

# Stat 565, HW 1

Artem Ivaniuk

February 2026

## Problem 1

### Description

Find three univariate time series of your interest. Plot the series, and comment on the pattern you observe.

### Solution

Code for plotting is in the attached Jupyter notebook at the end.

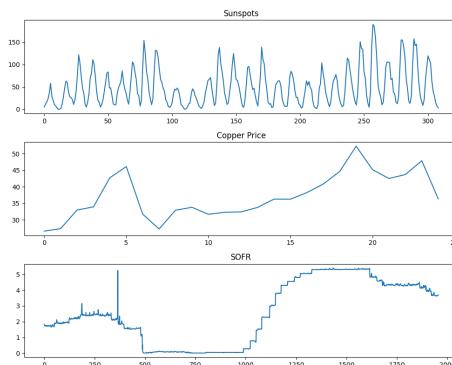


Figure 1: Three series: Sunspots, Copper, SOFR.

The sunspots time series, coming from the Python's statsmodels library's built-in datasets, exhibits clear cyclicity of periods going up and down. It is clearly not stationary, but has repetitiveness in the cycles once the lag is fixed.

The copper price time series is not stationary either but it does not exhibit any features of cyclicity. It has a slight upwards trend with minimal variance apart from moments of pattern reversion.

The SOFR time series is not stationary and shows large cyclical behavior with two cycles/periods of high rates, with a low period in the middle. Clearly not stationary but does not have a fixed lag dependency like the sunspots series.

## Problem 2

### Description

Let  $U_1$  and  $U_2$  be independent random variables with zero means and  $\text{Var}(U_1) = \text{Var}(U_2) = \sigma^2$ . Consider the time series  $x_t = U_1 \sin(2\pi\omega t) + U_2 \cos(2\pi\omega t)$ , where  $\omega \in [0, 1]$  is a fixed constant. Show that the autocovariance function of this process is given by  $\gamma_{t,t+h} := \text{Cov}(x_t, x_{t+h}) = \sigma^2 \cos(2\pi\omega h)$ . Find the autocorrelation function as well.

### Solution

Since  $\mathbb{E}(U_1) = \mathbb{E}(U_2) = 0$ , we have  $\mathbb{E}(x_t) = 0$ . Therefore,

$$\gamma_{t,t+h} = \text{Cov}(x_t, x_{t+h}) = \mathbb{E}(x_t x_{t+h}).$$

Then,  $x_{t+h} = U_1 \sin(2\pi\omega(t+h)) + U_2 \cos(2\pi\omega(t+h))$

Then,

$$\begin{aligned} x_t x_{t+h} &= (U_1 \sin(2\pi\omega t) + U_2 \cos(2\pi\omega t))(U_1 \sin(2\pi\omega(t+h)) + U_2 \cos(2\pi\omega(t+h))) \\ &= U_1^2 \sin(2\pi\omega t) \sin(2\pi\omega(t+h)) + U_2^2 \cos(2\pi\omega t) \cos(2\pi\omega(t+h)) \\ &\quad + U_1 U_2 [\sin(2\pi\omega t) \cos(2\pi\omega(t+h)) + \cos(2\pi\omega t) \sin(2\pi\omega(t+h))]. \end{aligned}$$

Taking expectation, with  $\mathbb{E}(U_1 U_2) = 0$  and  $\mathbb{E}(U_1^2) = \mathbb{E}(U_2^2) = \sigma^2$ , we obtain

$$\gamma_{t,t+h} = \sigma^2 [\sin(2\pi\omega t) \sin(2\pi\omega(t+h)) + \cos(2\pi\omega t) \cos(2\pi\omega(t+h))].$$

Using the identity  $\cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$ , with  $\alpha = 2\pi\omega(t+h)$  and  $\beta = 2\pi\omega t$ , we get  $\gamma_{t,t+h} = \sigma^2 \cos(2\pi\omega h)$ .  $\square$

Finally, the autocorrelation function is

$$\rho(h) = \frac{\gamma_{t,t+h}}{\gamma_{t,t}} = \frac{\sigma^2 \cos(2\pi\omega h)}{\sigma^2} = \cos(2\pi\omega h).$$

## Problem 3

### Description

Let  $\{w, t \in \mathbb{Z}\}$  be iid  $N(0, 1)$ , and consider the time series

$$x_t = w_t w_{t-1}, \quad y_t = x_t^2$$

1. Find the mean, autocovariance, and autocorrelation functions of  $x_t$
2. Simulate  $x_t$  of length  $T = 500$ , give the time series plot, comment
3. Find the mean, autocovariance, and autocorrelation functions of  $y_t$
4. Simulate  $y_t$  of length  $T = 500$ , give the time series plot, comment

## Solution

Let  $\{w_t : t \in \mathbb{Z}\}$  be i.i.d.  $N(0, 1)$ . Then,  $\mathbb{E}(w_t) = 0, \mathbb{E}(w_t^2) = 1, \mathbb{E}(w_t^4) = 3$ .

