Jie Wang Date : Oct 3rd, 2024

Exercises

1. (1 point) Discuss the primary benefits of detrending a time series before further analysis or modeling.

2. (1 point) Outline the key advantages of removing the seasonal component from a time series prior to processing or forecasting.

3. (6 points) Consider a time series $\mathcal{X} = (X_1, X_2, \ldots\}$ defined by the following recursion:

$$X_k = \alpha \cos\left(\frac{2\pi k}{T}\right) + \beta \sin\left(\frac{2\pi k}{T}\right) + \phi X_{k-1} + \epsilon_k + \theta \epsilon_{k-1} \quad \text{with } \epsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), .$$

   Answer the following questions:

   (a) Write a Python script to generate and plot a sample path of length $L = 1000$ using the following parameters: $\alpha = 2$, $\beta = 0.5$, $T = 50$, $\phi = 0.9$, $\theta = 0.7$, $\sigma_\epsilon^2 = 1$, and $Y_0 = 0$.

   (b) Plot the autocorrelation function (ACF) of the generated sample path.

   (c) Perform a seasonal adjustment of the sample path generated in the previous step using the `stl`. Ensure that the seasonal component being removed is strictly periodic (e.g., compute the average of all seasons). Plot the residual signal after the seasonal adjustment.

   (d) Plot the ACF of the residuals.

   (e) Does the seasonal adjustment aids in the visual analysis of the ACF? Justify your answer.

4. (5 points) Given two continuous random variables X and Y with a joint density function $f_{X,Y}(x, y)$, the conditional density is defined as:

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

   (a) Use the definition of conditional density to prove that:

$$f_{Y_1 Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2|Y_1}(y_2|y_1).$$

   (b) Let us define the set of random variable $\mathcal{Y}_{\leq k} = \{Y_k, Y_{k-1}, Y_{k-2}, \ldots\}$ and the related information set $\mathcal{F}_k = \{Y_k = y_k, Y_{k-1} = y_{k-1}, Y_{k-2} = y_{k-2}, \ldots\}$. Using the definition of conditional density, show how to prove:

$$f_{Y_1 Y_2|\mathcal{Y}_{\leq 0}}(y_1, y_2|\mathcal{F}_0) = f_{Y_1|\mathcal{Y}_{\leq 0}}(y_1|\mathcal{F}_0) f_{Y_2|\mathcal{Y}_{\leq 1}}(y_2|\mathcal{F}_1).$$

   (c) Using similar steps in a recursive manner, show how to prove the following identity:

$$f_{Y_1 Y_2 Y_3|\mathcal{Y}_{\leq 0}}(y_1, y_2, y_3|\mathcal{F}_0) = f_{Y_1|\mathcal{Y}_{\leq 0}}(y_1|\mathcal{F}_0) f_{Y_2|\mathcal{Y}_{\leq 1}}(y_2|\mathcal{F}_1) f_{Y_3|\mathcal{Y}_{\leq 2}}(y_3|\mathcal{F}_2)$$

5. (4 points) Consider the following stochastic process:

$$Y_{k+1} = \phi Y_k + \epsilon_{k+1} + \theta \epsilon_k, \quad \text{with } \epsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), \ |\phi| < 1, \text{ and } Y_0 = y_0.$$

   Using the *causal factorization* in (1), provide a closed-form expression for:

$$f_{Y_1, \ldots, Y_L|\mathcal{Y}_{\leq 0}}(y_1, \ldots, y_L|\mathcal{F}_0).$$

Exercises

1. (1 point) Discuss the primary benefits of detrending a time series before further analysis or modeling.

Cite : Page 9

① Enhance stationarity ⇒ by stabilizing the mean

the trend may induce a time-varying mean, which violates the WSS requeirement for time series analysis

② reduce bias from trend

Detrending makes seasonal patterns / cyclical fluctuations more apparent ⇒ This enables more precise forecasting of short term movements & improve interpretability

③ Reduce the risks of misleading correlations

After detrending, we can observe the residual noise properties more clearly ,

Overall, detreading leads to more relicble, interpretable & accurate analysis when modeling Time series

2. (1 point) Outline the key advantages of removing the seasonal component from a time series prior to processing or forecasting.

