Time Series - Homework

Théo Druilhe & Killian Steunou

February 28, 2024

Contents

1 Problem I

Take a prince index, the inflation, the GDP, the short-term interest rate, and the daily consumption of electricity of a developed country. You could take them from FRED or Mark Watson's website or any other source.

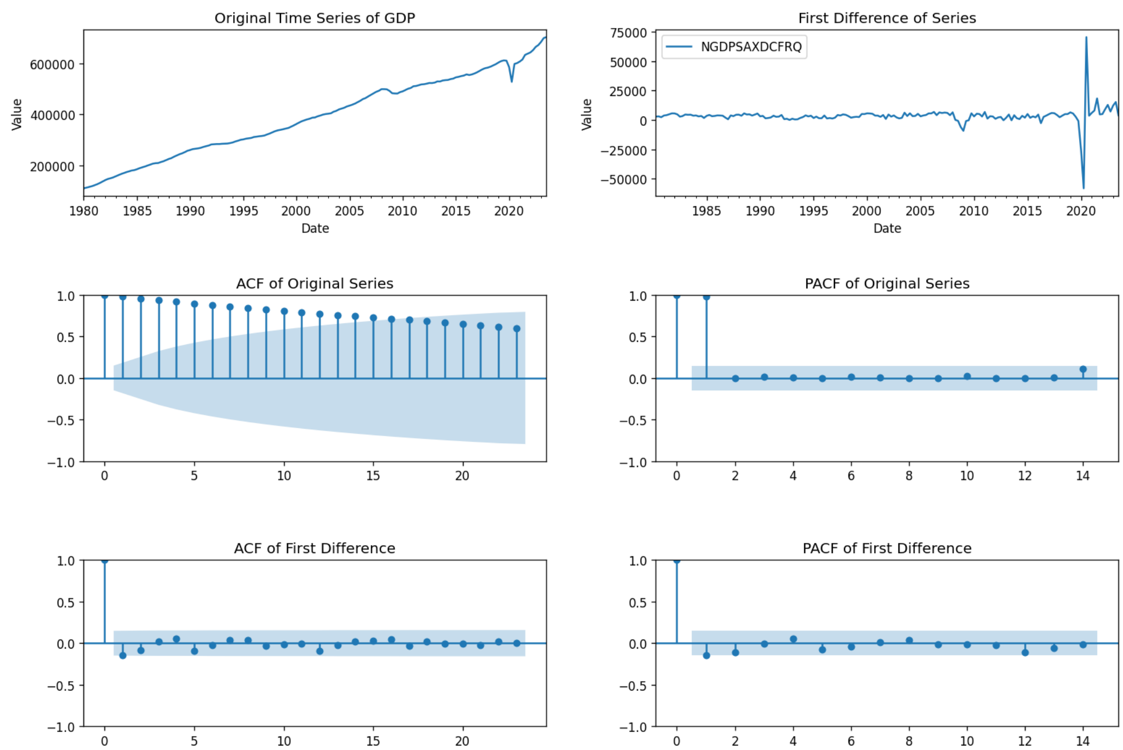
1.1 1.

We take the consumer price index, the inflation, the GDP, the short-term interest rate, and the daily consumption of electricity of France

1.1.1 GDP

```
            NGDPSAXDCFRQ
DATE
1980-01-01      108704.8
1980-04-01      111516.3
```

Plots




```
Descriptive Statistics for Original Series of GDP:
         NGDPSAXDCFRQ
count      175.000000
mean    390268.246857
```

```
std    157220.306558
min    108704.800000
25%    267353.750000
50%    388213.100000
75%    523864.650000
max    705014.700000

Descriptive Statistics for First Difference of GDP Series:
        NGDPSAXDCFRQ
count    174.000000
mean    3427.068391
std     7788.247567
min   -58242.500000
25%     2142.525000
50%     3587.750000
75%     4785.425000
max    70520.200000

ADF Test Statistic for Original Series of GDP: 0.5078161945075979
p-value for Original Series of GDP: 0.9851087370340792
ADF Test Statistic for First Difference of GDP: -10.949354335782362
p-value for First Difference of GDP: 8.878567484572884e-20
```

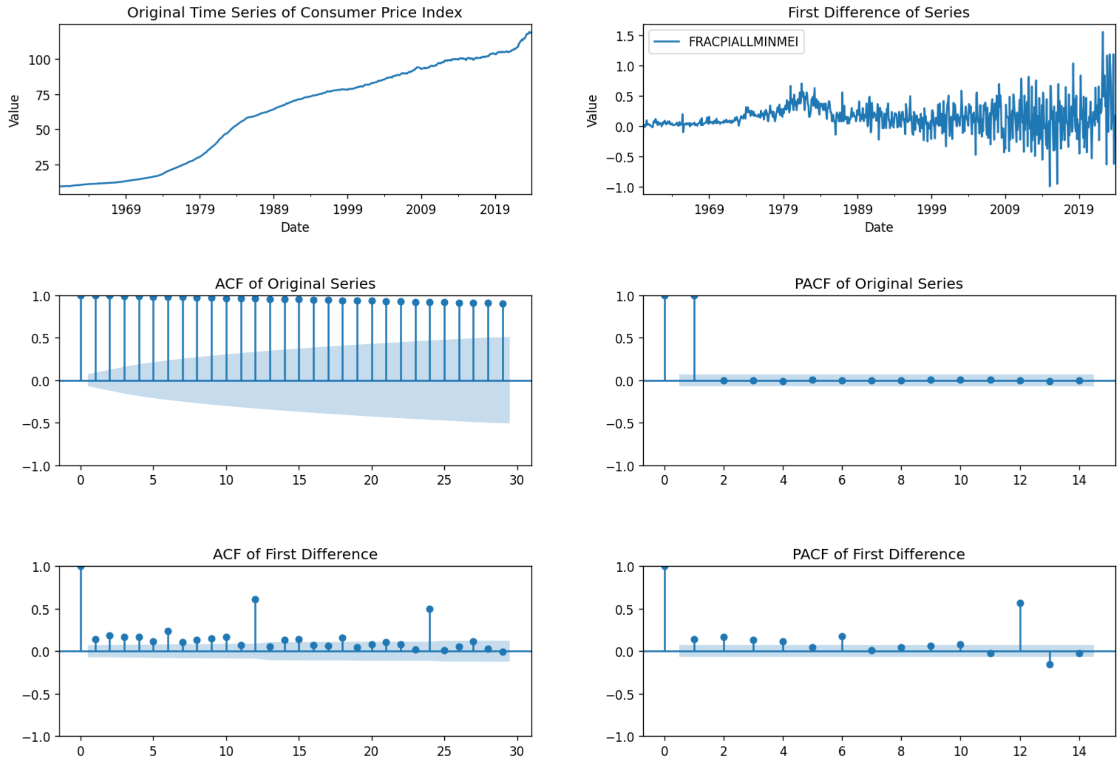
Comments

- The plot of the time series strongly suggest an increasing trend over time
- The ADF test statistic for the original series is positive, and the p-value is significantly greater than 0.05 (the typical threshold for statistical significance in such tests). This indicates that we fail to reject the null hypothesis, suggesting that the original GDP time series is non-stationary. This means that the series may have a trend, seasonality, or other structures that evolve over time rather than being constant.
- For the first-differenced series, the ADF test statistic is significantly negative, and the p-value is extremely low (practically zero for all practical purposes). This strongly suggests rejecting the null hypothesis in favor of the alternative hypothesis that the differenced series is stationary. This indicates that once we account for the changes between consecutive observations (by differencing), the resultant time series does not have any time-dependent structure like trends or seasonality.

