1: Is the following AR(2) process:

$$X_t = 1.92 - 1.1X_{t-1} + 0.18X_{t-2} + \epsilon_t,\ \epsilon_t \sim N(0,1)$$

covariance stationary? If so, calculate its mean and all auto-covariances.

The process is covariance stationary if all the roots of its characteristic equation are outside of the unit circle i.e. their absolute value is strictly larger than 1.
The characteristic equation is:

$$1 + 1.1z - 0.18z^2$$

Its roots are:

$$z_1 = \frac{55+\sqrt{4825}}{18}; z_2 = \frac{55-\sqrt{4825}}{18}$$

$|z_1| \approx 6.91; |z_1| \approx 0.8 < 1$
$\implies$ the process is not covariance stationary

To convince us further this in fact the case, let's try to simulate this process (*Code section 1.1*).
The series grows without bound and indeed they intuitively look non-stationary (variance across time is not preserved, the series is not homoscedastic).

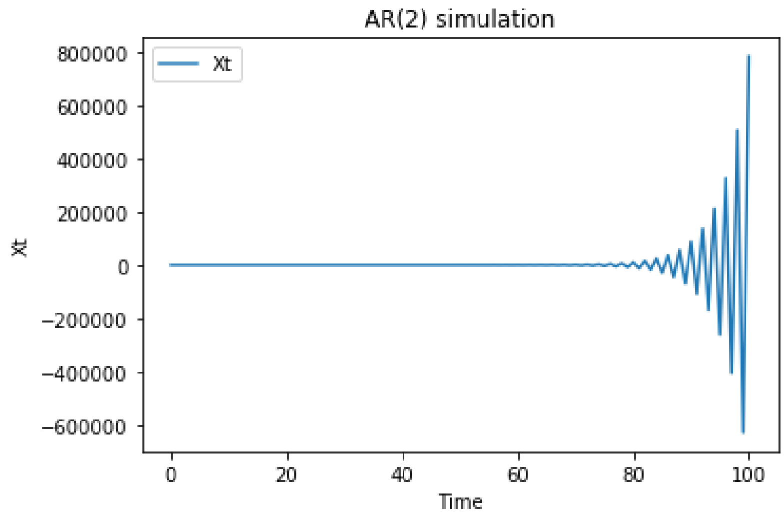

Figure 1: Simulation of the AR(2) process in Problem 1

2: Carry out a detailed discussion of the conditions on β_1 and β_2 under which the AR(2) time series

$$X_t = \alpha + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \epsilon_t,\ \epsilon_t \sim N(0,\sigma^2)$$

is covariance stationary

An AR(2) process can be expressed by

$$X_t - \mu = +\beta_1(X_{t-1} - \mu) + \beta_2(X_{t-2} - \mu) + \epsilon_t \tag{1}$$

Multiply X_{t-k} on both sides of (1) and take its expectation:

$$\Gamma_k = \beta_1 \Gamma_{k-1} + \beta_2 \Gamma_{k-2}$$

Let x_1 and x_2 be the roots of $\phi(x) = 1 - \beta_1 x - \beta_2 x^2$.

Based on this recursion formula, we have the general form for Γ_k

$$\Gamma_k = a(\frac{1}{x_1})^k + b(\frac{1}{x_2})^k \tag{2}$$

If we plug Γ_0 and Γ_1 in, we can get values of a and b.

To stop Γ_k from blowing up in (2), we must have x_1 and x_2 on or outside the unit root in the complex plane.

