Forecasting Commercial Electricity Consumption in the Agriculture Sector of Pakistan: Time Series Analysis

By

Muhammad Abdullah Arshad

Data Analyst

Project Report

Date: 26 August 2025

Phone: +92 303 4292979

Email: malikabdullah5151@gmail.com

LinkedIn: www.linkedin.com/in/muhammad-abdullah-arshad-3966a5245

Contents

1 Introduction

Electricity plays a very important role in the social and economic development of every country. It is not only required for industries and households, but also for agriculture, which is the backbone of Pakistan's economy. Agriculture in Pakistan provides employment to a large part of the population, supports rural communities, and contributes to national income. For this sector, the availability of electricity is crucial. Farmers depend on electricity for running tube wells, irrigation systems, processing of crops, storage facilities, and in some cases for operating modern agricultural machinery. With the rapid increase in population and food demand, the dependence of agriculture on commercial electricity is continuously growing. However, the electricity supply in Pakistan is often unstable and demand is much higher than the available resources. This imbalance creates major challenges for both farmers and policy makers.

In this situation, it becomes very important to study and forecast the future demand of electricity consumption in the agricultural sector. Forecasting helps to understand the trend of usage over time and gives policymakers, electricity providers, and farmers an idea of how much energy will be required in the future. Time series analysis is one of the most useful statistical tools for this purpose. It allows researchers to examine past consumption patterns, seasonal variations, and long-term trends in order to make predictions. By applying time series models, future electricity needs can be estimated with better accuracy. This information can support the government in energy planning, ensure efficient distribution, and reduce electricity shortages. For Pakistan, where agriculture and energy are both key sectors, such forecasting is not only useful but necessary for sustainable development. Therefore, this study focuses on forecasting commercial electricity consumption in the agriculture sector of Pakistan using time series analysis techniques.

2 Data Description

The data used in this study is secondary in nature. It has been collected from the *Economic Survey of Pakistan*, which is one of the most reliable sources of official statistics on the economic situation of the country. The data covers a long time period starting from the year 1971 and ending in 2023. Thus, the dataset consists of more than five decades of information, which makes it suitable for time series analysis. Since the study focuses on the agricultural sector, the main variable of interest is the **commercial electricity consumption in the agriculture sector of Pakistan**. This variable reflects how much electricity has been used in the agricultural sector, which is highly relevant because electricity plays an important role in irrigation, tube wells, machinery operation, and other agricultural activities.

Using this long time series dataset is very important because the agricultural sector of Pakistan has undergone significant changes over the years. By analyzing the electricity consumption in agriculture, we can also gain an understanding of how the energy needs of the sector have evolved with the changing policies, technology adoption, and overall economic growth. The dataset is also continuous and annual in nature, which makes it appropriate for the application of different time series methods. In addition to the main variable, background information from the Economic Survey is also used to provide better interpretation and context for the findings of this study.

3 Methodology

In order to analyze the commercial electricity consumption in the agriculture sector, a systematic time series approach has been applied. The first step of the methodology is to conduct Exploratory Data Analysis (EDA). In this step, the data is summarized and visualized through descriptive statistics, graphs, and tables. A time plot of the series is drawn to observe the long-term trend, seasonal patterns, or any sudden fluctuations in electricity consumption over the years. This initial exploration helps in understanding the overall behavior of the data.

After visualization, the next step is to check the stationarity of the time series. Stationarity means that the statistical properties of the series, such as mean and variance, remain constant over time. Since many time series models require stationarity, tests such as the Augmented Dickey-Fuller (ADF) test are used. If the data is found to be non-stationary, differencing is applied to make it stationary.

The Autoregressive Integrated Moving Average (ARIMA) model is used for the analysis. The ARIMA model is written as:

$$ARIMA(p,d,q): \quad \phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t$$

where Y_t is the time series, $\phi(B)$ is the autoregressive (AR) part with order p, $(1-B)^d$ represents the differencing to achieve stationarity, $\theta(B)$ is the moving average (MA) part with order q, and ε_t is the error term.

