École polytechnique fédérale de Lausanne

Assignement

Time Series

Authors:
Edouard Bueche 319539

Instructor:
Sofia Olhede

This is all my work

Exercise 1

i)

Given the stochastic process defined as:

$$X_t = \rho X_{t-1} + \epsilon_t, \tag{1}$$

where ϵ_t represents a mean-zero Gaussian white noise process with variance σ_ϵ^2 and $|\rho| < 1$. Additionally, it is given that $X_0 \sim N\left(0, \frac{\sigma_\epsilon^2}{1-\rho^2}\right)$ and is uncorrelated with ϵ_t for all $t \in \mathbb{N}$.

First, we compute the expected value of X_t:

$$\begin{aligned} E[X_t] &= E[\rho X_{t-1} + \epsilon_t] && (2)\\ &= \rho E[X_{t-1}] + E[\epsilon_t] && (3)\\ &= \rho E[X_{t-1}] = \cdots = \rho^t E[X_0], && (4) \end{aligned}$$

considering $E[\epsilon_t] = 0$. Since $E[X_0] = 0$, by recursively applying the expectation, we deduce that $E[X_t] = 0$ for all t.

The covariance of X_t is given by:

$$Var(X_t) = \rho^2 Var(X_{t-1}) + \sigma_\epsilon^2 = \rho^4 Var(X_{t-2}) + \sigma_\epsilon^2 \rho^2 + \sigma_\epsilon^2, \tag{5}$$

which you can continue until t=0 and get and geometric sum with reason ρ^2, multiplied by σ_ϵ^2 thus we obtain

$$Var(X_t) = \frac{\sigma_\epsilon^2}{1-\rho^2}$$

For $\tau > 0$, we consider the general covariance structure:

$$\text{Cov}(X_t, X_{t+\tau}) = \text{E}\left[(X_t - \text{E}[X_t])(X_{t+\tau} - \text{E}[X_{t+\tau}])\right] = \text{E}[X_t X_{t+\tau}]$$

Given $\text{E}[X_t] = 0$ for all t, we have:

$$\text{Cov}(X_t, X_{t+\tau}) = \text{E}\left[X_t\left(\rho^\tau X_t + \sum_{i=0}^{\tau-1} \rho^i \epsilon_{t+\tau-i}\right)\right] = \rho^\tau \text{E}[X_t^2] + \sum_{i=0}^{\tau-1} \rho^i \text{E}[X_t \epsilon_{t+\tau-i}]$$

Since X_t is uncorrelated with $\epsilon_{t+\tau-i}$ for $\tau > 0$, the expectation $\text{E}[X_t \epsilon_{t+\tau-i}] = 0$ for all i. Thus, the covariance simplifies to:

$$\text{Cov}(X_t, X_{t+\tau}) = \rho^\tau \text{E}[X_t^2]$$

Using again $E[X_0] = 0$ we get $\text{E}[X_t^2] = \text{Var}(X_t) = \frac{\sigma_\epsilon^2}{1-\rho^2}$, the covariance expression further simplifies to:

$$\text{Cov}(X_t, X_{t+\tau}) = \begin{cases} \frac{\sigma_\epsilon^2}{1-\rho^2} & : \tau = 0 \\ \rho^\tau \frac{\sigma_\epsilon^2}{1-\rho^2} & : \tau \neq 0 \end{cases}$$

Hence it depends only on the lag τ and not on t. It's indeed a second order stationary process

ii)

Let the stochastic process be defined as:

$$X_t = X_{t-1} + \epsilon_t, \tag{6}$$

where ϵ_t represents a mean-zero Gaussian white noise process with variance σ_ϵ^2. It starts with $X_0 = \epsilon_0$.

First, the expected value of X_t:

$$\begin{aligned} E[X_t] &= E[X_{t-1} + \epsilon_t] &(7)\\ &= E[X_{t-1}] + E[\epsilon_t] &(8)\\ &= E[X_{t-1}] = \cdots = E[X_0] = E[\epsilon_0] = 0, &(9) \end{aligned}$$

therefore

$$E[X_t] = 0, \forall t \in \mathbb{N}$$

The variance of X_t is:

$$\begin{aligned} Var(X_t) &= Var\left(\epsilon_0 + \sum_{i=1}^{t} \epsilon_i\right) &(10)\\ &= \sum_{i=0}^{t} Var(\epsilon_i) &(11)\\ &= (t+1)\sigma_\epsilon^2, &(12) \end{aligned}$$

due to the independence of ϵ_i terms.

