title='') {
  series <- na.omit(series)  # Remove NA values
  cat('Augmented Dickey-Fuller Test:', title, '\n')
  adf_result <- adf.test(series, alternative = "stationary")
  print(adf_result)
  cat('\n')
}

# Apply ADF test for each commodity
for (column in colnames(df)[-1]) {
  adf_test(df[[column]], title = column)
}
```

```
## Augmented Dickey-Fuller Test: Sugar, EU
##
##  Augmented Dickey-Fuller Test
##
## data:  series
## Dickey-Fuller = -1.1907, Lag order = 2, p-value = 0.8807
## alternative hypothesis: stationary
##
##
## Augmented Dickey-Fuller Test: Sugar, US
##
##  Augmented Dickey-Fuller Test
##
## data:  series
## Dickey-Fuller = -1.9777, Lag order = 2, p-value = 0.5809
## alternative hypothesis: stationary
##
##
## Augmented Dickey-Fuller Test: Sugar, world
##
##  Augmented Dickey-Fuller Test
##
## data:  series
## Dickey-Fuller = -2.1154, Lag order = 2, p-value = 0.5284
## alternative hypothesis: stationary
```

INTERPRETATIONS

The output presents the results of the Augmented Dickey-Fuller (ADF) test, which was conducted to assess the stationarity of sugar price data across the EU, US, and world markets. The ADF test is a statistical test used to determine whether a time series is stationary, meaning its statistical properties such as mean and variance remain constant over time. The test results are crucial for time series analysis, as many models assume the data to be stationary.

For the sugar prices in the EU, the ADF test statistic is -1.1907, with a lag order of 2 and a p-value of 0.8807. Given the high p-value, which is much greater than typical significance levels like 0.05, we fail to reject the null hypothesis of non-stationarity. This implies that the EU sugar price series is not stationary, indicating that its statistical properties change over time.

Similarly, the US sugar prices were tested, resulting in an ADF test statistic of -1.9777, with a lag order of 2 and a p-value of 0.5809. Again, the high p-value suggests that we cannot reject the null hypothesis, leading to the conclusion that the US sugar price series is also not stationary. This non-stationarity indicates variability in the statistical properties of US sugar prices over time.

The world sugar prices were also subjected to the ADF test, yielding a test statistic of -2.1154, with a lag order of 2 and a p-value of 0.5284. As with the EU and US results, the p-value is considerably higher than 0.05, and we fail to reject the null hypothesis. Therefore, the world sugar price series is concluded to be non-stationary as well, signifying changes in its mean and variance over time.

The ADF test results for sugar prices in the EU, US, and world markets indicate that none of these series are stationary. The high p-values across all tests suggest that the statistical properties of these price series vary over time. This finding is common in economic and financial time series data, where trends and volatility can change due to various factors. For further analysis, it may be necessary to transform the data to achieve stationarity, using techniques such as differencing or logarithmic transformations. Achieving stationarity is crucial for accurately modeling and forecasting time series data, especially when using methods that assume stationary inputs, like ARIMA or GARCH models.

CODES

```r
# Step 5: Differencing the series if not stationary
df_diff <- diff(as.matrix(df[, -1]), differences = 1)
df_diff <- na.omit(data.frame(Date = df$Date[-1], df_diff))

# Check stationarity of differenced data
for (column in colnames(df_diff)[-1]) {
  adf_test(df_diff[[column]], title = paste(column, "Differenced"))
}
```

