Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)
MAY–JUNE 2004

M3S8 (Solutions)

Time Series

seen ⇓

1. (a) $\{X_t\}$ is second-order stationary if $E\{X_t\}$ is a finite constant for all t, var$\{X_t\}$ is a finite constant for all t, and cov$\{X_t, X_{t+\tau}\}$, is a finite quantity depending only on τ and not on t. 3

sim. seen ⇓

(b) (i) Stationary (MA(2)!) $\mathsf{E}\{X_t\} = 0$,

$$\begin{aligned} s_\tau &= \mathsf{E}\{(\epsilon_t - 0.9\epsilon_{t-1})(\epsilon_{t+\tau} - 0.9\epsilon_{t+\tau-1})\} \\ &= \mathsf{E}\{\epsilon_t\epsilon_{t+\tau}\} - 0.9(\mathsf{E}\{\epsilon_t\epsilon_{t+\tau-1}\} + \mathsf{E}\{\epsilon_{t-1}\epsilon_{t+\tau}\}) + 0.81\mathsf{E}\{\epsilon_{t-1}\epsilon_{t+\tau-1}\}. \end{aligned}$$

So,

$$\text{var}\{X_t\} = s_0 = \sigma_\epsilon^2 + 0.81\sigma_\epsilon^2 = 1.81\sigma_\epsilon^2,$$

and

$$s_\tau = \begin{cases} 0.9\sigma_\epsilon^2 & |\tau| = 1; \\ 0 & |\tau| > 1, \end{cases}$$

none of which depend on t and so process is stationary. 4

(ii) For stationarity the roots of the characteristic equation $\Phi(z)$ in the defining equation $\Phi(B)X_t = \epsilon_t$ must lie **outside** the unit circle.

$$\begin{aligned} X_t &= \frac{9}{4}X_{t-1} - \frac{9}{8}X_{t-2} + \epsilon_t \\ (1 - \frac{9}{4}B + \frac{9}{8}B^2)X_t &= \epsilon_t \\ (1 - \frac{3}{4}B)(1 - \frac{3}{2}B)X_t &= \epsilon_t \end{aligned}$$

roots are $4/3$ and $2/3$, so process is **non-statationary** as $|2/3| < 1$. 4

unseen ⇓

(c) (i)

$$\begin{aligned} \text{var}\{X_t\} = s_0 &= \mathsf{E}\{X_t^2\} = \mathsf{E}\{(\alpha X_{t-1} + \epsilon_t + \alpha\epsilon_{t-1})(\alpha X_{t-1} + \epsilon_t + \alpha\epsilon_{t-1})\} \\ &= \alpha^2\mathsf{E}\{X_{t-1}^2\} + 2\alpha^2\mathsf{E}\{X_{t-1}\epsilon_{t-1}\} + \mathsf{E}\{\epsilon_t^2\} + \alpha^2\mathsf{E}\{\epsilon_{t-1}^2\} \\ s_0(1-\alpha^2) &= 2\alpha^2\mathsf{E}\{(\alpha X_{t-2} + \epsilon_{t-1} + \alpha\epsilon_{t-2})\epsilon_{t-1}\} + \sigma_\epsilon^2(1+\alpha^2) \\ s_0(1-\alpha^2) &= 2\alpha^2\sigma_\epsilon^2 + \sigma_\epsilon^2(1+\alpha^2) \\ s_0 &= \frac{\sigma_\epsilon^2(1+3\alpha^2)}{(1-\alpha^2)}. \end{aligned}$$

5

(ii) Process is stationary as $|\alpha| < 1$, multiply defining equation by X_{t-1} and take expectations

$$\mathsf{E}\{X_{t-1}X_t\} = \alpha\mathsf{E}\{X_{t-1}X_{t-1}\} + \mathsf{E}\{X_{t-1}\epsilon_t\} + \alpha\mathsf{E}\{X_{t-1}\epsilon_{t-1}\}$$

