PARAMETER AND STATE ESTIMATION

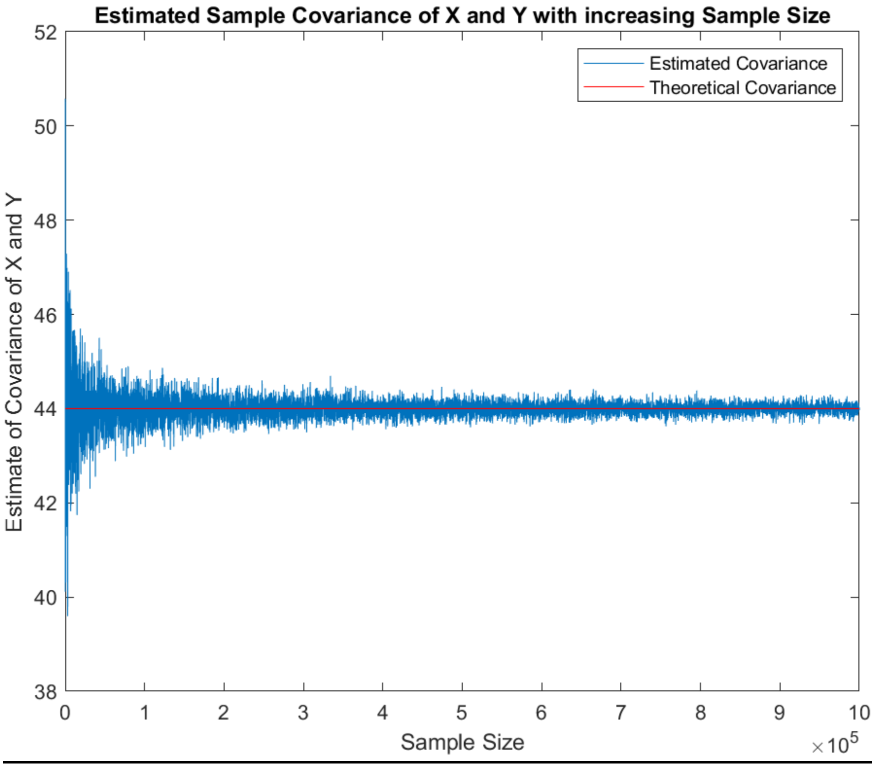
ASSIGNMENT 2

G PRASHANT (BS17B011)

Using Simulation to show the convergence of Covariance

The sample covariances of Random Variables X and Y were estimated with varying sample sizes N, from 10^2 to 10^6, in steps of 100. As calculated in the previous page, the theoretical value of the covariance is 44.

The plot of the estimated sample covariance against the sample size, along with the marked theoretical value is shown below:



From the plot, it can be observed that as the sample size increases, the sample covariance estimate (blue) of random variables X and Y tends to the theoretical value of 44 (red line), however with some fluctuations. Hence, as we keep increasing N, or when $N \to \infty$, $\widehat{\sigma_{yx}} \to 44$.

G. PRASHANT
BS17B011

CH5115: Parameter And State Estimation

ASSIGNMENT 2

Question 1

Part (a)

The random process $\{v[k]\}$ is given as

$$v[k] = A\cos^2(2\pi f k + \phi)$$

where only A is a random variable with 0 mean and unit variance.

For the random process to be covariance stationary,

$$E(v[k]) = \mu_k = \mu \quad \forall\ k > 0$$

and
$$\text{Cov}(v[k], v[k-l]) = \sigma_{vv}[k, k-l] = \sigma_{vv}(l) \quad \forall\ k > l$$

Calculating $E(v[k])$,

$$E(v[k]) = \int_{-\infty}^{\infty} v[k]\, p(A)\, dA \text{, where } p(A) \text{ is the distribution of } A$$

$$= \int_{-\infty}^{\infty} A\cos^2(2\pi f k + \phi)\, p(A)\, dA$$

$$= \cos^2(2\pi f k + \phi)\, E(A)$$

$$= 0$$

$$\therefore\ E(v[k]) = 0 \quad \forall\ k.$$

Calculating $\sigma_{vv}[K, K-\ell]$

$$\sigma_{vv}[K, K-\ell] = E\left((v[K]-\mu_K)(v[K-\ell]-\mu_{K-\ell})\right)$$

$$= E\left(v[K]\, v[K-\ell]\right) \qquad \text{since } \mu_K = 0 \;\forall K$$

$$= \int_{-\infty}^{\infty} A^2 \cos^2(2\pi f K+\phi) \cos^2(2\pi f(K-\ell)+\phi)\, p(A)\, dA$$

$$= \cos^2(2\pi f K+\phi) \cos^2(2\pi f(K-\ell)+\phi)\; E(A^2)$$

$$\text{since } \quad E(A^2) = Var(A^2) + (E(A))^2 = 1$$

$$\sigma_{vv}[K, K-\ell] = \cos^2(2\pi f K+\phi) \cos^2(2\pi f(K-\ell)+\phi)$$

$$\neq \sigma_{vv}(\ell)$$

It can be noticed that even though the process is mean stationary with $\mu_K = 0 \;\forall K$, the autocovariance function is not independent of K.

