Division of Mathematical Statistics

KTH Matematik

EXAM IN SF2943 TIME SERIES ANALYSIS
FRIDAY AUGUST 18 2017 KL 08:00–13:00.

Teacher: Pierre Nyquist, 08 – 790 6633.

Examiner: Henrik Hult, 08 – 790 6911.

Allowed aids: Pocket calculator, "Formulas and survey, Time series analysis" by Jan Grandell, without notes.

Any notation introduced must be explained and defined. Arguments and computations must be detailed so that they are easy to follow.

Problem 1

Figure 1 shows the number of sunspots for each year between 1770 and 1869 (100 observations), the so-called Wölfer sunspot numbers. The observed time series $\{X_t\}_{t=1}^{100}$ has sample mean $\bar{x} = 47.11$

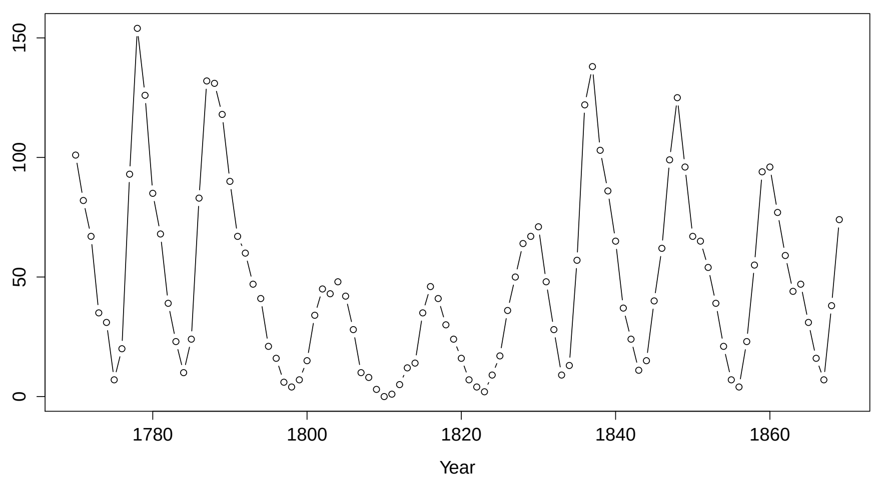

Figure 1: Wölfer sunspot numbers for Problem 1

and sample autocovariances and sample partial autocorrelations are given in Table 1 and shown in Figure 2.
Amongst AR(p) or MA(q) models, propose a suitable stationary time series model for the mean-corrected series

$$Y_t = X_t - \bar{x}, \quad t = 1, \dots, 100,$$

including specifying the model order p or q. For the proposed model, estimate the relevant model parameters, including the variance σ^2 of the underlying noise sequence.

h	0	1	2	3	4	5	6	7
$\hat{\gamma}(h)$	1385.9	1117.8	593.8	97.3	-233.4	-369.0	-294.3	-60.5
$\widehat{\text{PACF}}(h)$	-	0.807	-0.635	0.083	-0.061	-0.006	0.180	0.105

