3. Feladat

December 4, 2024

0 Előkészületek

0.1 Szükséges könyvtárak importálása

```
%reset -f
import pandas as pd
import numpy as np
import statsmodels.api as sm
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from scipy import stats
from statsmodels.tsa.holtwinters import SimpleExpSmoothing
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.stats.diagnostic import acorr_ljungbox
from statsmodels.graphics.gofplots import qqplot
```

0.2 Adatok beolvasása

```
# Oszlopok definiálása
cols = ['Idő', 'Érték']

# Adatok beolvasása string-ként
with open('data/bead3.csv', 'r', encoding='latin-1') as file:
    lines = file.readlines()

# Az első sor elhagyása és értékek átalakítása
data = [list(map(float, line.strip().strip('"').split(','))) for line in lines[1:
 ↪]]

# DataFrame létrehozása
df = pd.DataFrame(data, columns=cols)
```

1 Determinisztikus modell illesztése

1.1 Kód és eredmények

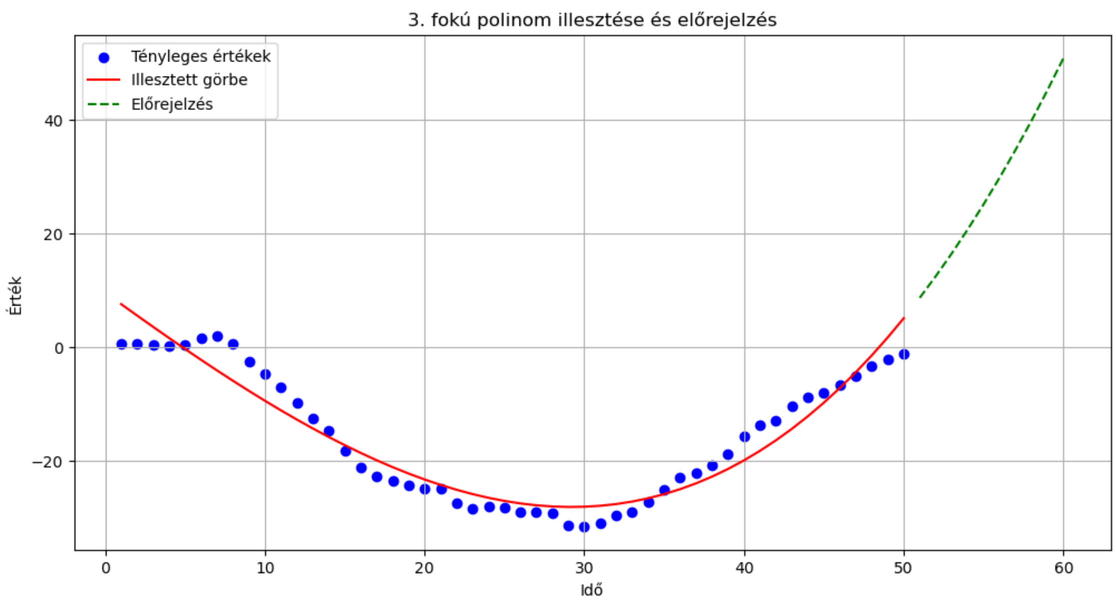
```
# Modellek összehasonlítása
max_degree = 10
selected_degree = 3  # a kiválasztott fokszám
results = []

# Különböző fokszámú modellek összehasonlítása
for degree in range(0, max_degree):
   X = np.vander(df['Idő'], degree + 1)
   model = sm.OLS(df['Érték'], X).fit()
   r2 = r2_score(df['Érték'], model.fittedvalues)

   results.append({
       'Fokszám': degree,
       'R²': r2,
       'AIC': model.aic,
       'BIC': model.bic
   })

# Eredmények kiíratása
results_df = pd.DataFrame(results)
print("\nModellek összehasonlítása:")
print(results_df)

# Kiválasztott fokszámú polinom illesztése
X = np.vander(df['Idő'], selected_degree + 1)
model = sm.OLS(df['Érték'], X).fit()

# Eredmények kiíratása
print(f"\n{selected_degree}. fokú polinom illesztése:")
print(model.summary().tables[0])
print(model.summary().tables[1])

# Reziduálisok vizsgálata
residuals = model.resid

# 1. Várható érték vizsgálata
resid_mean = np.mean(residuals)
resid_std = np.std(residuals, ddof=selected_degree+1)
t_stat = resid_mean / (resid_std/np.sqrt(len(residuals)))
p_value_mean = 2 * stats.t.cdf(-abs(t_stat), len(residuals)-1)

# 2. Normalitás vizsgálata (Shapiro-Wilk teszt)
shapiro_stat, shapiro_p = stats.shapiro(residuals)
```

```
# 3. Függetlenség vizsgálata (Ljung-Box teszt)
lb_stat = sm.stats.diagnostic.acorr_ljungbox(residuals)

# 4. Homoszkedaszticitás vizsgálata (Breusch-Pagan teszt)
bp_test = sm.stats.diagnostic.het_breuschpagan(residuals, model.model.exog)

print("\nHibatagok vizsgálata - eredmények:")
print("-" * 50)
print(f"1. Várható érték vizsgálata:")
print(f"Átlag (várható érték becslése): {resid_mean:.6f}")
print(f"t-statisztika: {t_stat}")
print(f"p-érték: {p_value_mean}")

print(f"\n2. Normalitás vizsgálata (Shapiro-Wilk):")
print(f"Teszt statisztika: {shapiro_stat:.6f}")
print(f"p-érték: {shapiro_p:.6f}")

print(f"\n3. Függetlenség vizsgálata (Ljung-Box):")
print("Lag  Teszt statisztika   p-érték")
print("-" * 35)
for lag, row in lb_stat.iterrows():
    print(f"{lag:2d}   {row['lb_stat']:14.6f}   {row['lb_pvalue']:.6e}")

print(f"\n4. Homoszkedaszticitás vizsgálata (Breusch-Pagan):")
print(f"Teszt statisztika: {bp_test[0]:.6f}")
print(f"p-érték: {bp_test[1]:.6f}")

# Előrejelzés a következő 10 időpontra
future_points = np.arange(len(df) + 1, len(df) + 11)
X_future = np.vander(future_points, selected_degree + 1)