The variance of X_t explicitly depends on t hence it isn't stationary.

iii)

Let X_t be the following stochastic process:

$$X_t = Y_t - Y_{t-1}, \quad \forall t \in \mathbb{N}_0, \tag{13}$$

where

$$Y_t = \mu_t + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2}, \quad \forall t \in \mathbb{Z}, \tag{14}$$

and $\mu_t = a_0 t + a_1$. Here, ϵ_t represents a mean-zero Gaussian white noise process with variance σ_ϵ^2.

First, we calculate the expected value of X_t:

$$\begin{aligned} E[X_t] &= E[Y_t - Y_{t-1}] &(15)\\ &= E[\mu_t + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} - (\mu_{t-1} + \epsilon_{t-1} + \epsilon_{t-2} + \epsilon_{t-3})] &(16)\\ &= E[a_0 t + a_1 - (a_0(t-1) + a_1) + \epsilon_t - \epsilon_{t-3}] &(17)\\ &= a_0 + E[\epsilon_t] - E[\epsilon_{t-3}] &(18)\\ &= a_0, &(19) \end{aligned}$$

since $E[\epsilon_t] = 0$ for all t.

The variance of X_t is given by:

$$\begin{aligned} Var(X_t) &= Var(a_0 + \epsilon_t - \epsilon_{t-3}) && (20)\\ &= Var(\epsilon_t) + Var(\epsilon_{t-3}) && (21)\\ &= 2\sigma_\epsilon^2, && (22) \end{aligned}$$

assuming ϵ_t are independent with constant variance σ_ϵ^2.

As we can rewrite X_t as $a_0 + \epsilon_t - \epsilon_{t-3}$, we can se that:

$$\text{Cov}(X_t, X_{t+\tau}) = \begin{cases} 2\sigma_\epsilon^2 & : \tau = 0 \\ -\sigma_\epsilon^2 & : \tau = \pm 3 \\ 0 & : \text{otherwise} \end{cases}$$

and this doesn't depend on t, hence we can assert that X_t is a second order stationary process.

iv)

Consider the stochastic process U_t defined as follows:
$U_0 \sim N\left(0, \frac{\sigma_\epsilon^2}{1-\theta^2}\right)$, and $U_t = -\theta U_{t-1} + \epsilon_t, \forall t \in \mathbb{N}$, with $|\theta| < 1$ and ϵ_t being a Gaussian white noise process with variance σ_ϵ^2.
Note that you can write $U_t = (-\theta)^t U_0 + \sum_{i=1}^t \epsilon_i (-\theta)^{t-i}$

The expected value of U_t is:

$$E[U_t] = E[-\theta U_{t-1} + \epsilon_t] = -\theta E[U_{t-1}] + E[\epsilon_t]$$

Since ϵ_t is a Gaussian white noise with mean zero ($E[\epsilon_t] = 0$) we get $E[U_t] = (-\theta)^t E[U_0] = 0$

The covariance at lag τ is given by:

$$\text{Cov}(U_t, U_{t+\tau}) = \text{Cov}((-\theta)^t U_0 + \sum_{i=1}^{t} \epsilon_i (-\theta)^{t-i}, (-\theta)^{t+\tau} U_0 + \sum_{i=1}^{t+\tau} \epsilon_i (-\theta)^{t+\tau-i})$$

It doesn't simplify as well as in point i) because we don't know if the error terms are correlated with U_0 or not
If they are correlated then this process isn't stationary as the equation above, when expanded will depend on t.

If otherwise they are uncorrelated then we have as in i):

$$\text{Cov}(U_t, U_{t+\tau}) = \begin{cases} \frac{\sigma_\epsilon^2}{1-\theta^2} & : \tau = 0 \\ (-\theta)^\tau \frac{\sigma_\epsilon^2}{1-\theta^2} & : \tau \neq 0 \end{cases}$$

Now we know that the U_t is second order stationary but what about X_t?