Now we show x_1 and x_2 can not be on the unit circle by contradiction:

Assume $z = a + bi$ is on the unit circle, which means $a^2 + b^2 = 1$. We plug it in $\phi(x)$:

$$1 - \beta_1(a + bi) - \beta_2(a^2 - b^2 + 2abi) = 0$$

which equals to

$$1 - \beta_1 a - \beta_2(a^2 - b^2) = 0, \qquad \beta_1 b + 2ab\beta_2 = 0$$

1) when $b = 0$, $z = \pm 1$, we have $(1 - \beta_2^2) = \beta_1^2$
2) when $b \neq 0$, we have $\beta_1 = -2a\beta_2$ and $a + 2a^2\beta_2 - \beta_2(2a^2 - 1) = 0$, so we have $\beta_2 = -1$

Recall that

$$\Gamma_0 = \frac{(1 - \beta_2)\sigma^2}{(1 + \beta_2)(1 - \beta_1 - \beta_2)(1 + \beta_1 - \beta_2)} \tag{3}$$

So under both 1) and 2), Γ_0 is undefined in (3) and the roots of $\phi(x)$ cannot be on the unit circle.

Therefore, the roots of $\phi(x)$ must lie outside of the unit circle. □

Now, we want to find conditions for the coefficients β_1 and β_2 of an AR(2) process that determine the series to be covariance stationary.

We have just proved that the roots of the characteristic equation $\phi(x) = 1 - \beta_1 x - \beta_2 x^2$ must lie outside the unit circle.
If we make the substitution for one of the roots $x^{-1} = \lambda$ we can express the characteristic equation as:

$$\lambda^2 - \beta_1 \lambda - \beta_2 = 0 \tag{4}$$

We call this the inverse characteristic equation and note that the roots of this equation must lie inside the unit circle for the series to be stationary since $|x| > 1 \iff |\lambda| < 1$.

The roots of this equation are given by:

$$|\lambda_{1,2}| = \left| \frac{\beta_1 \pm \sqrt{\beta_1 + 4\beta_2}}{2} \right| < 1$$

$$\implies |\lambda_1 \lambda_2| = |\beta_2| < 1 \tag{5}$$

Considering the definition of stationarity, we need to have a finite mean, for an AR(2) process:

$$\mu = \alpha + \beta_1 \mu + \beta_2 \mu$$
$$\implies \mu = \frac{\alpha}{1 - \beta_1 - \beta_2}$$
$$\implies 1 - \beta_1 - \beta_2 \neq 0$$

$$\implies \beta_1 + \beta_2 \neq 1 \tag{6}$$

Additional conditions for β_1 and β_2 can be determined using the Yule-Walker equations. Since Γ_0 represents the variance of X_t, the expression in (3) has to be positive.

We know $(1 - \beta_2)$ is greater than zero since $|\beta_2| < 1$.

Therefore, the denominator $(1 + \beta_2)(1 - \beta_1 - \beta_2)(1 + \beta_1 - \beta_2)$ needs to be greater than zero.

The factor $(1 + \beta_2)$ is greater than zero, we need $(1 - \beta_1 - \beta_2)(1 + \beta_1 - \beta_2)$ to be also greater than zero.

The only way for this to happen is for both factors to be greater than zero, it is not possible to have the case of both being negative since:

$$(1 - \beta_1 - \beta_2) < 0 \quad \text{and} \quad (1 + \beta_1 - \beta_2) < 0$$
$$\implies \beta_1 > (1 - \beta_2) \quad \text{and} \quad \beta_1 > -(1 - \beta_2)$$

So we have for both positive factors:

$$(1 - \beta_1 - \beta_2) > 0 \quad \text{and} \quad (1 + \beta_1 - \beta_2) > 0$$

$$\implies \beta_1 + \beta_2 < 1 \quad \text{and} \quad \beta_2 - \beta_1 < 1 \tag{7}$$

To sum up, we have three conditions for stationarity of the AR(2) process given by equations (5),(6) and (7):

- $|\beta_2| < 1$
- $\beta_1 + \beta_2 < 1$
- $\beta_2 - \beta_1 < 1$

The inequalities define a triangular area where the series AR(2) will be stationary depending on the values of β_1 and β_2:

