Time Series Foundamentals

Kaiwen Zhou

Contents

1 Topic: Mean Squared Prediction

Suppose $\{X_t\}$ is a stationary process with mean μ, and auto-covariance function $\gamma_X(h)$. Our goal is to predict: X_{n+h} using observations $X_1, X_2, \ldots, X_n$.

We define the **Best Linear Predictor** to be the one minimizes the Mean Squared Prediction Error (MSPE):

$$\widehat{X}_{n+h} = E\left[X_{n+h} \mid X_1, \cdots, X_n\right] = P_n X_{n+h} := a_0^* + a_1^* X_n + \cdots + a_n^* X_1 = a_0 + \mathbf{a}_n^\top \mathbf{x}_n = \arg\min_{a_0, \mathbf{a}_n} \mathbb{E}\left[\left(X_{n+h} - a_0 - \mathbf{a}_n^\top \mathbf{x}_n\right)^2\right] = \arg\min_{a_0, \mathbf{a}_n} L(a_0, \mathbf{a}_n)$$

where $\mathbf{a}_n = (a_1, \ldots, a_n) \in \mathbb{R}^n$, $\mathbf{x}_n = (x_n, \ldots, x_1) \in \mathbb{R}^n$, P in P_n stands for "Predictor" and n suggests that we are using n most recent observations.

To simplify our computation, we define $\mathbb{1} = (1, \ldots, 1) \in \mathbb{R}^n$, $\boldsymbol{\gamma}_n(h) = \begin{bmatrix} \gamma_X(h) \\ \gamma_X(h+1) \\ \vdots \\ \gamma_X(h+n-1) \end{bmatrix} \in \mathbb{R}^n$, $\Gamma_n = \begin{bmatrix} \gamma_X(0) & \gamma_X(1) & \cdots & \gamma_X(n-1) \\ \gamma_X(1) & \gamma_X(0) & \cdots & \gamma_X(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_X(n-1) & \gamma_X(n-2) & \cdots & \gamma_X(0) \end{bmatrix} \in \mathbb{R}^{n \times n}$.

Taking the partial derivatives of L, we obtain

$$0 = \frac{\partial L}{\partial a_0} = \mathbb{E}\left[\frac{\partial}{\partial a_0}(X_{n+h} - a_0 - \mathbf{a}_n^\top \mathbf{x}_n)^2\right] = \mathbb{E}\left[-2(X_{n+h} - a_0 - \mathbf{a}_n^\top \mathbf{x}_n)\right] = 2(a_0 - \mu - \mathbf{a}_n^\top \mathbb{1}) \Longrightarrow a_0^* = \mu(1 - \mathbf{a}_n^\top \mathbb{1})$$

and

$$\begin{aligned}
\mathbf{0} = \frac{\partial L}{\partial \mathbf{a}_n} &= \frac{\partial}{\partial \mathbf{a}_n} \mathbb{E}\left[(X_{n+h} - a_0 - \mathbf{a}_n^\top \mathbf{x}_n)^2\right] \\
&= \mathbb{E}\left[\frac{\partial}{\partial \mathbf{a}_n}\left(\mathbf{a}_n^\top \mathbf{x}_n \mathbf{x}_n^\top \mathbf{a}_n - 2X_{n+h}\mathbf{x}_n^\top \mathbf{a}_n + 2a_0 \mathbf{x}_n^\top \mathbf{a}_n + (\text{ terms without } \mathbf{a}_n)\right)\right] \\
&= 2\mathbb{E}[\mathbf{x}_n \mathbf{x}_n^\top \mathbf{a}_n - X_{n+h}\mathbf{x}_n + a_0 \mathbf{x}_n] \\
&= 2\mathbb{E}[\mathbf{x}_n \mathbf{x}_n^\top]\mathbf{a}_n - 2\mathbb{E}[X_{n+h}\mathbf{x}_n] + 2\mathbb{E}[a_0 \mathbf{x}_n] \\
&= 2(\Gamma_n + \mu^2 \mathbb{1}\mathbb{1}^\top)\mathbf{a}_n - 2(\boldsymbol{\gamma}_n(h) + \mu^2 \mathbb{1}) + 2a_0 \mu \mathbb{1} \\
a_0 = \mu(1 - \mathbf{a}_n^\top \mathbb{1}) \Longrightarrow &= 2(\Gamma_n + \mu^2 \mathbb{1}\mathbb{1}^\top)\mathbf{a}_n - 2(\boldsymbol{\gamma}_n(h) + \mu^2 \mathbb{1}) + 2\mu(1 - \mathbf{a}_n^\top \mathbb{1})\mu \mathbb{1} \\
&= 2\Gamma_n \mathbf{a}_n - 2\boldsymbol{\gamma}_n(h) \\
\Longrightarrow \mathbf{a}_n^* &= \Gamma_n^{-1} \boldsymbol{\gamma}_n(h)
\end{aligned}$$

