INDIAN INSTITUTE OF TECHNOLOGY MADRAS

Department of Chemical Engineering

CH5350 : Applied Time Series Analysis
Solutions to Assignment #5

1 Least Squares Estimation

1.1

Assumptions

1. $e[k] \in \mathcal{N}(0, \sigma_e^2)$
2. $x[k]$ is uncorrelated with noise $e[k]$
3. $x[k]$ is deterministic

Given model $y[k] = \mathbf{x}^T[k]\theta + e[k]$. The prediction error is given by

$$\boldsymbol{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{P}\mathbf{y} = \mathbf{P}^{\perp}\mathbf{y}$$

The SSE is given by

$$SSE = tr(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$$

Taking expectation on both sides

$$\mathrm{E}(SSE) = \mathrm{E}(tr(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')) = tr(\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')) = tr(\mathbf{P}^{\perp}E(\mathbf{y}\mathbf{y}')\mathbf{P}^{\perp})$$

Since, $x[k]$ is deterministic

$$\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon})' = tr(\mathbf{P}^{\perp}\mathrm{E}(\mathbf{e}\mathbf{e}')\mathbf{P}^{\perp}) = \sigma_e^2 tr({\mathbf{P}^{\perp}}^2)$$

The $P^{\perp}$ is given by

$$\begin{aligned}
tr(P^{\perp^2}) &= tr(I_{N\times N} - P_{N\times N}) \\
&= tr(I) - tr(\Phi(\Phi^{'}\Phi)^{-1}\Phi^{'}) \\
&= N - tr(\Phi^{'}\Phi(\Phi^{'}\Phi)^{-1}) \quad \because \quad tr(AB) = tr(BA) \\
&= N - P
\end{aligned}$$

Hence

$$\mathrm{E}(SSE) = (N-P)\sigma_e^2$$

We know $\hat{\sigma}_e^2 = \dfrac{SSE}{N-p}$.

$$\mathrm{E}(\hat{\sigma}_e^2) = (N-p)\sigma_e^2/(N-p) = \sigma_e^2$$

Hence, the predictor to estimate variance of noise is an unbiased estimator.

1.2

A general AR(p) process of non-zero mean is written as

$$y[k] - \mu = -d_1(y[k-1] - \mu) - ... - d_p(y[k-p] - \mu) + e[k]$$

In Yule-Walker method of estimation, we first calculate sample mean

$$\hat{\mu} = \frac{1}{N}\sum_{k=0}^{N-1} y[k]$$

and estimate auto-correlation functions using the mean of the process and rest is same. In least square estimation, to handle non-zero mean a column vector of 1 is added to regressor matrix $i.e$

$$\varphi = [1 \;\; y[k-1] \;\; y[k-2]...y[k-p]]$$

The least square estimate of mean for AR(p) model is

$$\mu_{ls} = \frac{1}{N}\sum_{k=p}^{N-1}$$

This is the difference between Y-W estimator and least square estimator in handling non-zero mean of process while estimation. **To obtain estimate of mean in least square**

estimate, we need less number of samples, where as in YW method all samples are necessary

As $N \to \infty$ both YW and LS estimators gives same estimate of mean.

1.3

The code for solving the question is

```
# Simulate an AR model
y <- arima.sim (n =10000 , list (ar=c (-1.1 , 0.28) ,sd =1))
# Add mean to data
y = y +3
# perform YW method of estimation
modyw=ar.yw(y,demean=TRUE)
x1=y[1:9998]
x2=y[2:9999]
x3=y[3:10000]
# Perform ls method of estimation
modls=lm(x3~I(x2)+I(x1))
```

The estimates obtained using both methods are given in Table 1

Table 1: Estimates obtained using different methods

Parameter	YW method	OLS
$\hat{d}_1$	-1.104	-1.1055
$\hat{d}_2$	0.2879	0.2887
$\hat{\mu}$	3.039	3.1744

It can clearly be seen that the mean obtained from Least Squares and Yule-Walker are almost the same.