We have, from **(c)(i)** that $\mathsf{E}\{X_{t-1}\epsilon_{t-1}\} = \alpha\sigma_\epsilon^2$, and using the fact that $\mathsf{E}\{X_tX_{t-\tau}\} = s_\tau$, we have

$$\begin{aligned} s_1 &= \alpha s_0 + \alpha\sigma_\epsilon^2 \\ \Rightarrow \rho_1 = \frac{s_1}{s_0} &= \alpha + \alpha\frac{1-\alpha^2}{1+3\alpha^2} = \frac{\alpha(1+3\alpha^2) + \alpha(1-\alpha^2)}{1+3\alpha^2} = \frac{2\alpha(1+\alpha^2)}{1+3\alpha^2}. \end{aligned}$$

4

method seen ⇓

2. (a) (i) **A:** $\pi_1 = 0.89, \pi_2 = 0.1, \pi_3 = 0.$ [2]

B:

$$
\begin{aligned}
(1-0.1B)\epsilon_t &= (1-B)X_t \\
\epsilon_t &= (1-B)(1-0.1B)^{-1}X_t \\
\epsilon_t &= (1-B)(1+0.1B+(0.1B)^2+(0.1B)^3+\ldots)X_t \\
\epsilon_t &= (1+(0.1-1)B+(0.1^2-0.1)B^2+(0.1^3-0.1^2)B^3+\ldots)X_t \\
\epsilon_t &= (1-0.9B-(0.1)0.9B^2-(0.1)^2 0.9B^3-\ldots)X_t \\
X_t &= 0.9X_t+0.09X_{t-1}+0.009X_{t-2}+\ldots+\epsilon_t.
\end{aligned}
$$

So $\pi_1 = 0.9, \pi_2 = 0.09, \pi_3 = 0.009.$ [4]

(ii) For **B:** $\pi_k = (0.1)^{k-1}0.9.$ [1]

(b) (i) The three properties of an LTI filter are:

[**1**] Scale-preservation:

$$L\{\{\alpha x_t\}\} = \alpha L\{\{x_t\}\}.$$

[**2**] Superposition:

$$L\{\{x_{t,1}+x_{t,2}\}\} = L\{\{x_{t,1}\} + L\{\{x_{t,2}\}\}.$$

[**3**] Time invariance:
If

$$L\{\{x_t\}\} = \{y_t\}, \quad \text{then} \quad L\{\{x_{t+\tau}\}\} = \{y_{t+\tau}\}.$$

Where τ is integer-valued, and the notation $\{x_{t+\tau}\}$ refers to the sequence whose t-th element is $x_{t+\tau}$. [3]

(ii) Model **A**, we have $L\{\{X_t\}\} = X_t - 0.89X_{t-1} - 0.1X_{t-2}$ so that $L\{\{X_t\}\} = \{\epsilon_t\}$.

$$
\begin{aligned}
L\left\{\{e^{i2\pi ft}\}\right\} &= e^{i2\pi ft} - 0.89e^{i2\pi f(t-1)} - 0.1e^{i2\pi f(t-2)} \\
&= e^{i2\pi ft}\left(1-0.89e^{-i2\pi f} - 0.1e^{-i4\pi f}\right),
\end{aligned}
$$

giving,

$$G(f) = 1 - 0.89e^{-i2\pi f} - 0.1e^{-i4\pi f}.$$

Since,

$$S_\epsilon(f) = |G(f)|^2 S_A(f) \quad \text{and} \quad S_\epsilon(f) = \sigma_\epsilon^2,$$

we have

$$S_A(f) = \frac{\sigma_\epsilon^2}{|1-0.89e^{-i2\pi f} - 0.1e^{-i4\pi f}|^2}.$$

[5]

Model **B**: similarly, we have $L\{\{X_t\}\} = \sum_{k=0}^{\infty} \pi_k X_{t-k}$, giving

$$S_B(f) = \frac{\sigma_\epsilon^2}{|1-\sum_{k=1}^{\infty}\pi_k e^{-i2\pi fk}|^2} = \frac{\sigma_\epsilon^2}{|1-\sum_{k=1}^{\infty}(0.1)^{k-1}0.9e^{-i2\pi fk}|^2}$$

[3]

