Lecture Notes #2

Time series analysis (STAT 5140/4140)

1 Simple linear regression review

Consider two random variables X and Y. Assume that the response variable is Y and the explanatory variable is X. For simplicity assume $E(X) = E(Y) = 0$. We would like to predict Y using X using a linear predictor.

$$\hat{Y} = b_1 X$$

Choose b_1 that minimizes Mean Square Error (MSE) $E(Y - \hat{Y})^2$. Optimal value of b_1 is:

$$\begin{aligned} b_1^* &= \frac{Cov(X,Y)}{Var(X)} \\ &= \frac{\sigma_{xy}}{\sigma_x^2} \\ &= \rho_{xy}\frac{\sigma_y}{\sigma_x} \end{aligned} \tag{1}$$

Note that to estimate b_1^* it is sufficient to know about the first and second order moments. Knowledge of higher order moments may require to fit a non-linear models. Now lets use Y to predict X. That is switch the role of the response and explanatory variables. The linear mode is given by:

$$\hat{X} = b_2 Y$$

Similarly the optimum value of b_2 that will minimize MSE is:

$$\begin{aligned} b_2^* &= \frac{Cov(X,Y)}{Var(Y)} \\ &= \frac{\sigma_{xy}}{\sigma_y^2} \\ &= \rho_{xy}\frac{\sigma_x}{\sigma_y} \end{aligned} \tag{2}$$

Using equations (1) and (2) we get:

$$\rho_{xy}^2 = b_1^* b_2^* \tag{3}$$

This implies correlation has no direction. That is given ρ_{xy}^2 one can't determine b_1^* and b_2^*.

1.0.1 Lurking Variable

When according the correlation two random variables are related it may happen that in real life the relation between those two variables may not make sense.

Example Consider: X=Number of firefighters at a fire and Y= Cost of the damage due to the fire.

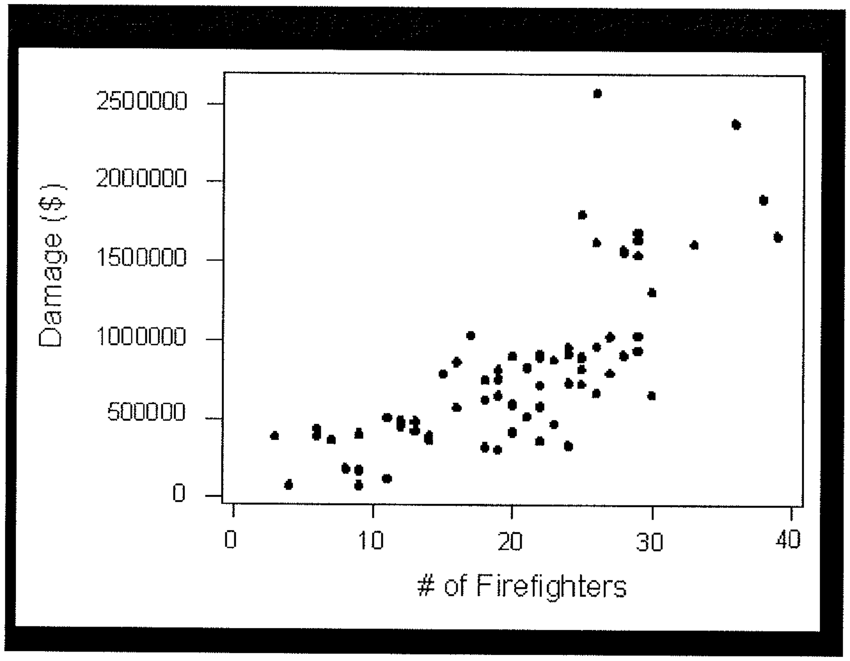

Figure 1: Scatter plot 1

As from the plot there is a positive correlation between X and Y. But the correlation does not make any sense in real life. Obviously more fire fighters do not mean more damage. To explain the correlation we need another variable Z=size of the fire. Now it is clear that X and Z and Y and Z are related. As a result it may appear that X and Y are related. In such cases we call the variable Z is a lurking variable.

It could happen X and Y are not related directly but they are indirectly related. To measure direct correlation between X and Y removing the effect of Z we compute partial covariance,$\sigma_{xy \cdot z}$, and partial correlation, $\rho_{xy \cdot z}$,between X and Y.

Simple correlation, ρ_{xy} measures direct dependency between X and Y. Where as partial correlation measures dependency in a indirect pathway.