Therefore, the random process $v[K] = A\cos^2(2\pi f K+\phi)$ is $\boxed{\text{not covariance stationary}}$.

Part (b)

Given: a random walk process

$$v[K] = v[K-1] + e[K]$$

where $e[K]$ is a white noise stationary process.

Calculating $E(v[K])$

$$E(v[K]) = E[v[K-1]) + E[e[K])$$

Since $e[K]$ is white noise, $E(e[K]) = 0$

Hence, $E(v[K]) = E(v[K-1]) = E(v[0]) = \mu_v$

Hence, mean is invariant of time

Calculating variance,

$$E((v[K]-\mu_v)^2) = E(v[K]^2) - \mu_v^2$$

$$= E(v[K-1]^2) + E(e[K]^2)$$

$$- 2E(v[K-1]e[K]) - \mu_v^2$$

Since the noise $e[K]$ is independent of signal $v[K-1]$,

$E(v[K-1]e[K]) = E(v[K-1])E(e[K]) = 0$.

Also, let $var(e[K]) = E(e[K]^2) = \sigma_e^2$

This gives,

$$var(v[K]) = E(v[K-1]^2) + \sigma_e^2 - \mu_v^2$$

$$\cancel{= E(v[K-2]^2) + 2\sigma_e^2 - \mu_v^2}$$

$$= var(v[K-1]) + \sigma_e^2$$

$$= var(v[K-2]) + 2\sigma_e^2$$

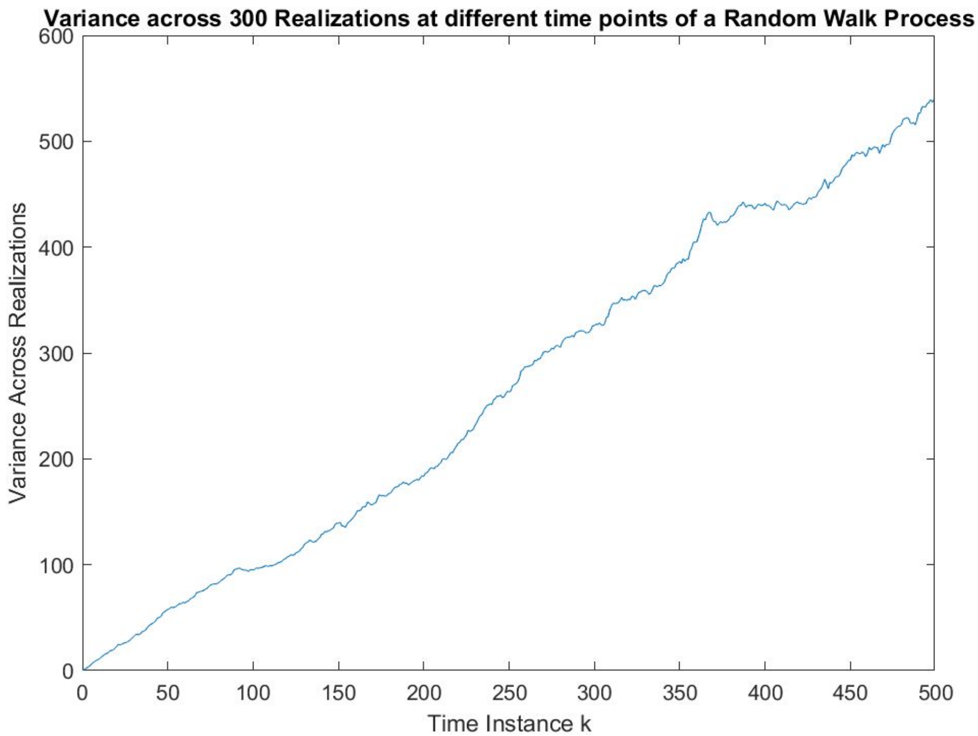
simplifying, $\boxed{var(v[K]) = var(v[0]) + K\sigma_e^2.}$

The variance of $v[K]$ is not independent of K, and this implies that the random walk process is variance-nonstationary.