Therefore, the **Best Linear Predictor** is given by

$$P_n X_{n+h} := a_0^* + \mathbf{a}_n^{*\top} \mathbf{x}_n \text{ where } \begin{cases} a_0^* &= \mu(1 - \mathbf{a}_n^{*\top} \mathbb{1}) \\ \mathbf{a}_n^* &= \Gamma_n^{-1} \boldsymbol{\gamma}_n(h) \end{cases} \tag{1}$$

Problem 1.1. *Show that Mean Squared Predictive Error is obtained by*

$$\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] = \gamma_X(0) - \mathbf{a}_n^{*\top} \boldsymbol{\gamma}_n(h)$$

Solution: From above, we get

$$\begin{aligned}
\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] &= \mathbb{E}\left[\left(X_{n+h} - a_0^* - \mathbf{a}_n^{*\top} \mathbf{x}_n\right)^2\right] \\
\mathbf{a}_n^{*\top} \mathbf{x}_n = \mathbf{x}_n^\top \mathbf{a}_n^* \Longrightarrow &= \mathbb{E}\left[X_{n+h}^2 + a_0^{*2} + \mathbf{a}_n^{*\top} \mathbf{x}_n \mathbf{x}_n^\top \mathbf{a}_n^* - 2a_0^* X_{n+h} - 2X_{n+h}\mathbf{a}_n^{*\top} \mathbf{x}_n + 2a_0^* \mathbf{a}_n^{*\top} \mathbf{x}_n\right] \\
&= \mathbb{E}\left[X_{n+h}^2\right] + a_0^{*2} + \mathbf{a}_n^{*\top} \mathbb{E}[\mathbf{x}_n \mathbf{x}_n^\top]\mathbf{a}_n^* - 2a_0^* \mathbb{E}[X_{n+h}] - 2\mathbf{a}_n^{*\top} \mathbb{E}[X_{n+h}\mathbf{x}_n] + 2a_0^* \mathbf{a}_n^{*\top} \mathbb{E}\left[\mathbf{x}_n\right] \\
&= \gamma_X(0) + \mu^2 + \mu^2(1 - \mathbf{a}_n^{*\top} \mathbb{1})^2 + \mathbf{a}_n^{*\top}(\Gamma_n + \mu^2 \mathbb{1}\mathbb{1}^\top)\mathbf{a}_n^* - 2\mu^2(1 - \mathbf{a}_n^{*\top} \mathbb{1}) - 2\mathbf{a}_n^{*\top}(\boldsymbol{\gamma}_n(h) + \mu^2 \mathbb{1}) + 2\mu^2(1 - \mathbf{a}_n^{*\top} \mathbb{1})\mathbf{a}_n^{*\top} \mathbb{1} \\
&= \gamma_X(0) + \mathbf{a}_n^{*\top} \Gamma_n \mathbf{a}_n^* - 2\mathbf{a}_n^{*\top} \boldsymbol{\gamma}_n(h) \\
\mathbf{a}_n^* = \Gamma_n^{-1} \boldsymbol{\gamma}_n(h) \Longrightarrow &= \gamma_X(0) + \mathbf{a}_n^{*\top} \Gamma_n \Gamma_n^{-1} \boldsymbol{\gamma}_n(h) - 2\mathbf{a}_n^{*\top} \boldsymbol{\gamma}_n(h) \\
&= \gamma_X(0) - \mathbf{a}_n^{*\top} \boldsymbol{\gamma}_n(h)
\end{aligned}$$

Hence, we can conclude that the Mean Squared Predictive Error is obtained by

$$\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] = \gamma_X(0) - \mathbf{a}_n^{*\top} \boldsymbol{\gamma}_n(h)$$

Remark 1.2.