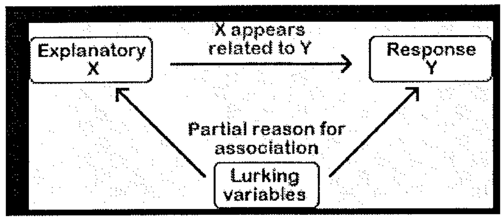

Figure 2:

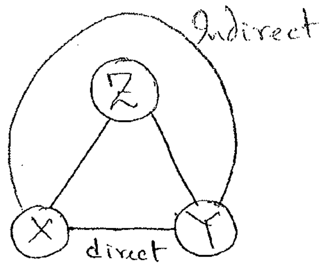

Figure 3:

To compute partial covariance first we need to remove the effect of the Z variable. Compute covariance between conditional random variables $X|Z$ and $Y|Z$. First mode X using Z and Y using Z.

$$\hat{X} = b_1^* Z \quad \text{and} \quad \hat{Y} = b_2^* Z \tag{4}$$

Where b_1^* and b_2^* are given by:

$$b_1^* = \frac{\sigma_{xz}}{\sigma_z^2} \tag{5}$$

$$b_2^* = \frac{\sigma_{yz}}{\sigma_z^2} \tag{6}$$

Now compute the residuals $\epsilon_{xz} = X - \hat{X}$ and $\epsilon_{yz} = Y - \hat{Y}$. This residuals are the part of X and Y that we can't explain using Z. So partial covariance would be the covariance between ϵ_{xz} and ϵ_{yz}.

$$\begin{aligned}
\sigma_{xy\cdot z} &= Cov(\epsilon_{xz}, \epsilon_{yz}) \\
&= Cov(X - \hat{X}, Y - \hat{Y}) \\
&= Cov(X - b_1^* Z, Y - b_2^* Z) \\
&= E((X - b_1^* Z, Y - b_2^* Z)) \text{ assuming } E[X] = E[Y] = E[Z] = 0 \\
&= E(XY) - b_2^* E(XZ) - b_1^* E(YZ) + b_1^* E(XZ) b_2^* E(XZ) \\
&= \{E(XY) - E(X)E(Y)\} - b_2^* \{E(XZ) - E(X)E(Z)\} - b_1^* \{E(YZ) - E(Y)E(Z)\} + b_1^* b_2^* \sigma_z^2 \\
&= \sigma_{xy} - b_2^* \sigma_{xz} - b_1^* \sigma_{yx} + b_1^* b_2^* \sigma_z^2 \\
&= \sigma_{xy} - \frac{\sigma_{yz}}{\sigma_z^2} \sigma_{xz} - \frac{\sigma_{xz}}{\sigma_z^2} \sigma_{yx} + \frac{\sigma_{yz}}{\sigma_z^2} \frac{\sigma_{xz}}{\sigma_z^2} \sigma_z^2 \\
&= \sigma_{xy} - \frac{\sigma_{xz} \sigma_{yz}}{\sigma_z^2} \qquad (7)
\end{aligned}$$

Similarly the partial correlation is given by:

$$\rho_{xy\cdot z} = \frac{\sigma_{xy} - \sigma_{xz}\sigma_{yz}}{\sqrt{1 - \rho_{zz}^2}\sqrt{1 - \rho_{yz}^2}} \qquad (8)$$

In R *pcor* function can be used to compute partial correlation. For a random vector $(X_1, \ldots, X_n)$ with a covariance matrix Σ we can compute the partial covariance matrix using the steps below:

1. Compute Σ^{-1}.
2. Normalize Σ^{-1} to get the corrosponding correlation matrix.
3. Compute $D = diag(\Sigma^{-1})$.
4. Find $P = -D^{-1}\Sigma^{-1}D$. Then P is the partial covariance matrix.

In time series analysis we will use partial auto-correlation functions to determine the order of an AR model.In R once can use *Cov2Cor* to normalize a covariance matrix.Similar to the partial correlation we can define Semi-partial correlation as $Corr(X, Y|Z)$, $Corr(X|Z, Y)$. In R *Spcor* can be used to copute semi-partial correlation. The function are available in *ppcor* package.

