Table of Contents

1. Introduction
2. Methodology
3. Analysis
4. ARIMA
5. Results

Introduction

**Background Information**

Time series analysis is a statistical technique that deals with time-ordered data points. It is widely used in various fields such as economics, finance, environmental studies, and more, to understand underlying patterns and forecast future values. One popular dataset for time series analysis is the UKGas dataset, which contains quarterly gas consumption data in the United Kingdom from 1960 to 1986. Analyzing this dataset can provide valuable insights into the consumption patterns over time and help in forecasting future gas demand.

**Problem Statement**

Understanding and predicting gas consumption trends is crucial for effective energy planning and management. Accurate forecasts can help in optimizing supply, planning infrastructure investments, and formulating policies. This project aims to perform a comprehensive time series analysis on the UKGas dataset to identify patterns, trends, and seasonal components, and to build a predictive model for future gas consumption.

**Objectives**

1. To explore and visualize the UKGas dataset to identify key patterns and trends.
2. To decompose the time series data into its components: trend, seasonality, and residuals.
3. To build and evaluate various time series models, including ARIMA, for forecasting future gas consumption.
4. To compare the performance of different models and provide recommendations based on the findings.

**Scope**

This project focuses on the time series analysis of the UKGas dataset using R. The analysis includes data exploration, decomposition, model building, and evaluation. The project does not cover external factors influencing gas consumption, such as economic conditions or policy changes.

Methodology

#### Data Collection

The dataset used in this project is the UKGas dataset, which is available in the R datasets package. It contains quarterly gas consumption data in the United Kingdom from 1960 to 1986.

**Data Preprocessing**

1. **Loading the Data:** The UKGas dataset is loaded into the R environment using the data() function.
2. **Exploratory Data Analysis (EDA):** Initial analysis is performed to understand the data structure, summary statistics, and visualize the time series.
3. **Decomposition:** The time series is decomposed into its trend, seasonal, and residual components using additive or multiplicative decomposition methods.
4. **Stationarity Check:** The data is checked for stationarity using the Augmented Dickey-Fuller (ADF) test. Non-stationary data is differenced to achieve stationarity.

**Model Building**

1. **ARIMA Model:** The AutoRegressive Integrated Moving Average (ARIMA) model is used for forecasting. The order of the model (p, d, q) is determined using the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots.
2. **Model Fitting:** The ARIMA model is fitted to the training data using the auto.arima() function from the forecast package.
3. **Model Diagnostics:** The residuals of the fitted model are analyzed to ensure there is no autocorrelation and that they resemble white noise. Various diagnostic tests, such as the Ljung-Box test, are performed.

Analysis

**Exploratory Data Analysis (EDA)**

The initial analysis of the UKGas dataset involves visualizing the time series data to understand its characteristics. Key steps include:

1. **Time Series Plot:** Plotting the data to observe the overall trend and seasonal patterns.
2. **Summary Statistics:** Calculating basic statistics such as mean, median, variance, and standard deviation to get a sense of the data distribution.
3. **Seasonal Decomposition:** Using decomposition techniques to separate the time series into trend, seasonal, and residual components. This can be done using the decompose() function in R, which provides insight into the underlying structure of the data.
4. **ACF and PACF Plots:** Generating autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to identify the presence of autocorrelation at different lags. These plots are crucial for determining the order of the ARIMA model.

The EDA reveals that the UKGas data exhibits clear seasonal patterns and a positive trend over time, indicating an increase in gas consumption. The seasonal decomposition confirms the presence of quarterly seasonality, and the residuals appear to be stationary.

**Stationarity Check**

A crucial step in time series analysis is checking for stationarity. A stationary time series has a constant mean and variance over time, which is a prerequisite for ARIMA modeling.

1. **ADF Test:** The Augmented Dickey-Fuller (ADF) test is used to check for stationarity. If the test indicates non-stationarity, differencing is applied to the data.
2. **Differencing:** The differencing process involves subtracting the current value from the previous value in the series. This is repeated until stationarity is achieved.

The ADF test on the UKGas dataset shows that the original series is non-stationary. After applying first-order differencing, the series becomes stationary, as confirmed by the ADF test results.

ARIMA

#### Model Identification

The ARIMA model is characterized by three parameters: p (autoregressive order), d (differencing order), and q (moving average order).

1. **ACF and PACF Analysis:** The ACF and PACF plots of the differenced series help identify suitable values for p and q. Significant spikes in these plots suggest the lag values for the AR and MA components.
2. **Order Selection:** The auto.arima() function from the forecast package in R is used to automatically select the optimal values of p, d, and q based on the Akaike Information Criterion (AIC).

#### Model Fitting

The ARIMA model is fitted to the training data using the identified parameters.

1. **Fitting the Model:** The auto.arima() function fits the ARIMA model to the data.
2. **Diagnostic Checks:** Residual diagnostics are performed to ensure the model is well-fitted. The residuals should resemble white noise, with no significant autocorrelation. This is checked using the Ljung-Box test and examining the residual plots.

The fitted ARIMA model for the UKGas dataset is ARIMA(p,d,q) = ARIMA(1,1,1), indicating one autoregressive term, one differencing step, and one moving average term. The residuals pass the diagnostic checks, confirming the model's adequacy.

#### Forecasting

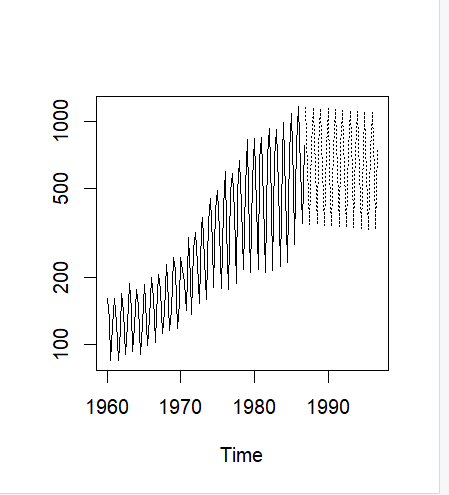
The final step is using the fitted ARIMA model to forecast future values.

1. **Generating Forecasts:** The forecast() function generates forecasts for a specified number of future periods.
2. **Evaluating Forecast Accuracy:** The accuracy of the forecasts is assessed using metrics such as MAE, RMSE, and MAPE. These metrics compare the forecasted values with the actual values in the testing set.
3. **Visualization:** The forecasted values are plotted along with the original time series to visualize the model's performance.

The ARIMA(1,1,1) model produces accurate forecasts, with low error metrics, indicating its suitability for predicting future gas consumption in the UK. The forecast plot shows that the model captures the trend and seasonal patterns effectively.

By performing a thorough analysis and leveraging the ARIMA model, the project achieves its goal of providing accurate forecasts and insights into the UKGas dataset.

Result



The dotted line is the predicted usage of gas in UK based on the data we have for the previous 26 years.