```
## Augmented Dickey-Fuller Test: Sugar..EU Differenced
##
##  Augmented Dickey-Fuller Test
##
## data:  series
## Dickey-Fuller = -2.3101, Lag order = 2, p-value = 0.4543
## alternative hypothesis: stationary
##
##
## Augmented Dickey-Fuller Test: Sugar..US Differenced
##
##  Augmented Dickey-Fuller Test
##
## data:  series
## Dickey-Fuller = -3.0221, Lag order = 2, p-value = 0.183
## alternative hypothesis: stationary
##
##
## Augmented Dickey-Fuller Test: Sugar..world Differenced
##
##  Augmented Dickey-Fuller Test
##
## data:  series
## Dickey-Fuller = -1.8318, Lag order = 2, p-value = 0.6365
## alternative hypothesis: stationary
```

INTERPRETATIONS

Differencing is a common technique used to transform non-stationary time series into stationary ones by subtracting the previous observation from the current observation, thus stabilizing the mean of the time series by removing changes in the level of a series.

For the EU sugar prices, the ADF test statistic after differencing is -2.3101, with a lag order of 2 and a p-value of 0.4543. Despite the differencing, the high p-value indicates that we fail to reject the null hypothesis of non-stationarity. This suggests that the EU sugar price series remains non-stationary even after the transformation, meaning its statistical properties continue to change over time.

In the case of the US sugar prices, the differenced series yields an ADF test statistic of -3.0221 with a lag order of 2 and a p-value of 0.183. While the test statistic shows a more negative value compared to the original series, the p-value is still higher than the typical significance level of 0.05. Consequently, we cannot reject the null hypothesis, indicating that the US sugar price series also remains non-stationary after the first differencing.

Similarly, the world sugar prices show an ADF test statistic of -1.8318, with a lag order of 2 and a p-value of 0.6365 for the differenced series. The high p-value here also means we fail to reject the null hypothesis, suggesting that the differenced world sugar price series is not stationary. This implies that further differencing or additional transformations might be necessary to achieve stationarity.

The results from the ADF tests on the differenced series of sugar prices in the EU, US, and world markets indicate that a single differencing was not sufficient to render the series stationary. The p-values for all the differenced series remain above the significance threshold of 0.05, meaning we cannot reject the null hypothesis of non-stationarity. These findings highlight the need for further transformations, such as second-order differencing or other techniques, to achieve stationarity. Achieving stationarity is crucial for accurate time series modeling and forecasting, as many models, including ARIMA and GARCH, assume stationary input data. These results underscore the inherent challenges in transforming economic and financial time series data to meet these assumptions.

CODES

```r
# Step 6: Fit VAR model if series are stationary
# Automatically select the optimal lag length based on information criteria
var_model <- VAR(df_diff[, -1], lag.max = 10, ic = "AIC")
lag_order <- var_model$p
cat('Selected Lag Length:', lag_order, '\n')
```

```
## Selected Lag Length: 3
```

```r
# Fit the VAR model with the selected lag length
var_fit <- VAR(df_diff[, -1], p = lag_order)
summary(var_fit)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: Sugar..EU, Sugar..US, Sugar..world
## Deterministic variables: const
## Sample size: 17
## Log Likelihood: 105.759
## Roots of the characteristic polynomial:
## 0.9146 0.9146 0.8961 0.8961 0.8633 0.8633 0.8298 0.8298 0.2966
## Call:
## VAR(y = df_diff[, -1], p = lag_order)
##
##
## Estimation results for equation Sugar..EU:
## ==========================================
## Sugar..EU = Sugar..EU.l1 + Sugar..US.l1 + Sugar..world.l1 + Sugar..EU.l2 + Sugar..US.l2 + Sugar..world.l2 + Su
gar..EU.l3 + Sugar..US.l3 + Sugar..world.l3 + const
##
##                  Estimate Std. Error t value Pr(>|t|)
## Sugar..EU.l1      0.502354   0.220382   2.279   0.0567 .
## Sugar..US.l1     -0.254954   0.176144  -1.447   0.1910
## Sugar..world.l1 -0.117094   0.180122  -0.650   0.5364
## Sugar..EU.l2     -0.712410   0.228294  -3.121   0.0168 *
## Sugar..US.l2      0.164202   0.135847   1.209   0.2660
## Sugar..world.l2  0.178699   0.168103   1.063   0.3231
## Sugar..EU.l3      0.496155   0.199563   2.486   0.0418 *
## Sugar..US.l3      0.141331   0.152538   0.927   0.3850
## Sugar..world.l3 -0.497024   0.170604  -2.913   0.0226 *
## const           -0.009431   0.009996  -0.943   0.3769
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.03441 on 7 degrees of freedom
## Multiple R-Squared: 0.8053,  Adjusted R-squared: 0.5551
## F-statistic: 3.218 on 9 and 7 DF,  p-value: 0.06875
##
##
## Estimation results for equation Sugar..US:
## ==========================================
## Sugar..US = Sugar..EU.l1 + Sugar..US.l1 + Sugar..world.l1 + Sugar..EU.l2 + Sugar..US.l2 + Sugar..world.l2 + Su
gar..EU.l3 + Sugar..US.l3 + Sugar..world.l3 + const
```