Table 1: ACVF and PACF estimates for the Wölfer sunspot numbers

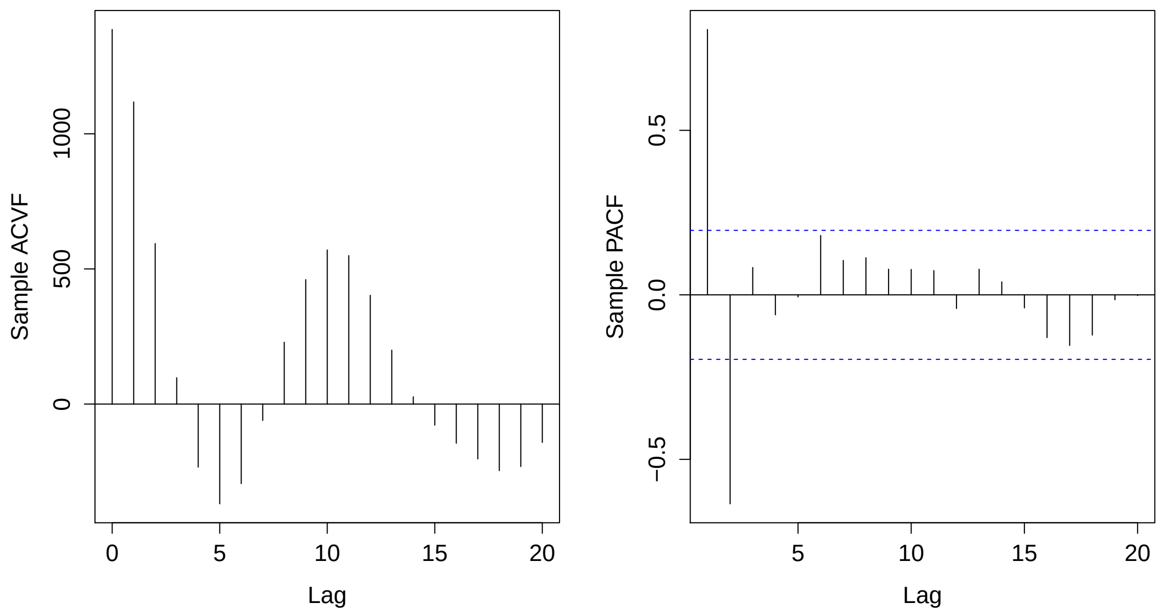

Figure 2: Sample ACVF and PACF for the Wölfer sunspot numbers

Problem 2

Consider the ARCH(1) process

$$X_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim IID(0,1),$$
$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2,$$

where $\alpha_0 \geq 0$ and $0 < \alpha_1 < 1$; the process is stationary and e_t and $X_{t-1}, X_{t-2}, \dots$ are independent for all t.

a) Compute the best linear predictor $\hat{X}_t$ of X_t using all previous values $X_{t-1}, X_{t-2}, \dots$, and the mean-squared error (MSE) associated with this predictor.

b) Your friend suggests that, because of the way X_t and h_t are defined, you should compute the best linear predictor $\hat{Y}_t$ of X_t^2, using all previous values $\{X_s\}_{s<t}$, and then use $\hat{X}_t = \sqrt{\hat{Y}_t}$ as your predictor for X_t. Compute this predictor and the associated MSE. How does it compare to the MSE in (a)?

Problem 3

The six figures A-F (in Figure 3) show a realisation, the sample autocorrelation function (based on a sample size of 200) and the spectral density for an AR(2) process and the corresponding three plots for an MA(2) process.

a) Which figures are realisations, sample ACFs and spectral densities, respectively? Motivate your answer.

b) Group the figures into two triplets, each containing a realisation, a sample ACF and a spectral density, such that the figures in each triplet correspond to the same time series. Which triplet corresponds to the MA(2) process, and which to the AR(2) process? Motivate your answers.