2 MLE Estimation

2.1

Given $x[1], x[2]$ fit an AR(1) model of form

$$x[k] = -d1x[k-1] + e[k]$$

The joint distribution of two samples is given by

$$f(x[1], x[2]) = f(x[1])f(x[2]|x[1])$$

Assuming the samples will follow Gaussian distribution,

$$f = \frac{1}{\sqrt{2\pi}\sigma_{x[1]}} \exp\left(-\frac{1}{2}\left(\frac{x[1]-\mu_{x[1]}}{\sigma_{x[1]}}\right)^2\right) \times \frac{1}{\sqrt{2\pi}\sigma_{x[2]}} \exp\left(-\frac{1}{2}\left(\frac{x[2]-\mu_{x[2]}}{\sigma_{x[2]}}\right)^2\right)$$

where

$$\begin{aligned}
\mu_{x[1]} &= 0 \\
\sigma^2_{x[1]} &= \frac{\sigma_e^2}{1-d_1^2} \\
\mu_{x[2]|x[1]} &= -d_1 x[1] \\
\sigma^2_{x[2]|x[1]} &= \sigma_e^2
\end{aligned}$$

Substituting values in above equation

$$\begin{aligned}
f &= \frac{1}{\sqrt{2\pi}\frac{\sigma_e}{\sqrt{1-d_1^2}}} \exp\left(-\frac{1}{2}\left(\frac{x[1]^2}{\frac{\sigma_e^2}{1-d_1^2}}\right)\right) \times \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{1}{2}\left(\frac{x[2]+d_1x[1]}{\sigma_e}\right)^2\right) \\
&= \frac{\sqrt{1-d_1^2}}{2\pi\sigma_e^2} \exp\left(-\frac{1}{2\sigma_e^2}\left(x^2[1] - d_1^2x^2[1] + x^2[2] + d_1^2x^2[1] + 2d_1x[1]x[2]\right)\right)
\end{aligned}$$

The negative of log likelihood function is given by

$$\begin{aligned}
L &= -\ln f \\
&= -0.5\ln(1-d_1^2) + ln(2\pi\sigma_e^2) + \frac{1}{2\sigma_e^2}\left(x^2[1] - d_1^2x^2[1] + x^2[2] + d_1^2x^2[1] + 2d_1x[1]x[2]\right)
\end{aligned}$$

The objective is to estimate $\hat{d}_1$ which maximises the above objective function. The assumption to be made is σ_e^2 is known. Hence $\hat{d}_1$ is obtained by

$$\begin{aligned} \frac{\partial L}{\partial d_1} &= 0 \\ \implies \frac{-\hat{d}_1}{1-d_1^2} + \frac{1}{2\sigma_e^2}(2x[1]x[2]) &= 0 \\ \implies \hat{d}_1 = \frac{x[1]x[2]}{\frac{\sigma_e^2}{1-d_1^2}} \end{aligned}$$

which is nothing but

$$d_1 = \frac{x[1]x[2]}{\sigma_{x[1]}^2}$$

The least square estimate for the process is obtained as

$$\hat{d_{l^ls}} = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y$$

where $\Phi = x[1]$. Hence $\hat{d_{l^ls}}$ is obtained as

$$\hat{d_{l^ls}} = \frac{x[2]x[1]}{x^2[1]}$$

In both cases, the estimate of d_1 is same. For AR models both least square and MLE methods gives same estimates.

3

The data is shown in Figure 1.