(c) $\pi_k = (0.1)^k 0.9$ so $\pi_k \approx 0$ for large k, also π_1 and π_2 are similar for both models, so although these models have different formulation (with one being non-stationary!) they are in fact quite similar in terms of their spectral shapes. [2]

seen ⇓

3. (a)

$$\mathsf{E}\{\widehat{s}_\tau^{(p)}\} = \frac{1}{N}\sum_{t=1}^{N-|\tau|}\mathsf{E}\{X_t X_{t+|\tau|}\} = \frac{1}{N}\left((N-|\tau|)s_\tau\right) = \left(1-\frac{|\tau|}{N}\right)s_\tau.$$

Hence $\widehat{s}_\tau^{(p)}$ is biased for s_τ. 1

Reasons for prefering the biased estimator:

1. For many stationary processes of practical interest
$$\text{mse}\{\widehat{s}_\tau^{(p)}\} < \text{mse}\{\widehat{s}_\tau^{(u)}\}.$$
2. If $\{X_t\}$ has a purely continuous spectrum we know that $s_\tau \to 0$ as $|\tau|\to\infty$. It therefore makes sense to choose an estimator that decreases nicely as $|\tau|\to N-1$ (i.e. choose $\widehat{s}_\tau^{(p)}$).
3. We know that the acvs must be positive semidefinite, the sequence $\{\widehat{s}_\tau^{(p)}\}$ has this property, whereas other unbiased estimators may not. 2

(b)

$$\mathsf{E}\{\widehat{S}^{(p)}(f)\} = \mathsf{E}\{|J(f)|^2\} \qquad \text{where} \quad J(f) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} X_t e^{-i2\pi f t}, \quad |f|\le\frac{1}{2}.$$

We know from the spectral representation theorem that there is an orthogonal increments process $Z(f)$ s.t. $\mathsf{E}\{|dZ(f)|^2\} = S(f)\,df$ and

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi f' t}\,dZ(f'),$$

so that,

$$\begin{aligned} J(f) &= \sum_{t=1}^{N}\left(\int_{-1/2}^{1/2}\frac{1}{\sqrt{N}}e^{i2\pi f' t}\,dZ(f')\right)e^{-i2\pi f t} \\ &= \int_{-1/2}^{1/2}\sum_{t=1}^{N}\frac{1}{\sqrt{N}}e^{-i2\pi(f-f')t}\,dZ(f') \end{aligned}$$

We find that,

$$\begin{aligned} \mathsf{E}\{\widehat{S}^{(p)}(f)\} &= \mathsf{E}\{|J(f)|^2\} = \mathsf{E}\{J^*(f)J(f)\} \\ &= \mathsf{E}\left\{\int_{-1/2}^{1/2}\sum_{t=1}^{N}\frac{1}{\sqrt{N}}e^{i2\pi(f-f')t}\,dZ^*(f')\int_{-1/2}^{1/2}\sum_{s=1}^{N}\frac{1}{\sqrt{N}}e^{-i2\pi(f-f'')s}\,dZ(f'')\right\} \\ &= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\sum_{t=1}^{N}\frac{1}{\sqrt{N}}e^{i2\pi(f-f')t}\sum_{s=1}^{N}\frac{1}{\sqrt{N}}e^{-i2\pi(f-f'')s}\mathsf{E}\{dZ^*(f')\,dZ(f'')\} \\ &= \int_{-1/2}^{1/2}\mathcal{F}(f-f')S(f')\,df', \end{aligned}$$

by the orthogonality of the increments process, and where $\mathcal{F}$ is Féjer's kernel defined by

$$\mathcal{F}(f) = \left|\sum_{t=1}^{N}\frac{1}{\sqrt{N}}e^{-i2\pi f t}\right|^2.$$

6

unseen ⇓

(c) Dataset 1 Periodogram B [periodic - spectral peak]
Dataset 2 Periodogram C [slowly varying - more power at low frequencies]
Dataset 3 Periodogram A [rapidly varying - more power at high frequencies]

3

(d) The periodogram of processes with a high dynamic range (defined as $10\log_{10}\{\max_f S(f)/\min_f S(f)\}$) can suffer from bias. 1 seen

Periodogram B could be biased at high frequencies due to leakage from the spectral peak. 1