1.1.2 Price index

```
            FRACPIALLMINMEI
DATE
1960-01-01         9.797457
1960-02-01         9.820050
1960-03-01         9.820050
1960-04-01         9.835111
1960-05-01         9.812519
```

Plots




```
Descriptive Statistics for Original Series of Consumer Price Index:
        FRACPIALLMINMEI
count       767.000000
mean         61.222645
std          34.226714
min           9.797457
25%          23.256459
50%          70.440000
75%          91.805000
max         118.890000
```

```
Descriptive Statistics for First Difference of Consumer Price Index Series:
        FRACPIALLMINMEI
count       766.000000
mean          0.141557
std           0.237738
min          -1.000000
25%           0.030031
50%           0.110271
75%           0.250000
max           1.550000
```

```
ADF Test Statistic for Original Series of Consumer Price Index:
-0.1684460330366885
p-value for Original Series of Consumer Price Index: 0.9421797365834444
ADF Test Statistic for First Difference of Consumer Price Index:
-3.214646766493181
p-value for First Difference of Consumer Price Index: 0.019152367449662137
```

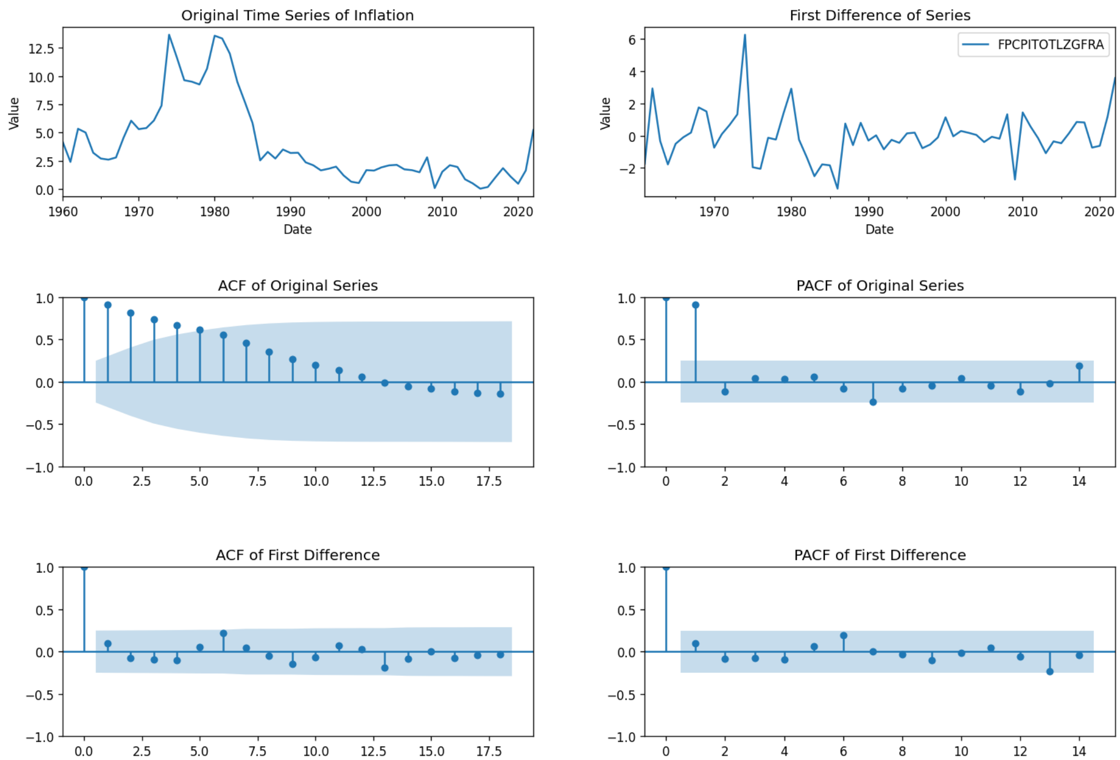
Comments

- We can see an increasing trend over time,
- The first difference plot is very noisy
- The ADF test statistic for the original CPI series is close to zero and negative, but the p-value is very high (much greater than 0.05). This means there's strong evidence that the series is non-stationary, likely containing a time-dependent structure such as a trend.
- For the first-differenced series, the ADF test statistic is significantly more negative, and the p-value is below the typical 0.05 threshold for statistical significance. This suggests that the differenced series is stationary, indicating that differencing has effectively removed the time-dependent structure from the data.

1.1.3 Inflation

```
            FPCPITOTLZGFRA
DATE
1960-01-01        4.139936
1961-01-01        2.400461
```

Plots




```
Descriptive Statistics for Original Series of Inflation:
        FPCPITOTLZGFRA
count        63.000000
mean          4.057410
std           3.658403
min           0.037514
25%           1.665320
50%           2.602001
75%           5.364399
max          13.649317

Descriptive Statistics for First Difference of Inflation Series:
        FPCPITOTLZGFRA
count        62.000000
mean          0.017459
std           1.524360
min          -3.292574
25%          -0.620838
50%          -0.122463
75%           0.729150
max           6.268716
```

```
ADF Test Statistic for Original Series of Inflation: -1.6208899976681996
p-value for Original Series of Inflation: 0.4722627369159214
ADF Test Statistic for First Difference of Inflation: -6.6493731626703925
p-value for First Difference of Inflation: 5.167278407629681e-09
```