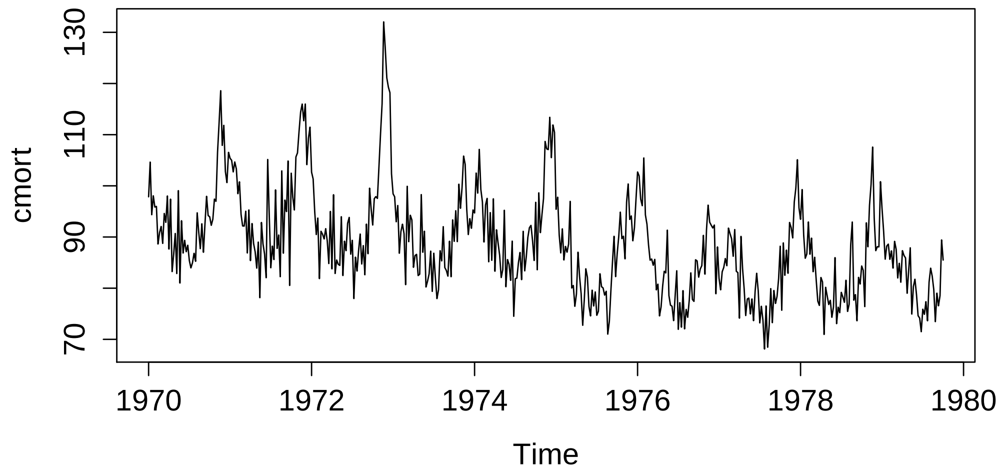

Figure 1: Cardiovascular data

Using YW method an AR(2) model is fitted for the data and estimates of parameters are given by Coefficients:

	ar1	ar2
	0.4339	0.4376
s.e.	0.04	0.04

The estimate of variance of $e[k]$, $\hat{\sigma}_e^2 = 32.84$.

Using OLS method an AR(2) model is fitted for the data and estimates of parameters are given by Coefficients:

	ar1	ar2
	0.4286	0.4418
s.e.	0.039	0.038

The estimate of variance of $e[k]$, $\hat{\sigma}_e^2 = 32.317$. The parameter estimates almost similar because we are estimating AR model. In least square estimation the standard errors in parameters are btained as 0.039 and 0.038. The asymptotic standard errors are calculated using formula derived in class as 0.039 and 0.039. Standard error in parameter estimates are same in both cases which suggests that this is the best model.

4

4.1

The data is shown in Figure 2

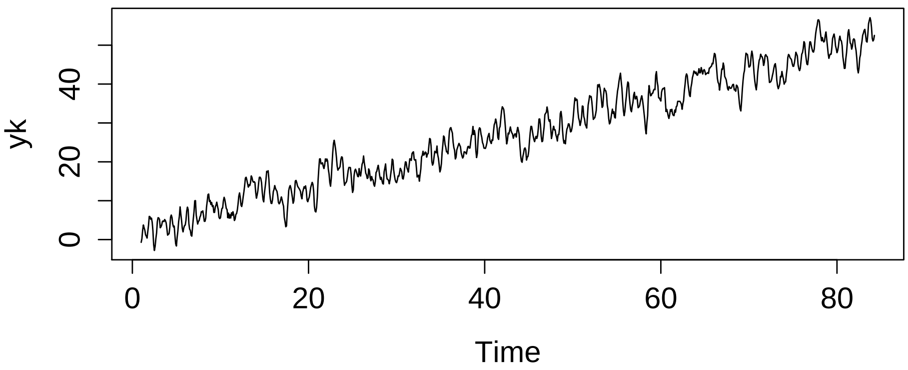

Figure 2: Cardiovascular data

Clearly, the data has some seasonal component. The data is decomposed using stl command and is shown in Figure 3

