STA2202 - Time Series Analysis - Assignment 2 - THEORY

Luis Correia - Student No. 1006508566

May 27th 2020

Submission instructions: Submit *three separate files* to A2 on Quercus - the deadline is 11:59PM on Tuesday, June 2.
- A PDF file with your Theory part answers.
- A PDF file with your Practice part report.
- A CSV file with your Practice part forecasts.

Theory

Question 1

1. Consider two discrete random variables X, Y with joint probabilities given by the contingency table:

$P(X,Y)$	$Y=-1$	$Y=0$	$Y=1$
$X=-1$	.05	.10	.15
$X=0$	.15	.15	.10
$X=+1$	.15	.00	.15

(a) [2 marks] Find the *Minimum Mean Square Error* (MMSE) predictor of Y given X, i.e. the conditional expectation $g(X) = \mathbb{E}[Y|X]$, and the MSE it achieves, i.e. $\mathbb{E}[(Y - g(X))^2]$.

(b) [2 marks] Find the *Best Linear Predictor* (BLP) of Y given X, i.e. $Y = a + bX$, for the BLP coefficients a, b, and the MSE it achieves.
(Note: This is an example where the MMSE predictor and the BLP are different.)

{*Solution.*}

item (a)

In this question we will calculate the MMSE Predictor $g(X) = E\big[Y\big|X\big]$ for all combinations of variables.

Calculating the Marginals we have the following table:

$P(X,Y)$	$Y=-1$	$Y=0$	$Y=1$	$P(X)$
$X=-1$	.05	.10	.15	.30
$X=0$	.15	.15	.10	.40
$X=+1$	.15	.00	.15	.30
$P(Y)$	.35	.25	.40	1.00

Now calculating the MMSE Predictor for each value of X

- For $X = -1$ we have $g(X) = E\big[Y\big|X = -1\big] = \sum_{y=-1}^{1} yP(Y = y|X = -1)$

$$g(X) = (-1)\frac{0.05}{0.30} + (1)\frac{0.15}{0.30} = \frac{0.10}{0.30} = \frac{1}{3}$$

- For $X = 0$ we have $g(X) = E\big[Y\big|X = 0\big] = \sum_{y=-1}^{1} yP(Y = y|X = 0)$

$$g(X) = (-1)\frac{0.15}{0.40} + (1)\frac{0.10}{0.40} = \frac{-0.05}{0.40} = -\frac{1}{8}$$

- For $X = 1$ we have $g(X) = E\big[Y\big|X = 1\big] = \sum_{y=-1}^{1} yP(Y = y|X = 1)$

$$g(X) = (-1)\frac{0.15}{0.30} + (1)\frac{0.15}{0.30} = \frac{0}{0.30} = 0$$

For each combination of $Y/g(X)$ we will calculate $MSE = E\big[\big(Y - g(X)\big)^2\big]$, then we have:

MSE	$g(X) = 1/3$	$g(X) = -1/8$	$g(X) = 0$
$Y = -1$	1.7777	0.1111	0.4444
$Y = 0$	0.7656	0.0156	1.2656
$Y = +1$	1.0000	0.0000	1.0000

Then multiplying each column/row for its respective probability in contingency table and summing up each term we obtain the expected Mean Square Error for $g(X)$, which is:

$$\implies MSE = 0.71042 \tag{1}$$

item (b)

In order to calculate the BLP of Y, lets first state the *Moment Generating Function* of Y, $m_Y(t)$, which will be useful when calculating the 1st and 2nd moments of Y to derive the estimates of parameters a and b of $g(X)$ and its MSE.

From the contingency table, we have that the $m_Y(t)$ is given by:

$$m_Y(t) = 0.25 + 0.35e^{-t} + 0.4e^{t} \tag{2}$$

Calculating the 1st and 2nd moments we have that:

$$\implies E(Y) = m'_Y(0) = 0.05 \tag{3}$$

and

$$\implies E(Y^2) = m''_Y(0) = 0.75 \tag{4}$$

Now, calculating the BLP, for each value of X we have 03 different BLPs, as follows:

- For $X = -1$ we have $g(X) = a + b \times (-1) = a - b$

- For $X = 0$ we have $g(X) = a + b \times (0) = a$
- For $X = 1$ we have $g(X) = a + b \times (1) = a + b$