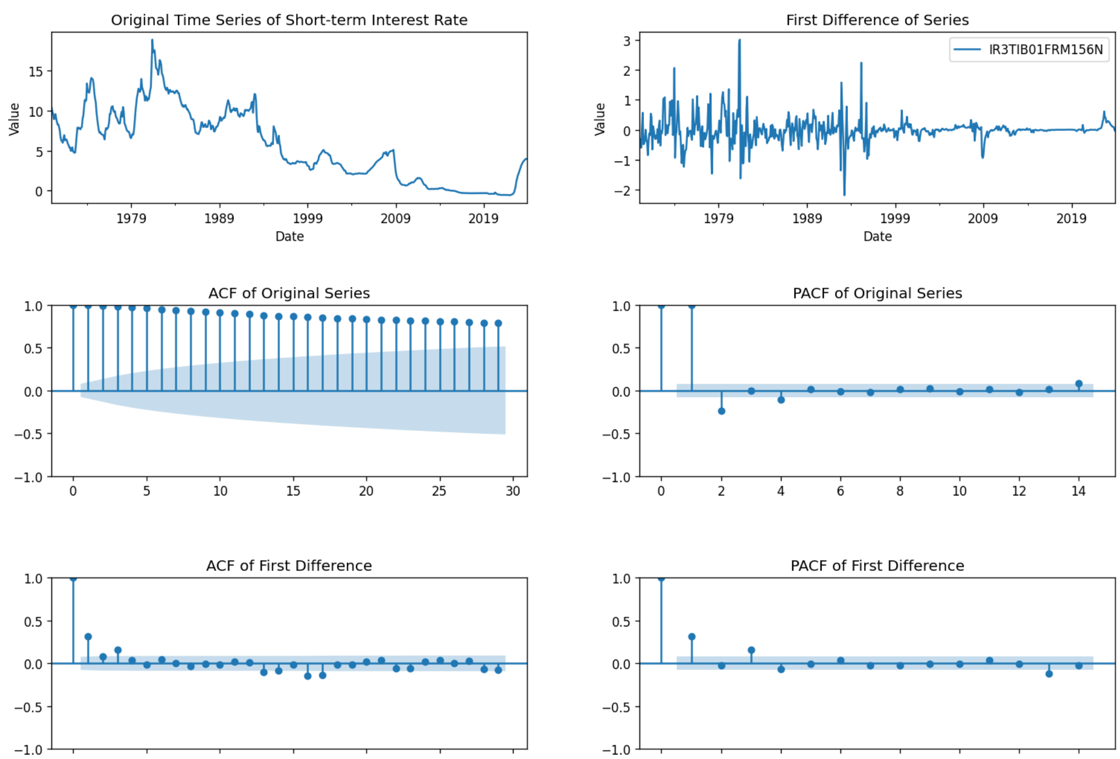
Comments

- The plot of the time series suggests it is stationary
- However the ADF test statistic for the original inflation series is negative but not enough to conclude stationarity, as indicated by the high p-value (greater than 0.05). This suggests that the original series is likely non-stationary
- For the first-differenced series, the ADF test statistic is much more negative, and the p-value is extremely low, practically zero. This strongly indicates that the first-differenced series is stationary, meaning that differencing the data has effectively removed the time-dependent structure.

1.1.4 Short-term interest rate

```
            IR3TIB01FRM156N
DATE
1970-01-01            10.35
1970-02-01             9.86
```

Plots



```
Descriptive Statistics for Original Series of Short-term Interest Rate:
        IR3TIB01FRM156N
count       648.000000
mean          5.510994
std           4.468329
min          -0.582000
25%           1.592175
50%           4.773900
75%           9.097875
max          18.920000

Descriptive Statistics for First Difference of Short-term Interest Rate Series:
        IR3TIB01FRM156N
count       647.000000
mean         -0.009923
std           0.416375
min          -2.192400
25%          -0.139650
50%          -0.005100
75%           0.077100
max           3.020000

ADF Test Statistic for Original Series of Short-term Interest Rate:
-1.8325783008835552
p-value for Original Series of Short-term Interest Rate: 0.3644115820582855
ADF Test Statistic for First Difference of Short-term Interest Rate:
-10.863717124734935
p-value for First Difference of Short-term Interest Rate: 1.4249110392897714e-19
```

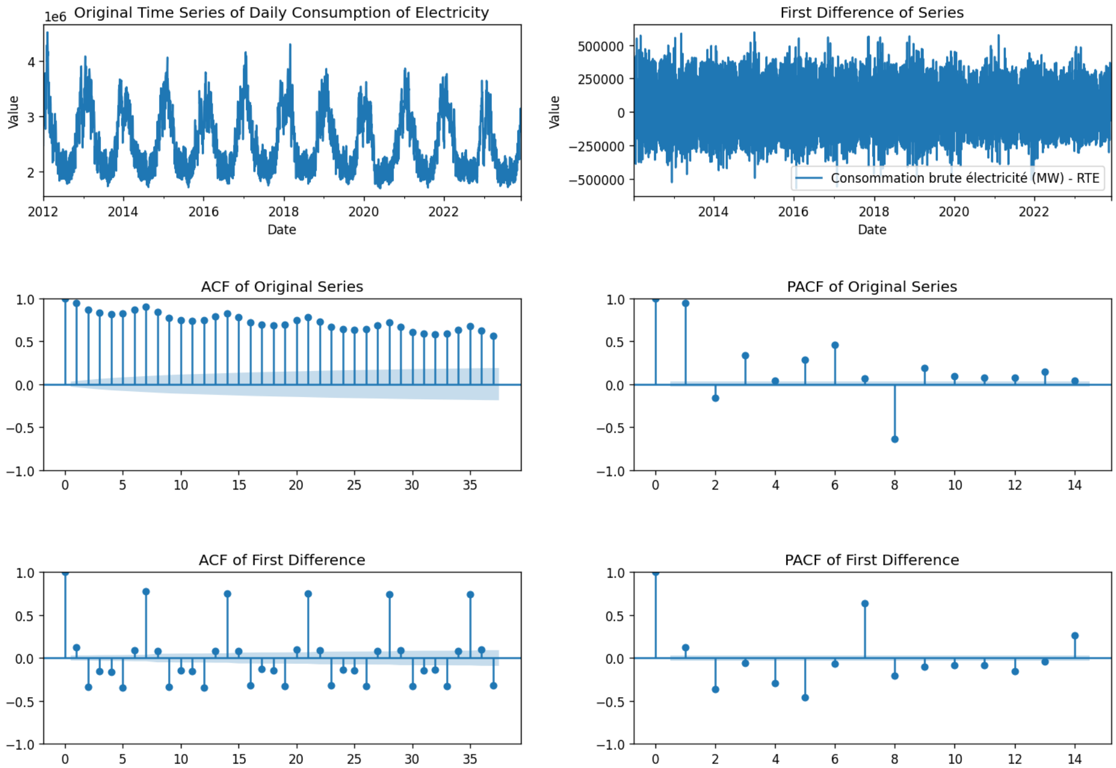
Comments

- The time series does not look stationary on the plot
- The ADF test statistic for the original series of Short-term Interest Rate is negative, yet the p-value is significantly above 0.05. This suggests that we cannot reject the null hypothesis of non-stationarity, implying the original series likely contains a time-dependent structure
- For the first-differenced series, the ADF test statistic is deeply negative, and the p-value is extremely low, practically zero. This strongly suggests rejecting the null hypothesis in favor of the alternative hypothesis, indicating the first-differenced series is stationary.