The model is estimated by identifying the appropriate values of p, d, and q using the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. In this study, the analysis is performed using the Python programming language, which provides powerful libraries for statistical modeling. Specifically, the `auto_arima` function from the `pmdarima` library is employed. This function automatically selects the best combination of parameters (p,d,q) by testing different models and minimizing information criteria such as AIC and BIC. This approach ensures efficiency, accuracy, and avoids the subjectivity involved in manual parameter selection.

After estimating the best-fitting ARIMA model, diagnostic checks are performed to ensure that the residuals behave like white noise. Finally, the selected ARIMA model is used for forecasting future electricity consumption in the agriculture sector of Pakistan. Forecasts provide useful insights for policymakers to plan energy allocation and improve agricultural productivity in the future.

4 Data Analysis and Results

4.1 Import Required Libararies

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.stats.stattools import jarque_bera
```

```
import scipy.stats as stats
```

4.2 Import and Explore and Data

[2]:
```
df = pd.read_csv("C:\\Users\\abdul\\Downloads\\Agriculture Electricity␣
↪Consumption.csv")
ts = df['Electricity, Agricultural (Gwh)']
df["Fiscal Year"] = df["Fiscal Year"].astype(int)
df.head()
```

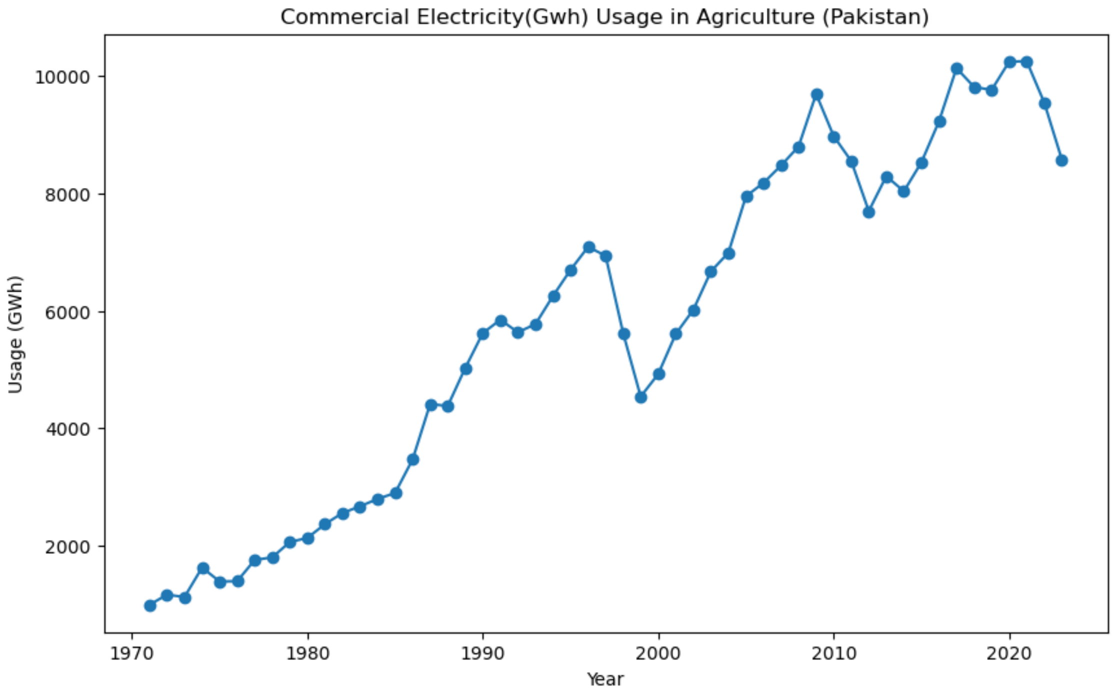
[2]:

	Fiscal Year	Electricity, Agricultural (Gwh)
0	1971	997
1	1972	1170
2	1973	1131
3	1974	1631
4	1975	1395

4.3 Exploratory Data Analysis (EDA) and Visualization of Time Series

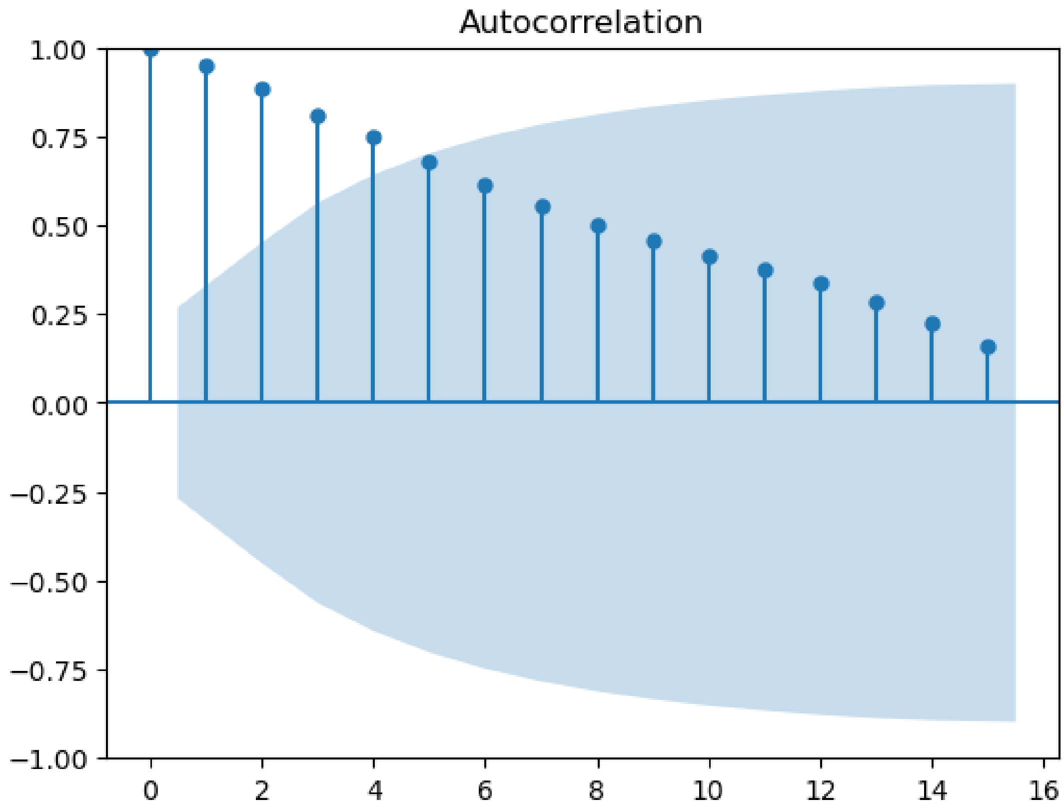
[3]:
```
print("Mean:", ts.mean())
print("Median:", ts.median())
print("Standard Deviation:", ts.std())
print("Minimum:", ts.min())
print("Maximum:", ts.max())