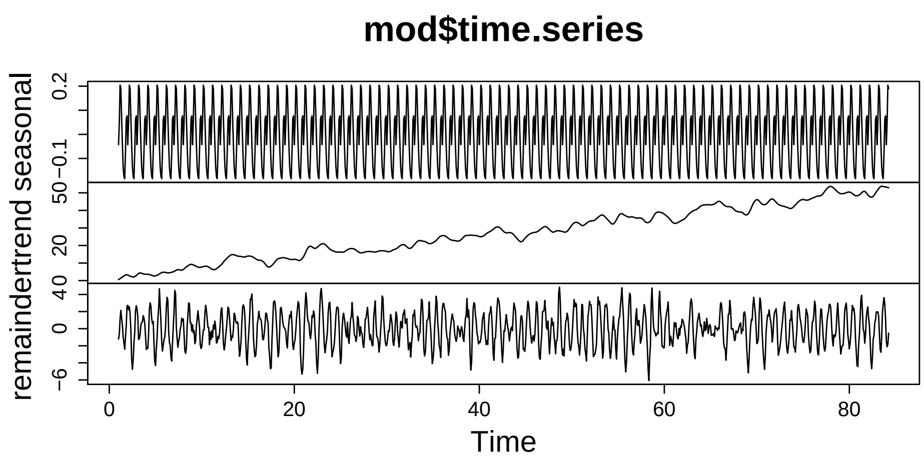

Figure 3: SARIMA data

The acf of remainder series is shown in Figure 4

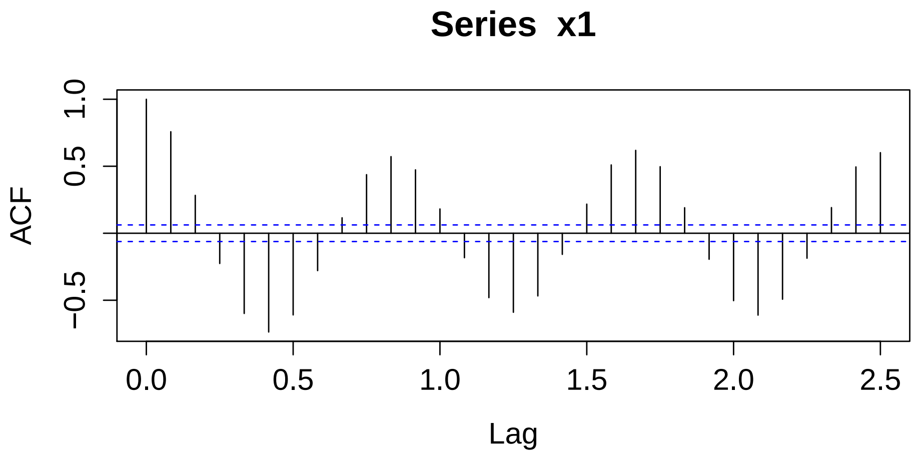

Figure 4: decomposed data

Clearly, the signal is periodic with a period of 10 samples. Hence a sum of cosine and sine signal is fitted using lm command. The acf of residuals is shown in Figure 5