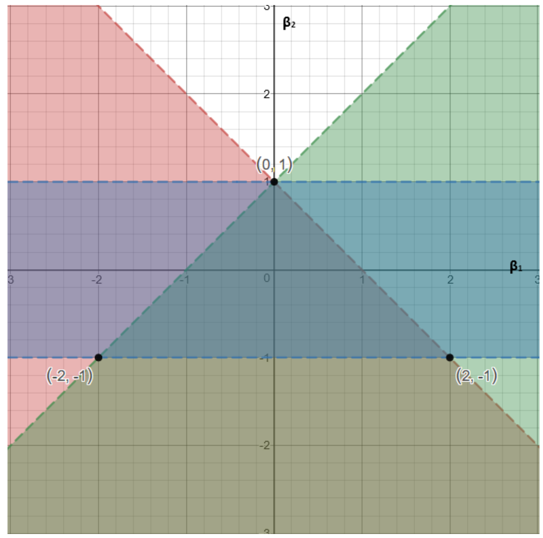

Figure 2: Area where AR(2) is stationary

3: Consider the time series of daily returns on two ETFs tracking broad market indices: SPY and IWV over the last 10 years, and let X_t denote the difference of these returns. Try to model X_t as an AR(p) time series model, and discuss the results.

We retrieve daily adjusted close data for the two ETFs from Yahoo! Finance, calculate daily returns and set our series to be the difference SPY - IWV (*Code section 3.1*)