In the above expression, if $var(v[0]) = 0$, then the variance of $v[k]$ becomes

$$var(v[k]) = k\sigma_e^2$$

MATLAB simulation of Gaussian White Noise (Mean = 0; Variance = 1)



From the above plot, it is evident that the variance of the random process is not stationary with respect to time, it rather increases with time instance k. As derived above, it scales linearly with k.

Question 2

Part (a)

Given $y[k] = y^*[k] + e[k]$, $e[k] \sim WN(0, \sigma_e^2)$

We assume $y[k]$ is stationary process (wide sense)

$E(y[k]) = E(y^*[k]) + 0$

$var(y[k]) = \sigma_y^2 = \sigma_{y^*}^2[k] + \sigma_e^2 + 2E(y^*[k]e[k])$

Since $\sigma_{eu}[l] = 0 \ \forall l$ and $u[k]$ is the driving force for y^*, we can say that $e[k]$ is completely uncorrelated to $y^*[k]$ as well, i.e. $\sigma_{ey^*}[l] = 0 \ \forall l$.

Hence, $\sigma_y^2 = var(y^*[k]) + \sigma_e^2 + 0$

$= \sigma_{y^*}^2 + \sigma_e^2$

This implies both mean and variance of y^* are time invariant.

$u[k] \sim WN(0, \sigma_u^2)$

We are given

$$y^*[k] = \frac{b_2^0 q^{-2}}{1 + f_1^0 q^{-1}} u[k]$$

which gives,

$$y^*[k] + f_1^0 y^*[k-1] = b_2^0 u[k-2] \quad -(1)$$

Squaring equation (1),

$$y^*[k]^2 + f_1^2 y^*[k-1]^2 + 2f_1^0 y^*[k] y^*[k-1] = b_2^{o2} u[k-2]^2$$
$$-(2)$$

Taking expectation on both sides of equation (2)

$$E(y^*[k]^2) + f_1^{o2} E(y^*[k-1]^2) + 2f_1 E(y^*[k] y^*[k-1])$$
$$= b_2^{o2} E(u[k-2]^2)$$
$$= b_2^{o2} \sigma_u^2 \quad -(3)$$

To find $E(y^*[k])$ we take expectation on (1)

$$E(y^*[k]) + f_1^o E(y^*[k-1]) = b_2^o E(u[k-2])$$
$$= 0$$

As $E(y^*[k]) = E(y^*[k-1]) = \mu_{y^*}$

$$\mu_{y^*}(1+f_1^o) = 0$$
$$\Rightarrow \mu_{y^*} = 0$$

Rewriting equation (3) in terms of variances and autocovariance, we get.

$$\sigma_{y^*}^2 + f_1^{o2} \sigma_{y^*}^2 + 2f_1^o \sigma_{y^*y^*}[1] = b_2^{o2} \sigma_u^2$$

or $$\sigma_{y^*}^2 (1 + f_1^{o2}) + 2f_1^o \sigma_{y^*y^*}[1] = b_2^{o2} \sigma_u^2 \quad -(4)$$

Now, we multiply $y^*[k-1]$ with equation (1).

$$y^*[k] y^*[k-1] + f_1^o y^*[k-1]^2 = b_2^o u[k-2] y^*[k-1] \quad -(5)$$

Taking expectation of equation (5)

$$\sigma_{y^*y^*}[1] + f_1^o \sigma_{y^*}^2 = \sigma_{y^*u}[1]. \quad -(6)$$

Now, in order to derive $\sigma_{y*u}[1]$, we generate two equations by ~~gener~~ multiplying $u[k]$, and $u[k-1]$ to equation (1), and take expectation of them. The equations are as follows:

$$\sigma_{y*u}[0] + f_1^\circ \sigma_{y*u}[1] = 0 \quad -(7)$$

$$\sigma_{y*u}[1] + f_1^\circ \sigma_{y*u}[0] = 0 \quad -(8)$$

Solving (7) and (8), we get

$$\sigma_{y*u}[0] = \sigma_{y*u}[1] = 0$$

Going back to equation (6)

$$\sigma_{y*y*}[1] + f_1^\circ \sigma_{y*}^2 = 0$$

or $$\sigma_{y*y*}[1] = -f_1^\circ \sigma_{y*}^2 \quad -(9)$$

Substituting in equation (4).

$$\sigma_{y*}^2 (1 + f_1^{\circ 2}) - 2 f_1^{\circ 2} \sigma_{y*}^2 = b_2^{\circ 2} \sigma_u^2$$

$$\Rightarrow \sigma_{y*}^2 = \frac{b_2^{\circ 2} \sigma_u^2}{1 - f_1^{\circ 2}}$$

