UNIVERSITY OF THE PHILIPPINES MINDANAO
College of Science and Mathematics
Department of Mathematics, Physics, and Computer Science

AMAT 132 (Introductory Forecasting) — Exercise 4: Time Series Properties

Saintjoy Mandalinao 2023-11993

General Objectives:
This exercise aims to help students:
1. Define an autocorrelation function (ACF) and PACF.
2. Transform a non-stationary series into an AR model.
3. Identify and estimate parameters of an AR model.
4. Define residuals and compute Akaike Information Criterion (AIC) for model selection.
5. Define and compute the Hurst Coefficient.

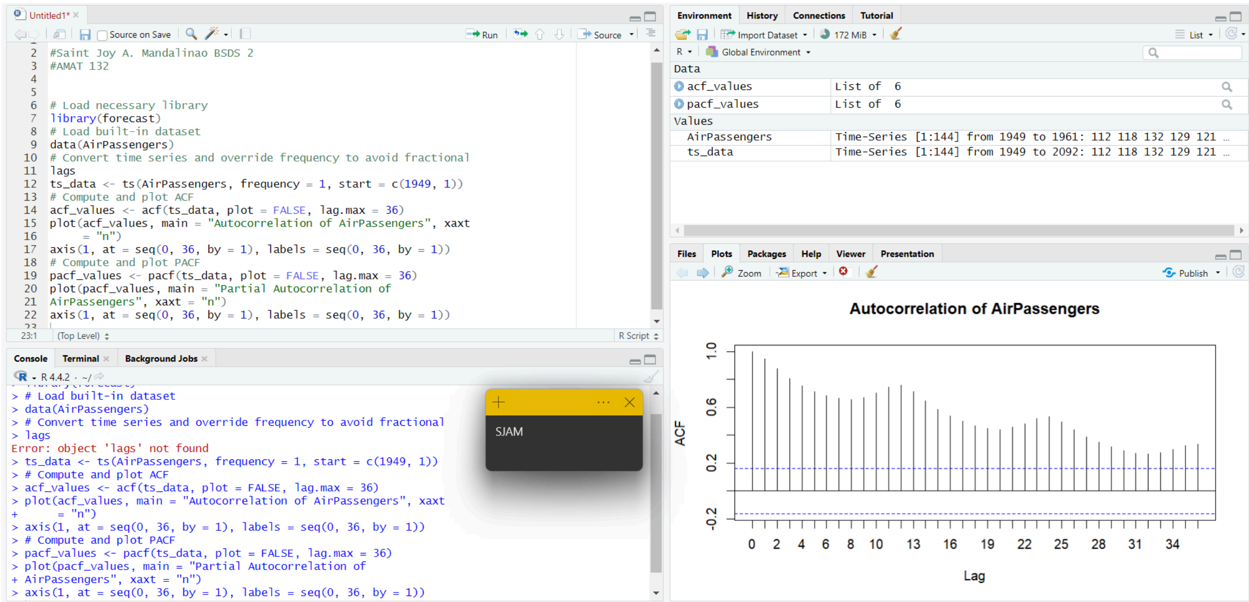
Exercise 4 a: Understanding Autocorrelation and PACF