We know from lecture that the prediction errors for BLP must be *uncorrelated* with the variables used in the prediction, i.e., we must have

$$E\big[\big(Y - g(X)\big)X\big] = 0, \forall X \in \{-1, 0, 1\} \tag{5}$$

Using the 1st moment of Y given by (3) and applying the result (5) for all possible values of X we have a system of equations that will permit estimate the values of a and b, as follows:

For $X = -1$

$$\begin{aligned}
E\big[\big(Y - (a - b)\big)(-1)\big] &= 0 \\
E\big[\big((a - b) - Y\big)\big] &= 0 \\
(a - b) - E\big(Y\big) &= 0 \\
(a - b) &= 0.05
\end{aligned}$$

and for $X = 1$

$$\begin{aligned}
E\big[\big(Y - (a + b)\big)(1)\big] &= 0 \\
E\big[\big(Y - (a + b)\big)\big] &= 0 \\
E\big(Y\big) - (a + b) &= 0 \\
(a + b) &= 0.05
\end{aligned}$$

$$\implies a = 0.05 \quad \text{and} \quad b = 0 \tag{6}$$

Then we have that (6) implies $g(X) = 0.05$.

Now, calculating the MSE of $g(X)$ and using the results (3), (4) and (6) we have that:

$$\begin{aligned}
MSE &= E\big[\big(Y - g(X)\big)^2\big] \\
&= E\big[\big(Y^2 - 2 \times (0.05)Y + (0.05)^2\big)\big] \\
&= E\big(Y^2\big) - 2 \times (0.05)E\big(Y\big) + 0.0025 \\
&= 0.75 - 0.10 \times (0.05) + 0.0025 \\
&= 0.7475
\end{aligned}$$

This concludes Question 1.

Question 2

2. Consider the AR(1) model $X_t = \phi X_{t-1} + W_t,\ \ W_t \sim \mathrm{WN}(0, \sigma_w^2)$.

(a) [3 marks] Find the covariance between the 1- & 2-step-ahead BLP errors, i.e. find

$$\mathrm{Cov}\left[(X_{n+1} - X_{n+1}^n)(X_{n+2} - X_{n+2}^n)\right]$$

as a function of (ϕ, σ_w^2).
(Note: this should be *non-zero*; generally the different-step-ahead forecasts will be correlated.)

(b) [3 marks] Find the covariance between the subsequent 1-step-ahead BLP errors, i.e. find $\mathrm{Cov}\left[(X_n - X_n^{n-1})(X_{n+1} - X_{n+1}^n)\right]$ as a function of (ϕ, σ_w^2).
(Note: These are similar to the model residuals *given perfect knowledge of the parameters.*)

{*Solution.*}

item (a)

We know that, for $AR(1)$ given by $X_t = \phi X_{t-1} + W_t$, with $W_t \sim WN(0, \sigma_W^2)$, the 1-Step-Ahead BLP estimator is given by

$$X_{n+1}^n = \phi X_n,\ \ n \geq 1 \tag{7}$$

and the 2-Step-Ahead BLP estimator is given by

$$X_{n+2}^n = \phi^2 X_n,\ \ n \geq 1 \tag{8}$$

Additionally, the ACVF is given by

$$\gamma(h) = \frac{\sigma_W^2 \phi^h}{1 - \phi^2},\ \ |\phi| < 1 \tag{9}$$

Applying (7) and (8) to the Covariance of both 1&2-Step-Ahead BLP Errors, we have:

$$\begin{aligned}
&Cov\big[\big(X_{n+1} - X_{n+1}^n\big), \big(X_{n+2} - X_{n+2}^n\big)\big] \\
&\qquad = Cov\big[\big(X_{n+1} - \phi X_n\big), \big(X_{n+2} - \phi^2 X_n\big)\big] \\
&\qquad = Cov\big(X_{n+1}, X_{n+2}\big) - \phi^2 Cov\big(X_{n+1}, X_n\big) - \phi Cov\big(X_n, X_{n+2}\big) + \phi^3 Cov\big(X_n, X_n\big) \\
&\qquad = \gamma(1) - \phi^2 \gamma(1) - \phi \gamma(2) + \phi^3 \gamma(0)
\end{aligned}$$