1.1.5 Electricity Daily Consumption

```
            Consommation brute électricité (MW) - RTE
Date
2012-01-01                                   2472864
2012-01-02                                   2912789
```

Plots




```
Descriptive Statistics for Original Series of Daily Consumption of Electricity:
        Consommation brute électricité (MW) - RTE
count                                4.352000e+03
mean                                 2.561585e+06
std                                  5.241475e+05
min                                  1.692364e+06
25%                                  2.172954e+06
50%                                  2.388416e+06
75%                                  2.957269e+06
max                                  4.516388e+06
```

```
Descriptive Statistics for First Difference of Daily Consumption of Electricity
Series:
        Consommation brute électricité (MW) - RTE
count                                 4351.000000
mean                                   139.878419
std                                 180429.329967
min                                -568887.000000
25%                                -110721.000000
50%                                  -7151.000000
```

```
75%                                       69545.500000
max                                      595023.000000

ADF Test Statistic for Original Series of Daily Consumption of Electricity:
-4.597050777489022
p-value for Original Series of Daily Consumption of Electricity:
0.000130796102098127
ADF Test Statistic for First Difference of Daily Consumption of Electricity:
-13.977161135377264
p-value for First Difference of Daily Consumption of Electricity:
4.2413199932288366e-26
```

Comments

- This time series shows a cyclic seasonal trend, which is normal since the electricity consumption vary along with the season.
- The ADF test statistic for the original series is significantly negative, and the p-value is well below the common threshold of 0.05. This indicates that we can reject the null hypothesis of the presence of a unit root, suggesting the series is stationary. Even though electricity consumption typically exhibits seasonal cycles, the test results imply that, aside from seasonality, the series does not have a time-dependent structure like a trend that would make it non-stationary.
- For the first-differenced series, the ADF test statistic is extremely negative, and the p-value is practically zero. This further confirms the absence of a unit root in the differenced series, indicating it is stationary. The first differencing would typically be used to remove any linear trend or seasonality, but given the original series already appears stationary, this step emphasizes the removal of any subtle non-stationary components that might not be immediately apparent.
- While the series exhibits short-term fluctuations, its statistical properties like the mean and variance do not change over time.
- This shows the difference between a strictly stationary process and a weakly stationary process (like this one that may look non-stationary but is not)

1.2 2.

Find the best AR(p) process for each of the variables.

To find the optimal p, we can start by a visual inspection: we count the number of lags in the PACF plots that are far away from 0.

- 1 for GDP
- 1 for Consumer Price Index
- 3 for Inflation
- 2 for the Short-term Interest Rate
- 5 for the Electricity Consumption

Note that this is only a visual method, we will find the optimal p analytically.

1.2.1 GDP

```
The best AR(p) model is AR(0).
```

An AR(0) model being the best fit for a time series, suggests that the current value of the series is not significantly influenced by its past values. In other words, the series is essentially behaving as if each observation is independent of the others, and the best predictor of the series at any point in time is the mean of the series, without taking into account any past information.

1.2.2 Price Index

```
The best AR(p) model is AR(4).
```

1.2.3 Inflation

```
The best AR(p) model is AR(0).
```

1.2.4 Short-term Interest Rate

```
The best AR(p) model is AR(4).
```

1.2.5 Electricity Daily Consumption

```
The best AR(p) model is AR(5).
```

1.3 3.

Provide 1-step and 3-step ahead forecasts for each variable.

1.3.1 GDP

```
1-step ahead forecast: 2023-10-01    390268.246857
Freq: QS-OCT, dtype: float64
3-steps ahead forecast: 2023-10-01    390268.246857
2024-01-01    390268.246857
2024-04-01    390268.246857
Freq: QS-OCT, dtype: float64
```

1.3.2 Price Index

```
1-step ahead forecast: 2023-12-01    118.247831
Freq: MS, dtype: float64
3-steps ahead forecast: 2023-12-01    118.247831
2024-01-01    118.338633
2024-02-01    118.419696
Freq: MS, dtype: float64
```

1.3.3 Inflation

```
1-step ahead forecast: 2023-01-01    4.05741
Freq: AS-JAN, dtype: float64
```

```
3-steps ahead forecast: 2023-01-01    4.05741
2024-01-01    4.05741
2025-01-01    4.05741
Freq: AS-JAN, dtype: float64
```

1.3.4 Short-term Interest Rate

```
1-step ahead forecast: 2024-01-01    3.937935
Freq: MS, dtype: float64
3-steps ahead forecast: 2024-01-01    3.937935
2024-02-01    3.949534
2024-03-01    3.952501
Freq: MS, dtype: float64
```

1.3.5 Electricity Daily Consumption

```
1-step ahead forecast: 2023-12-01    2.931038e+06
Freq: D, dtype: float64
3-steps ahead forecast: 2023-12-01    2.931038e+06
2023-12-02    2.929467e+06
2023-12-03    2.953747e+06
Freq: D, dtype: float64
```

2 Problem II

```
         Date Close/Last    Volume     Open     High      Low
0  02/22/2024    $785.38  86509970  $750.25  $785.75  $742.20
1  02/21/2024    $674.72  69029810  $680.06  $688.88  $662.48
```

We will observe only the Close/Last variable.

```
            Close/Last
Date
2024-02-22      785.38
2024-02-21      674.72
```

2.1 1.

Explain why it is better to work with (daily) log-returns.

It is often more convenient to work with daily logarithmic returns rather than raw price data because they have several desirable statistical properties that make them easier to analyze and model. Specifically:

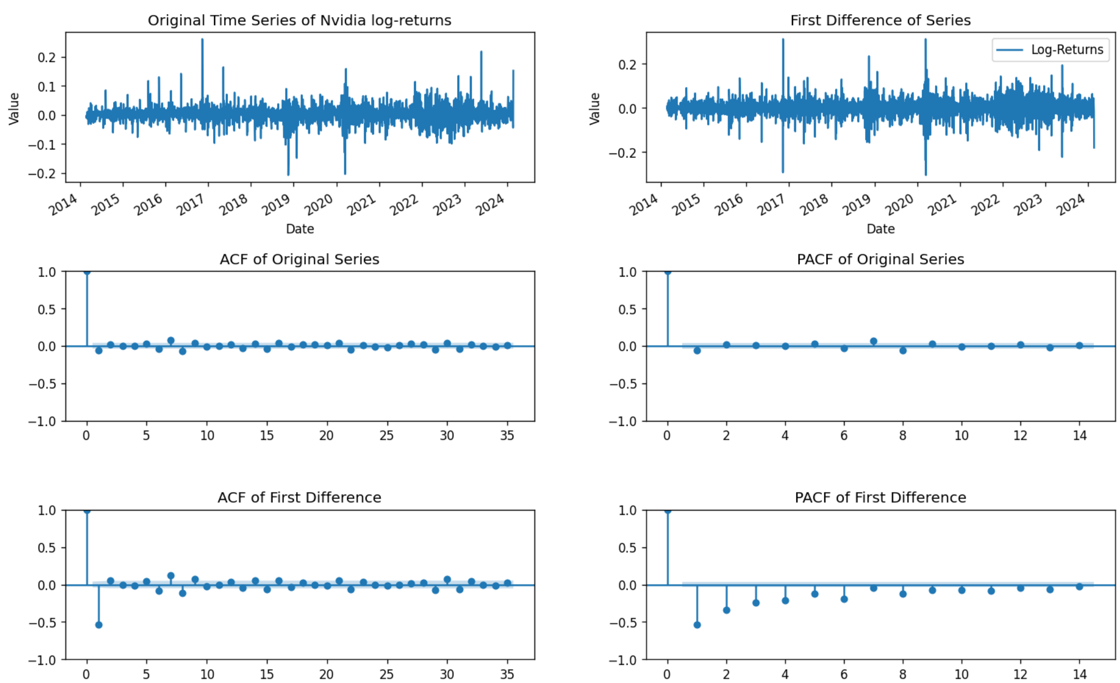
- Log-returns are more likely to exhibit stationary properties. Since log-returns tend to stabilize variance and exhibit less skewness and kurtosis compared to simple returns, they are more amenable to modeling with linear time series techniques, such as ARMA models.
- Financial time series often embody multiplicative processes, where prices today are a result of a percentage change applied to yesterday's prices. Log-returns directly represent these