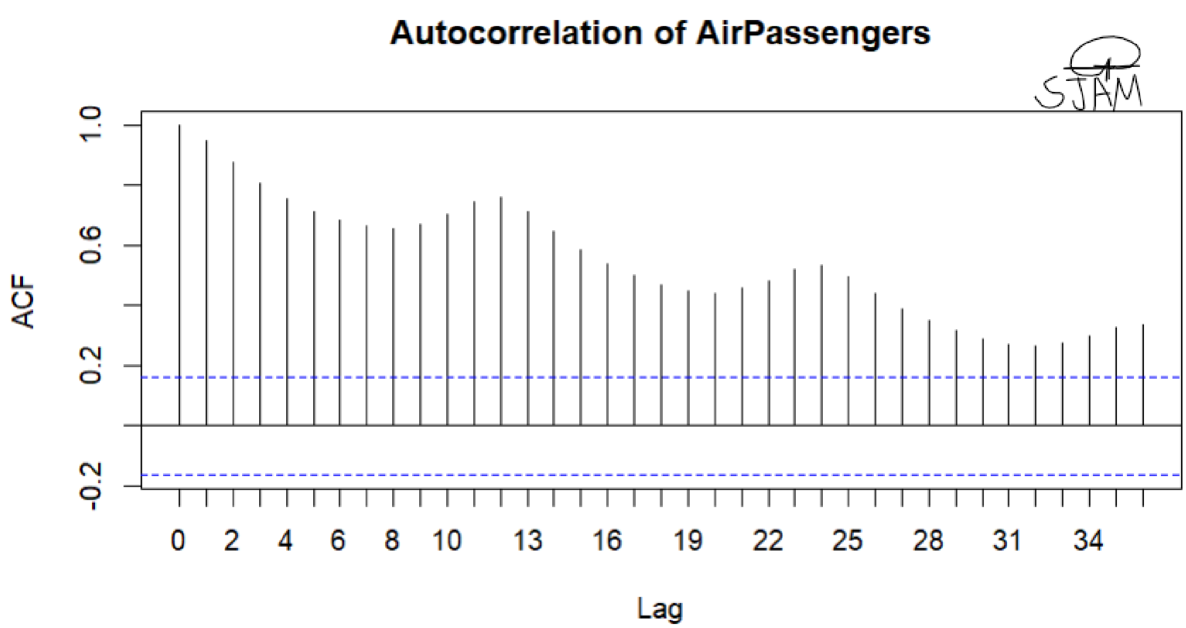
Objective:
Learn how to analyze autocorrelation and partial autocorrelation functions (PACF) of a time series.

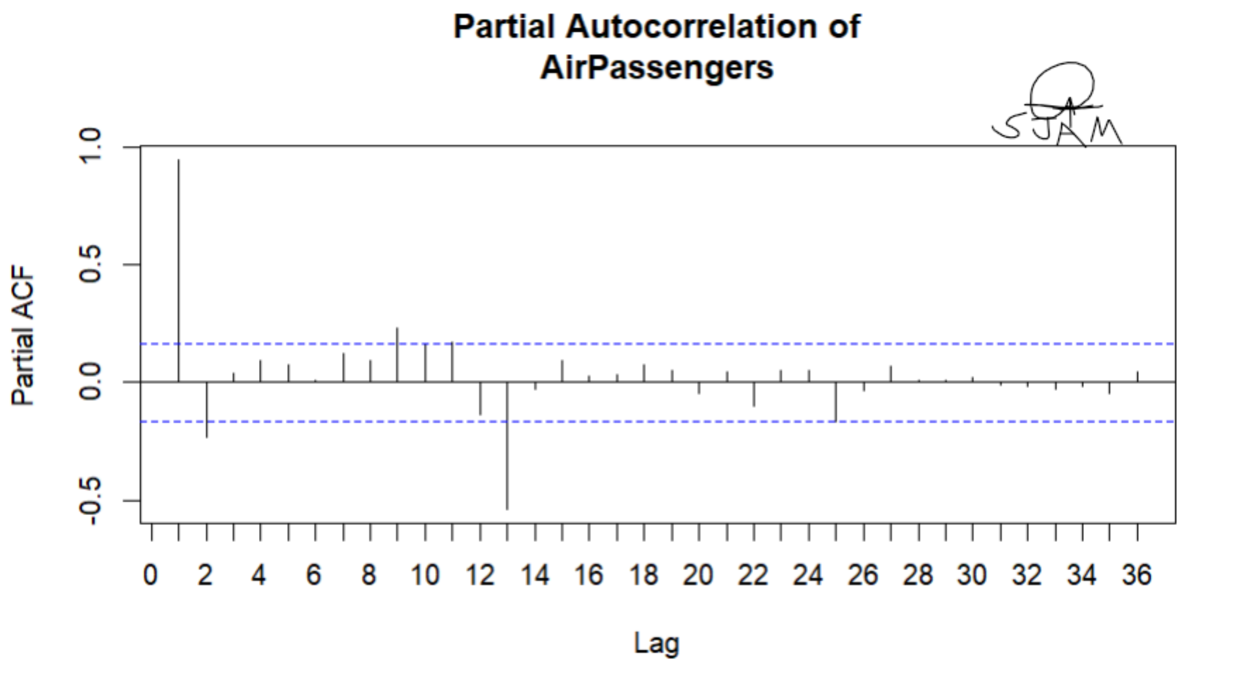
Instructions:
1. Load the AirPassengers dataset in R.
2. Plot the Autocorrelation Function (ACF) to identify serial correlation.
3. Plot the Partial Autocorrelation Function (PACF) to determine relevant lags.
4. Interpret the results:
   a. Identify significant lags in the ACF and PACF plots.
   b. Discuss how ACF and PACF help in selecting AR model parameters.

