1. Let $\boldsymbol{a} = (a_1, \ldots, a_k)^T$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)^T$, $\mathbf{1} = (1, \ldots, 1)^T$, $\boldsymbol{W} = (W_1, \ldots, W_k)^T$ and I_k be the identity matrix of order k, the MSE of $\hat{s}$ is

$$
\begin{aligned}
Q(\boldsymbol{a}) := E(\hat{s} - s)^2 &= E(\boldsymbol{\eta}^T\boldsymbol{a} - s)^T(\boldsymbol{\eta}^T\boldsymbol{a} - s) \\
&= \boldsymbol{a}^T E(\boldsymbol{\eta}\boldsymbol{\eta}^T)\boldsymbol{a} - 2\boldsymbol{a}^T E\boldsymbol{\eta} s + Es^2 \\
&= \boldsymbol{a}^T E(s\mathbf{1} + \boldsymbol{W})(s\mathbf{1} + \boldsymbol{W})^T\boldsymbol{a} - 2\boldsymbol{a}^T E(s\mathbf{1} + \boldsymbol{W})s + \sigma_s^2 \\
&= \sigma_s^2\boldsymbol{a}^T\mathbf{1}\mathbf{1}^T\boldsymbol{a} + \sigma_W^2\boldsymbol{a}^T\boldsymbol{a} - 2\sigma_s^2\boldsymbol{a}^T\mathbf{1} + \sigma_s^2.
\end{aligned}
$$

Letting $\frac{\partial Q(\boldsymbol{a})}{\partial \boldsymbol{a}} = 0$, we have

$$
\begin{aligned}
\boldsymbol{a} &= \sigma_s^2(\sigma_s^2\mathbf{1}\mathbf{1}^T + \sigma_W^2 I_k)^{-1}\mathbf{1} \\
&= \frac{\sigma_s^2}{\sigma_W^2 + k\sigma_s^2}\mathbf{1}.
\end{aligned}
$$

The resultant MSE is $\frac{\sigma_s^2\sigma_W^2}{\sigma_W^2 + k\sigma_s^2}$.

2. (1)

$$
\begin{aligned}
\sigma_n^2 &= E(X_t - \boldsymbol{a}^T\boldsymbol{X_n})(X_t - \boldsymbol{a}^T\boldsymbol{X_n}) \\
&= EX_t^2 - 2\boldsymbol{a}^T E(X_t\boldsymbol{X_n}) + \boldsymbol{a}^T E(\boldsymbol{X_n}\boldsymbol{X_n}^T)\boldsymbol{a}
\end{aligned}
$$

Letting $\frac{\partial \sigma_n^2(\boldsymbol{a})}{\partial \boldsymbol{a}} = 0$, we have $\boldsymbol{a} = \Gamma_n^{-1}E(X_t\boldsymbol{X_n})$, which yields

$$
\sigma_n^2 = EX_t^2 - E(X_t\boldsymbol{X_n})^T\Gamma_n^{-1}E(X_t\boldsymbol{X_n})
$$

By the stationarity of $\{X_n\}$, we have

$$
\Gamma_{n+1} = \begin{bmatrix} EX_t^2 & E(X_t\boldsymbol{X_n}^T) \\ E(X_t\boldsymbol{X_n}^T) & \Gamma_n \end{bmatrix}.
$$

Thus,

$$
\begin{aligned}
det(\Gamma_{n+1}) &= \begin{vmatrix} EX_t^2 - E(X_t\boldsymbol{X_n}^T)\Gamma_n^{-1}E(X_t\boldsymbol{X_n}) & E(X_t\boldsymbol{X_n}^T) \\ 0 & \Gamma_n \end{vmatrix} \\
&= det(\Gamma_n) \cdot [EX_t^2 - E(X_t\boldsymbol{X_n}^T)\Gamma_n^{-1}E(X_t\boldsymbol{X_n})].
\end{aligned}
$$

That is $\sigma_n^2 = \frac{det(\Gamma_{n+1})}{det(\Gamma_n)}$.

(2)

$$
\begin{aligned}
\lim_{n\to\infty}\sigma_n^2 &= \lim_{n\to\infty}\frac{det(\Gamma_{n+1})}{det(\Gamma_n)} \\
&= \exp(\lim_{n\to\infty}\log(det(\Gamma_{n+1})) - \log(det(\Gamma_n))) \\
&= \exp\left(\lim_{n\to\infty}\frac{\log(det(\Gamma_{n+1})) - \log(det(\Gamma_n))}{(n+1) - n}\right)
\end{aligned}
$$

By Stolz's theorem, $\lim_{n\to\infty}\sigma_n^2 = \exp(\lim_{n\to\infty}\frac{1}{n}\log(det(\Gamma_n))$

3. (1) Since $\{W(n)\}$ are i.i.d, the innovation of $X(n)$ is

$$X(n) - L(X(n)|X(n-1),\dots) = -\sum_{k=1}^{n}\binom{k+2}{2}X(n-k) + W(n) - \left(-\sum_{k=1}^{n}\binom{k+2}{2}X(n-k)\right)$$
$$= W(n).$$

(2) Since the equation holds clearly when $n=0$, the proof completes using mathematical induction.