percentage changes in a way that is additive, which aligns with the multiplicative nature of asset price movements.

- Financial time series often exhibit volatility clustering, where high-volatility events tend to be followed by high-volatility events, and low by low. Log-returns, due to their normalization effect, can make such patterns more evident and easier to model.
- Log-returns can be summed over time to obtain the cumulative return over a period, facilitating the analysis of returns over different time scales. This property is particularly useful in time series analysis, where we might need to aggregate daily returns into monthly or annual returns, or vice versa, to study seasonal effects, long-term trends, or to adjust the granularity of the analysis.

2.2 2.

```
            Close/Last  Log-Returns
Date
2014-02-26      4.6775    -0.001602
2014-02-25      4.6850    -0.009031
2014-02-24      4.7275          NaN
```




```
Descriptive Statistics for Original Series of Nvidia log-returns:
        Log-Returns
count   2516.000000
mean       0.002032
std        0.029205
min       -0.207712
```

```
25%      -0.011638
50%       0.002305
75%       0.016337
max       0.260876

Descriptive Statistics for First Difference of Nvidia log-returns Series:
        Log-Returns
count  2515.000000
mean     -0.000064
std       0.042508
min      -0.304967
25%      -0.020777
50%       0.000680
75%       0.021626
max       0.311400

ADF Test Statistic for Original Series of Nvidia log-returns:
-17.875207085858545
p-value for Original Series of Nvidia log-returns: 3.0169842823385467e-30
ADF Test Statistic for First Difference of Nvidia log-returns:
-16.925330807453236
p-value for First Difference of Nvidia log-returns: 9.869256218552703e-30
```

```
{'Mean': 0.0020321030029959684,
 'Variance': 0.0008529043167838703,
 'Skewness': 0.2708059885238492,
 'Kurtosis': 8.010386699674818,
 'ACF': array([ 1.        , -0.06412865,  0.02140111,  0.00276055, -0.0044299 ,
        0.02450937]),
 'PACF': array([ 1.        , -0.06415414,  0.01737389,  0.00525184, -0.00432397,
        0.02395187])}
```

- The time series looks stationary, which is conformed by the ADF test and its very low p-value
- The average daily log-return is positive (0.2%), indicating that, on average, the value of NVIDIA's stock has grown each day during the period analyzed. This positive mean suggests an overall upward trend in the stock price.
- The variance indicates the variability or volatility of the daily log-returns from the mean. A variance of this magnitude (0.09%) suggests that there has been moderate volatility in NVIDIA's stock prices during the period.
- A skewness value of 0.27 indicates a slight right skew, meaning there are more instances of returns exceeding the mean than falling below it. However, the skew is relatively mild, suggesting that the distribution of returns is not far from normal in terms of asymmetry.
- A kurtosis of 8.01 is significantly higher than 3, indicating a high likelihood of extreme returns (both positive and negative) compared to a normal distribution. This implies that investors could occasionally experience unusually large gains or losses.
- The ACF values suggest that there is very little autocorrelation in the log-returns, with the first lag showing a slight negative autocorrelation (-0.06) and the subsequent lags hovering around zero. This lack of strong autocorrelation suggests that past returns are not a reliable

predictor of future returns, a characteristic consistent with the efficient market hypothesis for stock prices.

- The PACF values also indicate minimal correlation, with values close to zero after the first lag. This further supports the notion that returns are largely unpredictable based on past values alone.

2.3 3.

Find the best AR model of the data.

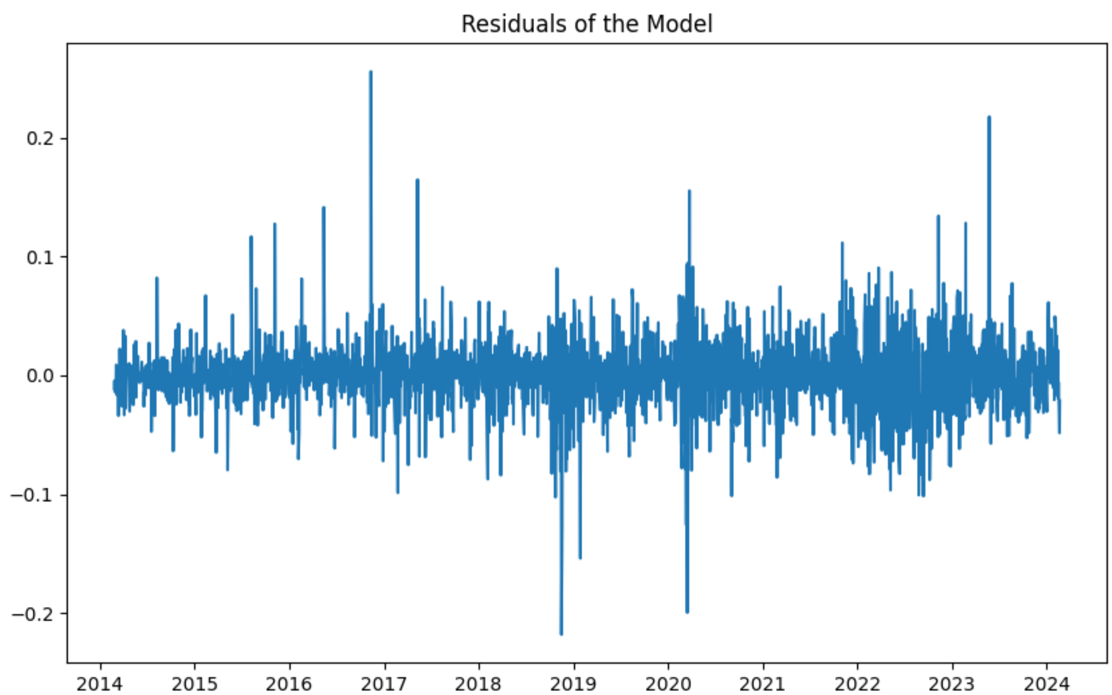
```
The best AR model of the log-returns is AR(1)
```

2.4 4.