Discussion Questions:
1. What does the ACF plot tell you about serial correlation?
2. How does the PACF plot help identify the order of an AR model?
3. What is the effect of seasonality on autocorrelation?

Codes and results







Answers:

1. The Autocorrelation Function (ACF) plot shows strong serial correlation in the AirPassengers dataset, as the autocorrelation values decline gradually instead of dropping off quickly. This indicates that future values are highly dependent on past values, meaning the time series will more than likely be non-stationary and will need differencing. The fact that there is a notable spike at several lags also indicates that there is a seasonality, that is, the data displays periodic fluctuations across time. The gradual deterioration of autocorrelations also confirms that the series is long-term dependent and has trends.

2. The **Partial Autocorrelation Function (PACF) plot** helps identify the order of an **AutoRegressive (AR) model** by showing the **direct correlation** between a time series and its past values while controlling for intermediate lags. The PACF plot is also a useful instrument in the identification of the correct order of an AR model by indicating the correlation between a time series and its lagged variables after controlling for the impact of intermediate lags. In this plot, the PACF spikes sharply at lag 1 and then lower or zero values at later lags. This type of pattern would mean that an AR(1) model would be suitable since the PACF "cuts off" at lag 1. On the contrary, the occurrence of large spikes at lags 2 or 3 before tapering off would indicate the potential use of an AR(2) or AR(3) model.

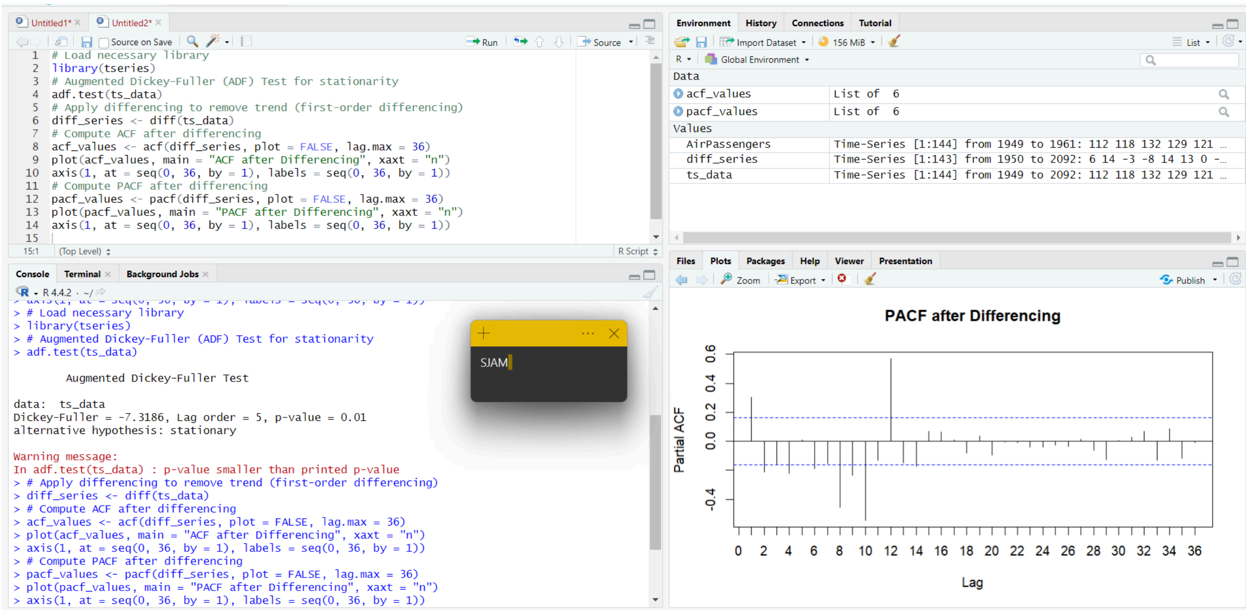
3. Seasonality plays a significant influence on autocorrelation, producing periodic patterns which cause large peaks at seasonal lags in the ACF graph. In this data, the cyclical nature of the strong correlations shows the presence of strong seasonal dependencies, i.e., previous values for the same season (e.g., airline passenger numbers in July of previous years) directly impact the current values. This would indicate that an ordinary ARIMA model might be inadequate, A standard ARIMA model can be unsuitable for seasonality time series since it assumes stationarity, which is violated by seasonal data because of repeated patterns. Normal differencing can eliminate short-term trends but will not remove the seasonal fluctuations, resulting in the autocorrelations at seasonal lags persisting. ARIMA also does not have explicit seasonal terms, hence cannot capture long-term dependencies from past seasonal cycles, which causes underfitting. To handle such problems, Seasonal ARIMA (SARIMA) or seasonal differencing is required, as they include seasonal autoregressive, moving average, and differencing terms that effectively describe and predict seasonal patterns.