Define $X_t = \sigma_t U_t, \forall t \in \mathbb{N}_0$, where σ_t is a deterministic non-negative function of t.

Using linearity of the expectation, the fact that σ_t is deterministic and $E[U_t] = 0$ for all t,

$$E[X_t] = E[\sigma_t U_t] = \sigma_t E[U_t] = 0$$

The covariance between X_t and $X_{t+\tau}$ is

$$\text{Cov}(X_t, X_{t+\tau}) = E[\sigma_t U_t \cdot \sigma_{t+\tau} U_{t+\tau}] = \sigma_t \cdot \sigma_{t+\tau} E[U_t U_{t+\tau}] = \sigma_t \cdot \sigma_{t+\tau} \cdot \gamma_\tau$$

If σ_t is a constant function $= c$ then $\text{Cov}(X_t, X_{t+\tau}) = c^2 \cdot \gamma_\tau$, thus is second order stationary but otherwise the covariance of $X_t, X_{t+\tau}$ depends on t and thus its not second order stationary.

Exercise 2

1

Assume we have a time series $X_1, X_2, \ldots, X_n$ that can be modeled as an AR(1) process, i.e.,

$$X_t = \phi X_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots, n, \tag{23}$$

where ϵ_t represents the error term at time t.

Recall that the forward least squares estimator for the AR(1) process is given by:

$$\phi_F = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=1}^{n-1} X_t^2}. \tag{24}$$

and that the Yule-Walker estimator, which is derived from the autocorrelation function, for an AR(1) model is:

$$\phi_{YW} = \frac{\sum_{t=2}^{n}(X_t - \overline{X})(X_{t-1} - \overline{X})}{\sum_{t=1}^{n-1}(X_t - \overline{X})^2}, \tag{25}$$

To derive the Yule-Walker estimator, we start by multiplying equation 23 by X_{t-k} and we get:

$$X_t X_{t-k} = \phi X_{t-1} X_{t-k} + \epsilon_t X_{t-k}$$

Taking expectations on both side yields that $\gamma_k = \gamma_{k-1} \cdot \phi$, thus $\phi = \frac{\gamma_1}{\gamma_0}$

The autocovariances γ_0 and γ_1 are defined as:

$$\gamma_0 = \frac{1}{n} \sum_{t=1}^{n} X_t X_t$$

$$\gamma_1 = \frac{1}{n-1} \sum_{t=2}^{n} X_t X_{t-1}$$

thus the Yule-Walker estimator is

$$\phi_{YW} = \frac{\gamma_1}{\gamma_0} = \frac{n}{n-1} \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=1}^{n} X_t^2}$$

The Forward least square estimator

$$\hat{\phi}_{FWL} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2}$$

by playing around with the sum we get:

$$\hat{\phi}_{FWL} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=1}^{n-1} X_t^2} \cdot \frac{\sum_{t=1}^{n} X_t^2}{\sum_{t=1}^{n} X_t^2} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=1}^{n} X_t^2} \cdot \frac{\sum_{t=1}^{n} X_t^2}{\sum_{t=1}^{n-1} X_t^2}$$

We recognize the Yule-Walker Estimator thus we replace

$$\phi_{YW} \cdot \frac{n-1}{n} \cdot (1 + \frac{X_n^2}{\sum_{t=1}^{n-1} X_t^2})$$

Let $a_n = \dfrac{n-1}{n}$ and $b_n = 1 + \frac{X_n^2}{\sum_{t=1}^{n-1} X_t^2}$.
The limit as n goes to infinity makes both these limits converge to 1 (quickly for a_n and it depends on the data for b_n)
They are efficiency equivalent estimators.

$$\phi_F = \phi_{YW} \text{ when n is large enough.} \tag{26}$$

2

I used R studio to do the following computations: The AR(1) model parameters estimated using **Yule-Walker** estimation are as follows:

Method	ϕ	σ_ϵ^2
Yule-Walker	0.61893	0.8453908
Forward Least Squares	0.6722094	0.7886105