Provide diagnostics.

2.4.1 Residual Analysis

Evaluating the residuals of the model to check if they behave like white noise.



The residuals look like a white noise process.

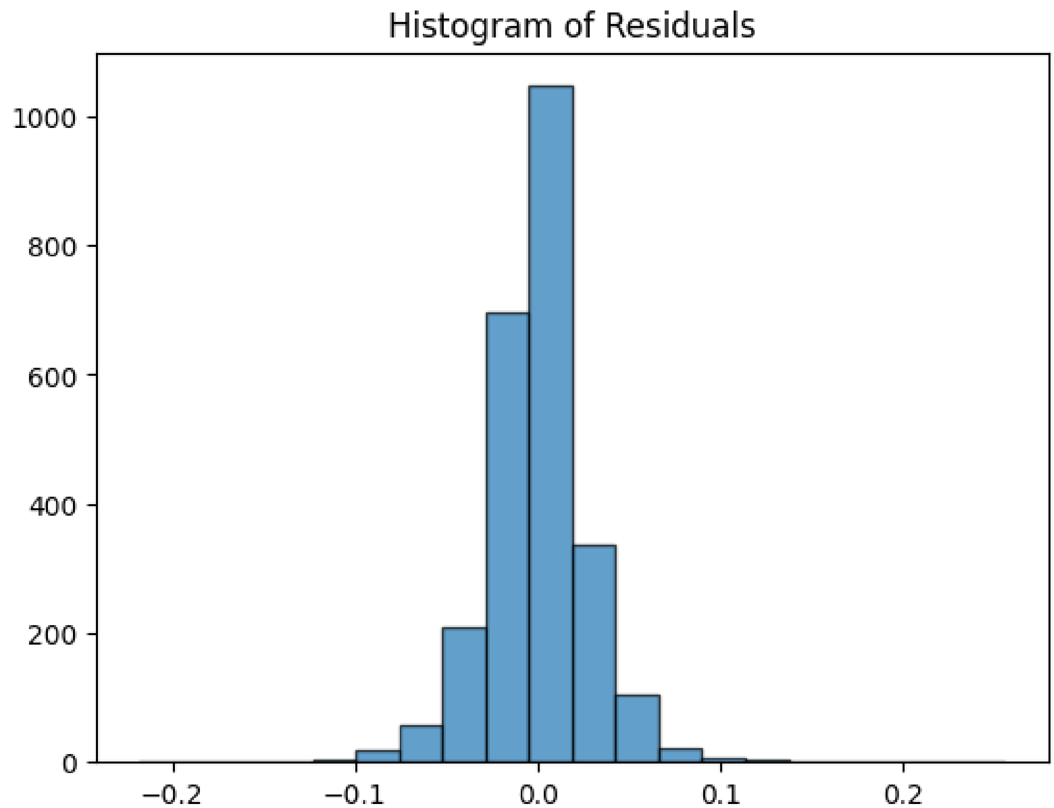
2.4.2 Stationarity of Residuals

```
ADF Statistic: -17.811618
p-value: 0.000000
Critical Values:
```

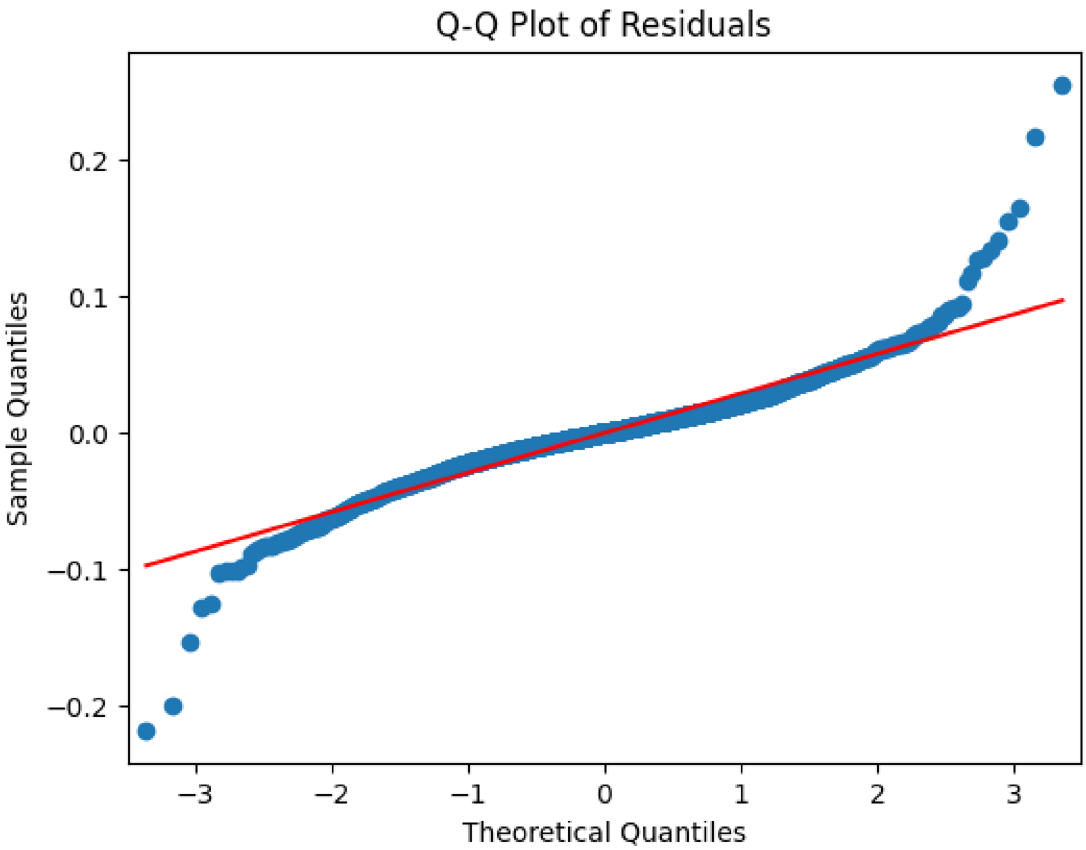
```
        1%: -3.433
        5%: -2.863
        10%: -2.567
The residuals are stationary, suggesting the model has captured underlying
trends and seasonality adequately.
```

2.4.3 Normality of Residuals

Histogram of residuals



Q-Q plot of residuals



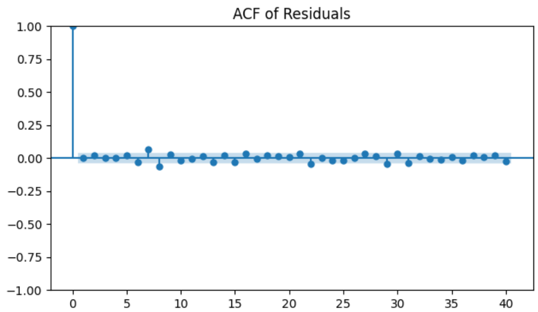
- The lower tail of the plot shows that the sample quantiles are higher than the theoretical quantiles of the normal distribution, indicating that the data has heavier tails than a normal distribution (i.e., it has more extreme negative values than expected under normality).
- The upper tail of the plot deviates above the line, which suggests that the data also has heavier tails on the positive side, meaning there are more extreme positive values than would be expected under normality.