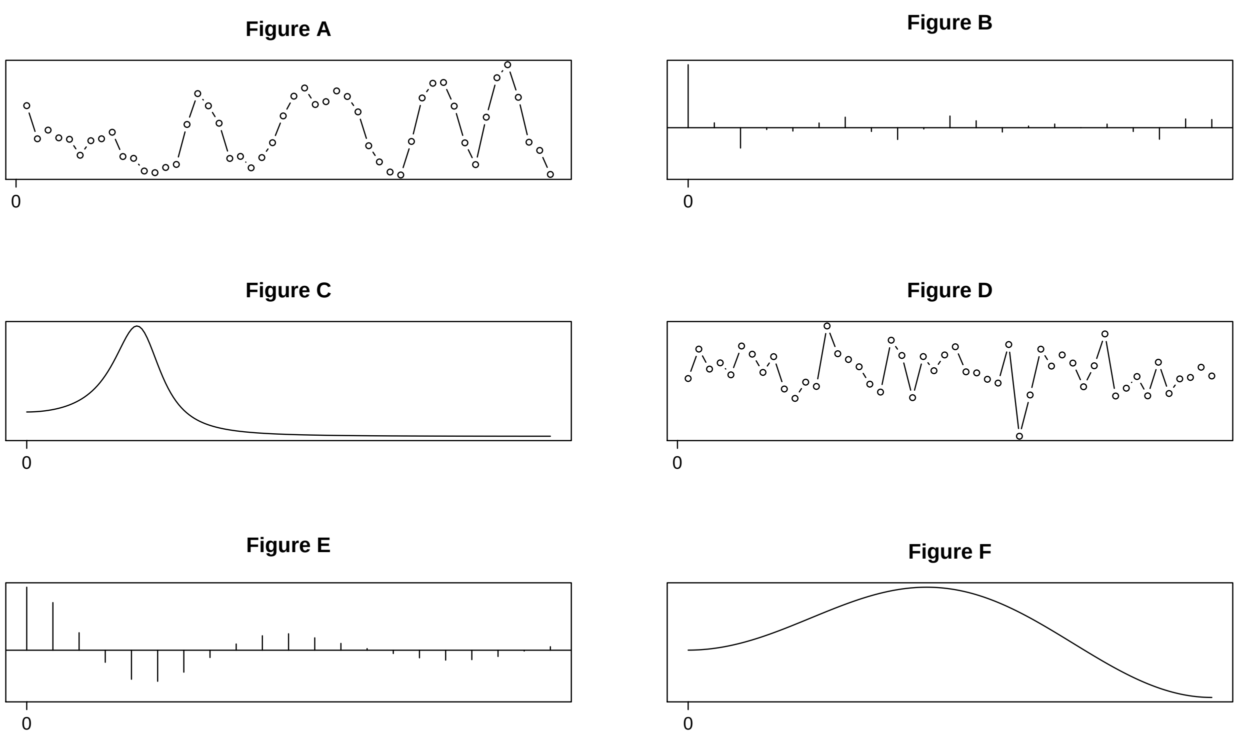

Figure 3: Figures for Problem 3

Problem 4

Let $\{X_t\}$ be a stationary time series with zero mean, (absolutely summable) ACVF γ_X and spectral density f_X. For some $\alpha \in (0,1)$ and nonnegative integer k, the time series $\{Y_t\}$ is given by

$$Y_t = \alpha X_t + (1-\alpha)X_{t-k}.$$

a) Show that $\{Y_t\}$ is stationary and compute its ACVF.

b) Compute the spectral density f_Y of $\{Y_t\}$.

c) If $\{X_t\} \sim \mathrm{WN}(0,\sigma^2)$, $k=3$, show that f_Y is periodic and determine the period.

Problem 5

Consider a sample of size n from $\{Z_t\} \sim IID(0, \sigma^2)$.

a) It is claimed that, for a sufficiently large n and given lag h (small enough compared to n), the sample acvf $\hat{\gamma}_Z(h)$ is asymptotically normal with mean 0 and variance σ^4/n. Make this claim *plausible.*

b) Describe in detail how the asymptotic normality in (a) can be used to design a test, based on the sample autocorrelation function

$$\hat{\rho}_Z(\cdot) = \frac{\hat{\gamma}_Z(\cdot)}{\hat{\gamma}_Z(0)},$$

for whether or not a time series $\{X_t\}$ is i.i.d. noise.

Hint: For Part (a), Theorem 1 (given below) might be of use.

Theorem 1 (Central limit theorem for m-dependent sequences) *If $\{X_t\}$ is a strictly stationary m-dependent sequence of random variables with mean zero and autocovariance function $\gamma(\cdot)$, such that $v_m = \gamma(0) + 2\sum_{j=1}^{m} \gamma(j) \neq 0$, then*

$$\sqrt{n}\bar{X}_n \to N(0, v_m), \quad n \to \infty$$

where the convergence is in distribution.

Good luck!