plt.figure(figsize=(10,6))
plt.plot(df["Fiscal Year"], df["Electricity, Agricultural (Gwh)"], marker='o')


plt.title("Commercial Electricity(Gwh) Usage in Agriculture (Pakistan)")
plt.xlabel("Year")
plt.ylabel("Usage (GWh)")
plt.show()
```

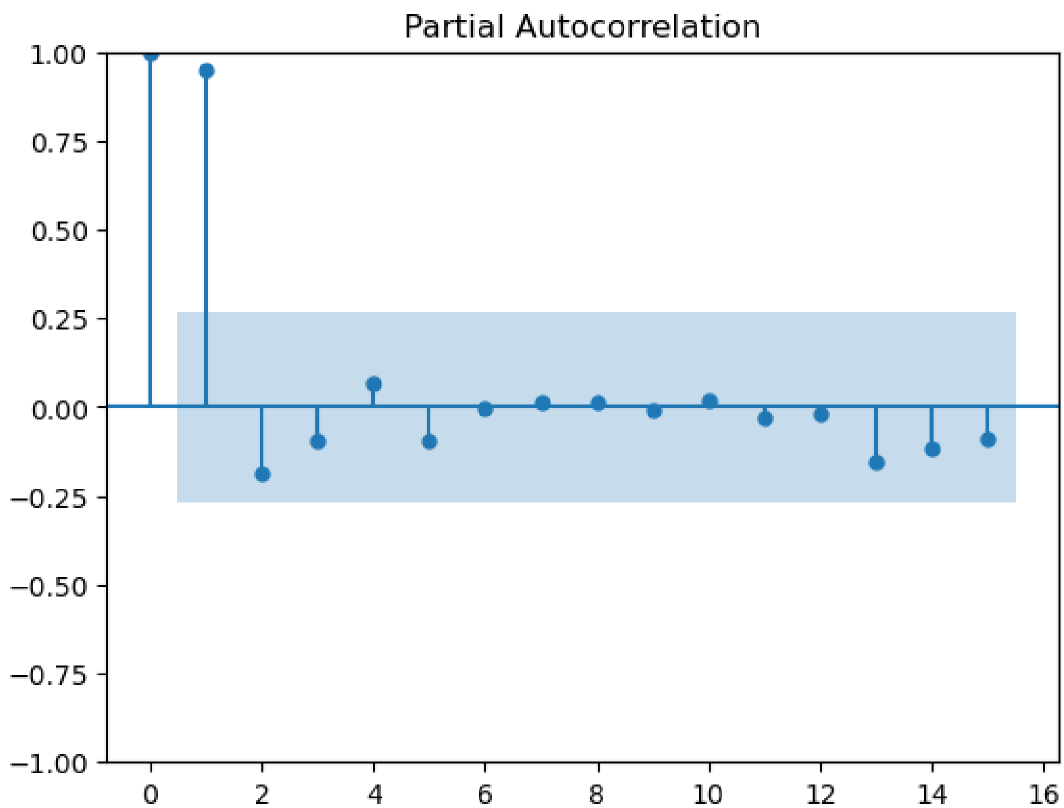
```
Mean: 5791.622641509434
Median: 5847.0
Standard Deviation: 2957.636557913814
Minimum: 997
Maximum: 10247
```



4.4 ACF and PACF

[4]:
```
plot_acf(ts, lags=15)
plt.show()

plot_pacf(ts, lags=15)
plt.show()
```


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
2
4
6
8
10
12
14
16



4.5 Checking Stationarity

[5]:
```
result = adfuller(ts)
print("ADF Statistic:", result[0])
print("p-value:", result[1])
if result[1] > 0.05:
    print("Series is Non-Stationary. Applying differencing...")
    tsdiff = ts.diff().dropna()
else:
    print("Series is Stationary. No differencing needed.")
```