Using (9) in this result we obtain then:

$$\begin{aligned}
&Cov\big[\big(X_{n+1} - X_{n+1}^n\big), \big(X_{n+2} - X_{n+2}^n\big)\big] \\
&\qquad = \frac{\sigma_W^2 \phi}{1 - \phi^2} - \phi^2 \frac{\sigma_W^2 \phi}{1 - \phi^2} - \phi \frac{\sigma_W^2 \phi^2}{1 - \phi^2} + \phi^3 \frac{\sigma_W^2}{1 - \phi^2} \\
&\qquad = \frac{\sigma_W^2 \phi}{1 - \phi^2}(1 - \phi^2) \\
&\qquad = \sigma_W^2 \phi
\end{aligned}$$

Therefore, the Covariance between 1&2-Step-Ahead BLP Errors is given by

$$Cov\big[\big(X_{n+1} - X_{n+1}^n\big), \big(X_{n+2} - X_{n+2}^n\big)\big] = \sigma_W^2 \phi \tag{10}$$

item (b)

Proceeding the same way in the previous item and adapting the results from (7) to the present case, we have that

$$
\begin{aligned}
&Cov\big[\big(X_n - X_n^{n-1}\big), \big(X_{n+1} - X_{n+1}^{n}\big)\big] \\
&\qquad = Cov\big[\big(X_n - \phi X_{n-1}\big), \big(X_{n+1} - \phi X_n\big)\big]\big] \\
&\qquad = Cov\big(X_n, X_{n+1}\big) - \phi Cov\big(X_n, X_n\big) - \phi Cov\big(X_{n-1}, X_{n+1}\big) + \phi^2 Cov\big(X_{n-1}, X_n\big) \\
&\qquad = \gamma(1) - \phi\gamma(0) - \phi\gamma(2) + \phi^2\gamma(1)
\end{aligned}
$$

Using (9) we obtain then:

$$
\begin{aligned}
&Cov\big[\big(X_n - X_n^{n-1}\big), \big(X_{n+1} - X_{n+1}^{n}\big)\big] \\
&\qquad = \frac{\sigma_W^2\phi}{1-\phi^2} - \phi\frac{\sigma_W^2}{1-\phi^2} - \phi\frac{\sigma_W^2\phi^2}{1-\phi^2} + \phi^2\frac{\sigma_W^2\phi}{1-\phi^2} \\
&\qquad = 0
\end{aligned}
$$

Therefore, the Covariance between the subsequent 1-Step-Ahead BLP Errors is given by

$$
Cov\big[\big(X_n - X_n^{n-1}\big), \big(X_{n+1} - X_{n+1}^{n}\big)\big] = 0 \tag{11}
$$

This concludes Question 2.

Question 3

3. [5 marks; **STA2202 (grad) students ONLY**] *Forecasting with estimated parameters*: Let $x_1, x_2, \ldots, x_n$ be a sample of size n from a causal $AR(1)$ process, $x_t = \phi x_{t-1} + w_t$. Let $\hat{\phi}$ be the Yule–Walker estimator of ϕ.

(a) Show $\hat{\phi} - \phi = O_p(n^{-1/2})$. See Appendix A for the definition of $O_p(.)$.

(b) Let x_{n+1}^n be the one-step-ahead forecast of x_{n+1} given the data $x_1, \ldots, x_n$, based on the known parameter, ϕ, and let $\hat{x}_{n+1}^n$ be the one-step-ahead forecast when the parameter is replaced by $\hat{\phi}$. Show $x_{n+1}^n - \hat{x}_{n+1}^n = O_p(n^{-1/2})$.

(Note: the *estimated* BLPs $\hat{X}_{n+m}^n$ based on the fitted parameters $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}_w^2)$ are less accurate than the *theoretical* BLPs based on the true parameters. This questions shows that for AR(1) 1-step-ahead predictions, their difference is bounded in probability at the usual rate of $1/\sqrt{n}$.)

{*Solution.*}

item (a)

As we saw in Lecture 6, that Yule-Walker estimate $\hat{\phi}_{YW}$ is a consistent estimator of the real parameter ϕ i.e., $E(\hat{\phi}_{YW}) = \phi$.