Hence, $$\sigma_y^2 = \sigma_{y*}^2 + \sigma_e^2$$

$$\therefore \boxed{\sigma_y^2 = \frac{b_2^{\circ 2} \sigma_u^2}{1 - f_1^{\circ 2}} + \sigma_e^2}$$

To determine $\sigma_{yy}[1]$

or

$$\sigma_{yy}[1] = \sigma_{y*y*}[1] + 0$$

From equation (9)

$$\sigma_{y*y*}[1] = \frac{-f_1^\circ {b_2^\circ}^2 \sigma_u^2}{1-f_1^2}$$

Hence,

$$\boxed{\sigma_{yy}[1] = \frac{-f_1^\circ {b_2^\circ}^2 \sigma_u^2}{1-f_1^2}}$$

To determine $\sigma_{yu}[1]$

$$\sigma_{yu}[1] = \sigma_{y*u}[1] + 0$$

From equations (7) and (8), $\sigma_{y*u} = 0$

Hence,

$$\boxed{\sigma_{yu}[1] = 0}$$

To determine $\sigma_{yu}[2]$,

$$\sigma_{yu}[2] = \sigma_{y*u}[2] + 0$$

Multiplying equation (1) with $u[K-2]$,

$$y^*[K]u[K-2] + f_1^\circ y^*[K-1]u[K-2] = b_2^\circ u[K-2]^2$$

Taking expectation,

$$\sigma_{y*u}[2] + f_1^\circ \sigma_{y*u}[1] = b_2^\circ \sigma_u^2$$

$$\Rightarrow \sigma_{y*u}[2] = b_2 \sigma_u^2$$

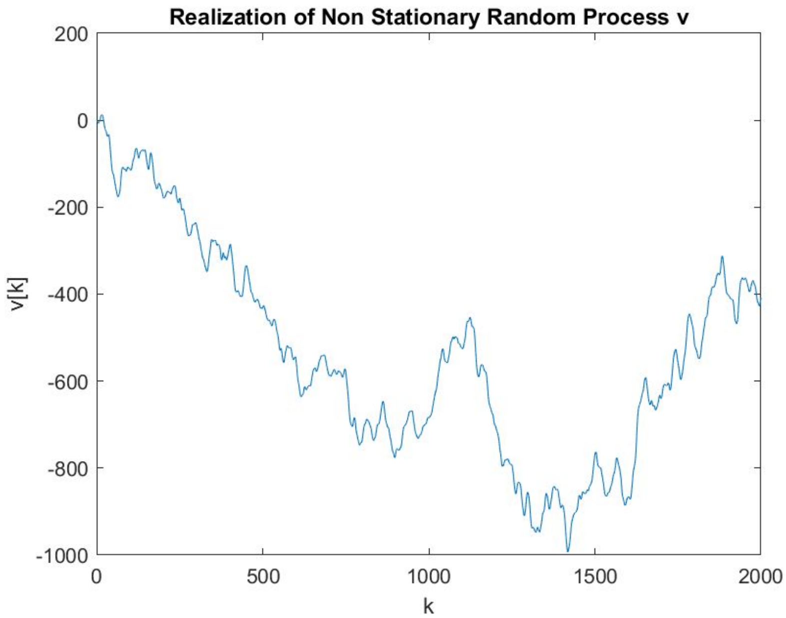
Hence,

$$\boxed{\sigma_{yu}[2] = b_2^\circ \sigma_u^2}$$

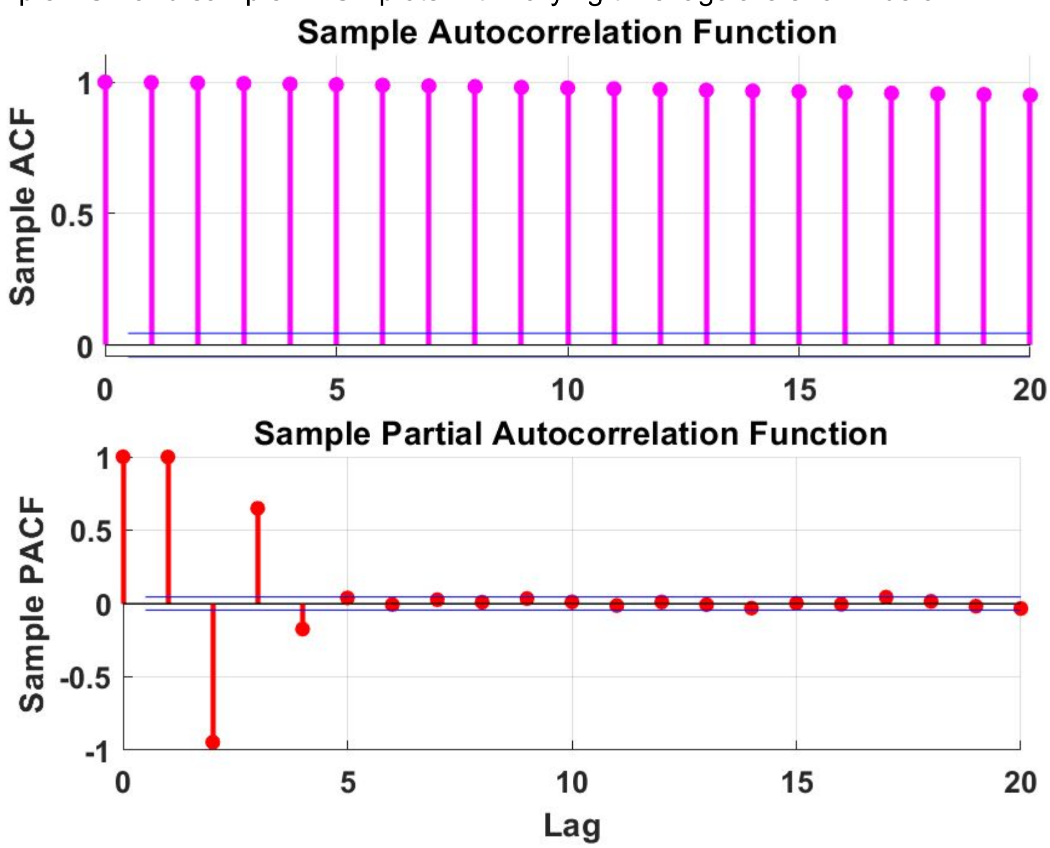
Question 3

Part (a) - Integrating Features

The plot of the given data for the process $\{v[k]\}$ is shown below:



The plot clearly indicates that the process is nonstationary and quite similar to a random walk.