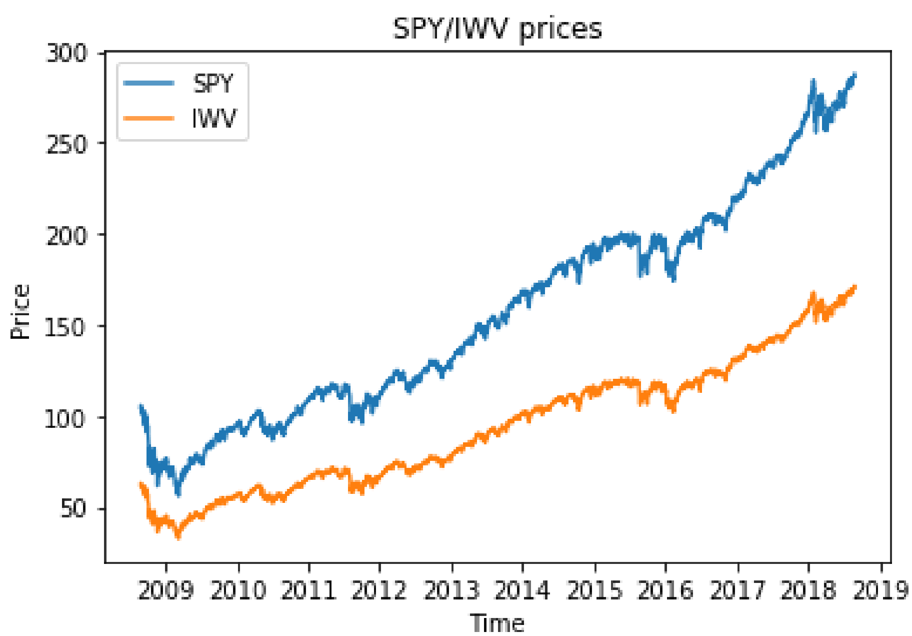

Figure 3: ETF prices: SPY and IWV

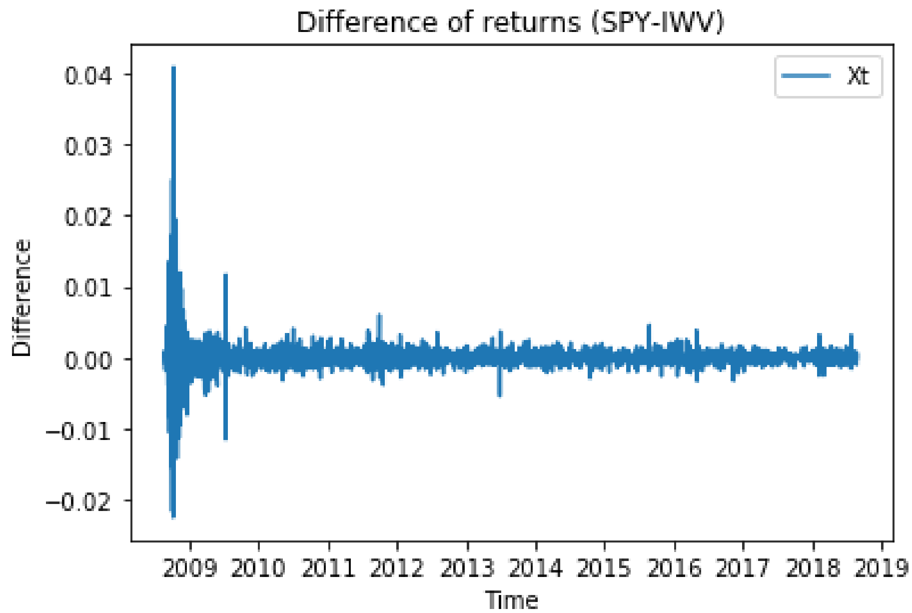

Figure 4: Difference of returns

First, we need to confirm the resulting series is covariance stationary. From the plot, it certainly looks more or less so with the exception of a seemingly larger than average variance around 2008-2009. We can look at the definition of stationarity and perform simple tests over the mean, variance and correlation of the series as an initial diagnostic tool. We know that, for a series to be stationary:

- the expected value for any t is a constant μ.
- the variance does not change through time.
- the auto-covariance only depends on the distance between the two points.

We can split the data into several sections and calculate the mean and variance of each group to compare (*Code section 3.2*)

	Mean	**Variance**
Section 1	-0.000004	0.000010
Section 2	-0.000007	0.000001
Section 3	0.000011	0.000001

It seems that, in regards to the first and second conditions, this series can be stationary.

For the third condition, we can look at the auto-correlation and partial auto-correlation plots. The auto-correlation (ACF) and partial auto-correlation function (PACF) plots hint that the series might not be stationary. There are significant changes in the values of correlation even beyond lag 20 (*Code section 3.3*).