To counter bias we can use a technique known as tapering: We form the product $\{h_t X_t\}$ where $\{h_t\}$ is a sequence of real-valued constants called a data taper. Define

$$J(f) = \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t} \qquad |f| \leq 1/2.$$

By the spectral representation theorem,

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi f' t}\, dZ(f'),$$

so that,

$$\begin{aligned} J(f) &= \sum_{t=1}^{N} h_t \left(\int_{-1/2}^{1/2} e^{i2\pi f' t}\, dZ(f') \right) e^{-i2\pi f t} \\ &= \int_{-1/2}^{1/2} \sum_{t=1}^{N} h_t e^{-i2\pi (f-f')t}\, dZ(f') \\ &= \int_{-1/2}^{1/2} H(f-f')\, dZ(f'), \end{aligned}$$

where,

$$H(f) = \sum_{t=1}^{N} h_t e^{-i2\pi f t}$$

We define our direct spectral estimator as,

$$\widehat{S}^{(d)}(f) = |J(f)|^2 = \left| \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t} \right|^2.$$

Then,

$$|J(f)|^2 = J^*(f) J(f) = \int_{-1/2}^{1/2} H^*(f-f')\, dZ^*(f') \int_{-1/2}^{1/2} H(f-f'')\, dZ(f''),$$

and hence by the orthogonality of the increments process,

$$\begin{aligned} \mathsf{E}\{\widehat{S}^{(d)}(f)\} &= \mathsf{E}\{|J(f)|^2\} \\ &= \int_{-1/2}^{1/2} |H(f-f')|^2 S(f')\, df' = \int_{-1/2}^{1/2} \mathcal{H}(f-f') S(f')\, df', \end{aligned}$$

where $\mathcal{H}(f) = |H(f)|^2$, we take $\sum_{t=1}^{N} h_t^2 = 1$.

The shape of $\mathcal{H}(f)$ determines the bias properties of the estimator - we choose a taper whose associated $\mathcal{H}(f)$ has low sidelobes to reduce sidelobe leakage and thus reduce the bias due to sidelow leakage. 6

method seen ⇓

4. (a)

$$\begin{aligned} \mathsf{E}\{X_t X_{t-\tau}\} &= \mathsf{E}\{X_{t-\tau}(\phi X_{t-6} + \epsilon_t)\} \\ &= \phi \mathsf{E}\{X_{t-\tau} X_{t-6}\} + \mathsf{E}\{X_{t-\tau}\epsilon_t\} \end{aligned}$$

When $\tau = 0$, we have

$$s_0 = \phi s_6 + \sigma_\epsilon^2 \qquad (1)$$

Now $\mathsf{E}\{X_{t-\tau}\epsilon_t\} = 0$ for all $\tau > 0$. So,

$$s_\tau = \phi s_{\tau-6} \quad \text{for } \tau > 0$$

Consider s_τ, for $\tau = 1, 2, 3, 4, 5$, we have

$$s_1 = \phi s_5; \quad s_2 = \phi s_4; \quad s_3 = \phi s_3; \quad s_4 = \phi s_2; \quad s_5 = \phi s_1$$

as $|\phi| < 1$ we must have $s_1 = \ldots = s_5 = 0$ (as $\phi \neq 0$)

$$s_{6k} = \phi s_{6k-6} = \phi s_{6(k-1)} \quad k = 1, 2, \ldots$$

We have,

$$s_6 = \phi s_0, \qquad (2)$$

thus, from equation (**??**),

$$s_6 = \phi(\phi s_6 + \sigma_\epsilon^2) \text{ and } s_6 = \frac{\phi \sigma_\epsilon^2}{1 - \phi^2}$$

and

$$s_0 = \frac{\sigma_\epsilon^2}{1 - \phi^2},$$

s_0 is non-zero and therefore s_{6k} is non-zero, for $k = 1, 2, \ldots$ 4

(b) (i) The Y-W estimators are obtained from **(a)** by replacing the acvs in equations (**??**) and (**??**) with their biased estimators, giving

$$\widehat{\phi} = \frac{\widehat{s}_6}{\widehat{s}_0} = \frac{\sum_{t=1}^{N-6} X_t X_{t+6}}{\sum_{t=1}^{N} X_t^2}$$

and 3

$$\widehat{\sigma_\epsilon^2} = \widehat{s}_0 - \widehat{\phi}\widehat{s}_6 = \frac{1}{N}\sum_{t=1}^{N} X_t^2 - \frac{1}{N}\frac{(\sum_{t=1}^{N-6} X_t X_{t+6})^2}{\sum_{t=1}^{N} X_t^2}$$