predictions = model.predict(X_future)

print("\nElőrejelzések:")
for i, pred in enumerate(predictions):
   print(f"{future_points[i]}. időpont: {pred:.2f}")

# Ábrázolás
plt.figure(figsize=(12, 6))
plt.scatter(df['Idő'], df['Érték'], color='blue', label='Tényleges értékek')
plt.plot(df['Idő'], model.fittedvalues, color='red', label='Illesztett görbe')
plt.plot(future_points, predictions, color='green', linestyle='--',␣
 ↪label='Előrejelzés')
plt.xlabel('Idő')
plt.ylabel('Érték')
plt.title(f'{selected_degree}. fokú polinom illesztése és előrejelzés')
plt.legend()
```

```
plt.grid(True)
plt.show()
```

```
Modellek összehasonlítása:
   Fokszám        R²         AIC         BIC
0        0  0.000000  387.062930  388.974953
1        1  0.062344  385.844343  389.668389
2        2  0.902997  274.412442  280.148511
3        3  0.921054  266.113215  273.761307
4        4  0.986175  180.999703  190.559818
5        5  0.987419  178.285939  189.758077
6        6  0.988357  176.408735  189.792896
7        7  0.988364  178.381243  193.677427
8        8  0.994452  143.342890  160.551097
9        9  0.992497  156.440981  171.737165

3. fokú polinom illesztése:
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Érték   R-squared:                       0.921
Model:                            OLS   Adj. R-squared:                  0.916
Method:                 Least Squares   F-statistic:                     178.9
Date:                Wed, 04 Dec 2024   Prob (F-statistic):           2.30e-25
Time:                        20:18:41   Log-Likelihood:                -129.06
No. Observations:                  50   AIC:                             266.1
Df Residuals:                      46   BIC:                             273.8
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0006      0.000      3.244      0.002       0.000       0.001
x2             0.0064      0.016      0.412      0.682      -0.025       0.038
x3            -2.0320      0.342     -5.935      0.000      -2.721      -1.343
const          9.5845      2.036      4.707      0.000       5.486      13.683
==============================================================================

Hibatagok vizsgálata - eredmények:
--------------------------------------------------
1. Várható érték vizsgálata:
Átlag (várható érték becslése): 0.000000
t-statisztika: 3.9543447227832004e-13
p-érték: 0.999999999999686

2. Normalitás vizsgálata (Shapiro-Wilk):
Teszt statisztika: 0.971069
```

```
p-érték: 0.255692

3. Függetlenség vizsgálata (Ljung-Box):
Lag  Teszt statisztika   p-érték
-----------------------------------
 1       36.706257   1.373379e-09
 2       57.588516   3.124732e-13
 3       66.733136   2.135808e-14
 4       68.558895   4.572428e-14
 5       68.644769   1.961309e-13
 6       70.969881   2.585887e-13
 7       75.831668   9.719502e-14
 8       83.001311   1.214121e-14
 9       92.734283   4.596171e-16
10      102.056297   2.110987e-17

4. Homoszkedaszticitás vizsgálata (Breusch-Pagan):
Teszt statisztika: 16.254783
p-érték: 0.001005

Előrejelzések:
51. időpont: 8.67
52. időpont: 12.46
53. időpont: 16.46
54. időpont: 20.69
55. időpont: 25.14
56. időpont: 29.81
57. időpont: 34.72
58. időpont: 39.86
59. időpont: 45.24
60. időpont: 50.86
```