Yule-Walker & Forward Least Squares Estimators

Exercise 3

Consider the autoregressive process of order 2 (AR(2)) defined by:

$$X_t = \frac{1}{2} X_{t-1} + a_2 X_{t-2} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \quad t \in \mathbb{Z},$$

where $a_2 \in \mathbb{R}$.
To ensure the process is stationary, the roots of the characteristic equation must lie outside the unit circle in the

complex plane.
For a general AR(2) model $X_t = a_1 X_{t-1} + a_2 X_{t-2} + \epsilon_t$, the sum of the coefficients $a_1 + a_2$ provides insight into the stationarity of the process.
Specifically, for the process to be stationary, these coefficients, when added, should not exceed 1 in absolute value, so in our case we have:

$$\left|\frac{1}{2} + a_2\right| < 1.$$

So for now we have that $a_2 \in (-\frac{3}{2}, \frac{1}{2})$
Let's see what happend when $a_2 = 0$, the characteristic equation of this AR(1) process will be thus

$$1 - \frac{1}{2}z = 0 \implies z = 2$$

so it will be stationary.
Now if $a_2 \neq 0$, we have :

$$1 - \frac{1}{2}z - a_2 z^2 = 0 \implies a_2 z^2 + \frac{1}{2}z - 1 = 0.$$

The roots are given by the quadratic formula

$$z = \frac{-\frac{1}{2} \pm \sqrt{\left(\frac{1}{2}\right)^2 - 4a_2(-1)}}{2a_2} = \frac{-\frac{1}{2} \pm \sqrt{\frac{1}{4} + 4a_2}}{2a_2}.$$

Disjonction on the discriminant $\Delta = \frac{1}{4} + 4a_2$:
If $\Delta = 0$ then $a_2 = \frac{-1}{16}$ and thus $z = 4$ and the process is stationary.

Let's suppose now that $\Delta < 0$(it means that $a_2 < \frac{-1}{16}$ then $z_{1,2} = \frac{-1}{4a_2} \pm i\frac{\sqrt{-(\frac{1}{4} + 4a_2)}}{2a_2}$, we can therefore compute the module squared of that:

$$|z_{1,2}|^2 = \frac{1}{16a_2^2} - \frac{\frac{1}{4} + 4a_2}{4a_2^2} = -\frac{1}{a_2}$$

If we want that to be greater than 1, as the discriminant is negatif, a_2 also is negativ

$$-\frac{1}{a_2} > 1$$

$$-1 < a_2$$

We found indeed that a_2 can't be more than $\frac{1}{2}$ so the final range for a_2 is:

$$a_2 \in (-1, \frac{1}{2})$$

Exercise 4

i)

The least squares objective function, which we aim to minimize, is defined as:

$$S(\beta) = \sum_{t=0}^{N-1} \left(Y_t - (\beta_0 + \beta_1 t + \beta_2 t^2)\right)^2$$

We vectorize the model as follows:

Here, $\mathbf{X}$ is the design matrix with each row corresponding to the values $1, t, t^2$ for each time t, and $\mathbf{Y}$ is the vector of observed values Y_t.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ 1 & 3 & 3^2 \\ 1 & 4 & 4^2 \\ \vdots & \vdots & \vdots \\ 1 & N-1 & (N-1)^2 \end{bmatrix}$$

$\boldsymbol{\beta}$ is the vector of parameters $[\beta_0, \beta_1, \beta_2]^{\mathrm{T}}$.
$\boldsymbol{\mu}$ is the vector of all μ_t values for $t = 0, \ldots, N-1$. That allows us to write

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

$W_t = \epsilon_t + 0.5\epsilon_{t-1} + 0.5\epsilon_{t-2} + 0.25\epsilon_{t-3}$, with $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.

Given the noise component structure $W_t = \epsilon_t + 0.5\epsilon_{t-1} + 0.5\epsilon_{t-2} + 0.25\epsilon_{t-3}$, we aim to construct a matrix E that encapsulates ϵ_t, ϵ_{t-1}, and ϵ_{t-2} for each time step t, and a coefficient vector C. The goal is to facilitate the computation of part of W through matrix multiplication.