① Enhance stationarity ⇒ by stabilizing the mean

the season may induce a time-varying mean, which violates the WSS requeirement for time series analysis

② Obsecure the underlying patterns

③ improve comparability across time periods

④ Reduce the risks of misleading correlations

⇒ All in all, turn data into more stable to understand time seriess

3. (6 points) Consider a time series $\mathcal{X} = (X_1, X_2, \ldots\}$ defined by the following recursion:

$$X_k = \alpha \cos\left(\frac{2\pi k}{T}\right) + \beta \sin\left(\frac{2\pi k}{T}\right) + \phi X_{k-1} + \epsilon_k + \theta \epsilon_{k-1} \quad \text{with } \epsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2),.$$

Answer the following questions: All the code are attached below

(a) Write a Python script to generate and plot a sample path of length $L = 1000$ using the following parameters: $\alpha = 2$, $\beta = 0.5$, $T = 50$, $\phi = 0.9$, $\theta = 0.7$, $\sigma_\epsilon^2 = 1$, and $Y_0 = 0$.

(b) Plot the autocorrelation function (ACF) of the generated sample path.

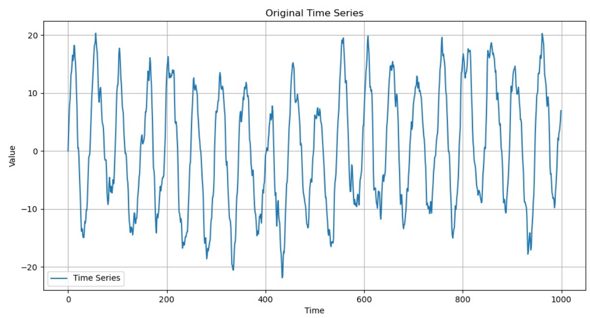
(c) Perform a seasonal adjustment of the sample path generated in the previous step using the `stl`. Ensure that the seasonal component being removed is strictly periodic (e.g., compute the average of all seasons). Plot the residual signal after the seasonal adjustment.

np. random. seed (538) here

(d) Plot the ACF of the residuals.

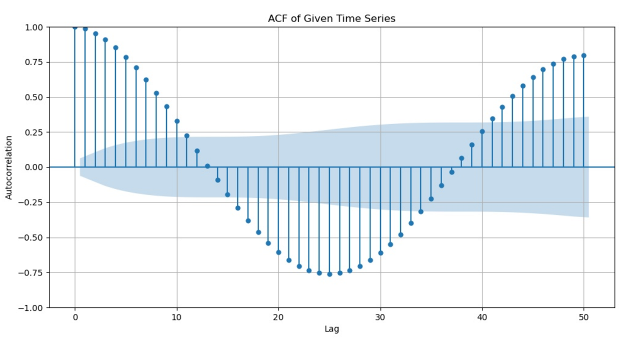
(e) Does the seasonal adjustment aids in the visual analysis of the ACF? Justify your answer.

1(a): here's the generated time series:



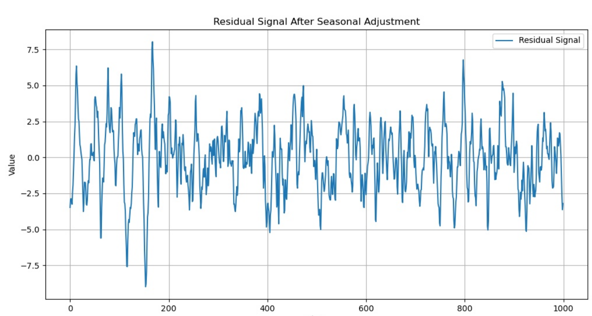

```python
# a) Write a Python script to generate and plot a sample path of length L = 1000 using the following parameters
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.seasonal import STL

# set random seed as 538
np.random.seed(538)
# Define the parameters
alpha = 2
beta = 0.5
T = 50
phi = 0.9
theta = 0.7
sigma_sq = 1
X_0 = 0

# Generate the time series
L = 1000
X = np.zeros(L)
X[0] = X_0
epsilon = np.random.normal(0, np.sqrt(sigma_sq), L)

for k in range(1, L):
    X[k] = alpha * np.cos(2 * np.pi * k / T) + beta * np.sin(2 * np.pi * k / T) \
    + phi * X[k-1] + epsilon[k] + theta * epsilon[k-1]

# Plot the time series
plt.figure(figsize=(12, 6))
plt.plot(X, label='Time Series')
plt.title('Original Time Series')
plt.xlabel('Time')
plt.ylabel('Value')
plt.legend()
plt.grid(True)
plt.savefig('T2_a.png')
plt.close()
```