3

(ii) The forward least squares estimator is given by

$$\widetilde{\phi} = \left(F^\top F\right)^{-1}(F^\top X_F) = \left(\sum_{t=7}^{N} X_t^2\right)^{-1} \sum_{t=1}^{N-6} X_t X_{t+6} = \frac{\sum_{t=1}^{N-6} X_t X_{t+6}}{\sum_{t=7}^{N} X_t^2}$$

$\boxed{4}$

$$\begin{aligned}
\widetilde{\sigma_\epsilon^2} &= \frac{(X_F - F\widetilde{\phi})^\top (X_F - F\widetilde{\phi})}{N-6-1} = \frac{1}{N-7}\sum_{t=7}^{N}(X_t - \widetilde{\phi}X_{t-6})^2 \\
&= \frac{1}{N-7}\left(\sum_{t=7}^{N} X_t^2 - 2\frac{\sum_{t=1}^{N-6} X_t X_{t+6}}{\sum_{t=7}^{N} X_t^2}\sum_{t=7}^{N} X_t X_{t-6} + \frac{\left(\sum_{t=1}^{N-6} X_t X_{t+6}\right)^2}{\left(\sum_{t=7}^{N} X_t^2\right)^2}\sum_{t=7}^{N} X_{t-6}^2\right) \\
&= \frac{1}{N-7}\left(\sum_{t=7}^{N} X_t^2 - 2\frac{(\sum_{t=1}^{N-6} X_t X_{t+6})^2}{\sum_{t=7}^{N} X_t^2} + \frac{\left(\sum_{t=1}^{N-6} X_t X_{t+6}\right)^2}{\left(\sum_{t=7}^{N} X_t^2\right)^2}\sum_{t=7}^{N} X_{t-6}^2\right)
\end{aligned}$$

$\boxed{4}$

(iii) as N increases, the value of $\sum_{t=7}^{N} X_{t-6}^2$, $\sum_{t=7}^{N} X_t^2$ and $\sum_{t=1}^{N} X_t^2$ all become closer (X_t is stationary), and $N/(N-7) \to 1$ so the the least squares and Yule-Walker estimators become closer in value (in fact they are asymptotically equivalent). $\boxed{2}$

seen $\Downarrow$

5. (a) We want to minimize,

$$\begin{aligned}
\mathsf{E}\{(X_{t+l} - X_t(l))^2\} &= \mathsf{E}\left\{\left(\sum_{k=0}^{\infty} \psi_k \epsilon_{t+l-k} - \sum_{k=0}^{\infty} \delta_k \epsilon_{t-k}\right)^2\right\} \\
&= \mathsf{E}\left\{\left(\sum_{k=0}^{l-1} \psi_k \epsilon_{t+l-k} + \sum_{k=0}^{\infty} [\psi_{k+l} - \delta_k] \epsilon_{t-k}\right)^2\right\} \\
&= \sigma_\epsilon^2 \left\{\left(\sum_{k=0}^{l-1} \psi_k^2\right) + \sum_{k=0}^{\infty} (\psi_{k+l} - \delta_k)^2\right\}.
\end{aligned}$$

The first term is independent of the choice of the $\{\delta_k\}$ and the second term is clearly minimized by choosing $\delta_k = \psi_{k+l}, k = 0, 1, 2, \ldots$. 4

method seen $\Downarrow$

(b) (i) We have $X_t = \Psi(B)\epsilon_t \Rightarrow \epsilon_t = \Psi^{-1}(B)X_t$, and so

$$\begin{aligned}
X_t(l) &= \sum_{k=0}^{\infty} \psi_{k+l} \epsilon_{t-k} = \Psi^{(l)}(B)\epsilon_t \quad [= \delta(B)\epsilon_t] \\
&= \Psi^{(l)}(B)\Psi^{-1}(B)X_t = G^{(l)}(B)X_t
\end{aligned}$$