1. For a linear Gaussian process $\{X_t\}$, take $X_{n+h} \mid X_1, \ldots, X_n = P_n X_{n+h} + \epsilon = a_0^* + \mathbf{a}_n^{*\top} \mathbf{x}_n + \epsilon$, then the residual term ϵ must be Gaussian. Then, we have

$$\mathbb{E}[X_{n+h} \mid X_1, \ldots, X_n - P_n X_{n+h}] = \mu - a_0^* + \mu \mathbf{a}_n^{*\top} \mathbb{1} = 0 \xrightarrow{a_0^* = \mu(1 - \mathbf{a}_n^{*\top} \mathbb{1})} \epsilon \sim \mathcal{N}(0, \mathrm{MSPE}(P_n X_{n+h})) = \mathcal{N}(0, \gamma_X(0) - \mathbf{a}_n^{*\top} \boldsymbol{\gamma}_n(h))$$

   by our computation; thus, the predictive distribution is also Gaussian:

$$X_{n+h} \mid X_1, \ldots, X_n \sim \mathcal{N}\left(P_n X_{n+h}, \mathrm{MSPE}\left(P_n X_{n+h}\right)\right) = \mathcal{N}\left(a_0^* + \mathbf{a}_n^{*\top} \mathbf{x}_n, \gamma_X(0) - \mathbf{a}_n^{*\top} \boldsymbol{\gamma}_n(h)\right)$$

   where $a_0^* = \mu(1 - \mathbf{a}_n^{*\top} \mathbb{1})$ and $\mathbf{a}_n^* = \Gamma_n^{-1} \boldsymbol{\gamma}_n(h)$.

2. For large n, Γ_n^{-1} is obtained through recursive algorithms. As an example look at Durbin-Levinson Algorithm

△

2 Topic: MLE for AR and MA processes

Direct MLE

Suppose that $\{X_t\}$ is a Gaussian time series (process) with mean μ and autocovariance function $\gamma_X(i, j)$. Let $\mathbf{x}_n = (X_1, \ldots, X_n) \in \mathbb{R}^n$. Let Γ_n denote the covariance matrix $\Gamma_n = Cov\left(\mathbf{x}_n, \mathbf{x}_n\right)$, and assume that Γ_n is nonsingular. Since any subset of a Gaussian process are jointly Gaussian, the likelihood function for given observation $\mathbf{x}_n$ (a single obseravtion) is

$$L\left(\Psi\right) = f(\mathbf{x}_n \mid \Psi) = \frac{1}{\sqrt{(2\pi)^n |\Gamma_n|}} \exp\left(-\frac{1}{2}(\mathbf{x}_n - \mu)^\top \Gamma_n^{-1} (\mathbf{x}_n - \mu)\right).$$

where Ψ is the parameter set, e.g.for the ARMA model, $\Psi_{ARMA(p,q)} = \left(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma^2\right)$.

To simplify our computation, we could consider the log-likelihood function:

$$\log L(\Psi) = -\frac{n}{2} \log(2\pi) + \frac{1}{2} \log |\Gamma_n|^{-1} - \frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu})^\top \Gamma_n^{-1} (\mathbf{x}_n - \boldsymbol{\mu})$$

Then, the values of parameters Ψ which maximizes the Log-likelihood function is the Maximum Likelihood Estimate of Ψ. That is,

$$\hat{\Psi}_{MLE} = \arg\max_{\Psi} \log L(\Psi)$$

Conditional MLE

For the ease of implementation, we can also use the conditional MLE. Using the fact that

$$f(\mathbf{x}_n \mid \Psi) = f(x_1, \ldots, x_n \mid \Psi) = f(x_n \mid x_1, \ldots, x_{n-1}, \Psi) f(x_1, \ldots, x_{n-1} \mid \Psi) \overset{\cdots}{=\!=} f(x_n \mid x_1, \ldots, x_{n-1}, \Psi) f(x_{n-1} \mid x_1, \ldots, x_{n-2}, \Psi) \cdots f(x_1 \mid \Psi)$$