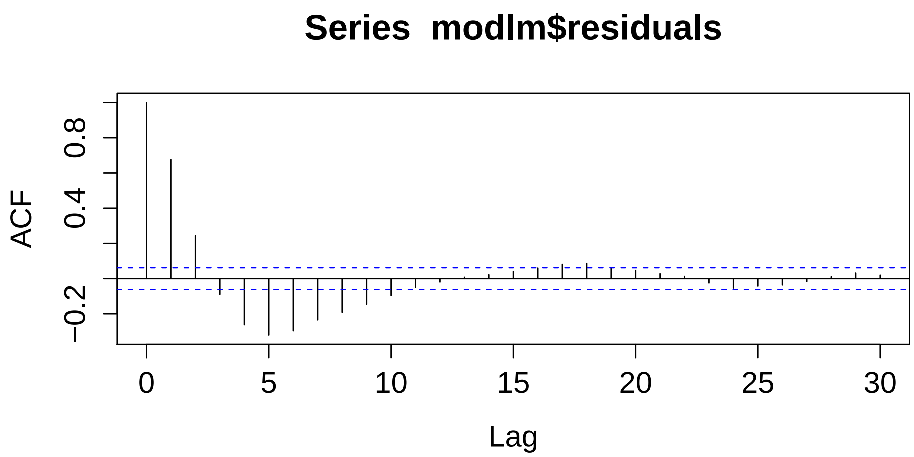

Figure 5: ACF of residuals

The pacf of residuals is shown in Figure 6

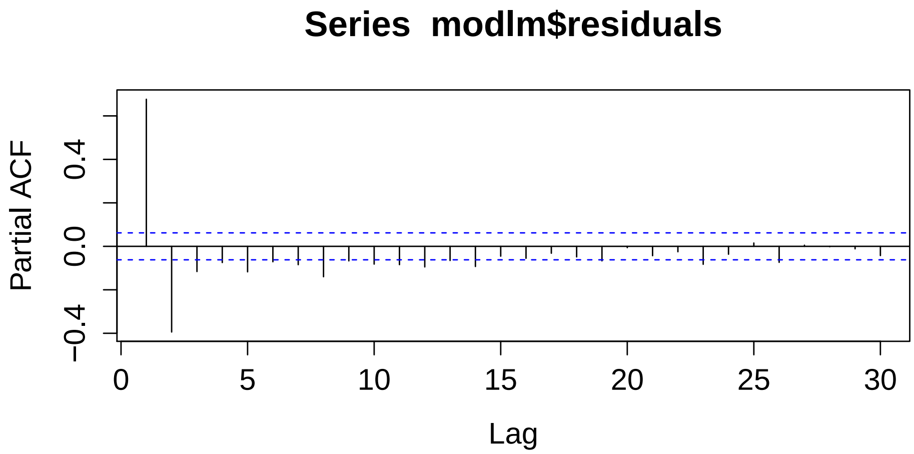

Figure 6: PACF of residuals

The pacf plot conforms that the process is of AR(2) model. Fitting an AR(2) model gives the coefficients as

arima(data1, order = c(2, 0, 1))

Coefficients:

	ar1	ar2
	-0.9434	-0.3940
s.e.	0.029	0.029

The ACF of residuals is shown in Figure 7

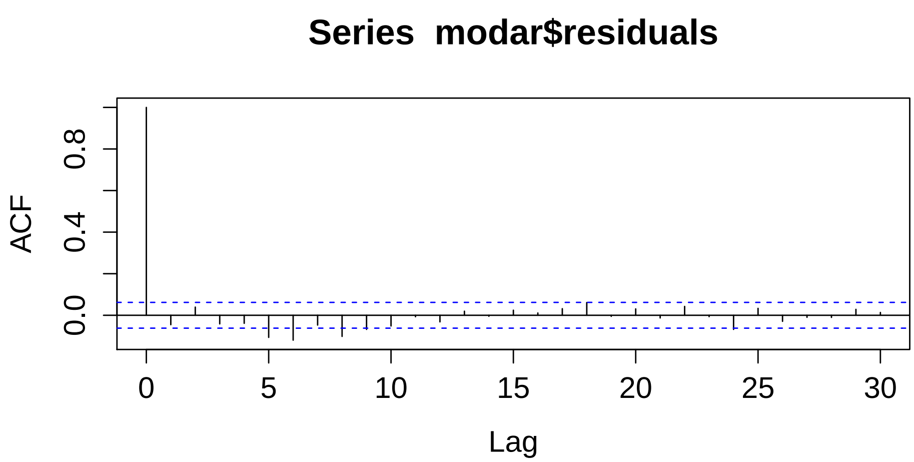

Figure 7: ACF of residuals

Clearly the ACF plot resembles like a white noise process. Hence the process is modelled as

$$x[k] = 2.22\sin(0.2\pi k) + v[k] \quad \text{where} \quad v[k] = -0.9434v[k-1] - 0.394v[k-2] + e[k]$$