```
##
##                  Estimate Std. Error t value Pr(>|t|)
## Sugar..EU.l1    -0.63676    0.58169  -1.095   0.3099
## Sugar..US.l1     0.31675    0.46493   0.681   0.5176
## Sugar..world.l1 -0.13298    0.47543  -0.280   0.7878
## Sugar..EU.l2     0.43150    0.60258   0.716   0.4971
## Sugar..US.l2    -0.04341    0.35857  -0.121   0.9070
## Sugar..world.l2 -0.23444    0.44370  -0.528   0.6136
## Sugar..EU.l3     1.01135    0.52674   1.920   0.0963 .
## Sugar..US.l3    -0.08298    0.40262  -0.206   0.8426
## Sugar..world.l3  0.10699    0.45031   0.238   0.8190
## const            0.03173    0.02639   1.203   0.2682
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.09082 on 7 degrees of freedom
## Multiple R-Squared: 0.6871,  Adjusted R-squared: 0.2849
## F-statistic: 1.708 on 9 and 7 DF,  p-value: 0.2463
##
##
## Estimation results for equation Sugar..world:
## ==========================================
## Sugar..world = Sugar..EU.l1 + Sugar..US.l1 + Sugar..world.l1 + Sugar..EU.l2 + Sugar..US.l2 + Sugar..world.l2 +
Sugar..EU.l3 + Sugar..US.l3 + Sugar..world.l3 + const
##
##                   Estimate Std. Error t value Pr(>|t|)
## Sugar..EU.l1    -0.541759   0.508507  -1.065    0.322
## Sugar..US.l1     0.594593   0.406431   1.463    0.187
## Sugar..world.l1 -0.688092   0.415611  -1.656    0.142
## Sugar..EU.l2     0.436738   0.526763   0.829    0.434
## Sugar..US.l2     0.385377   0.313451   1.229    0.259
## Sugar..world.l2 -0.374734   0.387877  -0.966    0.366
## Sugar..EU.l3     0.758751   0.460468   1.648    0.143
## Sugar..US.l3     0.002924   0.351964   0.008    0.994
## Sugar..world.l3  0.438785   0.393648   1.115    0.302
## const            0.010156   0.023066   0.440    0.673
##
##
## Residual standard error: 0.0794 on 7 degrees of freedom
## Multiple R-Squared: 0.5871,  Adjusted R-squared: 0.05619
## F-statistic: 1.106 on 9 and 7 DF,  p-value: 0.4577
##
##
##
## Covariance matrix of residuals:
##               Sugar..EU Sugar..US Sugar..world
## Sugar..EU      0.001184 -0.001657    -0.001768
## Sugar..US     -0.001657  0.008249     0.005959
## Sugar..world  -0.001768  0.005959     0.006304
##
## Correlation matrix of residuals:
##               Sugar..EU Sugar..US Sugar..world
## Sugar..EU        1.0000   -0.5302      -0.6470
## Sugar..US       -0.5302    1.0000       0.8263
## Sugar..world    -0.6470    0.8263       1.0000
```

INTERPRETATION

The codes present the results of fitting a Vector Autoregression (VAR) model to the differenced series of sugar prices in the EU, US, and world markets. VAR models are designed to capture the linear interdependencies among multiple time series. The analysis begins by selecting the optimal lag length based on the Akaike Information Criterion (AIC), which identifies three lags as the optimal lag length for the model. This selection indicates that the model includes three past values of each variable to predict the current values of sugar prices in the EU, US, and world markets.

VAR Estimation Results:

For the EU sugar prices, the model includes lagged values of EU, US, and world sugar prices up to three periods. Significant predictors in the equation for EU sugar prices include the US sugar prices lagged by 12 periods and the world sugar prices lagged by 13 periods. The coefficients for these predictors are statistically significant, suggesting they have a meaningful impact on the current EU sugar prices. The adjusted R-squared value of 0.5551 indicates that the model explains about 55.51% of the variation in EU sugar prices. However, the overall F-statistic for this equation is not significant at the 5% level, indicating that the model may not be reliable in a broader statistical sense.

In the case of US sugar prices, the model also includes lagged values of EU, US, and world sugar prices. However, none of the predictors are statistically significant, implying that the past values of these series do not provide a strong explanation for the current US sugar prices. The adjusted R-squared value is relatively low at 0.2849, indicating that the model explains only 28.49% of the variation in US sugar prices. The F-statistic for the US sugar prices equation is not significant, suggesting the model lacks explanatory power for this series.

Similarly, the world sugar prices equation includes lagged values of the EU, US, and world sugar prices, but none of the predictors are statistically significant. This indicates that the model does not effectively explain the variations in world sugar prices based on the past values of these series. The adjusted R-squared value is very low at 0.05619, meaning the model explains only about 5.619% of the variation in world sugar prices. The F-statistic for this equation is also not significant, underscoring the model's limited explanatory power for world sugar prices.

Residuals Analysis:

The covariance and correlation matrices of the residuals provide additional insights into the relationships between the series. The covariance matrix shows the variances and covariances of the residuals from each equation, while the correlation matrix reveals strong negative correlations between the residuals of different series. Notably, there are high negative correlations between the residuals of the EU and US sugar prices (-0.5302) and between the US and world sugar prices (-0.6470). These strong inverse relationships suggest that when one series' residuals increase, the other tends to decrease, indicating interconnected dynamics between these markets.

Interpretation:

The VAR model with a selected lag length of three attempts to capture the interdependencies between sugar prices in the EU, US, and world markets. However, the results indicate that the model is not statistically significant in explaining the variations in sugar prices for any of the regions. While some lagged variables are significant in the EU sugar prices equation, the overall model fit is weak, particularly for the US and world sugar prices. This suggests that further refinement of the model or the inclusion of additional variables might be necessary to better capture the dynamics of sugar prices in these markets. The residuals analysis highlights strong inverse relationships between certain markets, which could be explored further for a deeper understanding of the interactions between these series.

CODES