Then, the MLE is given by

$$\log L(\Psi) = \log f(x_1 \mid \Psi) + \sum_{i=2}^{n} \log f(x_i \mid x_1, \ldots, x_{i-1}, \Psi) \tag{2}$$

Similarly, the values of parameters Ψ which maximizes the Log-likelihood function is the Maximum Likelihood Estimate of Ψ. That is,

$$\hat{\Psi}_{MLE} = \arg\max_{\Psi} \log L(\Psi)$$

Remark 2.1.

1. We mostly use Conditional MLE in applications.
2. Since X_1 is the the first data point, the way it's picked is on its own. If X_1 is pick based on some distribution, we can surely keep the term $\log f(x_1 \mid \Psi)$ in the above equation and work with the exact MLE. If how X_1 is picked is unknown, we can treat it as deterministic, and drop the term $\log f(x_1 \mid \Psi)$ in the above equation.

△

Problem 2.2. *Using a programming language of your choice, repeat the MLE construction and simulation for AR and MA models as discussed this in the class.*

Solution:

(a) **Conditional MLE for AR(1) Process**

Consider observations $\{X_1, X_2, \ldots, X_n\}$ of a stationary Gaussian AR(1) process. For parameter vector is $\Psi = \left(\phi_0, \phi, \sigma^2\right)$, we have

$$X_t = \phi_0 + \phi X_{t-1} + Z_t, Z_t \sim IIDN\left(0, \sigma^2\right), \quad \mathbb{E}[X_t] = \frac{\phi_0}{1 - \phi}, Var\left(X_t\right) = \frac{\sigma^2}{1 - \phi^2} \text{ and } \gamma_X(h) = \frac{\phi^h \sigma^2}{1 - \phi^2}$$

The probability distribution for X_1 can be written as,

$$X_1 \sim \mathcal{N}\left(\frac{\phi_0}{1 - \phi}, \frac{\sigma^2}{1 - \phi^2}\right) \Longrightarrow f\left(x_1; \Psi\right) = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{1 - \phi^2}}} \exp\left\{\frac{-\left(x_1 - \frac{\phi_0}{1 - \phi}\right)^2}{2\frac{\sigma^2}{1 - \phi^2}}\right\}$$

Since $X_2 = \phi_0 + \phi X_1 + Z_2$, $X_2 \mid X_1 = x_1 \sim \mathcal{N}\left(\phi_0 + \phi x_1, \sigma^2\right)$. The conditional probability distribution for X_2 knowing X_1 can be written as,

$$f\left(x_2 \mid x_1; \Psi\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-\left(x_2 - \phi_0 - \phi x_1\right)^2}{2\sigma^2}\right\}$$

Similarly, we have $X_n \mid X_{n-1} = x_{n-1} \sim \mathcal{N}\left(\phi_0 + \phi x_{n-1}, \sigma^2\right)$ and

$$f\left(x_n \mid x_{n-1}; \Psi\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-\left(x_n - \phi_0 - \phi x_{n-1}\right)^2}{2\sigma^2}\right\}$$

Plug these in Equation (2), we obtain the Log-Likelihood function:

$$\begin{aligned}
\log L(\Psi) &= \log f(x_1 \mid \Psi) + \sum_{i=2}^{n} \log f(x_i \mid x_1, \ldots, x_{i-1}, \Psi) \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{\left(x_1 - \frac{\phi_0}{1-\phi}\right)^2}{2\frac{\sigma^2}{1-\phi^2}} + \sum_{i=2}^{n} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{\left(x_i - \phi_0 - \phi x_{i-1}\right)^2}{2\sigma^2} \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{\left(x_1 - \frac{\phi_0}{1-\phi}\right)^2}{2\frac{\sigma^2}{1-\phi^2}} - \frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log\left(\sigma^2\right) - \sum_{i=2}^{n} \frac{\left(x_i - \phi_0 - \phi x_{i-1}\right)^2}{2\sigma^2}
\end{aligned}$$