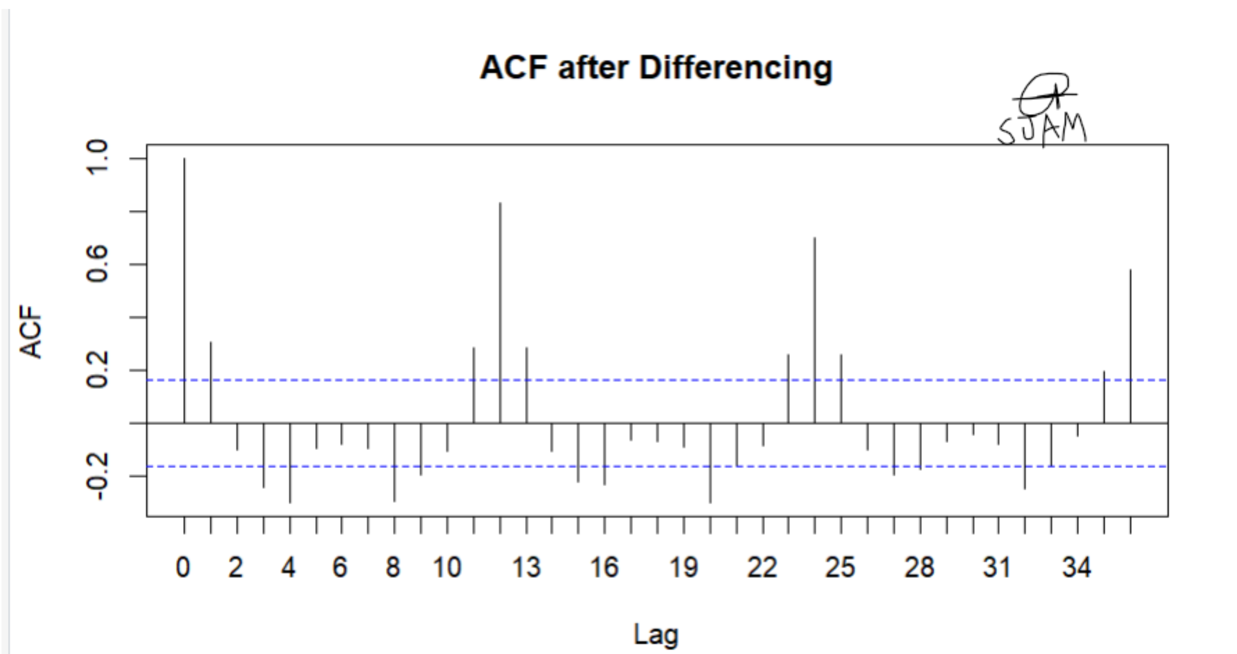
Exercise 4b: Stationarity Test and Transformation
Objective: Learn how to check for stationarity and apply transformations to make a time series stationary.