1.2 Értelmezés $\varepsilon = 0.05$ szignifikanciaszint mellett

1.2.1 Modellválasztás

Az AIC és BIC értékek alapján a 8. fokú polinom adná a legjobb illeszkedést, azonban a 3. fokú polinom mellett döntöttem az overfitting elkerülése végett.

1.2.2 Várható érték vizsgálata

H_0: E(ε) = 0
H_1: E(ε) $\neq$ 0
t-statisztika értéke: 0.0000
p-érték: 0.9999
Döntés: $0.9999 > 0.05$, tehát nem vetjük el H_0-t

1.2.3 Normalitás vizsgálata (Shapiro-Wilk teszt)

H_0: A hibatagok normális eloszlásúak
H_1: A hibatagok nem normális eloszlásúak
Teszt statisztika: 0.9711
p-érték: 0.2557
Döntés: $0.2557 > 0.05$, tehát nem vetjük el H_0-t

1.2.4 Függetlenség vizsgálata (Ljung-Box teszt)

H_0: A hibatagok függetlenek
H_1: A hibatagok autokorreláltak
A teszt minden vizsgált késleltetésre (1-10 lag) erősen szignifikáns autokorrelációt mutat
Döntés: Minden késleltetésre $p < 0.05$, tehát elvetjük H_0-t

1.2.5 Homoszkedaszticitás vizsgálata (Breusch-Pagan teszt)

H_0: A hibatagok homoszkedasztikusak
H_1: A hibatagok heteroszkedasztikusak
Teszt statisztika: 16.2548
p-érték: 0.0010
Döntés: $0.0010 < 0.05$, tehát elvetjük H_0-t

1.2.6 Összefoglaló értékelés

A hibatagok diagnosztikai vizsgálata alapján:
- A várható érték feltétel teljesül (az átlag gyakorlatilag 0).
- A normalitás feltétele teljesül (a hibatagok normális eloszlásúak).
- A függetlenség feltétele nem teljesül, erős pozitív autokorreláció van jelen.
- A homoszkedaszticitás feltétele nem teljesül, a hibatagok heteroszkedasztikusak.