```
ADF Statistic: -1.422423197095101
p-value: 0.5714802066153026
Series is Non-Stationary. Applying differencing...
```

4.6 Fit ARIMA Model

[6]:
```
from pmdarima import auto_arima
model_auto = auto_arima(
    ts,
    seasonal=False,
```

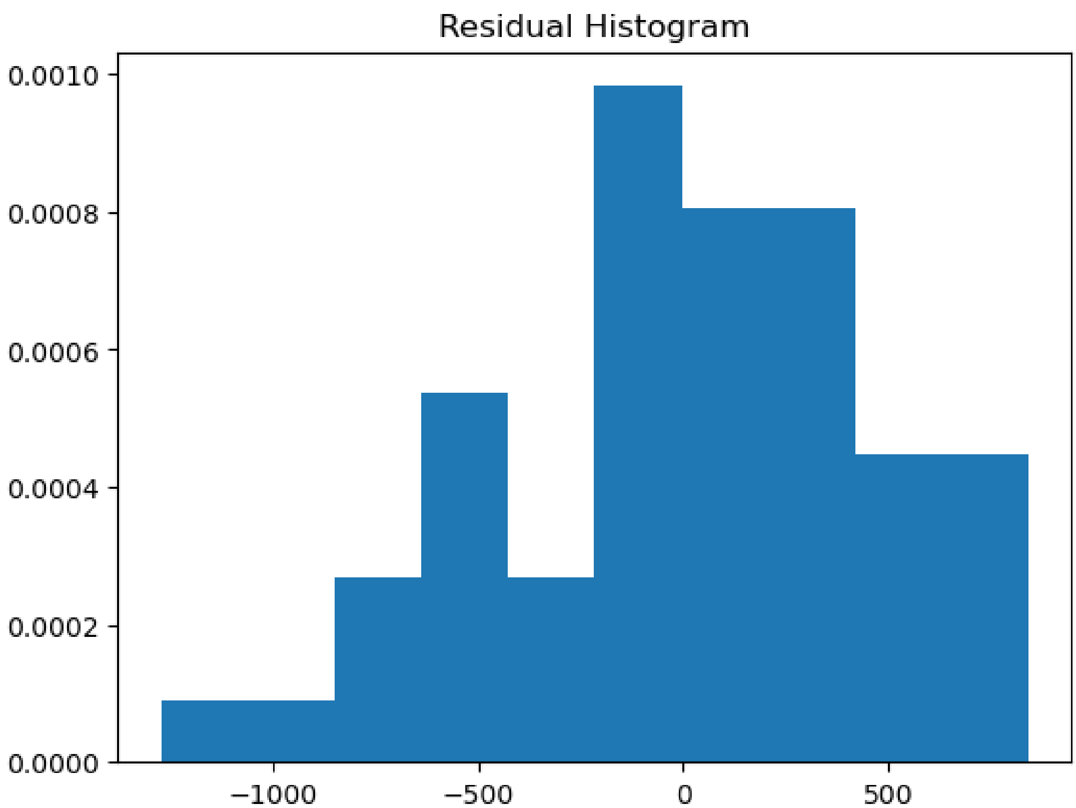
```
    stepwise=True,
    suppress_warnings=True
)
print(model_auto.summary())
```

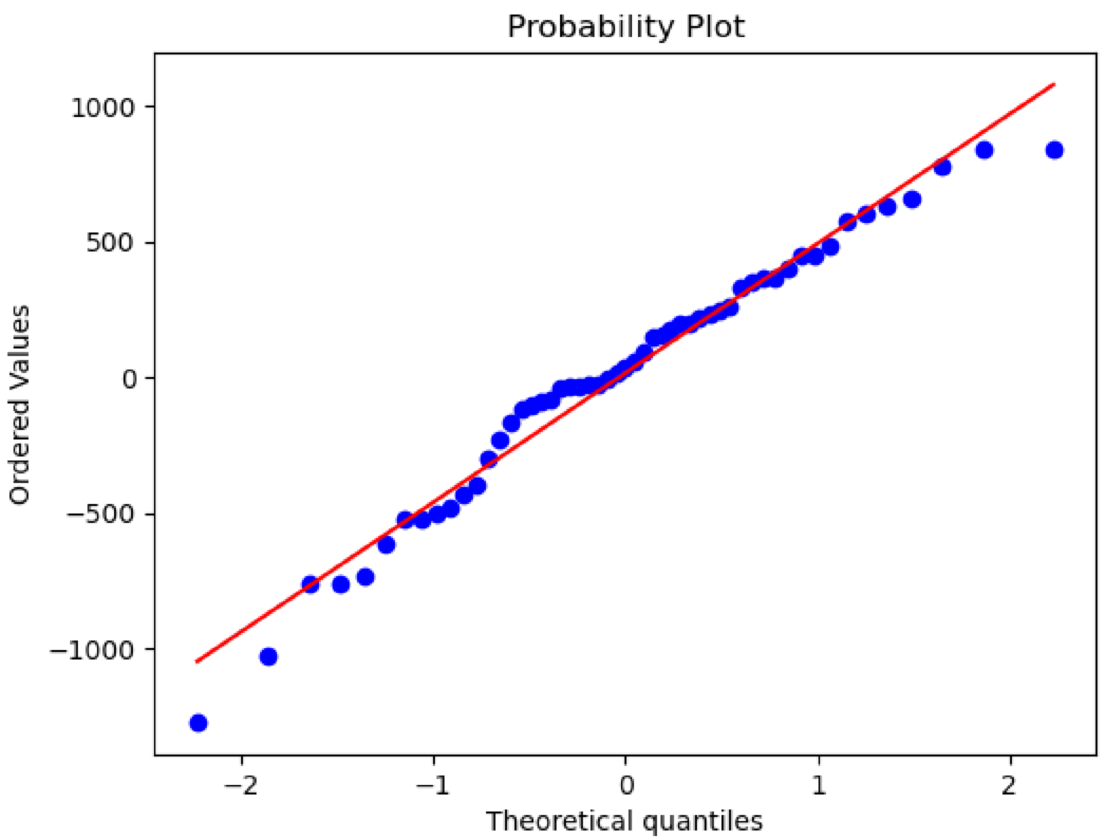
```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   53
Model:               SARIMAX(0, 1, 3)   Log Likelihood                -392.487
Date:                Tue, 26 Aug 2025   AIC                            794.973
Time:                        14:33:13   BIC                            804.730
Sample:                             0   HQIC                           798.714
                                 - 53
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept    156.3362     67.220      2.326      0.020      24.587     288.086
ma.L1          0.3281      0.136      2.405      0.016       0.061       0.595
ma.L2         -0.0130      0.121     -0.107      0.914      -0.251       0.225
ma.L3         -0.4123      0.207     -1.988      0.047      -0.819      -0.006
sigma2      2.068e+05   4.48e+04      4.614      0.000    1.19e+05    2.95e+05
================================================================================
===
Ljung-Box (L1) (Q):                   0.08   Jarque-Bera (JB):
2.71
Prob(Q):                              0.78   Prob(JB):
0.26
Heteroskedasticity (H):               2.96   Skew:
-0.56
Prob(H) (two-sided):                  0.03   Kurtosis:
3.05
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).
```

4.7 Diagnostic Checking

[7]:
```
import matplotlib.pyplot as plt
import scipy.stats as stats
from statsmodels.stats.stattools import jarque_bera