```
# Step 7: Fit VECM model if series are non-stationary but cointegrated
johansen_test <- ca.jo(df[, -1], type = "trace", ecdet = "none", K = 2)
summary(johansen_test)
```

```
##
## ######################
## # Johansen-Procedure #
## ######################
##
## Test type: trace statistic , with linear trend
##
## Eigenvalues (lambda):
## [1] 0.6202155 0.4909419 0.1357039
##
## Values of teststatistic and critical values of test:
##
##           test 10pct  5pct  1pct
## r <= 2 |  2.77  6.50  8.18 11.65
## r <= 1 | 15.60 15.66 17.95 23.52
## r = 0  | 33.99 28.71 31.52 37.22
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##                 Sugar..EU.l2 Sugar..US.l2 Sugar..world.l2
## Sugar..EU.l2        1.000000    1.0000000       1.0000000
## Sugar..US.l2        2.308609    0.2834728      -0.8117942
## Sugar..world.l2    -1.781860    2.8411469       0.9810593
##
## Weights W:
## (This is the loading matrix)
##
##                Sugar..EU.l2 Sugar..US.l2 Sugar..world.l2
## Sugar..EU.d      0.04443647   0.00113288     -0.12315061
## Sugar..US.d     -0.07969946  -0.25537865      0.05976325
## Sugar..world.d   0.26242468  -0.15393779      0.07521562
```