2 Exponenciális simítás alkalmazása

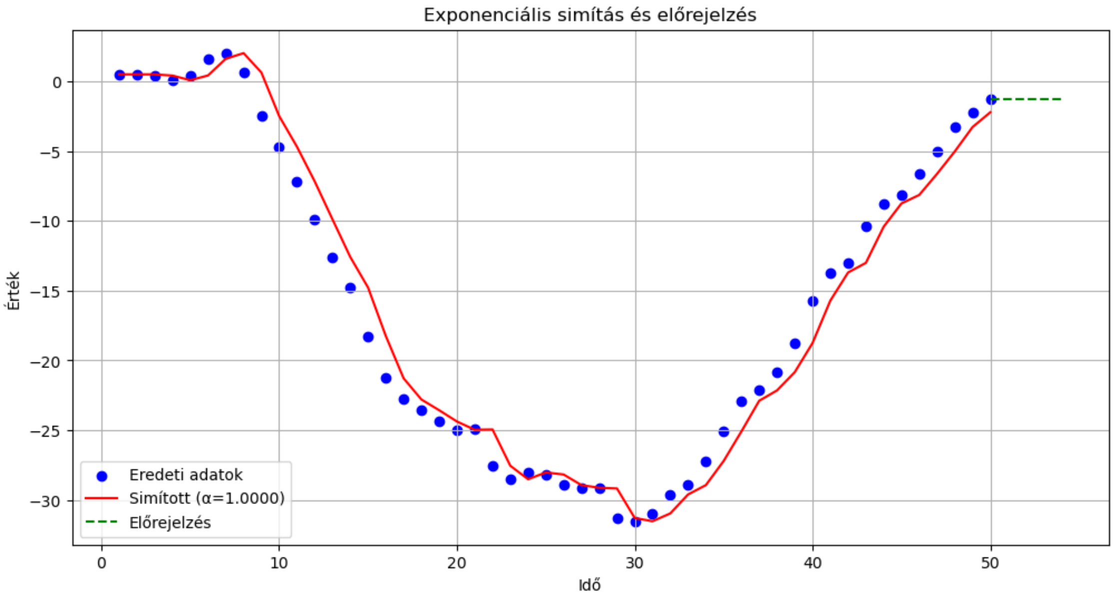
2.1 Kód és eredmények

```
# Exponenciális simítás (optimális alpha meghatározása)
model = SimpleExpSmoothing(df['Érték']).fit()
alpha = model.model.params['smoothing_level']

# Illesztett értékek és előrejelzések
fitted_values = model.fittedvalues
forecast = model.forecast(5)

# Illeszkedés vizsgálata
mae = mean_absolute_error(df['Érték'], fitted_values)
mse = mean_squared_error(df['Érték'], fitted_values)
rmse = np.sqrt(mse)

# Reziduálisok vizsgálata
residuals = model.resid

# 1. Várható érték vizsgálata
resid_mean = np.mean(residuals)
resid_std = np.std(residuals, ddof=1)
t_stat = resid_mean / (resid_std / np.sqrt(len(residuals)))
p_value_mean = 2 * stats.t.cdf(-abs(t_stat), len(residuals) - 1)

# 2. Normalitás vizsgálata (Shapiro-Wilk teszt)
shapiro_stat, shapiro_p = stats.shapiro(residuals)

# 3. Függetlenség vizsgálata (Ljung-Box teszt)
lb_stat = sm.stats.diagnostic.acorr_ljungbox(residuals)
```

```python
# 4. Homoszkedaszticitás vizsgálata (Breusch-Pagan teszt)
exog = sm.add_constant(fitted_values)
bp_test = sm.stats.diagnostic.het_breuschpagan(residuals, exog)

# Eredmények kiírása
print("\nIlleszkedési mutatók:")
print(f"MAE = {mae:.4f}")
print(f"MSE = {mse:.4f}")
print(f"RMSE = {rmse:.4f}")
print(f"Smoothing level (alpha) = {alpha:.4f}")

print("\nHibatagok vizsgálata - eredmények:")
print("-" * 50)
print("1. Várható érték vizsgálata:")
print(f"Átlag: {resid_mean:.6f}")
print(f"t-statisztika: {t_stat:.6f}")
print(f"p-érték: {p_value_mean:.6f}")

print("\n2. Normalitás vizsgálata (Shapiro-Wilk):")
print(f"Teszt statisztika: {shapiro_stat:.6f}")
print(f"p-érték: {shapiro_p:.6f}")

print(f"\n3. Függetlenség vizsgálata (Ljung-Box):")
print("Lag  Teszt statisztika   p-érték")
print("-" * 35)
for lag, row in lb_stat.iterrows():
    print(f"{lag:2d}   {row['lb_stat']:14.6f}   {row['lb_pvalue']:.6e}")

print("\n4. Homoszkedaszticitás vizsgálata (Breusch-Pagan teszt):")
print(f"Teszt statisztika: {bp_test[0]:.6f}")
print(f"p-érték: {bp_test[1]:.6f}")

print("\nElőrejelzések:")
for i, pred in enumerate(forecast, 1):
    print(f"{len(df) + i}. időpont: {pred:.2f}")

# Ábrázolás
plt.figure(figsize=(12, 6))
plt.scatter(df['Idő'], df['Érték'], color='blue', label='Eredeti adatok')
plt.plot(df['Idő'], fitted_values, 'r-', label=f'Simított (α={alpha:.4f})')

future_points = np.arange(len(df), len(df) + len(forecast))
plt.plot(future_points, forecast, 'g--', label='Előrejelzés')

plt.title('Exponenciális simítás és előrejelzés')
plt.xlabel('Idő')
plt.ylabel('Érték')
```

```python
plt.legend()
plt.grid(True)
plt.show()
```

```
Illeszkedési mutatók:
MAE = 1.3446
MSE = 2.7510
RMSE = 1.6586
Smoothing level (alpha) = 1.0000

Hibatagok vizsgálata - eredmények:
--------------------------------------------------
1. Várható érték vizsgálata:
Átlag: -0.035400
t-statisztika: -0.149436
p-érték: 0.881823

2. Normalitás vizsgálata (Shapiro-Wilk):
Teszt statisztika: 0.961779
p-érték: 0.105540

3. Függetlenség vizsgálata (Ljung-Box):
Lag  Teszt statisztika   p-érték
-----------------------------------
 1        34.645889   3.954729e-09
 2        59.005126   1.538862e-13
 3        82.658607   8.253867e-18
 4       104.418817   1.126498e-21
 5       118.862058   5.466764e-24
 6       131.665637   5.730822e-26
 7       144.778551   5.062839e-28
 8       151.797860   8.266950e-29
 9       155.033991   7.976957e-29
10       156.828963   1.461608e-28

4. Homoszkedaszticitás vizsgálata (Breusch-Pagan teszt):
Teszt statisztika: 0.487821
p-érték: 0.484901

Előrejelzések:
51. időpont: -1.25
52. időpont: -1.25
53. időpont: -1.25
54. időpont: -1.25
55. időpont: -1.25
```