Solutions

Problem 1

The sample PACF suggests that an AR(2) model is suitable for the mean-corrected data. Moreover, the sample ACVF does not contradict such a model (the oscillating behavior) whereas it is in stark contrast to an MA model of reasonable order - after normalizing and comparing to $1.96/\sqrt{100}$ (the IID case) we see that the sample ACF is non-zero for some lags h greater than 10, which would require an MA model of order $q > 10$.
Having settled on an AR(2) model, we can estimate the model parameters ϕ_1, ϕ_2 and σ^2 using the Yule-Walker equations. The estimated covariance matrix and vector are given by (Yule-Walker equations with $p = 2$)

$$\hat{\Gamma}_2 = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) \end{pmatrix} = \begin{pmatrix} 1385.9 & 1117.8 \\ 1117.8 & 1385.9 \end{pmatrix},$$

and

$$\hat{\gamma}_2' = \begin{pmatrix} \hat{\gamma}(1) & \hat{\gamma}(2) \end{pmatrix} = \begin{pmatrix} 1117.8 & 593.8 \end{pmatrix}.$$

The estimates of ϕ_1 and ϕ_2 are obtained from

$$\hat{\Gamma}_2\hat{\phi} = \hat{\gamma}_2,$$

which amounts to

$$\hat{\phi} = \hat{\Gamma}_2^{-1}\hat{\gamma}_2 = \begin{pmatrix} 1.319 & -0.635 \end{pmatrix}'.$$

Furthermore, the estimate of σ^2 is given by

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}'\hat{\gamma}_2 = 289.061.$$

Answer: The estimated model is

$$X_t - 1.319X_{t-1} + 0.635X_{t-2} = Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, 289.061).$$

Problem 2

a) The best linear predictor $\hat{X}_t$ of X_t, given $X_{t-1}, X_{t-2}, \dots$, has the form

$$\hat{X}_t = c_0 + c_1X_{t-1} + c_2X_{t-2} + \dots,$$

for constants $c_0, c_1, \dots$ such that the MSE $E[(X_t - \hat{X}_t)^2]$ is minimized. Recall that the ARCH(1) process is a white noise process. Indeed, we have$E[X_t] = 0$ and

$$\begin{aligned}\mathrm{Var}(X_t) &= E[h_t e_t^2]\\ &= \alpha_0 + \alpha_1 E[X_{t-1}^2]\\ &= \alpha_0 + \alpha_1 \mathrm{Var}(X_{t-1}).\end{aligned}$$

By stationarity $\mathrm{Var}(X_t) = \mathrm{Var}(X_{t-1})$ and we have

$$\mathrm{Var}(X_t) = \frac{\alpha_0}{1-\alpha_1}.$$

Moreover, for $h \neq 0$,

$$\begin{aligned}\mathrm{Cov}(X_t, X_{t-h}) &= E\left[X_t X_{t-h}\right]\\ &= E\left[\sqrt{h_t}e_t\sqrt{h_{t-h}}e_{t-h}\right]\\ &= E\left[\sqrt{\alpha_0 + \alpha_1 X_{t_1}^2}e_t\sqrt{\alpha_0 + \alpha_1 X_{t-h-1}^2}e_{t-h}\right]\\ &= E\left[\sqrt{\alpha_0 + \alpha_1 X_{t_1}^2}\sqrt{\alpha_0 + \alpha_1 X_{t-h-1}^2}e_{t-h}\right]E[e_t]\end{aligned}$$

where in the last step we have used that e_t is independent on the remaining quantities. Since $E[e_t] = 0$ the covariance between X_t and X_{t-h} is also 0 for all $h \neq 0$. Thus, $\{X_t\}$ is $\mathrm{WN}(0, \alpha_0/(1-\alpha_1))$. This implies that the best linear predictor is $\hat{X}_t = 0$, which gives the MSE

$$\begin{aligned}E[(X_t - \hat{X}_t)^2] &= E[X_t^2]\\ &= \frac{\alpha_0}{1-\alpha_1}.\end{aligned}$$

Answer: The best linear predictor is $\hat{X}_t = 0$ and has an MSE of $\alpha_0(1-\alpha_1)$.