(b):



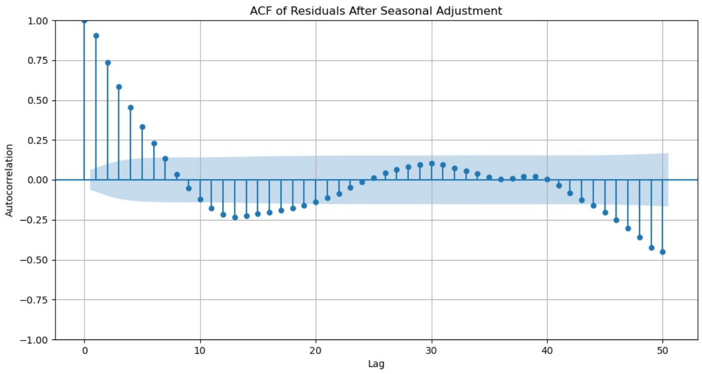

```python
# b) Plot the autocorrelation function (ACF) of the generated sample path.
plt.figure(figsize=(12, 6))
plot_acf(X, lags=50, ax=plt.gca()) # current axis
plt.xlabel('Lag')
plt.ylabel('Autocorrelation')
plt.title('ACF of Given Time Series')
plt.grid(True)
plt.savefig('T2_b.png')
plt.close()

# c) Perform a seasonal adjustment of the sample path generated in the previous st

# Custom STL, the average seasonal component
# seasonal_component = np.tile(np.mean(X.reshape(-1, T), axis=0), L // T + 1)[:L]
# adjusted_series = X - seasonal_component

# Standard STL decomposition on the time series
stl = STL(X, period=T)
result = stl.fit()

# Extract the residual component
residuals = result.resid

plt.figure(figsize=(12, 6))
plt.plot(residuals, label='Residual Signal')
plt.title('Residual Signal After Seasonal Adjustment')
plt.xlabel('Time')
plt.ylabel('Value')
plt.legend()
plt.grid(True)
plt.savefig('T2_c.png')
plt.close()
```

(c):



(d):

np. random . seed (538) here




```python
# d) Plot the ACF of the residuals.
plt.figure(figsize=(12, 6))
plot_acf(residuals, lags=50, ax=plt.gca())
plt.title('ACF of Residuals After Seasonal Adjustment')
plt.xlabel('Lag')
plt.ylabel('Autocorrelation')
plt.grid(True)
plt.savefig('T2_d.png')
plt.close()
```

(e): It's clearly that removing seasonality greatly helps the overall ACF visualization

① magnitude is much lower

$\Rightarrow$ less correlated to the periodic part of $\alpha \cos\left(\frac{2\pi k}{T}\right) + \beta \sin\left(\frac{2\pi k}{T}\right)$

4. (5 points) Given two continuous random variables X and Y with a joint density function $f_{X,Y}(x, y)$, the conditional density is defined as:

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}. \quad (1)$$

(a) Use the definition of conditional density to prove that:

$$f_{Y_1 Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2|Y_1}(y_2|y_1).$$

(b) Let us define the set of random variable $\mathcal{Y}_{\leq k} = \{Y_k, Y_{k-1}, Y_{k-2}, \ldots\}$ and the related information set $\mathcal{F}_k = \{Y_k = y_k, Y_{k-1} = y_{k-1}, Y_{k-2} = y_{k-2}, \ldots\}$. Using the definition of conditional density, show how to prove:

$$f_{Y_1 Y_2|\mathcal{Y}_{\leq 0}}(y_1, y_2|\mathcal{F}_0) = f_{Y_1|\mathcal{Y}_{\leq 0}}(y_1|\mathcal{F}_0) f_{Y_2|\mathcal{Y}_{\leq 1}}(y_2|\mathcal{F}_1).$$