2.2 Exponenciális simítás eredményei

2.2.1 Modell specifikációk

A modellben a SimpleExpSmoothing függvény által meghatározott $\alpha = 1.0000$ simítási paramétert használtuk.

2.2.2 Illeszkedési mutatók

MAE (Mean Absolute Error): 1.3446
Az átlagos abszolút hiba azt mutatja, hogy az előrejelzéseink átlagosan 1.3446 egységgel térnek el a tényleges értékektől.

MSE (Mean Squared Error): 2.7510
Az átlagos négyzetes hiba az előrejelzési hibák négyzetének átlaga, jelen esetben 2.7510. Ez a mutató érzékeny a nagyobb eltérésekre, mivel a hibákat négyzetre emeli.

RMSE (Root Mean Squared Error): 1.6586
A négyzetes átlaggyök hiba az MSE négyzetgyöke, ami az előrejelzési hibák átlagos nagyságát adja meg az eredeti mértékegységben.

Simítási paraméter (alpha): 1.0000
Az alpha értéke 1.0, ami azt jelenti, hogy a modell teljes mértékben az utolsó megfigyelésre támaszkodik az előrejelzés során. Ebben az esetben a modell nem simítja az adatokat, hanem minden előrejelzés az utolsó ismert érték lesz.

2.3 Hibatagok tulajdonságainak vizsgálata $\varepsilon = 0.05$ szignifikanciaszint mellett

2.3.1 Várható érték vizsgálata

H_0: E(ε) = 0
H_1: E(ε) $\neq$ 0
t-statisztika értéke: -0.1494
p-érték: 0.8818
Döntés: 0.8818 > 0.05, tehát nem vetjük el H_0-t

2.3.2 Normalitás vizsgálata (Shapiro-Wilk teszt)

H_0: A hibatagok normális eloszlásúak
H_1: A hibatagok nem normális eloszlásúak
Teszt statisztika: 0.9618
p-érték: 0.1055
Döntés: 0.1055 > 0.05, tehát nem vetjük el H_0-t

2.3.3 Függetlenség vizsgálata (Ljung-Box teszt)

H_0: A hibatagok függetlenek
H_1: A hibatagok autokorreláltak
A teszt minden vizsgált késleltetésre (1-10 lag) erősen szignifikáns autokorrelációt mutat
Döntés: Minden késleltetésre p < 0.05, tehát elvetjük H_0-t