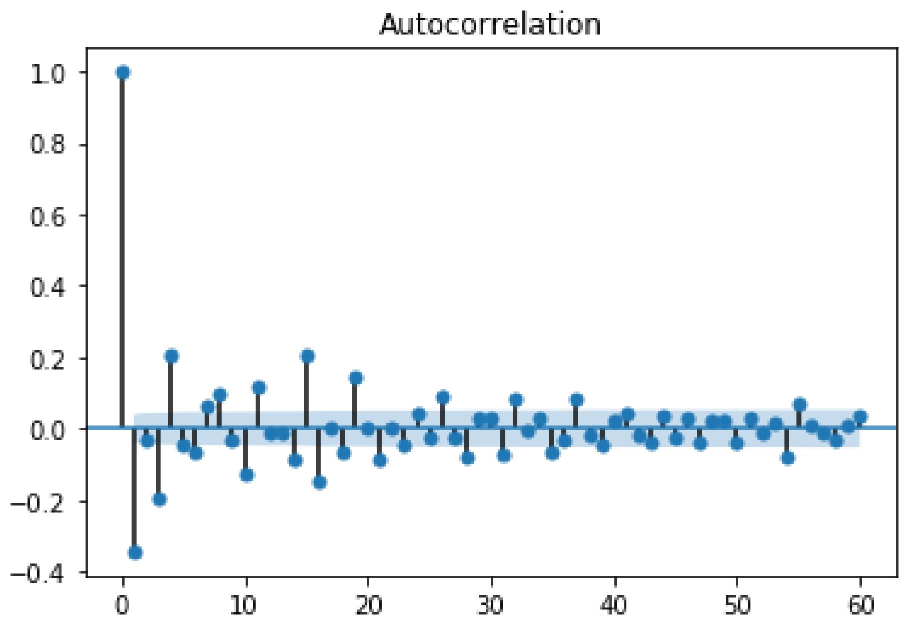

Figure 5: Auto-correlation plot

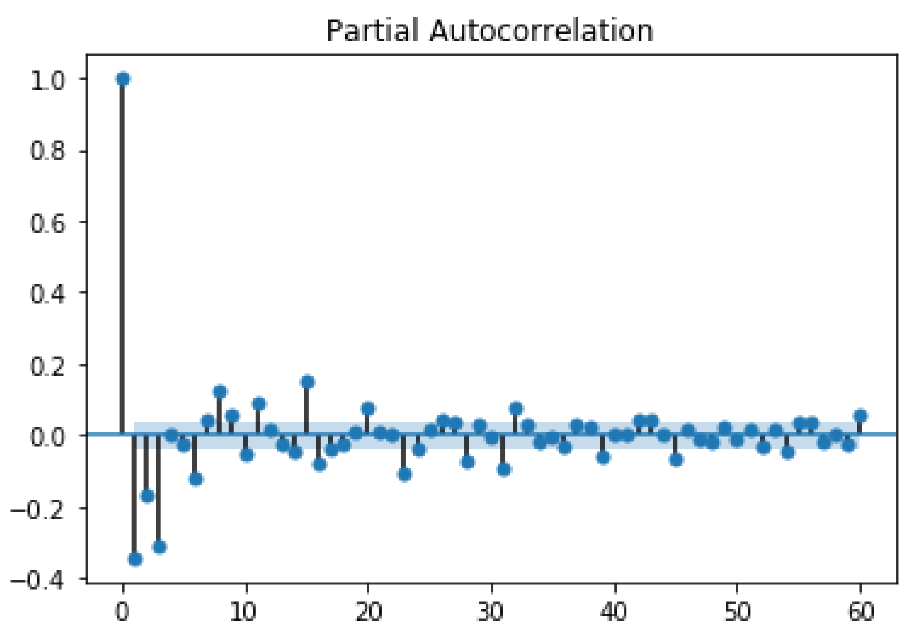

Figure 6: Partial auto-correlation plot

Another option to check for stationarity is to perform an statistical test like the augmented Dickey-Fuller (ADF) test.
The null hypothesis of the ADF test is that the time series is not stationary.
The alternate hypothesis is that the time series is stationary.
We interpret this result using the p-value from the test. A p-value below a threshold (normally 5% or 1%) suggests we reject the null hypothesis (stationary), otherwise a p-value above the threshold suggests we fail to reject the null hypothesis (non-stationary) (*Code section 3.4*).

ADF Statistic	-10.367593
p-value	0.000000
1%	-3.433
5%	-2.863
10%	-2.567

According to the above results, we can clearly assume that the series is stationary and we can try to fit this data to an AR process.
A way to do this is to decide on a maximum number of lags and calculate the AIC value for all of the possible models.
The parameter p resulting from the model with the minimum of the AIC values is chosen. Having chosen the maximum number of lags to be 20, the algorithm found that including most of the parameters would improve the AIC value (*Code section 3.5*).

Lags	AIC
1	-12.607
2	-12.635
3	-12.734
4	-12.734
5	-12.734
6	-12.748
7	-12.749
8	-12.765
9	-12.767
10	-12.769
11	-12.777
12	-12.777
13	-12.777
14	-12.778
15	-12.8
16	-12.805
17	-12.806
18	-12.806
19	-12.806
20	-12.81

Again, the auto-correlation (ACF) and partial auto-correlation function (PACF) plots are an useful tool to understand what might be happening.
The PACF will display the partial correlation which each of the lagged values and one would expect that, for an ideal AR(p) process, the significance of correlation values beyond p to disappear.
This is not the case when the graph is generated for the time series in study. We see there is an important amount of correlation with the previous 3 days but also that the importance of the lag values does not really fade when increasing the lag amount.
We can observe the same in the ACF plot, there are significant values of correlation with lag values like 4, 8 and 15 among others which might be an indication of the presence of weekly seasonality in the series.