1.  $\mathbb{E}(x_t) = \mathbb{E}(w_t)\mathbb{E}(w_{t-1}) = 0$ .

Since  $\mathbb{E}(x_t) = 0, \gamma_x(h) = \text{Cov}(x_t, x_{t+h}) = \mathbb{E}(x_t x_{t+h})$ .

For  $h = 0, \gamma_x(0) = \text{Var}(x_t) = \mathbb{E}(w_t^2 w_{t-1}^2) = \mathbb{E}(w_t^2)\mathbb{E}(w_{t-1}^2) = 1$ .

For  $h = \pm 1, \gamma_x(1) = \mathbb{E}(w_t w_{t-1} w_{t+1} w_t) = \mathbb{E}(w_t^2)\mathbb{E}(w_{t-1})\mathbb{E}(w_{t+1}) = 0$ , and similarly  $\gamma_x(-1) = 0$ .

For  $|h| \geq 2$ , the variables are independent, so  $\gamma_x(h) = 0$ .

Hence,

$$\gamma_x(h) = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0, \end{cases} \quad \rho_x(h) = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

2. Code is in the attached Jupyter notebook at the end.

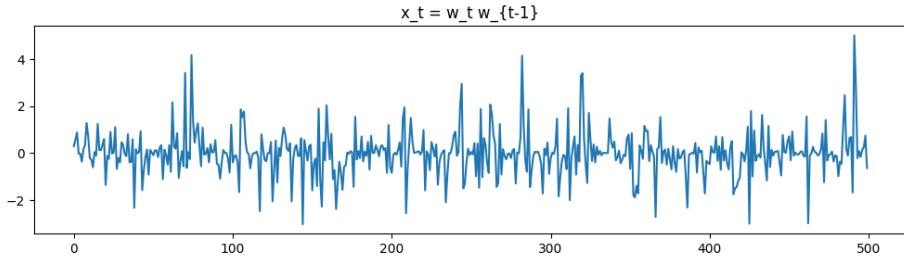


Figure 2: Simulation of  $x_t$

There is no cyclical or patterned behavior to this series, with clear mean of 0 and fairly regular variance in both directions. The only pattern that I would suspect from this is volatility clustering since the variance near large deviations from the mean tends to also be large. Could be a statistically insignificant observation though as a result of a single simulation instance.

3.  $\mathbb{E}(y_t) = \mathbb{E}(w_t^2)\mathbb{E}(w_{t-1}^2) = 1$ .

By definition,  $\gamma_y(h) = \text{Cov}(y_t, y_{t+h}) = \mathbb{E}(y_t y_{t+h}) - [\mathbb{E}(y_t)]^2$ .

For  $h = 0, \mathbb{E}(y_t^2) = \mathbb{E}(w_t^4 w_{t-1}^4) = \mathbb{E}(w_t^4)\mathbb{E}(w_{t-1}^4) = 9$ , so  $\gamma_y(0) = 9 - 1 = 8$ .

For  $h = \pm 1, \mathbb{E}(y_t y_{t+1}) = \mathbb{E}(w_t^4)\mathbb{E}(w_{t-1}^2)\mathbb{E}(w_{t+1}^2) = 3$ , hence  $\gamma_y(1) = \gamma_y(-1) = 3 - 1 = 2$ .

For  $|h| \geq 2$ , independence gives  $\gamma_y(h) = 0$ .

Thus,

$$\gamma_y(h) = \begin{cases} 8, & h = 0, \\ 2, & |h| = 1, \\ 0, & |h| \geq 2, \end{cases} \quad \rho_y(h) = \begin{cases} 1, & h = 0, \\ \frac{1}{4}, & |h| = 1, \\ 0, & |h| \geq 2. \end{cases}$$

4. Code is in the attached Jupyter notebook at the end. Please note that this was generated from the same instance of generated  $w_t$  as the  $x_t$  plot.