(c) Using similar steps in a recursive manner, show how to prove the following identity:

$$f_{Y_1 Y_2 Y_3|\mathcal{Y}_{\leq 0}}(y_1, y_2, y_3|\mathcal{F}_0) = f_{Y_1|\mathcal{Y}_{\leq 0}}(y_1|\mathcal{F}_0) f_{Y_2|\mathcal{Y}_{\leq 1}}(y_2|\mathcal{F}_1) f_{Y_3|\mathcal{Y}_{\leq 2}}(y_3|\mathcal{F}_2) \quad (4)$$

(a): with (1), let $\begin{cases} X = Y_2 \\ Y = Y_1 \end{cases}$, $\begin{cases} x = y_2 \\ y = y_1 \end{cases}$

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_2 Y_1}(y_2, y_1)}{f_{Y_1}(y_1)}$$

$$\Rightarrow f_{Y_2 Y_1}(y_2, y_1) = f_{Y_1}(y_1) \cdot f_{Y_2|Y_1}(y_2|y_1) \quad (2)$$

(b): For $\forall$ Y_2, Y_1, Y_0, with chain rule,

we have: $F_1 = \{F_0, Y_1\}$, as F_1 is observation before Y_2 (3)

So $f_{Y_1 Y_2|\mathcal{Y}\leq 0}(y_1, y_2|F_0) = f_{Y_1|\mathcal{Y}\leq 0}(y_1|F_0) \cdot \underline{f_{Y_2|\mathcal{Y}\leq 0}(y_2|y_1, F_0)}$

By (3) $\underline{f_{Y_2|\mathcal{Y}\leq 0}(y_2|y_1, F_0)} = f_{Y_2|\mathcal{Y}\leq 0}(y_2|F_1)$

$\Rightarrow f_{Y_1 Y_2|\mathcal{Y}\leq 0}(y_1, y_2|F_0) = f_{Y_1|\mathcal{Y}\leq 0}(y_1|F_0) \cdot f_{Y_2|\mathcal{Y}\leq 0}(y_2|F_1)$

Proved by chain rule

(c): $f_{Y_1 Y_2 Y_3|\mathcal{Y}\leq 0}(y_1, y_2, y_3|F_0)$

$= f_{Y_1|\mathcal{Y}\leq 0}(y_1|F_0) f_{Y_2|Y_1, \mathcal{Y}\leq 0}(y_2|y_1, F_0) f_{Y_3|Y_1, Y_2, \mathcal{Y}\leq 0}(y_3|y_2, y_1, F_0)$

$= f_{Y_1|\mathcal{Y}\leq 0}(y_1|F_0) \cdot f_{Y_2|\mathcal{Y}\leq 0}(y_2|F_1) \cdot f_{Y_3|\mathcal{Y}\leq 0}(y_3|F_2)$ by (2), (3)

Proved the (4) via recursive chain rule

5. (4 points) Consider the following stochastic process:

$$Y_{k+1} = \phi Y_k + \epsilon_{k+1} + \theta\epsilon_k, \quad \text{with } \epsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma^2),\ |\phi| < 1,\ \text{and } Y_0 = y_0.$$

Using the *causal factorization* in (1), provide a closed-form expression for:

$$f_{Y_1,\ldots,Y_L|\mathcal{Y}_{\le 0}}(y_1,\ldots,y_L|\mathcal{F}_0).$$

The given stochastic process is ARMA(1)

with $|\phi|<1$, we know it's stationary

As ARMA(1) has: Y_{k+1} depends on and only on Y_k and gaussian noise

each observation depends only on its immediate predecessor

$\therefore f_{Y_k|y_{\le k-1}}(y_k|F_{k-1}) = f_{Y_k|Y_{k-1}}(y_k|y_{k-1}) \Rightarrow$ $E[Y_k|Y_{k-1}] = E[\phi Y_k + \epsilon_{k+1} + \theta\epsilon_k | Y_{k-1}]$

$$= \phi Y_{k-1} + E[\varepsilon_k|Y_{k-1}] + \theta E[\varepsilon_{k-1}|Y_{k-1}]$$

$$= \phi Y_{k-1}$$

$$Var[Y_k|Y_{k-1}] = Var(\phi Y_{k-1} + \epsilon_{k+1} + \theta\epsilon_k | Y_{k-1}]$$

$$= Var[\phi Y_{k-1}|Y_{k-1}) + Var[\varepsilon_k|Y_{k-1}] + \theta^2 Var[\varepsilon_{k-1}|Y_{k-1}]$$

$$= 0 + \sigma^2 + \theta^2\sigma^2$$

$$= (1+\theta^2)\sigma^2$$

we have: $Y_{k+1}|Y_k, \varepsilon_k \sim N(\phi Y_k, (1+\theta^2)\sigma^2)$

$$\Rightarrow f_{Y_{k+1}|Y_k,\varepsilon_k}(y_{k+1}|y_k,\epsilon_k) = \frac{1}{\sqrt{2\pi\sigma^2(1+\theta^2)}} \exp\left(-\frac{(y_k - (\phi\cdot y_{k-1} + \theta\epsilon_{k-1}))^2}{2(1+\theta^2)\sigma^2}\right)$$