The sample ACF and sample PACF plots with varying time lags are shown below:



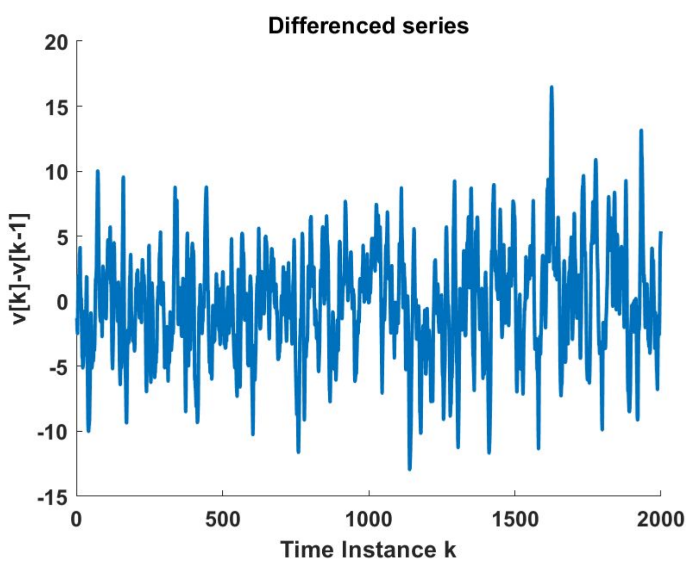
An integrating process has a long term memory, i.e., the value at one time instant is highly dependent on the previous time instants, even for a larger time lag. We can formulate two integrating factors based on the observation of the sample ACF and sample PACF plots:

- The sample ACF value is nearly closer to 1 for all represented lag values and the ACF decay is extremely slow with increasing lag. This suggests that even for a higher lag value, the observations are correlated, indicating a long-term memory.
- The PACF value at lag 1, is almost equal to unity, indicating a strong correlation between two observations with lag 1. Further, the PACF decays abruptly with increasing lag, indicating no direct correlation between two observations with higher lag. This behavior is in general observed in random walk processes.

Part (b) - Determination of suitable ARIMA model

Step 1:
Perform ADF test on v, to determine the presence of integrating or random walk effects
Result: p-value of 0.8455 (which is greater than 0.05) was obtained, and we do not reject the null hypothesis of the presence of unit root in the time series.