The matrix E is constructed to have $(N-1) \times 4$ dimensions, where each row corresponds to a time step t from 0 to $N-1$

$$E = \begin{bmatrix} \epsilon_0 & \epsilon_{-1} & \epsilon_{-2} & \epsilon_{-3} \\ \epsilon_1 & \epsilon_0 & \epsilon_{-1} & \epsilon_{-2} \\ \epsilon_2 & \epsilon_1 & \epsilon_0 & \epsilon_{-1} \\ \epsilon_3 & \epsilon_2 & \epsilon_1 & \epsilon_0 \\ \vdots & \vdots & \vdots & \vdots \\ \epsilon_{N-1} & \epsilon_{N-2} & \epsilon_{N-3} & \epsilon_{N-4} \end{bmatrix}$$

The coefficient vector C is defined as:

$$C = \begin{bmatrix} 1 \\ 0.5 \\ 0.5 \\ 0.25 \end{bmatrix}$$

Multiplication of E by C yields a vector that represents

$$W = \begin{bmatrix} \epsilon_0 + \frac{1}{2}\epsilon_{-1} + \frac{1}{2}\epsilon_{-2} + \frac{1}{4}\epsilon_{-3} \\ \epsilon_1 + \frac{1}{2}\epsilon_0 + \frac{1}{2}\epsilon_{-1} + \frac{1}{4}\epsilon_{-2} \\ \epsilon_2 + \frac{1}{2}\epsilon_1 + \frac{1}{2}\epsilon_0 + \frac{1}{4}\epsilon_{-1} \\ \epsilon_3 + \frac{1}{2}\epsilon_2 + \frac{1}{2}\epsilon_1 + \frac{1}{4}\epsilon_0 \\ \vdots \\ \epsilon_N + \frac{1}{2}\epsilon_{N-1} + \frac{1}{2}\epsilon_{N-2} + \frac{1}{4}\epsilon_{N-3} \end{bmatrix}$$

Thus it gives the vectorized version of W_t which allows us to write the first equation this way:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}$$

And this rewriting makes it possible to let :

$$S(\beta) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^2$$

Differentiating with respect to β and letting this $= 0$ gives :

$$(\mathbf{X}^T\mathbf{X})\hat{\beta} = \mathbf{X}^T\mathbf{Y}$$

The least squares estimator $\hat{\beta}$ minimizes $S(\beta)$, leading to:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

ii)

If Y_t had a diagonal covariance matrix, by the Gauss-Markov theorem, we would know that our least squares estimator $\hat{\beta}$ would be the Best Linear Unbiased Estimator (BLUE). This means it would have the smallest variance among all linear unbiased estimators of β, under the assumption that the error terms in the linear regression model are uncorrelated(no autocorrelation in the system and have equal variances(homoskedasticity).

iii)

When analyzing the covariance matrix of Y_t, denoted as Σ, for the process defined by $Y_t = \mu_t + W_t$, where $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2$ and W_t includes the stochastic error terms, it is essential to understand the roles of μ_t and W_t in the model.

The term μ_t represents the deterministic component of the model, a function of time t that is fully determined by the parameters β_0, β_1, and β_2. Since μ_t does not involve any randomness, it does not contribute to the variability (and hence the covariance) of the observed values Y_t.

In contrast, W_t represents the stochastic component of the model, incorporating randomness through the error terms ϵ_t and their structured dependencies. The variability in the observed Y_t arises entirely from this stochastic component, making W_t the sole contributor to the covariance matrix Σ.

Therefore, we can write

$$\Sigma = \text{Cov}(Y_t, Y_{t+k}) = \text{Cov}(W_t, W_{t+k})$$

The variance of W_t, considering the definition of W_t

$$W_t = \epsilon_t + 0.5\epsilon_{t-1} + 0.5\epsilon_{t-2} + 0.25\epsilon_{t-3}$$

and the fact that ϵ_t are i.i.d. normal variables with mean 0 and variance σ_ϵ^2, can be computed as follows:

$$\text{Var}(W_t) = \sigma_\epsilon^2 + (0.5^2 + 0.5^2 + 0.25^2)\sigma_\epsilon^2 = (1 + 0.25 + 0.25 + 0.0625)\,\sigma_\epsilon^2 = 1.5625\sigma_\epsilon^2$$

For $k > 0$, the covariance between W_t and W_{t+k} depends on the shared ϵ terms.