Example Consider two random variables

$$\begin{aligned}
X &= 2Z + 2W \\
Y &= Z + V
\end{aligned}$$

Assume $E(Z) = E(W) = E(V) = 0$, and $Cov(Z, W) = Cov(V, W) = Cov(V, Z) = 0$. Covariance between X and Y is given by:

$$\begin{aligned} Cov(X, Y) &= E\{(2Z + 3W)(Z + V)\} \\ &= 2E(Z^2) \\ &= 2\sigma_z^2 \end{aligned} \tag{9}$$

Direct covariance between X and Y is non zero. But form the construction of the random variable X and Y are not directly related. They are related via Z. Covariance between Y and Z is given by:

$$\begin{aligned} Cov(Y, Z) &= E\{(Z + W)(Z)\} \\ &= \sigma_z^2 \end{aligned} \tag{10}$$

Covariance between X and Z is given by:

$$\begin{aligned} Cov(X, Z) &= E\{(2Z + 3W)(Z)\} \\ &= 2\sigma_z^2 \end{aligned} \tag{11}$$

Partial covariance between X and Y:

$$\begin{aligned} \sigma_{xy \cdot z} &= \sigma_{xy} - \frac{\sigma_{xz}\sigma_{yz}}{\sigma_z^2} \\ &= 0. \end{aligned}$$

The partial correlation between X and Y is zero.

2 Random Signals (functions)

A discrete-time random signal is a function X_t whose characteristics can not be accurately describe by an existing mathematical function. At each point t it characterizes by probability distribution. A random signal can't be predicted accurately. Each observation of a random signal has some error or uncertainty. At any time point t the random signal can take several values but we only observe one realization. A random signal is a collection of all possible realization of X_t. Focus of a time series modeling is to infer the truth based on a single realization. For a random signal a particular realization is not in our interest. We rather focus on its statistical properties.

Example Assume that we have a time series temp. data at $6am$ every day or 3 years. Using that data we build a model. Assume that we came up with the best model. Can we use that model to predict temp. at noon? WHY?

Its is clear that the model can't not be used to predict temp. at noon. As the distribution of the random variable, temperature is a function of time. To avoid we define:

Definition *A time series* $\{X_t\}$ *is* strictly stationary *if the random vector* $(X_{t_1}, \ldots, X_{t_n})$ *and* $(X_{t_1+\tau}, \ldots, X_{t_n+\tau})$ *have the same joint distribution, for all set of indices* $(t_1, \ldots, t_n)$ *and for all* $\tau, n \in \mathbb{N}$. *That is:*

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+\tau}, \ldots, X_{t_n+\tau}), \forall\, n, \tau \in \mathbb{N}$$

Joint distribution is invariant of time, and it does not depend on:

1. When you start collecting data.
2. How many data you have collected.

A random process which is is not stationary is called Non-stationary.

Example Consider a random process:

$$X_t = ACos(\omega t + \phi), \text{ where } \phi \sim Uniform(0, \pi)$$

. This process is not stationary as:

$$\begin{aligned} E[X_t] &= A \int_0^{\pi} Cos(\omega t + \phi)\frac{1}{\pi} d\phi \\ &= \frac{A}{\pi} \int_0^{\pi} Cos(\omega t + \phi) d\phi \\ &= \frac{2A}{\pi} Sin(\omega t) \end{aligned}$$

As $E[X_t]$ is a function of time.

In real life it is very hard to prove that a process is stationary. As in real life we do not know about the density function of the process. We only have data from a random process. This definition is very strict in nature. It is hard to work with this definition in practice.

Definition *A random process* $\{X_t\}$ *is called weak stationary if:*

1. *The mean of the process is independent of time. That is* $\mu_X = E[X_t]$ *is not a function of* t.
2. *Variance of the process is finite. That is* $Var(X_t) < \infty$.

3. *The auto-covariance function for a laf h:*

$$\gamma(t, t+h) = Cov(X_t, X_{t+h})$$

is a function of the lag h only.

Remark 1. For a strict stationary process the random variables X_t are identically distributed. (from the definition choose n=1)

2. $(X_t, X_{t+h}) \stackrel{d}{=} (X_1, X_{1+h})$.