preds = model_auto.predict_in_sample()
```

```
# Calculate residuals
resid = ts - preds

# Histogram of residuals
plt.hist(resid, bins=10, density=True)
plt.title("Residual Histogram")
plt.show()

# Q-Q plot
stats.probplot(resid, dist="norm", plot=plt)
plt.show()

# Jarque-Bera test
jb_stat, jb_pvalue, skew, kurt = jarque_bera(resid)
print("Jarque-Bera p-value:", jb_pvalue)
if jb_pvalue > 0.05:
    print("Residuals are approximately normal.")
else:
    print("Residuals are not normal.")
```



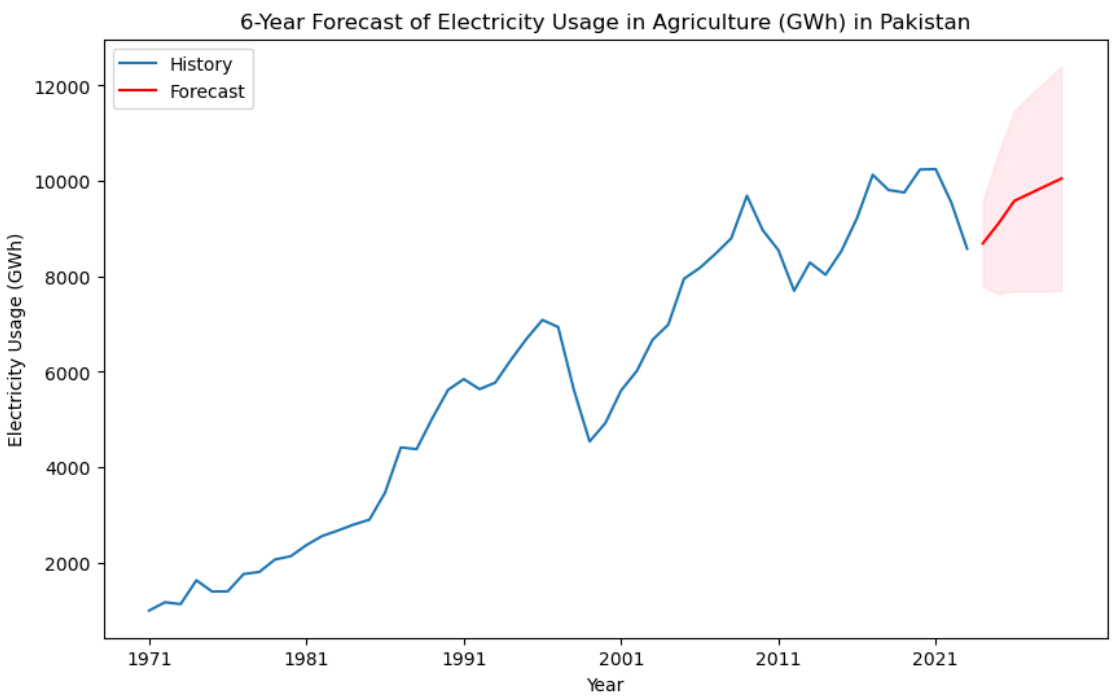