By Tchebycheff's inequality we also have that:

$$P\{|\hat{\phi}_{YW} - \phi| > \epsilon\} \leq \frac{E[(\hat{\phi}_{YW} - \phi)^2]}{\epsilon^2} \tag{12}$$

We also saw in class that the distribution of the M.S.E in the right side of (12) converges to a Normal distribution with variance equals to $\sigma_w^2 \Gamma_n^{-1}/n$.

$$\Longrightarrow P\{|\hat{\phi}_{YW} - \phi| > \epsilon\} \leq \frac{\sigma_w^2 \Gamma_n^{-1}}{n\epsilon^2} \to 0, \text{as } n \to \infty \tag{13}$$

Then we have that $\hat{\phi}_{YW} \xrightarrow{P} \phi$.

From this point, in order to find the rate of convergence, it follows from Tchebycheff's inequality that:

$$P\{\sqrt{n}|\hat{\phi}_{YW} - \phi| > \delta(\epsilon)\} \leq \frac{\sigma_w^2 \Gamma_n^{-1}/n}{\delta^2(\epsilon)/n} = \frac{\sigma_w^2 \Gamma_n^{-1}}{\delta^2(\epsilon)} \tag{14}$$

By doing $\epsilon = \sigma_w^2 \Gamma_n^{-1}/\delta^2(\epsilon)$ it follows that $\delta^2(\epsilon) = \sigma_w \Gamma_n^{-1/2}/\sqrt{\epsilon}$

$$\Longrightarrow \hat{\phi}_{YW} - \phi = O(n^{1/2}) \tag{15}$$

item (b)

Now let's consider X_{n+1}^n be the 1-Step-Ahead forecast for X_{n+1} and $\hat{X}_{n+1}^n$ as the 1-Step-Ahead forecast when the parameter is replaced by $\hat{\phi}$.

Using the same approach adopted in previous item, the result from (7) we have:

$$\left|X_{n+1} - \hat{X}_{n+1}^n\right| = \left|X_{n+1} - \phi X_n\right| \tag{16}$$

Using similar approach used in (12) and (13), we have that:

$$P\{\left|\hat{X}^n_{n+1} - X_{n+1}\right| > \epsilon\} \leq \frac{E[(X_{n+1} - \hat{X}^n_{n+1})^2]}{\epsilon^2} = \frac{MSE}{\epsilon^2} \tag{17}$$

Now, calculating the MSE for 1-Step-Ahead and using the result in (9) we have that:

$$\begin{aligned}
E\big[(X_{n+1} - \hat{X}^n_{n+1})^2\big] &= E\big[(X_{n+1} - \phi X_n)^2\big] \\
&= E\big(X^2_{n+1} - 2\phi X_{n+1}X_n + \phi^2 X^2_n\big) \\
&= E\big(X^2_{n+1}\big) - 2\phi E\big(X_{n+1}X_n\big) + \phi^2 E\big(X^2_n\big) \\
&= Var\big(X_{n+1}\big) - 2\phi Cov\big(X_{n+1}, X_n\big) + \phi^2 Var\big(X_n\big) \\
&= \gamma(0) - 2\phi\gamma(1) + \phi^2\gamma(0) \\
&= \frac{\sigma^2_W}{1-\phi^2} - 2\phi\frac{\sigma^2_W\phi}{1-\phi^2} + \phi^2\frac{\sigma^2_W}{1-\phi^2} \\
&= \frac{(1-\phi^2)\sigma^2_W}{1-\phi^2} \\
&= \sigma^2_W
\end{aligned}$$

$$\implies E\big[(X_{n+1} - \hat{X}^n_{n+1})^2\big] = \sigma^2_W \tag{18}$$

Substituting (18) in (17), we have that

$$P\{\left|X_{n+1} - \hat{X}^n_{n+1}\right| > \epsilon\} \leq \frac{\sigma^2_W}{\epsilon^2} \tag{19}$$

Using the same approach of convergence as in previous item, it follows from Tchebycheff's inequality that:

$$P\{\sqrt{n}\left|X_{n+1} - \hat{X}^n_{n+1}\right| > \delta(\epsilon)\} \leq \frac{\sigma^2_w/n}{\delta^2(\epsilon)/n} = \frac{\sigma^2_w}{\delta^2(\epsilon)} \tag{20}$$

By doing $\epsilon = \sigma^2_w/\delta^2(\epsilon)$ it follows that $\delta^2(\epsilon) = \sigma_w/\sqrt{\epsilon}$. Then we have:

$$\implies X_{n+1} - \hat{X}^n_{n+1} = O\big(n^{1/2}\big) \tag{21}$$

This concludes Question 3.