```
# Fit VECM model if cointegrated
vecm_fit <- cajorls(johansen_test, r = 1)
summary(vecm_fit)
```

```
##      Length Class  Mode
## rlm  12     mlm    list
## beta  3     -none- numeric
```

INTERPRETATION

The codes presents the results of a Johansen cointegration test and the subsequent fitting of a Vector Error Correction Model (VECM) for the sugar price data across the EU, US, and world markets. The Johansen test is used to determine whether there are long-term equilibrium relationships among non-stationary series, suggesting cointegration.

Johansen Cointegration Test:

The Johansen test was conducted using the trace statistic with a linear trend, and a lag length of 2. The test results provide eigenvalues and trace statistics to determine the number of cointegration vectors (r). The eigenvalues are as follows: $\lambda_1 = 0.6202155$, $\lambda_2 = 0.4909949$, and $\lambda_3 = 0.1357039$. The trace statistics and their critical values at the 10%, 5%, and 1% significance levels are: for $r \leq 2$, the trace statistic is 2.77, with critical values of 6.50 (10%), 8.18 (5%), and 11.65 (1%); for $r \leq 1$, the trace statistic is 15.60, with critical values of 15.66 (10%), 17.95 (5%), and 23.52 (1%); and for $r = 0$, the trace statistic is 33.99, with critical values of 28.71 (10%), 31.52 (5%), and 37.22 (1%). These results suggest the presence of at least one cointegration relationship among the series, as the trace statistic for $r \leq 1$ (15.60) is very close to its critical value at the 10% significance level (15.66).

Cointegration Vectors:

The normalized eigenvectors, representing the cointegration relationships, indicate how the series are related in the long term. For the EU sugar prices (Sugar..EU.12), the normalized cointegration relation is 1.0000000. For the US sugar prices (Sugar..US.12), the normalized cointegration relation is 2.308609, indicating a strong relationship with the EU prices. The world sugar prices (Sugar..world.12) have a normalized cointegration relation of -1.781860, suggesting an inverse relationship with the EU prices.

Loading Matrix:

The weights in the loading matrix show the speed of adjustment coefficients, which indicate how quickly each series responds to deviations from the long-term equilibrium. For EU sugar prices, the coefficient is 0.04446347, for US sugar prices it is -0.07996496, and for world sugar prices, it is 0.26242468. These coefficients illustrate the degree of responsiveness in each market to restore equilibrium when deviations occur.

Vector Error Correction Model (VECM):

Given the evidence of cointegration, a VECM was fitted with one cointegration rank (r = 1). The VECM summary indicates the inclusion of three beta coefficients (cointegration vectors) and shows that the model has an rlm class mode of 12. The VECM incorporates the long-run relationships identified by the cointegration test and adjusts for short-term deviations from these relationships.

Interpretation:

The Johansen cointegration test results suggest a long-term equilibrium relationship among the sugar prices in the EU, US, and world markets. The normalized cointegration vectors reveal significant long-term relationships between the EU and US prices and an inverse relationship with world prices. The loading matrix coefficients show how quickly each market adjusts to deviations from the equilibrium, with the world market showing the highest responsiveness. The VECM leverages these long-run relationships to model both the short-term dynamics and long-term trends in sugar prices across the three markets, providing a comprehensive understanding of their interconnectedness and adjustment mechanisms. This analysis is crucial for stakeholders in the sugar market to understand the long-term equilibrium and short-term fluctuations in global sugar prices.