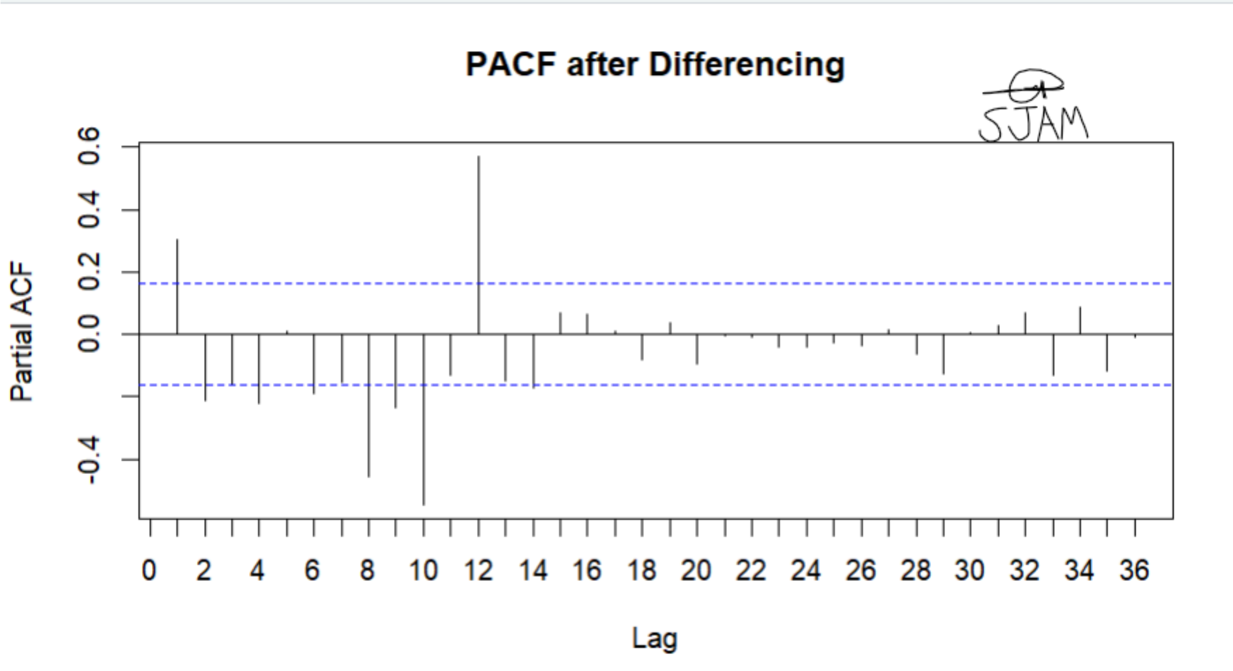
Instructions:
1. Check for stationarity using the Augmented Dickey-Fuller (ADF) test.
2. If the series is non-stationary, apply differencing to remove trends.
3. Compare the ACF and PACF plots before and after transformation.
4. Discuss how transformations affect the behavior of a time series.

Discussion Questions:
1. What does the ADF test result tell you about the stationarity of the data?
2. How did differencing change the time series?
3. Why is it important to make a series stationary before modeling?

Codes and results







Answers:

1. The Augmented Dickey-Fuller test result gives the statistical evidence as to whether or not the time series is stationarity. From the test above, the resulting Dickey-Fuller value is -7.3186 and the associated p-value = 0.01, and at a lag order of 5. With the p-value (0.01) less than 0.05, we reject the null hypothesis which is that the time series does not have a unit root, meaning it's not stationary. This outcome strongly indicates that the series is now stationary following differencing. The extreme negative Dickey-Fuller statistic of -7.3186 further supports this conclusion, as more negative values imply greater stationarity. This verifies that differencing was able to eliminate trends and stabilize the series so that it can be modeled using time series.