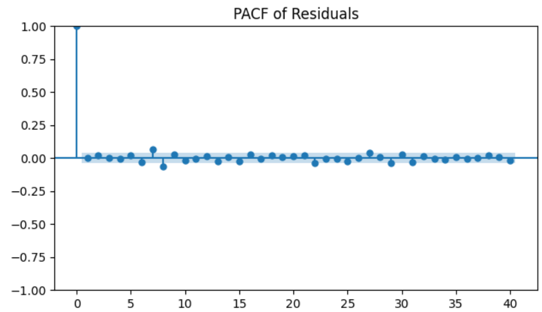
These deviations suggest that the residuals are not perfectly normally distributed, particularly with evidence of heavier tails, we can test it.

```
Jarque-Bera test p-value: 0.0
```

This p-value suggests that the residuals do not follow a normal distribution.

2.4.4 ACF and PACF of Residuals





It seems that there is no autocorrelation among the residuals. We can perform a test to assess this.

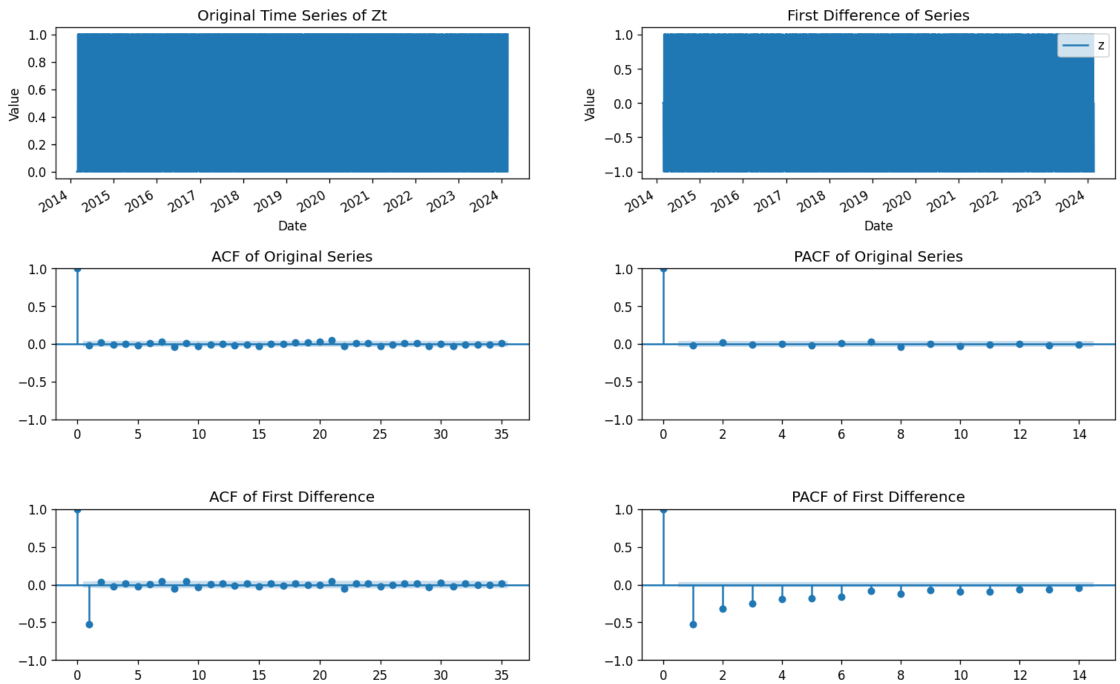
```
Ljung-Box test p-value: 0.001001256626280272
```

The very small p-value leads us to rejet the null hypothesis, i.e. we can conclude that there is significant evidence of autocorrelation in the residuals at one or more of the lags up to lag 10.

2.5 5.

Define the variable $z_t = 1$ when the return is positive, and $z_t = 0$. Do the same work (that you did for the return) for the process z_t.

```
            Close/Last  Log-Returns  z
Date
2024-02-22      785.38     0.151870  1
2024-02-21      674.72    -0.028923  0
2024-02-20      694.52    -0.044508  0
2024-02-16      726.13    -0.000620  0
2024-02-15      726.58    -0.016949  0
```




```
Descriptive Statistics for Original Series of Zt:
                  z
count  2517.000000
mean      0.541518
std       0.498372
min       0.000000
25%       0.000000
50%       1.000000
75%       1.000000
max       1.000000

Descriptive Statistics for First Difference of Zt Series:
                  z
count  2516.000000
mean     -0.000397
std       0.713126
min      -1.000000
25%      -1.000000
50%       0.000000
75%       1.000000
max       1.000000

ADF Test Statistic for Original Series of Zt: -51.34214803587893
p-value for Original Series of Zt: 0.0
```

```
ADF Test Statistic for First Difference of Zt: -16.806749823067154
p-value for First Difference of Zt: 1.208048375572964e-29
```

This time series is stationary.

2.5.1 Descriptive Statistics

```
{'Mean': 0.5415176797775129,
 'Variance': 0.2483749612334062,
 'Skewness': -0.16664621489002404,
 'Kurtosis': -1.9722290390628279,
 'ACF': array([ 1.        , -0.02374788,  0.02058922, -0.0134848 , -0.00115224,
        -0.01762376]),
 'PACF': array([ 1.        , -0.02375732,  0.0200525 , -0.01255691, -0.00217291,
        -0.01722938])}
```

54% of the log-returns are positive.

2.5.2 Best AR model

```
  0%|          | 0/1257 [00:00<?, ?it/s]
The best AR model of z is AR(1257)
```

- The process z_t is stationary.
- Since the ACF showed no autocorrelation in z_t, building an AR model is not relevant, the best process is the one that has the most lags…

3 Problem III

- model: $y_t = \beta x_{t-1} + u_t$ and $x_t = \rho x_{t-1} + v_t$, where u_t and v_t are error terms with specified covariance.
- x_0 should be drawn by the ergotic distribution of x_t. That is a $N(E[x_t], Var(x_t))$
  - $E[x_t] = E[\rho x_{t-1} + v_t]$

  Since x_{t-1} is independent of v_t, we can split the expectation:

  $$E[x_t] = \rho E[x_{t-1}] + E[v_t]$$

  Since the process is stationary, we assume that $E[x_t] = E[x_{t-1}] = \mu$. So, we rewrite the equation as:

  $$\mu = \rho\mu + E[v_t]$$

  Solving for μ, we get:

  $$\mu = \frac{E[v_t]}{1-\rho}$$

Since the process v_t is assumed to be zero-mean, then $E[v_t] = 0$, and $\mu = 0$.