2.3.4 Homoszkedaszticitás vizsgálata (Breusch-Pagan teszt)

H_0: A hibatagok homoszkedasztikusak
H_1: A hibatagok heteroszkedasztikusak
Teszt statisztika: 0.4878
p-érték: 0.4849
Döntés: 0.4849 > 0.05, tehát nem vetjük el H_0-t

2.4 Összefoglaló értékelés

A hibatagok diagnosztikai vizsgálata alapján:
- A várható érték feltétel teljesül (az átlag gyakorlatilag 0)
- A normalitás feltétele teljesül (a hibatagok normális eloszlásúak)
- A függetlenség feltétele nem teljesül, erős pozitív autokorreláció van jelen
- A homoszkedaszticitás feltétele teljesül (a szórás állandó)

3 Box-Jenkins modell

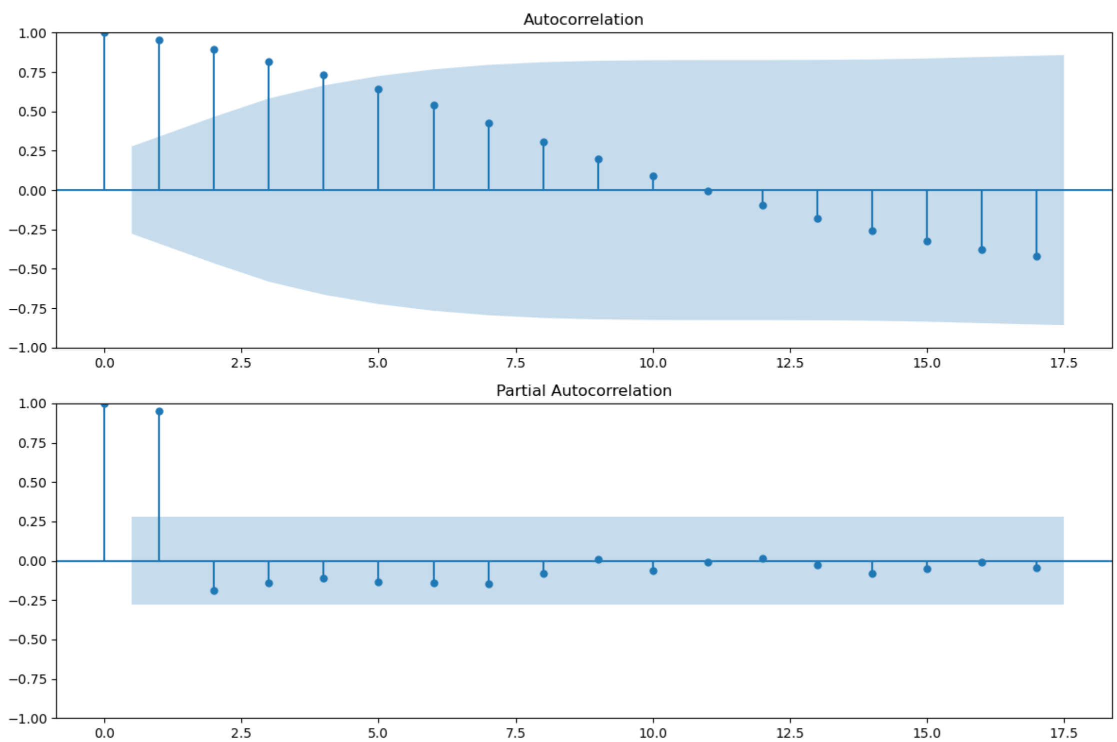
3.1 Kód és eredmények

```
# Idősor stacionaritásának vizsgálata (ADF teszt)
adf_result = adfuller(df['Érték'])
print('\nADF Teszt eredménye:')
print(f'ADF Statisztika: {adf_result[0]:.4f}')
print(f'p-érték: {adf_result[1]:.4f}')
```