4.2

The ACF of the data is shown in Figure 8

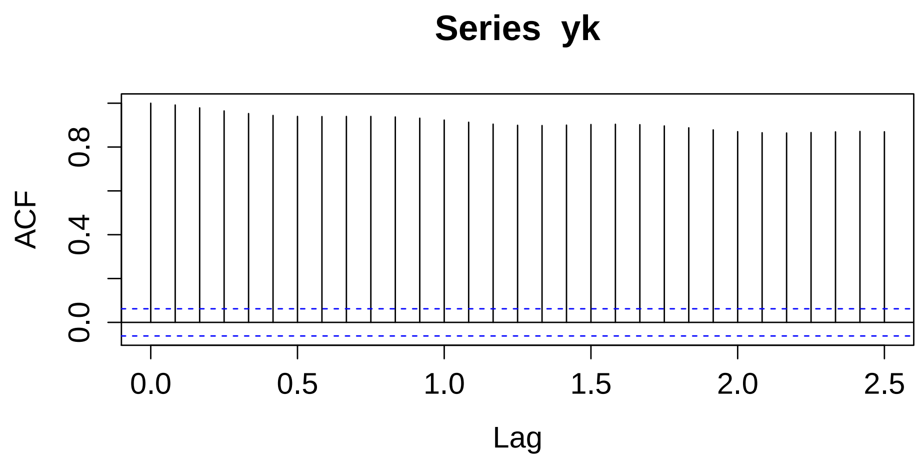

Figure 8: ACF of sarima data

Clearly, it shows non-stationarity. Hence the series is differenced and the acf of differenced series is shown in Figure 9

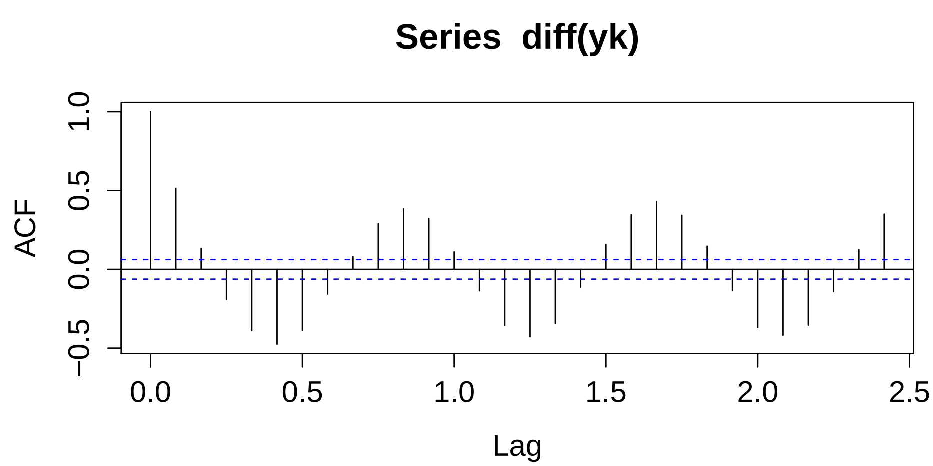

Figure 9: ACF of differenced series

It has seasonality of 10 samples. Hence the series is modelled using arima command in

R as

mod=arima(yk,order=c(1,0,0),seasonal=list(period=10,order=c(1,0,1)))

Coefficients:

	ar1	ar2	sar1	sma1
	1.4971	-0.5041	-0.5242	0.6545
s.e.	0.0276	0.027	0.1119	0.0982

The acf plot of residuals is shown in Figure 10

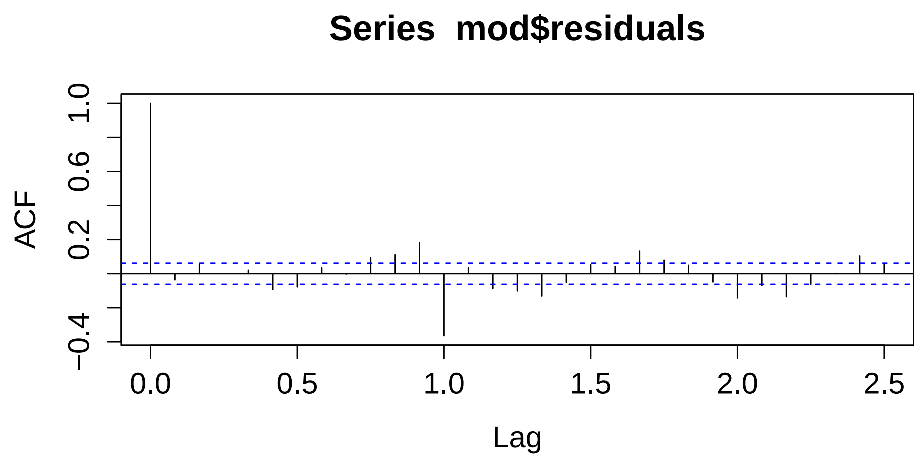

Figure 10: ACF of residuals

ACF plot almost resembles like that of white noise process. Hence the final model is