Now

$$X_t - \frac{1}{2}X_{t-1} = \epsilon_t \;\Rightarrow\; \left(1 - \frac{B}{2}\right) X_t = \epsilon_t$$

So that,

$$\Rightarrow \Psi(B) = \left(1 - \frac{B}{2}\right)^{-1} = 1 + \frac{B}{2} + \frac{B^2}{4} + \frac{B^3}{8} + \ldots$$

So $\delta_k = \psi_{k+l} = 2^{-(k+l)}$ and when $l = 1$, $\delta_k = 2^{-(k+1)}$ giving

$$\Psi^{(1)}(B) = \left(\frac{1}{2} + \frac{B}{4} + \frac{B^2}{8} + \ldots\right)$$

We have,

$$G^{(1)}(B) = \Psi^{(1)}(B)\Psi^{-1}(B) = \left(\frac{1}{2} + \frac{B}{4} + \frac{B^2}{8} + \ldots\right)\left(1 - \frac{B}{2}\right) = \frac{1}{2}$$

Giving

$$X_t(1) = G^{(1)}(B)X_t = \frac{1}{2}X_t$$

Similarly, when $l = 2$ we have $\delta_k = 2^{-(k+2)}$ and 4

$$G^{(2)}(B) = \Psi^{(2)}(B)\Psi^{-1}(B) = \left(\frac{1}{4} + \frac{B}{8} + \frac{B^2}{16} + \ldots\right)\left(1 - \frac{B}{2}\right) = \frac{1}{4}$$

Giving

$$X_t(2) = G^{(2)}(B)X_t = \frac{1}{4}X_t$$

2

seen ⇓

(ii) From **(a)(i)** we have

$$\sigma^2(l) = \sigma_\epsilon^2 \left\{ \left(\sum_{k=0}^{l-1} \psi_k^2 \right) + \sum_{k=0}^{\infty} (\psi_{k+l} - \delta_k)^2 \right\}$$

When $\delta_k = \psi_{k+l}$ the second term vanishes, and we have,

$$\sigma^2(l) = \mathsf{E}\{(X_{t+l} - X_t(l))^2\} = \sigma_\epsilon^2 \sum_{k=0}^{l-1} \psi_k^2,$$

3

Giving,

unseen ⇓

$$\begin{aligned} \sigma^2(1) &= \sigma_\epsilon^2 \psi_0^2 = \sigma_\epsilon^2 \\ \sigma^2(2) &= \sigma_\epsilon^2(\psi_0^2 + \psi_1^2) = \sigma_\epsilon^2 \left(1 + \frac{1}{4}\right) = \sigma_\epsilon^2 \frac{5}{4} \end{aligned}$$

2

(iii) We have

$$\begin{aligned} X_{t+1}(1) &= \sum_{k=0}^{\infty} \psi_{k+1} \epsilon_{t+1-k} \\ &= \psi_1 \epsilon_{t+1} + \psi_2 \epsilon_t + \psi_3 \epsilon_{t-1} + \ldots, \end{aligned}$$

but,

$$\begin{aligned} X_t(2) &= \sum_{k=0}^{\infty} \psi_{k+2} \epsilon_{t-k} \\ &= \psi_2 \epsilon_t + \psi_3 \epsilon_{t-1} + \psi_4 \epsilon_{t-2} + \ldots, \end{aligned}$$

and,

$$\begin{aligned} X_{t+1}(1) &= X_t(2) + \psi_1 \epsilon_{t+1} \\ &= X_t(2) + \psi_1 (X_{t+1} - X_t(1)) \\ &= X_t(2) + \frac{1}{2} (X_{t+1} - X_t(1)) \end{aligned}$$

Hence, to forecast X_{t+2} we can modify the 2-step ahead forecast at time t by producing an 1-step ahead forecast at time $t+1$ using X_{t+1} as it becomes available.
Note that we have

$$X_{t+1}(1) = \frac{1}{4} X_t + \frac{1}{2} \left(X_{t+1} - \frac{1}{2} X_t \right) = \frac{1}{2} X_{t+1}$$

as expected.

5