3. If $\{X_t\}$ is strict stationary process and $E(X_t^2) < \infty$ then the process is weak stationary.

4. An *iid* time series is weak stationary.

5. Weak stattionary does not imply strict stationary. See the example below:

Definition *For a random variable X the cumulative probability distribution function (CDF) defined as:*

$$F_X(x) = P(X \leq x)$$

Example Consider iid random variable $\{Z_t\}$ from $Normal(0, 1)$. Using Z_t construct X_t as:

$$X_t = \begin{cases} Z_t, & \text{if } t \text{ is even} \\ \frac{1}{\sqrt{2}}(Z_t^2 - 1) & \text{if } t \text{ is odd} \end{cases}$$

This process is weak stationary but not stationary. The mean of the process is given by:

$$E[X_t] = \begin{cases} E(Z_t) = 0, & \text{if } t \text{ is even} \\ E\{\frac{1}{\sqrt{2}}(Z_t^2 - 1)\} = 0 & \text{if } t \text{ is odd} \end{cases}$$

Also, the variance of the process is given by: [Note $Z_t^2 \sim \chi^2_{(1)}$]

$$Var[X_t] = \begin{cases} Var(Z_t) = 1, & \text{if } t \text{ is even} \\ Var\{\frac{1}{\sqrt{2}}(Z_t^2 - 1)\} = \frac{1}{2}Var[Z_t^2] = 1 & \text{if } t \text{ is odd} \end{cases}$$

Also you can prove $Cov(X_t, X_{t+h}) = 0$. This proves that the process $\{X_t\}$ is weak stationary. But this process is not a strict stationary process. As:

$$P[X_t \leq x_t] = \begin{cases} P[Z_t \leq z_t], & \text{if } t \text{ is even} \\ P\{\frac{1}{\sqrt{2}}(Z_t^2 - 1) = P\{-\sqrt{\sqrt{2}x_t + 1} \leq Z_t \leq \sqrt{\sqrt{2}x_t + 1}\} & \text{if } t \text{ is odd} \end{cases}$$

Now if $x_t = 0$ then,

$$P[X_t \leq 0] = \begin{cases} 0.5, & \text{if } t \text{ is even} \\ 0.68 & \text{if } t \text{ is odd} \end{cases}$$

So the cdf's are different for t odd and even. As a result the process is not strict stationary.

Example Consider the random process $X_t = Z_1 Cos(ct) + Z_2 Sin(ct)$, where $\{Z_t\} \sim N(0,1)$. This process is weak stationary.

$$E(X_t) = E\left\{Z_1 Cos(ct) + Z_2 Sin(ct)\right\}$$

$$= Cos(ct)\,E(Z_1) + Sin(ct)\,E(Z_2) = 0 \quad \text{(not a function of time)}$$

$$V(X_t) = V\left\{Z_1 Cos(ct) + Z_2 Sin(ct)\right\}$$

$$= Cos^2(ct)\,V(Z_1) + Sin^2(ct)\,V(Z_2) = \sigma^2 < \infty$$

$$Cov(X_t, X_{t+h})$$

$$= Cov\left(Z_1 Cos(ct) + Z_2 Sin(ct),\; Z_1 Cos(ct+ch) + Z_2 Sin(ct+ch)\right)$$

$$= Cos(ct)\,Cos(ct+ch)\,V(Z_1) + Sin(ct)\,Sin(ct+ch)\,V(Z_2)$$

$$= \sigma^2\left\{Cos(ct)\,Cos(ct+ch) + Sin(ct)\,Sin(ct+ch)\right\}$$

$$= \sigma^2 Cos(ct+ch-ct) = \sigma^2 Cos(ch) \quad \text{(function of } h \text{ only)}$$

So the process is stationary.

Example Let $\{Z_t\}$ be a Normal iid process with mean $\mu = 0$ and variance $\sigma^2 = 1$. Show that $X_t = Z_t Z_{t-1}$ is weak stationary.

$$E(X_t) = E(Z_t Z_{t-1}) = E(Z_t)\,E(Z_{t-1})$$
$$= 0 \quad \text{as they are independent}$$

$$V(X_t) = V(Z_t Z_{t-1})$$
$$= E(Z_t^2\, Z_{t-1}^2) - \left(E(Z_t Z_{t-1})\right)^2$$
$$= E(Z_t^2)\,E(Z_{t-1}^2) - 0 \quad [\text{As } Z_t^2 \text{ and } Z_{t-1}^2 \text{ are independent}]$$
$$= 1 \quad [Z_t^2 \sim \chi^2_{(1)}]$$

$$Cov(X_t, X_{t+h}) = 0 \quad \forall\, h \neq 0.$$

So $E(X_t)$ is not a function of time

$V(X_t) < \infty$ and not a fuction of time

$Cov(X_t, X_{t+h}) = 0 \quad \forall\, h \neq 0.$

the process in stationary.

In real life it is easy to work with weak stationary processes. *From now by stationary process we will mean weak stationary.*