(3) Let $\boldsymbol{X} = (X(1),\dots,X(10))^T$ and $\boldsymbol{W} = (W(0),W(1),\dots,W(10))^T$, we have

$$\boldsymbol{X} = A\boldsymbol{W},$$

where

$$A = \begin{bmatrix} -3 & 1 & 0 & 0 & \cdots & 0 \\ 3 & -3 & 1 & 0 & \cdots & 0 \\ -1 & 3 & -3 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Hence, $\Gamma := Cov(\boldsymbol{X}) = ACov(\boldsymbol{W})A^T = AA^T$

Solving $\Gamma\boldsymbol{a} = E(\boldsymbol{X}X(12)) = (0,\dots,0,-1,6)^T$ with respect to $\boldsymbol{a}$, we conclude from Property 1 (textbook p.146) that

$$L(X(12)|\boldsymbol{X}) = \boldsymbol{a}^T\boldsymbol{X}$$

The corresponding prediction error is $E(\boldsymbol{a}^T\boldsymbol{X} - X(12))^2$.

4. (1) The best linear predictions are

$$L(Y_{t+1}|Y_t,\dots) = L(40+\epsilon_{t+1} - 0.6\epsilon_t + 0.8\epsilon_{t-1}|\epsilon_t,\epsilon_{t-1}) = 40 - 0.6\epsilon_t + 0.8\epsilon_{t-1} = 35.6$$
$$L(Y_{t+2}|Y_t,\dots) = L(40+\epsilon_{t+2} - 0.6\epsilon_{t+1} + 0.8\epsilon_t|\epsilon_t) = 40 + 0.8\epsilon_t = 40 + 0.8\epsilon_t = 41.6$$

(2) Since

$$Y_{t+1} - L(Y_{t+1}|Y_t,\dots) = \epsilon_{t+1} \sim N(0,20)$$

and

$$Y_{t+2} - L(Y_{t+2}|Y_t,\dots) = \epsilon_{t+2} - 0.6\epsilon_{t+1} \sim N(0,27.2),$$

the 95% confidence intervals for Y_{t+1} and Y_{t+2} are respectively

$$35.6 \pm 1.96\sqrt{20},\ 41.6 \pm 1.96\sqrt{27.2}.$$

(3) According to the spectral density of MA processes, the spectral density of Y_t is

$$f(\lambda) = \frac{20}{2\pi}|1 - 0.6e^{-i\lambda} + 0.8e^{-2i\lambda}| = \frac{10}{\pi}(1 - 2.16\cos\lambda + 1.6\cos 2\lambda).$$

5. (1) Let $\gamma_k = Cov(X_t, X_{t-k}),\ k \in \mathbb{Z}$ and $\boldsymbol{X} = (X_t, X_{t-1})^T$, then

$$E(Y_{t+1}\boldsymbol{X}) = \frac{1}{2}E[(X_{t+1} + X_t)\boldsymbol{X}] = \frac{1}{2}(\gamma_0 + \gamma_1, (1+\phi_1)\gamma_1)^T$$

Solving $E(\boldsymbol{X}\boldsymbol{X}^T)\boldsymbol{a} = E(Y_{t+1}\boldsymbol{X})$, we have

$$\begin{aligned}\boldsymbol{a} &= (E(\boldsymbol{X}\boldsymbol{X}^T))^{-1}E(Y_{t+1}\boldsymbol{X}) \\ &= \frac{1}{2(\gamma_0^2 - \gamma_1^2)}(\gamma_0^2 + \gamma_0\gamma_1 - (1+\phi_1)\gamma_1^2, \phi_1\gamma_0\gamma_1 - \gamma_1^2)^T,\end{aligned}$$

from which the best linear prediction of Y_{t+1} is $L(Y_{t+1}|X_t, X_{t-1}) = \boldsymbol{a}^T\boldsymbol{X}$. Its MSE is

$$\begin{aligned}E(\boldsymbol{a}^T\boldsymbol{X} - Y_{t+1})^2 &= \frac{1}{4}E(-X_{t+1} + \frac{\gamma_0\gamma_1 - \gamma_1^2}{\gamma_0^2 + \gamma_1^2}X_t + \frac{\phi_1\gamma_0\gamma_1 - \gamma_1^2}{\gamma_0^2 + \gamma_1^2}X_{t-1})^2 \\ &= \frac{1}{4}\boldsymbol{b}^T E[(X_{t+1}, X_t, X_{t-1})^T(X_{t+1}, X_t, X_{t-1})]\boldsymbol{b}\end{aligned}$$

where $\boldsymbol{b} = (-1, \frac{\gamma_0\gamma_1 - \phi_1\gamma_1^2}{\gamma_0^2+\gamma_1^2}, \frac{\phi_1\gamma_0\gamma_1 - \gamma_1^2}{\gamma_0^2+\gamma_1^2})^T$.

Substituting in $\gamma_0 = \frac{1+2\phi_1\theta_1+\theta_1^2}{1-\phi_1^2}\sigma^2$, $\gamma_1 = \frac{\phi_1^2\theta_1+\phi_1\theta_1^2+\theta_1+\phi_1}{1-\phi_1^2}\sigma^2$ and $\gamma_2 = \frac{\phi_1^2+\phi_1^3\theta_1+\phi_1^2\theta_1+\phi_1\theta_1}{1-\phi_1^2}\sigma^2$ yields the final result.

(2) According to the spectral density of time-invariant linear filter, the spectral density of $\{Y_t\}$ is

$$\begin{aligned}f_Y(\lambda) &= |\frac{1}{2}(1 + e^{-i\lambda})|^2 f_X(\lambda) \\ &= \frac{\sigma^2}{8\pi} \cdot \frac{(1+\cos\lambda)(1 + 2\theta_1\cos\lambda + \theta_1^2)}{1 - 2\phi_1\cos\lambda + \phi_1^2}\end{aligned}$$