```
Jarque-Bera p-value: 0.3293629667999541
Residuals are approximately normal.
```

4.8 Forecasting

[8]:
```
N = len(ts)
ts.index = pd.Index(range(1971, 1971 + N), name="Year")
# Forecast steps
steps = 6
forecast, conf_int = model_auto.predict(n_periods=steps, return_conf_int=True)


last_year = ts.index[-1]
forecast_index = pd.Index(range(last_year + 1, last_year + steps + 1),␣
 ↪name="Year")


plt.figure(figsize=(10,6))
plt.plot(ts.index, ts.values, label="History")
plt.plot(forecast_index, forecast, label="Forecast", color='red')
```

```
plt.fill_between(forecast_index, conf_int[:, 0], conf_int[:, 1],
                 color='pink', alpha=0.3)


plt.title("6-Year Forecast of Electricity Usage in Agriculture (GWh) in␣
 ↪Pakistan")
plt.xlabel("Year")
plt.ylabel("Electricity Usage (GWh)")
plt.xticks(range(1971, last_year + steps + 1, 10))  # show ticks every 10 years
plt.legend()
plt.show()

forecast = np.array(forecast, dtype=float)
print(pd.Series(forecast, index=forecast_index))
```




```
Year
2024     8690.861328
2025     9111.018331
2026     9581.450825
2027     9737.787007
2028     9894.123189
2029    10050.459371
dtype: float64
```

5 Discussion and Conclusion

The forecasting results provide useful insights into the future trend of the studied variable over the period 2024 to 2029. Based on the analysis, the forecasted values are as follows: 2024 (8690.86), 2025 (9111.02), 2026 (9581.45), 2027 (9737.79), 2028 (9894.12), and 2029 (10050.46). These figures clearly indicate a steady upward movement over the coming years. The growth pattern, although gradual, reflects long-term sustainability and highlights the importance of preparing for changes that may arise in this period. Such predictions are important because they help researchers, government bodies, and other stakeholders to better understand the future direction and to make informed decisions.

The forecast suggests that the variable under study is expected to continue its increasing trend without major fluctuations. This provides a sense of stability, but at the same time, it also places responsibility on policymakers to design strategies that can make this growth meaningful and beneficial. For example, if the forecast is related to economic growth, it will be important to ensure that the benefits of this increase are distributed fairly across different regions and social groups. If the forecast relates to industrial or energy production, it may be necessary to adopt policies that promote efficiency, sustainability, and environmental protection. In either case, careful planning is needed to avoid future challenges such as resource shortages, inequality, or rising costs.

From a policy-making perspective, several recommendations can be made. First, continuous monitoring of the actual values against the forecast is essential to ensure that deviations are identified early and corrective measures are taken. Second, investment in innovation, technology, and infrastructure should be prioritized to sustain growth in the long run. Third, social policies must focus on improving access to services and opportunities, so that the positive impact of growth reaches all parts of society. Finally, environmental considerations must not be ignored; development should follow a sustainable path where natural resources are preserved for future generations.

In conclusion, the forecasted results show a promising future with a consistent upward trend from 2024 to 2029. This outlook offers both opportunities and responsibilities. By using these findings in policy-making, governments and organizations can better manage resources, improve long-term planning, and achieve balanced growth. The forecasts are not just numbers but a guide for strategic action, which if handled wisely, can contribute to progress and stability in the coming years.