- $\text{Var}[x_t] = \text{Var}[\rho x_{t-1} + v_t]$

Since x_{t-1} is independent of v_t, we can use the properties of variance to split the variance:

$$\text{Var}[x_t] = \rho^2 \text{Var}[x_{t-1}] + \text{Var}[v_t]$$

Since the process is stationary, we assume that $\text{Var}[x_t] = \text{Var}[x_{t-1}] = \sigma_x^2$. So, we rewrite the equation as:

$$\sigma_x^2 = \rho^2 \sigma_x^2 + \text{Var}[v_t]$$

Solving for σ^2, we get:

$$\sigma^2 = \frac{\text{Var}[v_t]}{1 - \rho^2}$$

3.1 1. Simulation

3.1.1 First Design: 1927-1996

```
Estimated beta: 0.17035156608813973
```

Interpretation: The value of β (0.17) is quite the same that the one we used to simulate the data (0.21)

Changing ρ to 0.5

```
Estimated beta: -0.32510176188149387
```

Changing ρ to 0

```
Estimated beta: -0.3584196017396177
```

Changing the value of ρ decrease significatively the value of our estimated β that is now negative around -0.35.

Changing σ_{uv} to 0

```
Estimated beta: 0.17035156608813973
```

Interpretation: it doesn't change much the value of our estimated β

Double the sample size

```
Estimated beta: 0.18187374198092954
```

Interpretation: doubling the sample size slightly increases the value of our estimated β, which is closer to the true value 0.21.

3.2 2. Forecasting

```
2-step ahead: OLS estimate of beta: 0.4114, Standard deviation: 0.0483
2-step ahead: Newey-West OLS estimate of beta: 0.4114, Standard deviation:
0.0503
4-step ahead: OLS estimate of beta: 0.2949, Standard deviation: 0.0547
4-step ahead: Newey-West OLS estimate of beta: 0.2949, Standard deviation:
0.0747
```

The standard deviations associated with the Newey-West estimates are higher than those of the OLS estimates. This increase in standard deviation for Newey-West adjustments reflects the method's more conservative approach in estimating uncertainty, accounting for autocorrelation and heteroskedasticity.

Now, to compare with the true analytical formula, we need to compute it.

Given the model equations:

$$y_t = \beta x_{t-1} + u_t,$$
$$x_t = \rho x_{t-1} + v_t,$$

where $[u_t, v_t]^T$ is distributed as $\mathcal{N}(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, we aim to derive the true analytical formula for the standard deviation of the OLS estimator, $\hat{\beta}$.

Step 1: Expressing $\hat{\beta}$

The OLS estimator for β is given by:

$$\hat{\beta} = \frac{\sum(x_{t-1} - \bar{x})(y_t - \bar{y})}{\sum(x_{t-1} - \bar{x})^2}.$$

For simplicity, assuming $\bar{x} = 0$ and $\bar{y} = 0$ leads to:

$$\hat{\beta} = \frac{\sum x_{t-1} y_t}{\sum x_{t-1}^2}.$$

Step 2: Variance of $\hat{\beta}$

Given that $y_t = \beta x_{t-1} + u_t$ and $x_t = \rho x_{t-1} + v_t$, we can write:

$$Var(\hat{\beta}) = E\left[\left(\frac{\sum x_{t-1} u_t}{\sum x_{t-1}^2}\right)^2\right].$$

Step 3: True Analytical Formula

Considering the autoregressive process of x_t and the error term variance, we approximate:

$$Var(\hat{\beta}) \approx \frac{\sigma_u^2}{n \cdot \frac{\sigma_v^2}{1-\rho^2}},$$

where n is the number of observations.

Then, we have \$Std() $\approx$ 0.0436\$

```
True Standard Deviation = 0.043588989435406726
```

- For the 2-step ahead forecast, both the OLS and Newey-West OLS standard deviations are close to the true standard deviation, with the OLS estimate being slightly closer. This shows that while both estimation methods provide reasonably accurate measures of dispersion, the OLS method might slightly underestimate the variability for this forecast horizon.
- For the 4-step ahead forecast, the standard deviation from the Newey-West OLS significantly exceeds the true standard deviation, indicating an overestimation of uncertainty. Meanwhile, the OLS estimate is closer but still over the true standard deviation, showing that both methods overestimate the variability, with Newey-West being more conservative.

3.3 3. Robert Shiller Data

```
        Date  S&P Compound  Dividends  Earnings    CPI  Date Fraction  \
0 1871-02-01          4.50       0.26       0.4  12.84        1871.13
1 1871-03-01          4.61       0.26       0.4  13.03        1871.21

   Long_interest_rate  Real price  Real Dividend Real_total_return_price  ...  \
0                5.32      107.25           6.20                  107.77  ...
1                5.33      108.27           6.11                  109.30  ...

   CAPE Unnamed: 13  TR CAPE  Unnamed: 15  Excess CAPE yield  \
0   NaN         NaN      NaN          NaN                NaN
1   NaN         NaN      NaN          NaN                NaN

   Montlhy total bond returns Real total bond returns  \
0                        1.0                     0.97
1                        1.0                     0.96

   10 year annualized stock real return  10 year annualized bond real return  \
0                                13.09%                                9.46%
1                                13.10%                                9.62%

  real 10 year excess annualized returns
0                                  3.62%
1                                  3.48%

[2 rows x 22 columns]

                            OLS Regression Results
==============================================================================
Dep. Variable:             Real price   R-squared:                       0.893
Model:                            OLS   Adj. R-squared:                  0.893
Method:                 Least Squares   F-statistic:                 1.530e+04
Date:                Wed, 28 Feb 2024   Prob (F-statistic):               0.00
Time:                        15:31:41   Log-Likelihood:                -12997.
No. Observations:                1829   AIC:                         2.600e+04
Df Residuals:                    1827   BIC:                         2.601e+04
Df Model:                           1
```

```
Covariance Type:            nonrobust
================================================================================
=
                   coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-
const          -542.1298     12.440    -43.580      0.000    -566.528
-517.732
Real Dividend    63.2054      0.511    123.697      0.000      62.203
64.208
==============================================================================
Omnibus:                      551.041   Durbin-Watson:                   0.023
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2144.602
Skew:                           1.425   Prob(JB):                         0.00
Kurtosis:                       7.474   Cond. No.                         43.9
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```