b) It is easily realized that the best linear predictor $\hat{Y}_t$ of X_t^2, based on all previous values $\{X_s\}_{s<t}$, is given by $h_t = \alpha_0 + \alpha_1 X_{t-1}^2$; one can for example either use the fact that this is the best predictor in the MSE sense, and since it linear in the previous values of the process also the best linear predictor, or treat X_t^2 as an AR(1) process on the squared residuals (from which the predictor is immediately obtained as for a "standard" AR process). The suggestion is therefore that we use $\hat{X}_t = \sqrt{h_t}$ as the predictor for x_t. We proceed to compute the associated MSE:

$$\begin{aligned}
E\left[\left(X_t - \hat{X}_t\right)\right] &= E\left[\left(\sqrt{h_t}e_t - \sqrt{h_t}\right)^2\right] \\
&= E\left[h_t e_t^2 - 2h_t e_t + h_t\right] \\
&= E\left[h_t\right] E[e_t^2] - 2E\left[h_t\right] E[e_t] + E\left[h_t\right] \\
&= 2E[h_t] \\
&= 2\left(\alpha_0 + \alpha_1 \frac{\alpha_0}{1-\alpha_1}\right) \\
&= 2\frac{\alpha_0}{1-\alpha_1}.
\end{aligned}$$

In the computations we have used the independence between e_t and all previous values $X_{t-1}, X_{t-2}, \ldots$ of the process. The conclusion is that the MSE for the proposed predictor is twice that of the best linear predictor.

Answer: The proposed predictor is $\hat{X}_t = \sqrt{h_t}$ and has twice the MSE of the predictor in (a).

Problem 3

a) We recognize that Figures C and F both have a continuous index on the x-axis, whereas the other four figures have a discrete index. This suggests that Figures C and F show spectral densities, since the spectral density is a function of a continuous variable (angular frequency or frequency). Next, recall that the ACF of a time series is maximal and equal to one at lag zero. This is consistent with Figures B and E, whereas Figures A and D are *not* maximal at index $x = 0$. Thus, Figures B and E show sample ACFs and Figures A and D realisations of the time series.

Answer: C and F show spectral densities, B and E ACFs, A and D realisations.

b) Studying the sample ACF plots, Figure B suggests zero correlation for lags ≥ 2, whereas Figure E shows a slowly decaying correlation). The latter is consistent with a slowly varying realisation, such as the one observed in Figure A. The negative correlation at lag 2 in Figure B is consistent with an oscillating realisation, such as that observed in Figure D.

Next, Figure C shows a spectral density with large contents of low frequencies: the peak is at roughly $\lambda = \pi/4$, which is consistent with the slowly varying realisation shown in Figure A (it suggests a period length of ≈ 8). Figure F on the other hand has a (wide) peak at about

$\lambda = \pi/2$, consistent with a period of length 4. This is precisely what is seen in Figure D, as well as in agreement with the negative correlation at lag 2 in Figure B. Thus the triplets are A-C-E and B-D-F.

The triplet B-D-F belongs to an MA(2) process (e.g., zero correlation for lags > 2) and the triplet A-C-E belongs to an AR(2) model (e.g., non-zero ACF for most lags, sharp peak in the spectral density).

Answer: The triplets are B-D-F (MA(2)) and A-C-E (AR(2)).

Remark: Both processes had parameters 1.3 *and* -0.7.

Problem 4

a) The mean of Y_t is

$$\begin{aligned}\mu_Y(t) &= E[Y_t]\\ &= E[\alpha X_t + (1-\alpha)X_{t-k}]\\ &= \alpha E[X_t] + (1-\alpha)E[X_{t-k}]\\ &= 0,\end{aligned}$$

since $\{X_t\}$ is a zero-mean process. Thus, the mean $\mu_Y(t)$ does not depend on t. Remains to check that the same holds for the autocovariance function associated with $\{Y_t\}$. For any integer $h > 0$, using the bilinearity of the covariance function and stationarity of $\{X_t\}$,

$$\begin{aligned}Cov(Y_t, Y_{t+h}) &= Cov\left(\alpha X_t + (1-\alpha)X_{t-k}, \alpha X_{t+h} + (1-\alpha)X_{t+h-k}\right)\\ &= \alpha^2\gamma_X(h) + \alpha(1-\alpha)\gamma_X(h-k) + \alpha(1-\alpha)\gamma_X(h+k) + (1-\alpha)^2\gamma_X(h).\end{aligned}$$