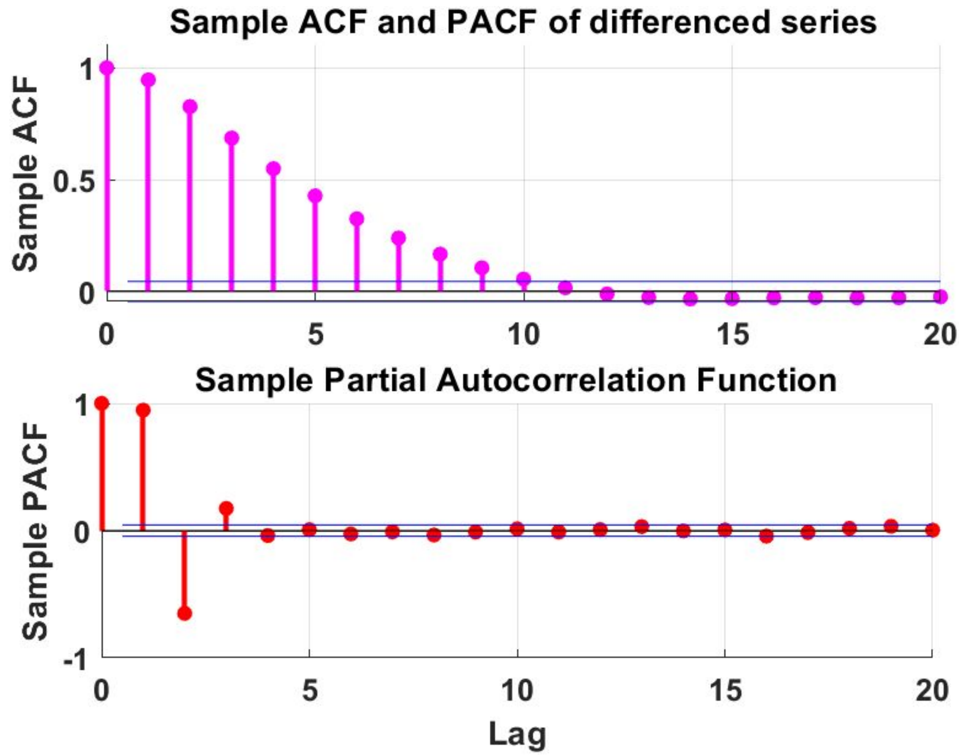
Step 2: Generate and visualize first order Difference Series
Generate the difference series of the random process v, and visualize it to determine if it has a stationary-like behavior.
Result:



From the above plot, we can say that the difference series might be a stationary process.

Step 3: Generate sample ACF and PACF plots for first-order Difference Series and ADF test
In order to confirm the AR and MA behavior of the difference series, we can observe the sample ACF and sample PACF plots of the difference series.

Result:



The sample ACF plot of the difference series decays exponentially, indicating autoregressive (AR) behavior, and the sample PACF value tails off to zero eventually, indicating a moving-average (MA) effect. To further confirm the stationary and non-integrating behavior of the difference series, an ADF test on the difference series was performed, and a p-value of 10^{-3} (which is less than 0.05), thereby rejecting the null hypothesis.

We notice that the process is ARIMA, with a first-order integrating factor i.e. ARIMA(P,1,M) model.

Step 4: Determining P and M of the ARIMA model

Multiple combinations of P and M were tried, some of them resulted in the parameter estimates being statistically insignificant. Two models that are of interest are the **ARIMA(2,1,1)** and **ARIMA(3,1,0)** as they are equally complex with the same number of parameters, all of which are statistically significant, as summarized below:

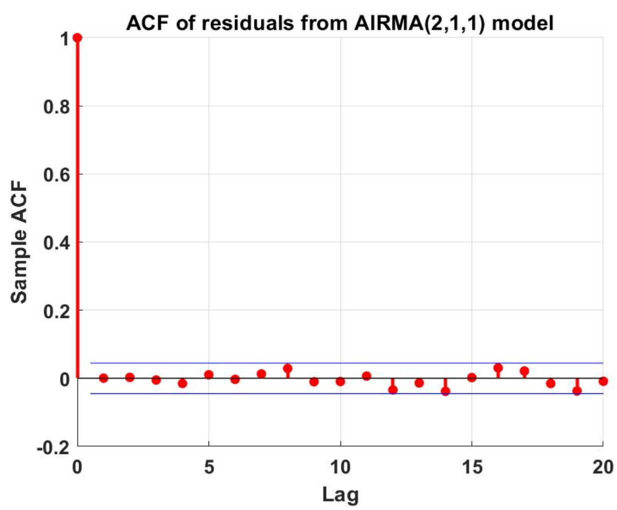
ARIMA(2,1,1) Model (Gaussian Distribution):

	Value	StandardError	TStatistic	PValue
Constant	0	0	NaN	NaN
AR{1}	1.4103	0.02846	49.554	0
AR{2}	-0.5069	0.027996	-18.106	2.8587e-73
MA{1}	0.27114	0.031749	8.5402	1.3394e-17
Variance	1.0057	0.030717	32.742	3.955e-235

