ADF Test

Diana Akolzina

May 26, 2023

1 Introduction

The ADF Test code performs analysis on a data file using time series techniques to determine the stationarity of a time series. The Augmented Dickey-Fuller (ADF) test is utilized, and the code provides summary statistics and test results including the ADF statistic, p-value, and critical values.

2 Prerequisites

Make sure you have the following libraries installed:

- numpy
- matplotlib
- pmdarima
- pandas
- statsmodels

3 Usage

Place the data file in the same directory as the code or provide the full path as a string in the pd.read-csv method. Run the code to analyze the data and obtain the ADF test results. Ensure that the data is formatted with a date column and a corresponding data column.

4 Code Explanation

The code begins by loading the data from the specified file using pandas. The date column is parsed as a datetime type and set as the index of the dataframe.

```python
df = pd.read_csv('data.csv')
df['DateTime'] = pd.to_datetime(df['DateTime'])
df = df.set_index('DateTime')
```

Next, the data is split into training and test sets, with the training set containing 80 percent of the data.

```python
train_size = int(len(df) * 0.8)
train, test = df[0:train_size], df[train_size:len(df)]
```

The ADF test is then performed on the training data to determine the stationarity of the time series. The ADF statistic, p-value, and critical values are calculated.

```python
result = adfuller(train['Data'])
adf_statistic = result[0]
p_value = result[1]
critical_values = result[4]
```

The summary statistics of the training data are printed, followed by the ADF test results.

```
1  print('Summary Statistics for the Training Data:')
2  print(train.describe())
3
4  print('\nAugmented Dickey-Fuller Test:')
5  print(f'ADF Statistic: {adf_statistic}')
6  print(f'p-value: {p_value}')
7  print('Critical Values:')
8  for key, value in critical_values.items():
9  print(f'\t{key}: {value}')
```

If the ADF statistic is lower than the critical value at a given significance level (e.g., 0.05), the null hypothesis of non-stationarity is rejected, indicating that the time series is stationary. Conversely, if the ADF statistic is higher, the null hypothesis is not rejected, suggesting that the time series is not stationary.

5   Result Interpretation

The ADF test helps in understanding the stationarity of the time series data. Stationarity is a desirable property for many time series analysis techniques as it allows for more reliable forecasting and analysis. One of these techniques is SARIMA.

By analyzing the ADF statistic, p-value, and critical values, you can determine whether the data exhibits stationarity or not. The summary statistics of the training data provide an overview of the data's distribution and central tendencies.