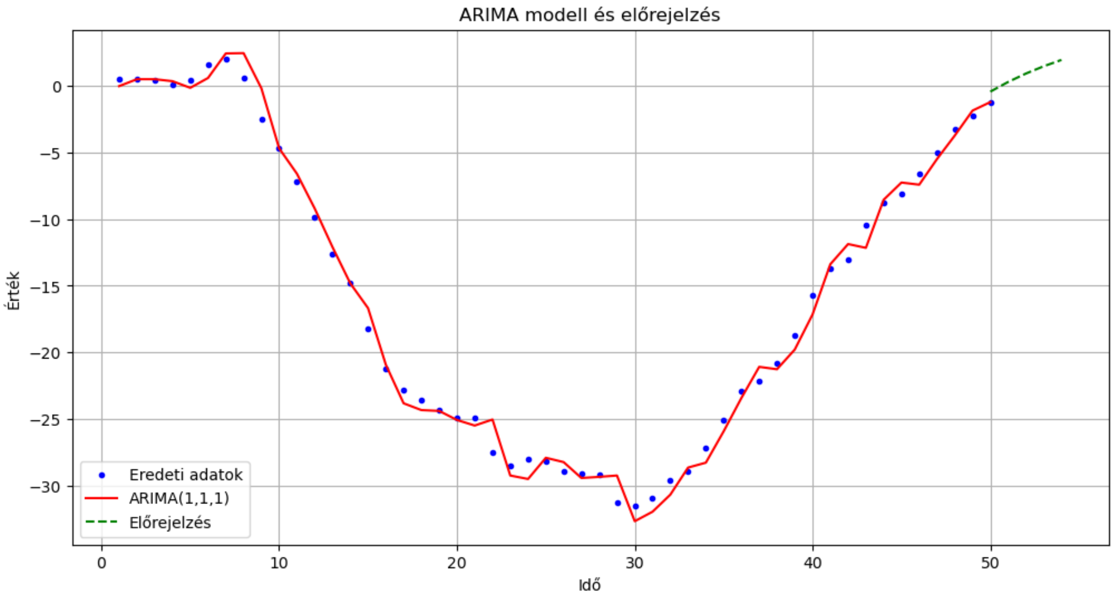
```python
# ACF és PACF ábrák a paraméterek meghatározásához
fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(12, 8))
plot_acf(df['Érték'], ax=ax1)
plot_pacf(df['Érték'], ax=ax2)
plt.tight_layout()
plt.show()

# ARIMA modell illesztése
p, d, q = 1, 1, 1
model = ARIMA(df['Érték'], order=(p, d, q))
results = model.fit()

# Illesztett értékek és előrejelzések
fitted_values = results.fittedvalues
forecast = results.forecast(steps=5)

# Reziduálisok vizsgálata
residuals = results.resid

# 1. Várható érték vizsgálata
resid_mean = np.mean(residuals)
resid_std = np.std(residuals, ddof=1)
t_stat = resid_mean / (resid_std/np.sqrt(len(residuals)))
p_value_mean = 2 * stats.t.cdf(-abs(t_stat), len(residuals)-1)

# 2. Normalitás vizsgálata
shapiro_stat, shapiro_p = stats.shapiro(residuals)

# 3. Függetlenség vizsgálata (Ljung-Box teszt)
lb_stat = sm.stats.diagnostic.acorr_ljungbox(residuals)

# 4. Homoszkedaszticitás vizsgálata
exog = sm.add_constant(fitted_values)
bp_test = sm.stats.diagnostic.het_breuschpagan(residuals, exog)

print('\nModell eredmények:')
print(results.summary().tables[0])
print(results.summary().tables[1])

print('\nHibatagok vizsgálata:')
print(f'Várható érték teszt p-érték: {p_value_mean:.4f}')
print(f'Shapiro-Wilk teszt p-érték: {shapiro_p:.4f}')
print(f"Ljung-Box teszt:")
print("Lag  Teszt statisztika   p-érték")
print("-" * 35)
for lag, row in lb_stat.iterrows():
```

```
    print(f"{lag:2d}   {row['lb_stat']:14.6f}   {row['lb_pvalue']:.6e}")
print(f'\nBreusch-Pagan teszt p-érték: {bp_test[1]:.4f}')

print('\nElőrejelzések:')
for i, pred in enumerate(forecast, 1):
    print(f'{len(df) + i}. időpont: {pred:.2f}')

# Ábrázolás
plt.figure(figsize=(12, 6))
plt.plot(df['Idő'], df['Érték'], 'b.', label='Eredeti adatok')
plt.plot(df['Idő'], fitted_values, 'r-', label=f'ARIMA({p},{d},{q})')
future_points = np.arange(len(df), len(df) + 5)
plt.plot(future_points, forecast, 'g--', label='Előrejelzés')
plt.title('ARIMA modell és előrejelzés')
plt.xlabel('Idő')
plt.ylabel('Érték')
plt.legend()
plt.grid(True)
plt.show()
```

```
ADF Teszt eredménye:
ADF Statisztika: -2.5644
p-érték: 0.1006
```



```
Modell eredmények:
                               SARIMAX Results
==============================================================================
Dep. Variable:                  Érték   No. Observations:                   50
Model:                 ARIMA(1, 1, 1)   Log Likelihood                 -68.798
Date:                Wed, 04 Dec 2024   AIC                            143.597
Time:                        20:18:42   BIC                            149.272
Sample:                             0   HQIC                           145.750
                                 - 50
Covariance Type:                  opg
==============================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8688      0.090      9.658      0.000       0.692       1.045
ma.L1         -0.2181      0.159     -1.372      0.170      -0.530       0.094
sigma2         0.9504      0.193      4.913      0.000       0.571       1.329
==============================================================================

Hibatagok vizsgálata:
Várható érték teszt p-érték: 0.8523
Shapiro-Wilk teszt p-érték: 0.0627
Ljung-Box teszt:
Lag  Teszt statisztika   p-érték
-----------------------------------
 1         0.212220   6.450331e-01
 2         3.389275   1.836658e-01
 3         3.694513   2.963968e-01
 4         7.022499   1.347040e-01
 5         7.781213   1.687128e-01
 6         7.867512   2.479714e-01
 7        13.683233   5.711086e-02
 8        13.708309   8.969225e-02
 9        15.190585   8.583220e-02
10        16.478098   8.673992e-02

Breusch-Pagan teszt p-érték: 0.5577

Előrejelzések:
51. időpont: -0.41
52. időpont: 0.31
53. időpont: 0.94
54. időpont: 1.49
55. időpont: 1.97
```