2. Differencing greatly changed the time series by eliminating trends and rendering it stationary. Prior to differencing, the ACF plot demonstrated persistent, strong autocorrelations for several lags, suggesting that previous values had a long-lasting effect on future values. This implied that the series was not stationary, with a trend towards an increase and perhaps seasonality as well. On differencing, the new ACF plot indicates that the majority of the autocorrelations are within the confidence interval except for seasonal peaks, implying that the series has stabilized with the passage of time. The PACF plot also indicates fewer significant lags, implying

short-run dependencies have been eliminated. Additionally, the ADF test result (Dickey-Fuller = -7.3186, p-value = 0.01) confirms that the series is now stationary, as the p-value is below 0.05, allowing us to reject the null hypothesis of a unit root. This means the time series now has a more consistent mean and variance, making it suitable for further analysis.

3. It is important to ensure that a time series is stationary for proper forecasting because most statistical models assume stable patterns in the data. A stationary series has a constant mean, variance, and autocovariance over time, which makes analysis more trustworthy and enhances the precision of parameter estimation. Without stationarity, variable relationships can change unpredictably, and this makes predictions unreliable. By differencing and checking stationarity using the ADF test and ACF/PACF plots, we establish a more stable data set which is more amenable to useful analysis and long-run forecasting.

Exercise 4c: Estimating an Autoregressive (AR) Model

Objective: Learn how to build an Autoregression (AR) model, estimate its parameters, and evaluate its performance.

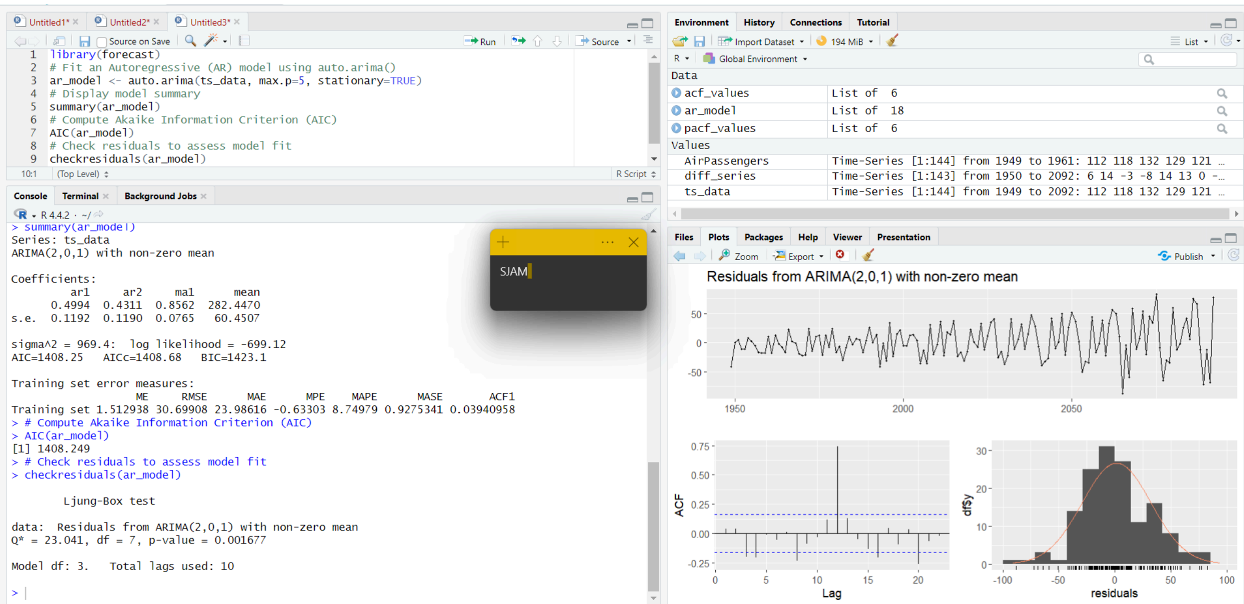
Instructions:

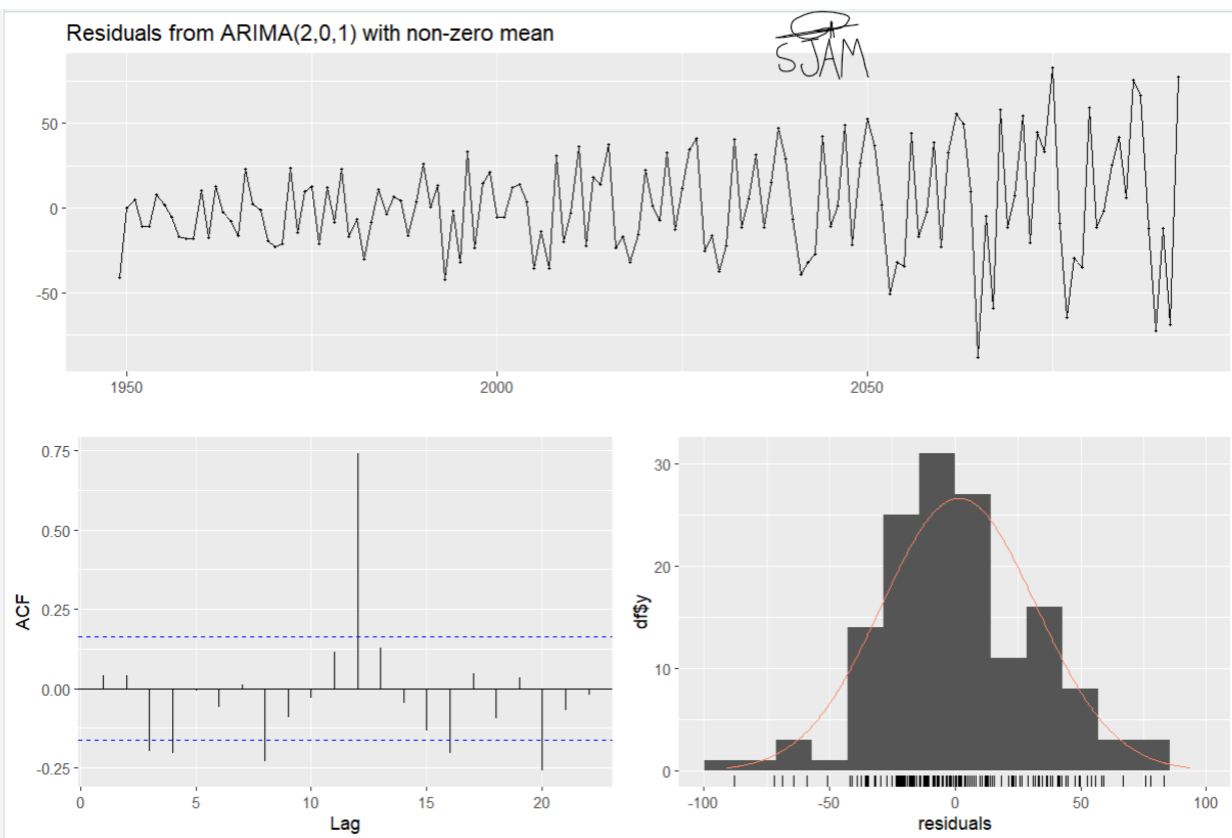
1. Fit an Autoregressive (AR) model to a stationary time series.
2. Use the Akaike Information Criterion (AIC) to select the best AR order. Unset
3. Analyze the model's residuals to assess its fit.
4. Interpret the estimated coefficients and discuss the validity of the model

Discussion Questions:

1. What AR order does the model select based on AIC?
2. How do residuals help assess the goodness of fit of an AR model?
3. How does the PACF plot relate to the AR model order?