This clearly does not depend on time t, but only on h (and k, which is fixed), and thus $\{Y_t\}$ is a stationary process. Moreover, we can easily find the ACVF from the last display:

$$\begin{aligned}\gamma_Y(h) &= \alpha^2\gamma_X(h) + \alpha(1-\alpha)\gamma_X(h-k) + \alpha(1-\alpha)\gamma_X(h+k) + (1-\alpha)^2\gamma_X(h)\\ &= \left(1 - 2\alpha + 2\alpha^2\right)\gamma_X(h) + \alpha(1-\alpha)\left(\gamma_X(h-k) + \gamma_X(h+k)\right).\end{aligned}$$

b) Since the ACVF γ_X is absolutely summable, the same thing holds for γ_Y. Therefore, the spectral density can be computed as ($\lambda \in (-\pi, \pi]$)

$$\begin{aligned}f_Y(\lambda) &= \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_Y(h)e^{-ih\lambda}\\ &= \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\left(\left(1-2\alpha+2\alpha^2\right)\gamma_X(h)e^{-ih\lambda} + \alpha(1-\alpha)\left(\gamma_X(h-k)+\gamma_X(h+k)\right)e^{-ih\lambda}\right)\\ &= \left(1-2\alpha+2\alpha^2\right)f_X(\lambda) + \frac{\alpha(1-\alpha)}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_X(h-k)e^{-ih\lambda} + \frac{\alpha(1-\alpha)}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_X(h+k)e^{-ih\lambda}\\ &= \left(1-2\alpha+2\alpha^2\right)f_X(\lambda) + \alpha(1-\alpha)f_X(\lambda)\left(e^{-ik\lambda+e^{ik\lambda}}\right)\\ &= \left(1-2\alpha+2\alpha^2+2\alpha(1-\alpha)\cos(k\lambda)\right)f_X(\lambda).\end{aligned}$$

c) First note that if $\{X_t\} \sim \mathrm{WN}(0, \sigma^2)$, then

$$f_X(\lambda) = \frac{\sigma^2}{2\pi},$$

and the spectral density f_Y simplifies to

$$f_Y(\lambda) = \frac{\sigma^2(1 - 2\alpha + 2\alpha^2 + 2\alpha(1-\alpha)\cos(k\lambda))}{2\pi}.$$

If $k = 3$ this simplifies further to

$$f_Y(\lambda) = \frac{\sigma^2(1 - 2\alpha + 2\alpha^2 + 2\alpha(1-\alpha)\cos(3\lambda))}{2\pi},$$

and since cosine is a periodic function with period 2π, f_Y has period $2\pi/3$.

Problem 5

a) Let $\bar{Z}_n$ denote the sample mean of the n observations. The sample acvf $\hat{\gamma}_Z$ is given by, for a given lag h,

$$\begin{aligned}\hat{\gamma}_Z(h) &= \frac{1}{n}\sum_{t=1}^{n-h}\left(Z_t - \bar{Z}_n\right)\left(Z_{t+h} - \bar{Z}_n\right) \\ &= \frac{1}{n}\sum_{t=1}^{n-h} X_t X_{t+h} - \frac{1}{n}\sum_{t=1}^{n-h} X_t \bar{X}_n - \frac{1}{n}\sum_{t=1}^{n-h} X_{t+h}\bar{X}_n + \frac{1}{n}\sum_{t=1}^{n-h}\bar{X}_n^2.\end{aligned}$$

Since $\{Z_t\}$ is a zero-mean process, for n large enough it should hold that $\bar{X}_n \approx 0$ (holds w.p. 1 as n goes to infinity). Thus we can make the crude approximation that $\bar{X}_n = 0$ in the previous display, which results in

$$\hat{\gamma}_Z(h) \approx \frac{1}{n}\sum_{t=1}^{n-h} X_t X_{t+h}.$$

Next, make a non-rigorous argument for

$$\sqrt{n}\frac{1}{n}\sum_{t=1}^{n-h} X_t X_{t+h} \to N(0, \sigma^4), \quad n \to \infty, \tag{1}$$

where the convergence is in distribution. This in turn implies that for n large we have the approximate distribution $N(0, \sigma^4/n)$ for $(1/n)\sum_{t=1}^{n-h} X_t X_{t+h}$, which is the claim stated in the problem formulation.