ARIMA(3,1,0) Model (Gaussian Distribution):

	Value	StandardError	TStatistic	PValue
Constant	0	0	NaN	NaN
AR{1}	1.675	0.022401	74.775	0
AR{2}	-0.92249	0.038389	-24.03	1.3491e-127
AR{3}	0.17444	0.021791	8.0051	1.1941e-15
Variance	1.0073	0.030829	32.674	3.638e-234

However, the AIC score of ARIMA(2,1,1) is lower than that of ARIMA(3,1,0). Hence, the former is more suitable to characterize this random process.
ACF of residuals from **ARIMA(2,1,1)** model is shown below:



Furthermore, the ARIMA(2,1,1) passed the Ljung-Box test with a p-value of 0.7668, indicating that the residuals are white noises.

Question 4

Part (a)

Given GWN process $y[k] \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known and there are N observations of $y[k]$.
GWN process is equivalent to IID.
The pdf of GWN is given by

$$f(y[k]|\mu) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y[k]-\mu)^2}{2\sigma^2}\right)$$

Let $\mathbf{y}$ denote the data of N observations

$$f(\mathbf{y}/\mu) = \prod_{k=1}^{N} f(y[k]|\mu)$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left(\sum_{k=1}^{N} -\frac{(y[k]-\mu)^2}{2\sigma^2}\right)$$

Let the log likelihood function $l(\mu|\mathbf{y})$

$$l(\mu|\mathbf{y}) = \frac{\sum_{k=1}^{N} -(y[k]-\mu)^2}{2\sigma^2} + N \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)$$

Now, the ML estimate of μ is

$$\hat{\mu}_{ML} = \arg\max_{\mu} \; l(\mu|\mathbf{y})$$

i.e. estimate of μ that maximizes log-likelihood of data

$$\frac{\partial s}{\partial \mu} = -\frac{N}{\sigma^2}$$

$$I(\mu) = -E\left(\frac{\partial s}{\partial \mu}\right) = \frac{N}{\sigma^2}$$

Hence, $\boxed{I(\mu) = \frac{N}{\sigma^2}}$

Part (b)

Given linear regression problem

$$Y = aX + b + \epsilon \quad , \quad \epsilon \sim N(0, \sigma_e^2)$$

or

$$y[K] = a\,x[K] + b + \epsilon[K] \quad \forall\ K = 1 \text{ to } N$$

let $\hat{y}[K] = a\,x[K] + b$

$$\Rightarrow y[K] = \hat{y}[K] + \epsilon[K]$$

Since $f(\epsilon[K]) = \frac{1}{\sqrt{2\pi}\,\sigma_e} \exp\left(\frac{-\epsilon[K]^2}{2\sigma_e^2}\right)$

and as $\hat{y}[K]$ is not a random variable,

$$f(y[K]; \hat{y}[K]) = \frac{1}{\sqrt{2\pi}\,\sigma_e} \exp\left(-\frac{(y[K] - \hat{y}[K])^2}{2\sigma_e^2}\right)$$

i.e. the mean gets shifted to $y[K]$

Let y denote data of N observations.

$$\hat{\mu}_{ML} = \underset{\mu}{\arg\max} \left(\frac{\sum_{K=1}^{N} -(y[K]-\mu)^2}{2\sigma^2} + N \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right)\right)$$

For μ to be maximum

$$\frac{d\,l(\hat{\mu}; y)}{d\mu} = 0.$$

$$\Rightarrow \frac{d}{d\hat{\mu}}\left(- \frac{\left(\sum_{K=1}^{N} y[K]^2 + N\hat{\mu}^2 - 2\hat{\mu}\sum_{K=1}^{N} y[K]\right)}{2\sigma^2}\right) + 0 = 0$$

$$\Rightarrow \quad -2N\hat{\mu} + 2\sum_{K=1}^{N} y[K] = 0$$

or
$$\boxed{\hat{\mu}_{ML} = \frac{\sum_{K=1}^{N} y[K]}{N} = \bar{y}_N}$$

$\bar{y}_N$ ↳ sample mean

The score function can be defined as

$$S(y;\mu) = \frac{\partial l(\mu; y)}{\partial \mu} = -\frac{\left(2\mu N - 2\sum_{K=1}^{N} y[K]\right)}{2\sigma^2}$$

$$= \frac{\sum_{K=1}^{N} y[K] - N\mu}{\sigma^2}$$

Fischer's information

$$I(\mu) = \mathrm{var}(S) = E\left[\left(\frac{\partial l}{\partial \mu}\right)^2\right] = -E\left[\frac{\partial^2 l}{\partial \mu^2}\right] = -E\left[\frac{\partial S}{\partial \mu}\right]$$

Here, we assume the process is iid as ϵ generation is iid.