Maximize this function to estimate the parameters. (Much easier to implement)

Example: Suppose we have AR(1) process $X_t = 0.9X_{t-1} + Z_t, Z_t \sim IIDN\left(0, 0.7^2\right)$. We simulate this process and pretend that we do not know the true values for ϕ and σ. Then, we use the conditional MLE to estimate the parameter vector $\Psi = \left(\phi, \sigma^2\right)$ By the above deduction, we get the log likelihood function is

$$\begin{aligned}
\log L(\Psi) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{x_1^2}{2\frac{\sigma^2}{1-\phi^2}} - \frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log\left(\sigma^2\right) - \sum_{i=2}^{n} \frac{\left(x_i - \phi x_{i-1}\right)^2}{2\sigma^2} \\
\text{treat } X_1 \text{ as deterministic } &= -\frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log\left(\sigma^2\right) - \sum_{i=2}^{n} \frac{\left(x_i - \phi x_{i-1}\right)^2}{2\sigma^2}
\end{aligned}$$

Then, we have

$$\hat{\Psi}_{MLE} = \arg\max_{\Psi} \log L(\Psi)$$

For the coding work on this problem, check the attached jupyter notebook `Time Series and Statistical Arbitrage HW2.ipynb`

(b) **Conditional MLE for MA(1) Process**

Consider observations $\{X_1, X_2, \ldots, X_n\}$ of a Gaussian MA(1) process. For parameter vector is $\Psi = \left(\theta_0, \theta_1, \sigma^2\right)$, we have

$$X_t = \theta_0 + \theta_1 Z_{t-1} + Z_t, \quad Z_t \sim IIDN\left(0, \sigma^2\right)$$

The probability distribution for X_t can be written as,

$$X_t \mid Z_{t-1} \sim \mathcal{N}\left(\theta_0 + \theta_1 z_{t-1}, \sigma^2\right) \Longrightarrow f\left(x_t \mid z_{t-1}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-\left(x_t - (\theta_0 + \theta_1 z_{t-1})\right)^2}{2\sigma^2}\right\}$$

Plug these in Equation (2), we obtain the Log-Likelihood function:

$$\begin{aligned}
\log L(\Psi) &= \log f(x_1 \mid \Psi) + \sum_{i=2}^{n} \log f(x_i \mid x_1, \ldots, x_{i-1}, \Psi) \\
\text{treat } X_1 \text{ as deterministic } &= -\frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log\left(\sigma^2\right) - \sum_{i=2}^{n} \frac{\left(x_t - \theta_0 - \theta_1 z_{t-1}\right)^2}{2\sigma^2}
\end{aligned}$$

Maximize this function to estimate the parameters. (Much easier to implement)

Example: Suppose we have MA(1) process $X_t = 0.5Z_{t-1} + Z_t, Z_t \sim IIDN\left(0, 0.5^2\right)$. We simulate this process and pretend that we do not know the true values for ϕ and σ. Then, we use the conditional MLE to estimate the parameter vector $\Psi = \left(\theta_1, \sigma^2\right)$ By the above deduction, we get the log likelihood function is

$$\log L(\Psi) = -\frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log\left(\sigma^2\right) - \sum_{i=2}^{n} \frac{\left(x_t - \theta_1 z_{t-1}\right)^2}{2\sigma^2}$$

Then, we have

$$\hat{\Psi}_{MLE} = \arg\max_{\Psi} \log L(\Psi)$$

For the coding work on this problem, check the attached jupyter notebook `Time Series and Statistical Arbitrage HW2.ipynb`

3 Topic: MLE for AR and MA processes

Problem 3.1. *For the time series model below:*

$$X_t = 0.9X_{t-1} + Z_t \quad Z_t \textit{ is } N\left(0, 0.7^2\right)$$

(I) Simulate a path with 1000 data points.

(II) Take the simulated path as the realized data and estimate the parameters using the MLE estimator.

(III) Repeat steps (I) and (II) for 100 times and plot the distributions of the parameters you have estimated. What are the 95% confidence levels around your mean estimated parameters?