where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ and assuming ϵ_t are independent, we calculate the covariances as follows:

Covariance for $k = 1$, Cov(W_t, W_{t-1}):

$$\text{Cov}(W_t, W_{t-1}) = \text{Cov}(\epsilon_{t-1}, 0.5\epsilon_{t-1}) + \text{Cov}(0.5\epsilon_{t-2}, 0.5\epsilon_{t-2}) + \text{Cov}(0.5\epsilon_{t-3}, 0.25\epsilon_{t-3}) = \frac{7}{8}\sigma_\epsilon^2$$

Covariance for $k = 2$, Cov(W_t, W_{t-2}):

$$\text{Cov}(W_t, W_{t-2}) = \text{Cov}(0.5\epsilon_{t-2}, \epsilon_{t-2}) + \text{Cov}(0.25\epsilon_{t-3}, 0.5\epsilon_{t-3}) = 0.5\sigma_\epsilon^2 + 0.125\sigma_\epsilon^2 = \frac{5}{8}\sigma_\epsilon^2$$

Covariance for $k = 3$, Cov(W_t, W_{t-3}):

$$\text{Cov}(W_t, W_{t-3}) = \text{Cov}(0.25\epsilon_{t-3}, \epsilon_{t-3}) = \frac{1}{4}\sigma_\epsilon^2$$

Covariance for $k = 4$, Cov(W_t, W_{t-4}):

$$\text{Cov}(W_t, W_{t-4}) = 0$$

This last result occurs because W_{t-4} does not share any ϵ_t terms with W_t, hence there is no overlap, and the covariance is zero.

The definition of W_t incorporates error terms ϵ_t up to a lag of 3 with decreasing weights. This means that any W_{t-k} where $k \geq 4$ will involve error terms that do not overlap with those in W_t.
Specifically, W_{t-4} would be influenced by ϵ_{t-4}, ϵ_{t-5},ϵ_{t-6} and ϵ_{t-7}, none of which are present in the expression for W_t.Thus it follows that:

$$\text{Cov}(W_t, W_{t-k}) = 0 \quad \text{for} \quad k \geq 4$$

In essence, the autocorrelation structure induced by the W_t process is limited to a finite window of the most recent four lags. Beyond this window, the process does not "remember" its past values, leading to zero covariance between W_t and W_{t-k} for $k \geq 4$.

$$\Sigma = \begin{bmatrix} 1.5625\sigma_\epsilon^2 & 0.875\sigma_\epsilon^2 & 0.625\sigma_\epsilon^2 & 0.25\sigma_\epsilon^2 & 0 & \cdots & 0 \\ 0.875\sigma_\epsilon^2 & 1.5625\sigma_\epsilon^2 & 0.875\sigma_\epsilon^2 & 0.625\sigma_\epsilon^2 & 0.25\sigma_\epsilon^2 & \ddots & \vdots \\ 0.625\sigma_\epsilon^2 & 0.875\sigma_\epsilon^2 & 1.5625\sigma_\epsilon^2 & 0.875\sigma_\epsilon^2 & 0.625\sigma_\epsilon^2 & \ddots & 0 \\ 0.25\sigma_\epsilon^2 & 0.625\sigma_\epsilon^2 & 0.875\sigma_\epsilon^2 & 1.5625\sigma_\epsilon^2 & \ddots & \ddots & 0.25\sigma_\epsilon^2 \\ 0 & 0.25\sigma_\epsilon^2 & 0.625\sigma_\epsilon^2 & \ddots & \ddots & 0.875\sigma_\epsilon^2 & 0.625\sigma_\epsilon^2 \\ \vdots & \ddots & \ddots & \ddots & 0.875\sigma_\epsilon^2 & 1.5625\sigma_\epsilon^2 & 0.875\sigma_\epsilon^2 \\ 0 & \cdots & 0 & 0.25\sigma_\epsilon^2 & 0.625\sigma_\epsilon^2 & 0.875\sigma_\epsilon^2 & 1.5625\sigma_\epsilon^2 \end{bmatrix}$$

This matrix is symmetric, with the diagonal elements representing the variance of W_t and the off-diagonal, we can clearly see that this matrix isn't diagonal at all, this is due to the lags in ϵ making W_t correlated for a few steps.

iv)

Given that U is the matrix of eigenvectors of the covariance matrix Σ, and Λ is the diagonal matrix of the corresponding eigenvalues, we define the transformation of the vector Y as $Z = U^T Y$. Here, we calculate the mean and covariance of Z.