Codes and results





1. Akaike Information Criterion (AIC) is a significant criterion for finding the best order for an AutoRegressive (AR) model by striking a balance between goodness of fit and model complexity. The model presented here is ARIMA(2,0,1), i.e., autoregressive order (p) is 2, therefore two previous observations are used to predict future values. AIC allows models to be compared, and the model having the lowest AIC value is generally preferred as it optimizes the trade-off between complexity and accuracy. If an AR model is selected based on AIC, it would imply that an AR(2) setup best fits the data while ensuring the risk of overfitting is minimized.

2. Residuals are important in determining the goodness of fit of an AR model. Residuals should ideally be like white noise, i.e., randomly scattered around zero with no apparent patterns or autocorrelation. If residuals contain structure, trends, or high levels of autocorrelation, it indicates that the model has not captured the underlying time series dynamics. In the residual diagnostics provided, the ACF of residuals has a huge spike at lag 10, which means that the model still has some unexplained patterns, indicating scope for improvement.

3. The Partial Autocorrelation Function (PACF) plot is an important tool in finding the right AR order (p) for an AR model. The PACF captures the direct relationship between a time series and its lagged values after removing the effects of intermediate lags. Generally, the AR order is chosen by the lag where the PACF plot truncates (i.e., where nonzero values fall to close to zero). If the PACF indicates high correlations at lags 1 and 2 but declines afterward, it means that an AR(2) model would be suitable, corresponding to the ARIMA(2,0,1) model chosen.

Exercise 4d: Computing the Hurst Coefficient
Objective: Measure the long-term memory of a time series using the Hurst Coefficient (H).
Instructions:
1. Compute the Hurst Exponent to analyze trends vs. mean-reverting behavior.
2. Interpret whether the series is persistent, mean-reverting, or random walk.

Hurst Coefficient Interpretation:
- $H = 0.5 \rightarrow$ Random Walk (no long-term memory).
- $H > 0.5 \rightarrow$ Persistent trend (long-term influence).

● H < 0.5 → Mean-reverting behavior.

Discussion Questions:

1. What does your Hurst coefficient value suggest about the time series?
2. How does the Hurst exponent help in selecting forecasting models?

Codes and results

```
install.packages("pracma")
library(pracma)
# Convert time series to numeric for Hurst exponent calculation
hurst_value <- hurstexp(as.numeric(ts_data))
print(hurst_value)

```

6:1 (Top Level) ÷ R Script

Console Terminal × Background Jobs ×

R 4.4.2 · ~/

```
The downloaded binary packages are in
	C:\Users\saint\AppData\Local\Temp\Rtmp4iwXMz\downloaded_packages
> library(pracma)
Warning message:
package ‘pracma’ was built under R version 4.4.3
> # Convert time series to numeric for Hurst exponent calculation
> hurst_value <- hurstexp(as.numeric(ts_data))
Simple R/S Hurst estimation:        0.8206234
Corrected R over S Hurst exponent:  1.018389
Empirical Hurst exponent:           0.7583141
Corrected empirical Hurst exponent: 0.7685706
Theoretical Hurst exponent:         0.5263576
> print(hurst_value)
$Hs
[1] 0.8206234

$Hrs
[1] 1.018389

$He
[1] 0.7583141

$Hal
[1] 0.7685706

$Ht
[1] 0.5263576
```

SJAM

Answers:

1. Based on the Hurst exponent values obtained from the calculation, the time series exhibits strong long-term dependence. The Simple R/S Hurst estimation (0.8206) and the Corrected empirical Hurst exponent (0.7686) suggest a persistent trend, as both values are significantly greater than 0.5. This indicates that the time series is not a random walk but instead follows a pattern where past trends are likely to continue in the future.

2. The high Empirical Hurst exponent (0.7583) and Corrected R over S Hurst exponent (1.0184) reflect intense memory effects, and the overall trend-following characteristic of the data outweighs. Although the Theoretical Hurst exponent (0.5264) is nearer to

randomness, the overall trend-following nature of the data is prevalent. This means that long-term dependencies should be included in forecasting models, since stationary or mean-reverting models (e.g., AR models) might not be as good, while trend-following models capture the series' long-range correlations more accurately.