(IV) Repeat (III) but use 5000 times instead of 100 times.

Use any programming language or algorithm you are comfortable with but we need to see the code.

Solution: Please check the attached jupyter notebook `Time Series and Statistical Arbitrage HW2.ipynb`

4 Application: Mean Squared Prediction and the Best Linear Predictor

Problem 4.1. *An MA(1) process is given by*

$$X_t = Z_t + \theta Z_{t-1} \quad \theta = 0.8, \quad Z_t \sim WN(0,1)$$

Our observations indicates that $\{X_1 = 3.2020, X_2 = 1.5625\}$.

What is the best linear prediction and MSPE for X_3*?*

Solution: From Equation (1), we know that

$$P_n X_{n+h} := a_0^* + \mathbf{a}_n^{*\top}\mathbf{x}_n \text{ where } \begin{cases} a_0^* &= \mu(1 - \mathbf{a}_n^{*\top}\mathbb{1}) \\ \mathbf{a}_n^* &= \Gamma_n^{-1}\boldsymbol{\gamma}_n(h) \end{cases} \xrightarrow{n=2,h=1} P_2 X_3 := a_0^* + \mathbf{a}_2^{*\top}\mathbf{x}_2 \text{ where } \begin{cases} a_0^* &= \mu(1 - \mathbf{a}_2^{*\top}\mathbb{1}) \\ \mathbf{a}_2^* &= \Gamma_2^{-1}\boldsymbol{\gamma}_2(1) \end{cases}$$

Since $\mathbb{E}[X_t] = 0 = \mu$ and $\gamma_X(h) = (1+\theta^2)\gamma_Z(h) + \theta\gamma_Z(h-1) + \theta\gamma_Z(h+1) = \begin{cases} (1+\theta^2)\sigma^2 & \text{if } h = 0 \\ \theta\sigma^2 & \text{if } h = \pm 1 \\ 0 & \text{otherwise} \end{cases}$, we have

$$\Gamma_2 = \begin{bmatrix} \gamma_X(0) & \gamma_X(1) \\ \gamma_X(1) & \gamma_X(0) \end{bmatrix} = \begin{bmatrix} (1+\theta^2)\sigma^2 & \theta\sigma^2 \\ \theta\sigma^2 & (1+\theta^2)\sigma^2 \end{bmatrix}, \boldsymbol{\gamma}_2(1) = \begin{bmatrix} \gamma_X(1) \\ \gamma_X(2) \end{bmatrix} = \begin{bmatrix} \theta\sigma^2 \\ 0 \end{bmatrix} \Longrightarrow \begin{cases} a_0^* &= 0 \\ \mathbf{a}_2^* &= \Gamma_2^{-1}\boldsymbol{\gamma}_2(1) = \frac{1}{1+\theta^2+\theta^4}\begin{bmatrix} \theta+\theta^3 \\ -\theta^2 \end{bmatrix} \overset{\theta=0.8}{=\!=} \begin{bmatrix} \frac{820}{1281} \\ -\frac{400}{1281} \end{bmatrix} \end{cases}$$

Therefore, the Best Linear Predictor for X_3 is

$$\hat{X}_3 = P_n X_{n+h} := a_0^* + \mathbf{a}_n^{*\top}\mathbf{x}_n = \frac{820}{1281}X_2 - \frac{400}{1281}X_1 = 1.5618$$

The corresponding MSPE is

$$MSPE = \mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] = \gamma_X(0) - \mathbf{a}_n^{*\top}\boldsymbol{\gamma}_n(h) \xrightarrow{n=2,h=1} = \gamma_X(0) - \mathbf{a}_2^{*\top}\boldsymbol{\gamma}_2(1) = 1.64 - \begin{bmatrix} \frac{820}{1281} & -\frac{400}{1281} \end{bmatrix}\begin{bmatrix} 0.8 \\ 0 \end{bmatrix} = \frac{656}{1281}$$

Hence, the Best Linear Predictor for X_3 is $\hat{X}_3 = 1.5618$ and the corresponding $MSPE = \frac{656}{1281}$.