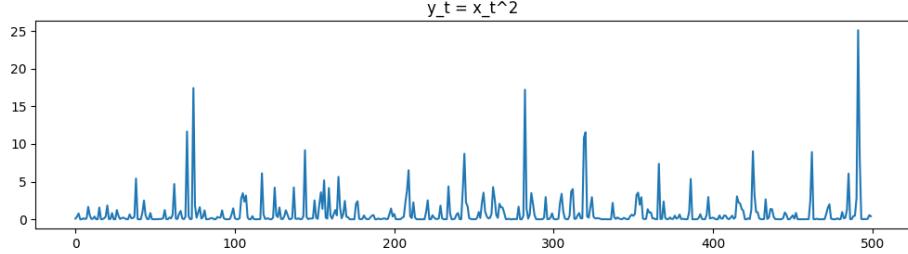


Figure 3: Simulation of  $y_t$

There is a clear pattern of having a mean of 0 but with large jumps of variance towards the positive (squared  $x_t$  cannot be negative). There also seem to be patterns of volatility clustering since the deviations from the mean tend to be large around large jumps of  $y_t$ , which would make sense since  $y_t = x_t = w_t^2 w_{t-1}^2$ , so there is a direct relation to the previous  $w_t$  value. At the same time, there is no cyclicity or other pattern to observe.

## Problem 4

### Description

For each of the three series you found for Homework 01, Problem 1, do the following. Note, if your series  $\{x_t\}$  does not look stationary, try to do the following for the process  $\{y_t\}$ , where  $y_t := x_t - x_{t-1}$

1. Plot the sample autocorrelation function
2. Test whether  $\rho_1$  is zero, using 95% level
3. Perform the Ljung-Box test, using a suitable number (your own choice) of lags

### Solution

Code is in the attached Jupyter notebook at the end, where I used Python's built-in statsmodel library and its implementation of ACF and Ljung-Box test. The sample ACF plots are on the next page, alongside a table of results of the  $\rho_1$  and Ljung-Box tests. Please note that since all three of my series did not seem to be stationary, I have implemented a method that used the  $y_t := x_t - x_{t-1}$  process with the  $x_t$  values of the original time series.

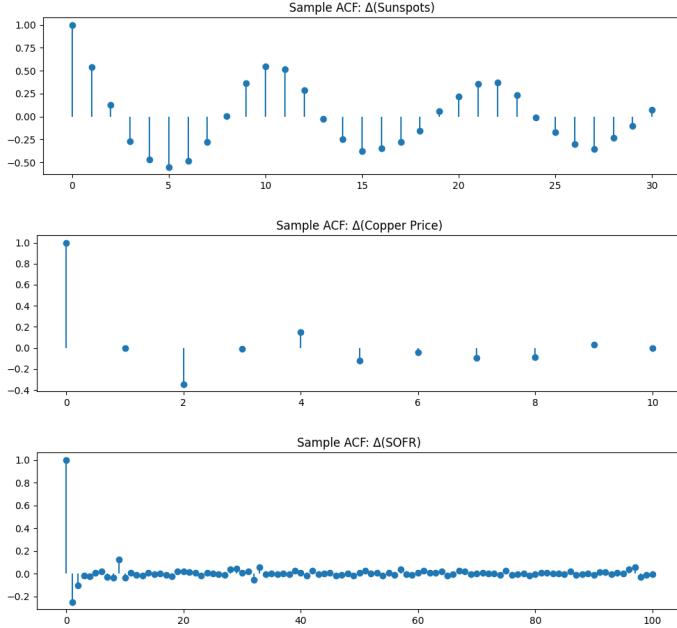


Figure 4: Sample ACFs of subtracted Sunspots, Copper Price, and SOFR series

Series	$n$	$r_1$	p-value ( $\rho_1$ )	Lags	Ljung–Box p-value
$\Delta(\text{Sunspots})$	308	0.541	0.000	15	$2.65 \times 10^{-138}$
$\Delta(\text{Copper Price})$	24	-0.005	0.982	5	0.468
$\Delta(\text{SOFR})$	1955	-0.251	0.000	50	$6.93 \times 10^{-22}$

To test whether  $\rho_1 = 0$ , I used a two-sided test at the 95% level based on the sample autocorrelation at lag 1. For the differenced Sunspots series, the sample autocorrelation  $r_1 \approx 0.54$  and the p-value is essentially zero, so I reject  $H_0 : \rho_1 = 0$ , indicating strong positive autocorrelation. For the differenced Copper Price series,  $r_1 \approx -0.005$  with a p-value of about 0.98, so I fail to reject  $H_0$ , suggesting no evidence of autocorrelation. For the differenced SOFR series,  $r_1 \approx -0.25$  with a p-value essentially zero, so I reject  $H_0$ , indicating significant negative autocorrelation.

For the Ljung–Box test I chose the number of lags based on the length of each series. For the differenced Sunspots series with  $n=308$ , using 15 lags, the p-value is effectively zero, so I reject the null hypothesis of no autocorrelation up to lag 15. For the differenced Copper Price series with  $n=24$ , using 5 lags, the p-value is about 0.47, so I fail to reject the null hypothesis, and the series is consistent with independent white noise. For the differenced SOFR series with  $n=1955$ , using 50 lags, the p-value is approximately  $6.9 \times 10^{-22}$ , leading to rejection of the null hypothesis, indicating serial dependence.