Mean of Z Considering $Y = \mu + W$, where μ represents the mean vector of Y and W denotes the stochastic component with a mean of zero:

$$E[Z] = E[U^T Y] = U^T E[Y]$$

Assuming $E[Y] = \mu$ and W has zero mean, the mean of Z is given by:

$$E[Z] = U^T \mu$$

The covariance of Z can be expressed as follows:

$$\text{Cov}(Z) = E\left[(Z - E[Z])(Z - E[Z])^T\right]$$

Substituting $Z = U^T Y$:

$$\text{Cov}(Z) = E\left[(U^T Y - U^T \mu)(U^T Y - U^T \mu)^T\right] = U^T E\left[(Y - \mu)(Y - \mu)^T\right] U$$

Given that $E\left[(Y - \mu)(Y - \mu)^T\right] = \Sigma$, the covariance matrix of Y, we have:

$$\text{Cov}(Z) = U^T \Sigma U$$

Since U contains the eigenvectors of Σ and Λ its eigenvalues, by the property of eigendecomposition:

$$\text{Cov}(Z) = \Lambda$$

The mean of Z, $E[Z]$, is obtained by transforming the mean vector of Y, μ, with U^T, resulting in $U^T\mu$. The covariance of Z, $\text{Cov}(Z)$, simplifies to Λ, the diagonal matrix of eigenvalues of Σ. This indicates that in the transformed space defined by U^T, the components of Z are uncorrelated with variances equal to the eigenvalues of the original covariance matrix Σ.

This transformation, therefore, diagonalizes the covariance matrix, turning potentially correlated variables Y into uncorrelated ones Z with variances given by the eigenvalues of Σ.

v)

Given that $Z = U^T Y$, where U is the matrix of eigenvectors of the covariance matrix Σ, and Λ is the diagonal matrix of the corresponding eigenvalues, we aim to calculate the least squares solution for β from Z. To achieve this, we first transform the design matrix X into $\tilde{X}$ using the transformation U^T exactly the same way that we crafted Z from Y, and then apply the least squares formula in this transformed space.

To align with the transformation applied to Y to obtain Z, we transform X as follows:

$$\tilde{X} = U^T X$$

The least squares estimate of β, denoted as $\hat{\beta}$, in the transformed space is given by the formula:

$$\hat{\beta} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T Z$$

Using the definition of Z and X we developp:

$$\hat{\beta} = ((U^T X)^T U^T X)^{-1}(U^T X)^T U^T Y = (X^T U U^T X)^{-1} X^T U U^T Y = (X^T X)^{-1} X^T Y$$

Which is exactly the least square estimator we had earlier.

Exercise 5

Consider the autoregressive moving average (ARMA) process specified by:

$$Z_t = 0.5Z_{t-1} + 0.1Z_{t-2} + \varepsilon_t + 0.5\varepsilon_{t-1}, \quad t = 2, 3, \dots \tag{27}$$

where $\varepsilon_t \sim N(0, \sigma^2)$.

(i) Characteristic Polynomial of the Autoregressive Part

The characteristic polynomial for the autoregressive (AR) part of the process can be obtained by considering the homogeneous equation:

$$0.1\lambda^2 - 0.5\lambda - 1 = 0 \tag{28}$$

(ii) Roots of the Characteristic Polynomial

The roots of the characteristic polynomial are found using the quadratic formula:

$$\lambda = \frac{-0.5 \pm \sqrt{(-0.5)^2 - 4 \cdot (-1) \cdot 0.1}}{2 \cdot 0.1} \tag{29}$$

Simplifying, we get:

$$\lambda = -2.5 \pm \frac{\sqrt{0.25 + 0.4}}{0.2} = -2.5 \pm 4.03 \tag{30}$$

It yields -6.53 and 1.53 For autoregressive AR processes, stationarity requires that the roots of the characteristic polynomial lie outside the unit circle. Therefore we know that the process is stationnary.