To make the convergence (1) plausible we want to appeal to Theorem 1 (stated above). More specifically, if we can show that $\{X_t X_{t+h}\}_{t=1,2,\ldots}$ is an m-dependent sequence, for some m, and white noise[1], then the theorem gives the desired convergence. First, note that for $s \neq t$ and $|s-t| \neq h$,

$$E\left[X_t X_t + h X_s X_{s+h}\right] = E[X_t]E[X_{t+h}]E[X_s]E[X_{s+h}] = 0,$$

so that $\text{Cov}(X_t X_{t+h}, X_s X_{s+h}) = 0$ for all $t \neq s$. Hence, this is a white noise sequence with variance

$$E\left[(X_t X_{t+h})^2\right] = E[X_t^2]E[X_{t+h}^2] = \sigma^4.$$

Second, we check whether the sequence is m-dependent for some m. This is indeed the case for $m = h$: because $\{X_t\}$ is an IID sequence, $X_t X_{t+h}$ and $X_s X_{s+h}$ are independent for every s such that $s \notin \{t, t-h, t+h\}$. That is, for s such that $|s-t| > h$, the variables are independent, which is precisely the definition of a h-dependent sequence.

For n large compared to h, $\hat{\gamma}_Z(h)$ can roughly be viewed as the sample mean of $n-h$ observations of $X_t X_{t+h}$. With the properties we have just established for this sequence, Theorem 1 states that this sample mean, scaled by $\sqrt{n}$, converges in distribution to a normal random variable with variance

$$v_h = \gamma(0) + 2\sum_{j=1}^{h} \gamma(j),$$

where γ is the covariance function of the "new" series $\{X_t X_{t+h}\}_t$. Since we have established that this is a white noise sequence,

$$\gamma(l) = \begin{cases} \sigma^4, & l = 0, \\ 0, & l \neq 0, \end{cases}$$

which in turn implies that $v_h = \sigma^4$. Thus, this non-rigorous argument using Theorem 1 suggests that

$$\sqrt{n}\hat{\gamma}_Z(h) \to N(0, \sigma^4), \quad n \to \infty,$$

which translate to $\hat{\gamma}Z(h)$ being asymptotically normal with mean 0 and variance σ^4/n, i.e., abusing notation,

$$\hat{\gamma}_Z(h) \approx N(0, \sigma^4/n)$$

(in terms of distributions). This completes part (a).

b) We start by making the observation that $\gamma_Z(0) = \sigma^2$ and

$$\hat{\gamma}_Z(0) = \frac{1}{n}\sum_{t=1}^{n}(Z_t - \bar{Z}_n)^2 \to \sigma^2, \quad n \to \infty.$$

[1] Recall that a q-dependence refers to dependence and independence of variables a certain distance apart, whereas white noise means uncorrelated, hence there is no conflict in having a q-dependent white noise sequence; this speaks to the difference between q-dependent and q-correlated sequences

For large n we can thus use the approximation $\hat{\gamma}_Z(0) \approx \sigma^2$, and similarly

$$\hat{\rho}_Z(h) \approx \frac{\hat{\gamma}_Z(h)}{\sigma^2}.$$

From part (a) we know that the numerator is approximately $N(0, \sigma^4/n)$-distributed, which implies that the ratio defining $\hat{\rho}_Z(h)$ is approximately $N(0, 1/n)$-distributed.

With the $N(0, 1/n)$-distribution "established", designing a test for checking whether a time series is i.i.d. noise, based on $\hat{\rho}_Z$, is straightforward - see the textbook (Section 1.6) or any book / course on introductory statistics for approximate hypothesis testing.