Hence,

$$f(y;\hat{y}) = \left(\frac{1}{\sqrt{2\pi}\sigma_e}\right)^N \exp\left(\sum_{K=1}^{N} -\frac{(y[K]-\hat{y}[K])^2}{2\sigma_e^2}\right)$$

$$= f(y;x,a,b)$$

Score function for a

$$S_a(y;x[K],a,b) = \frac{\partial l(y;x,a,b)}{\partial a}$$

where l is the log-likelihood function

$$l(y;x,a,b) = \log f(y;a,b,x)$$

$$= \sum_{K=1}^{N} -\frac{(y[K]-\hat{y}[K])^2}{2\sigma_e^2} + N\log\left(\frac{1}{\sqrt{2\pi}\sigma_e}\right)$$

$$\frac{\partial l(y;x,a,b)}{\partial a} = \sum_{K=1}^{N} \frac{\partial l}{\partial \hat{y}[K]} \cdot \frac{\partial \hat{y}[K]}{\partial a}$$

$$= \sum_{K=1}^{N} -\frac{(ax[K]^2 - y[K]x[K] + x[K]b)}{\sigma_e^2}$$

$$\frac{\partial^2 l(y;x,a,b)}{\partial a^2} = \sum_{K=1}^{N} \frac{-x[K]^2}{\sigma_e^2}$$

Hence,

$$I(a) = -E\left(\frac{\partial S}{\partial a}\right) = -E\left(\frac{\partial^2 l}{\partial a^2}\right)$$

$$= -E\left(\frac{-\sum_{K=1}^{N} x[K]^2}{\sigma_e^2}\right)$$

$$\boxed{I(a) = \sum_{K=1}^{N} \frac{x[K]^2}{\sigma_e^2}}$$

Also,

$$\frac{\partial l(y;x,a,b)}{\partial b} = \sum_{K=1}^{N} + \frac{(y[k] - a\,x[k] - b)}{\sigma_e^2}$$

$$\frac{\partial^2 l(y;x,a,b)}{\partial b^2} = -\frac{N}{\sigma_e^2}$$

Hence, $I(b) = -E\left(\frac{\partial S}{\partial b}\right) = -E\left(\frac{\partial^2 l}{\partial b^2}\right) = -E\left(\frac{-N}{\sigma_e^2}\right)$

$$\boxed{I(b) = \frac{N}{\sigma_e^2}}$$

Now, to determine $\hat{a}_{ML}$ and $\hat{b}_{ML}$ (MLE estimates) we need to solve for $\frac{\partial l}{\partial a} = \frac{\partial l}{\partial b} = 0$.

$$\frac{\partial l}{\partial a} = \sum_{K=1}^{N} - \frac{(a x[K]^2 - y[K]x[K] + b x[K])}{\sigma_e^2} = 0$$

$$\Rightarrow \hat{a}_{ML} \sum_{K=1}^{N} x[K]^2 - \sum_{K=1}^{N} y[K]x[K] + b\sum_{K=1}^{N} x[K] = 0$$

$$\hat{a}_{ML} = \frac{\sum_{K=1}^{N} y[K]x[K] - \hat{b}_{ML}\sum_{K=1}^{N} x[K]}{\sum_{K=1}^{N} x[K]^2} \longrightarrow (1)$$

Similarly,

$$\frac{\partial l}{\partial b} = \frac{\sum_{K=1}^{N} -(y[K] - ax[K] - b)}{\sigma_e^2} = 0$$

$$\text{or} \quad \hat{a}_{ML} = \frac{\sum_{K=1}^{N} y[K] - b}{\sum_{K=1}^{N} x[K]} = \frac{\bar{y} - \hat{b}_{ML}}{\bar{x}} \quad -(2)$$

$$\text{or} \quad \hat{b}_{ML} = \bar{y} - \hat{a}_{ML}\bar{x}$$

Solving (1) and (2)

$$\boxed{\begin{aligned} \hat{a}_{ML} &= \frac{\sum_{K=1}^{N} (y[K]-\bar{y})(x[K]-\bar{x})}{\sum_{K=1}^{N} (x[K]-\bar{x})^2} \\ &= \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2} \\ \hat{b}_{ML} &= \bar{y} - \hat{a}_{ML}\bar{x} \end{aligned}}$$

NOTE

$\bar{x}, \bar{y}$ represent sample mean.

Question 4

Part (b)

It was shown that the MLE estimates of the parameters a and b obtained through simulation and analytical approach matched. The plot of the likelihood function is shown below, and the MLE estimate is marked with a red dot.