By Causal factorization in (1),

we can have Density func as

$$f_{Y_1\ldots Y_L|y_{\le 0}}(y_1,\ldots,y_L|F_0) = \prod_{k=1}^{L} f_{Y_k|y_{\le k-1}}(y_k|F_{k-1})$$

$$= \left(2\pi\sigma^2(1+\theta^2)\right)^{-\frac{L}{2}} \prod_{k=1}^{L} \exp\left(-\frac{(y_k - (\phi\cdot y_{k-1} + \theta\epsilon_{k-1}))^2}{2(1+\theta^2)\sigma^2}\right)$$

where $F_0 = y_0$ as initial condition

6. (1 point) Based on a visual inspection, determine whether the temporal evolution of AAPL stock prices (Fig. 7-left) appears stationary. Justify your answer with observations.

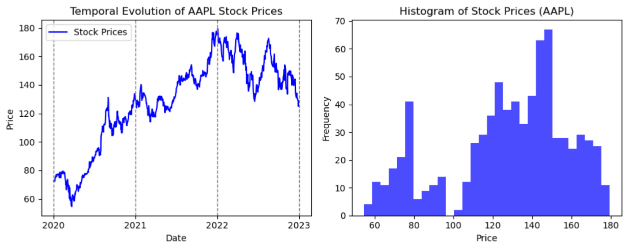


Figure 7: (Left) Histogram of AAPL stock prices. (Right) Histogram of log-transformed returns.

it's clearly a non-stationary time series,

As stock price has:

① trend ↗ when 2020 ~ 2022
↘ when 2022 ~ 2023

⇒ the mean shifts, being non-constant

② Variance: it's non consistent, as flunctuation appears frequently in (Right)

→ it's heteroskedasticity

③ Seasonal Pattern:

The flunctuation seems similar in each year, showing potential seasonality

7. (1 point) Evaluate whether the temporal evolution of AAPL log returns (Fig. 8-left) is stationary. Conduct a formal hypothesis test in Python and interpret the results based on the observed p-value.

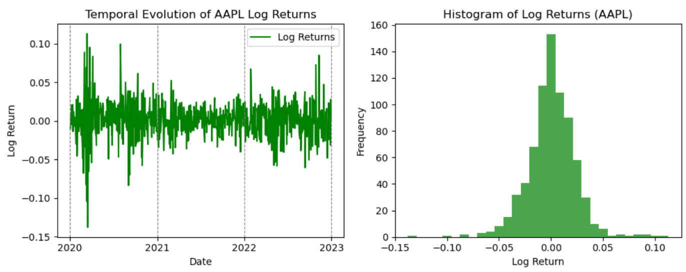


Figure 8: (Left) Time series of log returns. (Right) Histogram of log-returns.

As shown below the log return are stationary with p-value = 2,86e-14 << 0,05

```python
from statsmodels.tsa.stattools import adfuller

# Augmented Dickey-Fuller Test
adf, p_value, _, _,critical_values, _= adfuller(log_returns)

print(f'ADF Statistic: {adf}')
print(f'p-value: {p_value}')
print('Critical Values:')
for key, value in critical_values.items():
    print(f'   {key}: {value}')

if p_value < 0.05:
    print("The log returns are stationary (reject the null hypothesis of unit root).")
else:
    print("The log returns are not stationary (fail to reject the null hypothesis of unit root).")
```

[6] ✓ 0.0s

```
ADF Statistic: -8.749390973803425
p-value: 2.8638790488670384e-14
Critical Values:
   1%: -3.439146171679794
   5%: -2.865422101274577
   10%: -2.568837245865348
The log returns are stationary (reject the null hypothesis of unit root).
```