There are indications that an AR model would not fit the data very well, for now, and as a rough approximation, let's try to fit the data to an AR(3) model since the partial autocorrelation graph has highlighted the fact that there is strong correlation with these lagged values. The parameters found using maximum likelihood function optimization are (*Code section 3.6*):

$\Rightarrow$

constant	0.000002
L1	-0.453945
L2	-0.292082
L3	-0.308429

Compare with the parameters obtained for an AR(20) fit. The values beyond 3 are smaller in scale but significant (see, for example the L8 value).

constant	0.000003
L1	-0.439352
L2	-0.267585
L3	-0.332349
L4	-0.036363
L5	-0.007929
L6	-0.038296
L7	0.094632
L8	0.161325
L9	0.075421
L10	-0.009118
L11	0.071307
L12	0.020420
L13	-0.039292
L14	-0.012794
L15	0.093710
L16	-0.099706
L17	-0.027465
L18	-0.002121
L19	0.043944
L20	0.077198

4: (Bonus Problem) Show that the AR(1) time series model

$$X_t = \alpha + \beta X_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2)$$

with $0 < \beta < 1$ can be viewed as a result of discretization of the continuous time Ornstein-Uhlenbeck process:

$$dX_t = \lambda(\mu - X_t)dt + \gamma dW_t$$

where $\lambda, \gamma > 0$. Find the mapping between the parameters of these two models.

We solve the SDE and then match the parameters.

$$dX_t = \lambda(\mu - X_t)dt + \gamma dW_t$$

$\Rightarrow$

$$dX_t + \lambda X_t dt = \lambda\mu dt + \gamma dW_t$$

$\Rightarrow$

$$e^{\lambda t} dX_t + \lambda e^{\lambda t} X_t dt = \lambda\mu e^{\lambda t} dt + \gamma e^{\lambda t} dW_t$$

$\Rightarrow$

$$de^{\lambda t}X_t = \mu de^{\lambda t} + \gamma e^{\lambda t} dW_t$$

Integrate from X_t to X_{t+1}:

$$e^{\lambda(t+1)}X_{t+1} - e^{\lambda t}X_t = \mu(e^{\lambda(t+1)} - e^{\lambda t}) + \gamma \int_t^{t+1} e^{\lambda k}\, dW_k$$

Compare the above equation with AR model:

$$X_t = \alpha + \beta X_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$

Thus, we first know that :

$$\alpha = \mu(1 - e^{-\lambda}),\ \beta = e^{-\lambda}$$

Consider the term

$$\gamma e^{-\lambda(t+1)} \int_t^{t+1} e^{\lambda k}\, dW_k$$

$$= \gamma e^{-\lambda} \int_0^1 e^{\lambda k}\, dW_k$$

It is actually random distributed with mean 0 and variance

$$\gamma^2 e^{-2\lambda} \int_0^1 e^{2\lambda k}\, d_k$$
$$= \frac{\gamma^2(1 - \ e^{-2\lambda})}{2\lambda}$$

Thus,

$$\epsilon_t \sim N(0, \frac{\gamma^2(1 - \ e^{-2\lambda})}{2\lambda})$$
$$\sigma = \gamma\sqrt{\frac{(1 - \ e^{-2\lambda})}{2\lambda}}$$

The answer is :

$$\alpha = \mu(1 - e^{-\lambda}),\ \beta = e^{-\lambda}$$
$$\sigma = \gamma\sqrt{\frac{(1 - \ e^{-2\lambda})}{2\lambda}}$$