$$\hat{x}[k] = (1.3893x[k-1] - 0.557x[k-2]) \times (0.5242x[k-10] - 0.6545e[k-10])$$

5

5.1

Given the function

$$J = E((Y - g(x))^2)$$

The objective is to find $g(x)$ such that the value of J is minimum. *i.e*

$$\begin{aligned}
\frac{\partial J}{\partial g(x)} &= 0 \\
\frac{\partial}{\partial g(x)} E((Y - g(x))^2) &= 0 \\
\frac{\partial}{\partial g(x)} E(Y^2 + g(x)^2 - 2Y g(x)) &= 0 \\
E(2g(x) - 2Y) &= 0 \\
\implies \hat{g}(x) &= E(Y|X)
\end{aligned}$$

In class, we already studied that the conditional expectation is best predictor in mean square error sense. This gives proof for the result.

5.2

The joint Gaussian distribution of two variables X and Y having unconditional means μ_X and μ_Y is given by

$$f(x, y) = \frac{1}{2\pi \det(\sum_{xy})^{0.5}} \exp(0.5(x - \mu_X) \textstyle\sum_{xy}^{-1} (y - \mu_Y))$$

where

$$\textstyle\sum_{xy} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$$

To compute conditional expectation $E(Y|X)$, lets first compute marginal density function. Assume X follows Gaussian distribution

$$\begin{aligned}
f_Y(y|x) &= \frac{f(x, y)}{f(x)} \\
&= \frac{\frac{1}{\sqrt{2\pi} \det(\sum_{xy})^{0.5}} \exp(0.5(x - \mu_X) \sum_{xy}^{-1} (y - \mu_Y))}{\frac{1}{\sqrt{2\pi(\sigma_{xy})}} \exp(0.5(x - \mu_X)\sigma_{xy}^{-1})}
\end{aligned}$$

The conditional expectation $E(Y|X)$ is given by

$$E(Y|X) = \int_{-\infty}^{\infty} y f(Y|X) dy$$

Solving we get

$$E(Y|X) = \mu_y + \rho_{XY}\frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

Thus the conditional expectation is a linear function of X.

5.3

Given random variable

$$y = x^2 + z$$

Given x and z are independent zero-mean processes with unit variance. In part (a) we already proved that conditional expectation is best predictor in MSE sense. Hence the best predictor in MSE sense is

$$\hat{y} = x^2$$

The value of MSE is obtained as

$$\begin{aligned} MSE &= E((y - \hat{y})^2) \\ &= E((x^2 + z - x^2)^2) \\ &= E(z^2) \\ &= 1 \end{aligned}$$

5.4

Suppose a linear approximation is thought for y as

$$\hat{y} = \alpha x + \beta$$

The objective function will be

$$\begin{aligned} J &= E((y - \hat{y})^2) \\ &= E((x^2 + z - \alpha x + \beta)^2) \end{aligned}$$

To find values of α and β, we will take gradients and make it equal to zero. (Assuming x is normally distributed we get

$$\alpha = 0, \beta = 1$$

The value of MSE is given by

$$\begin{aligned}
E((y-\hat{y})^2) &= E(((x^2)+z-1)^2) \\
&= E(x^4+z^2-2zx^2+1-2x^2-2z) \\
&= E(x^4) = 3
\end{aligned}$$

MSE is high because the true process is non-linear and it don't have any linear approximation.