3.2 Értelmezés $\varepsilon = 0.05$ szignifikanciaszint mellett

3.2.1 ADF teszt eredménye

H_0: Az idősor nem stacionárius
H_1: Az idősor stacionárius
ADF Statisztika: -2.5644
p-érték: 0.1006
Döntés: $0.1006 > 0.05$, tehát nem vetjük el H_0-t, az idősor nem stacionárius, ezért szükséges differenciálnunk

3.2.2 Modell paraméterek

ARIMA(1,1,1) modellt illesztettünk, ahol:
- p = 1 (autoregresszív tag), mert a PACF ábrán az első késleltetés volt szignifikáns
- d = 1 (differenciálás rendje), mert az idősor nem stacionárius
- q = 1 (mozgóátlag tag), mert az ACF ábra az első késleltetésnél mutat szignifikáns értéket

3.2.3 Paraméterek szignifikanciája

- AR(1) tag: p-érték $= 0.000 < 0.05$, szignifikáns

- MA(1) tag: p-érték $= 0.170 > 0.05$, nem szignifikáns

3.2.4 Modell illeszkedési mutatók

AIC: 143.597
BIC: 149.272
Log Likelihood: -68.798

3.3 Hibatagok tulajdonságainak vizsgálata

3.3.1 Várható érték vizsgálata

p-érték: $0.8523 > 0.05$, tehát nem vetjük el H_0-t

3.3.2 Normalitás vizsgálata (Shapiro-Wilk teszt)

p-érték: $0.0627 > 0.05$, tehát nem vetjük el H_0-t

3.3.3 Függetlenség vizsgálata (Ljung-Box teszt)

Minden késleltetésre $p > 0.05$, tehát nem vetjük el H_0-t

3.3.4 Homoszkedaszticitás vizsgálata (Breusch-Pagan teszt)

p-érték: $0.5577 > 0.05$, tehát nem vetjük el H_0-t

3.4 Összefoglaló értékelés

- A modell diagnosztikája megfelelő:
  - A várható érték feltétel teljesül
  - A hibatagok normális eloszlásúak
  - A függetlenség feltétele teljesül
  - A homoszkedaszticitás feltétele teljesül
- Az AR(1) tag szignifikáns, míg az MA(1) tag nem
- Az előrejelzések növekvő trendet mutatnak