The University of British Columbia – Okanagan

DATA315

Assignment 4

Due on March 31. Please upload the file that contains your assignment to Canvas. Two data sets, `wolfRMNP` and `zx` are posted separately under the data tab on Canvas. Show your steps and R codes for your calculation.

Question 1

Which of the following autoregressive processes are stationary? In each case, identify the order.

(a) (2 points)

$$z_t = 0.4z_{t-3} + \varepsilon_t$$

(b) (2 points)

$$z_t = 0.7z_{t-1} - 1.2z_{t-2} + \varepsilon_t$$

(c) (2 points)

$$z_t = -0.2z_{t-1} - 1.2z_{t-3} + \varepsilon_t$$

(d) (2 points)

$$z_t = 0.7z_{t-1} - 0.2z_{t-2} + 0.1z_{t-3} + 0.4z_{t-4} + \varepsilon_t$$

(e) (2 points)

$$z_t = z_{t-1} - 2z_{t-2} + 3z_{t-3} + 4z_{t-4} + \varepsilon_t$$

Question 2

Find $E[z_t]$, $\text{Var}(z_t)$, ρ_1 and ρ_2 for each of the following models:

(a) (8 points) $z_t = -.2z_{t-2} + \varepsilon_t$, $\sigma_e^2 = 9$.

(b) (8 points) $z_t = -.8z_{t-1} + .1z_{t-2} + \varepsilon_t$, $\sigma_e^2 = 4$.

Question 3

Use the model $z_t = -0.7z_{t-1} + 0.2z_{t-2} - 0.1z_{t-3} - 0.4z_{t-4}$ to answer the following questions.

(a) (3 points) Find the Yule-Walker equations that could be used to find the autocorrelation function (acf) up to lag p.

(b) (9 points) Find ρ_1, ρ_2 and ρ_3

(c) (8 points) Find ρ_6 using the general solution of the difference equation

Question 4

To simulate an AR(p) process of length n, with parameters $\phi_1, \ldots, \phi_p$ and variance σ_e^2 in R, use

```r
ar.sequence <- arima.sim(n, model=list(ar=c(phi_1, ...,  phi_p)),
 sd=sigma_e)
```

(a) (2 points) Simulate four sets of time series of 13 values from

$$z_t = .8z_{t-1} + .1z_{t-2} + \varepsilon_t$$

where ε_t has variance 9. Obtain a trace plot of each time series.

(b) (18 points) Simulate another four sets of time series of 200 values from the same model. Plot the ACF and PACF. Compare with a plot of the corresponding theoretical values. (This means that you would need to calculate the first 23 lag theoretical ACF and PACF values. You will need to solve the appropriate difference equation for this. See the General AR chapter for details.)

(c) (3 points) Simulate another four sets time series of 1000 values from the same model. Obtain a trace plot of each time series.

(d) (6 points) Re-simulate another four sets time series of 1000 values from the same model. Compute the sample mean, variance, ACF and PACF (for lags 1 through 4) for each of the simulated series, and compare with the corresponding theoretical ACF and PACF values.

Question 5

Consider the following model: $z_t = .2z_{t-1} - .3z_{t-2} + .1z_{t-3} + .4z_{t-4} + \varepsilon_t$, $\sigma_e^2 = 4$.

(a) (3 points) At what point do you expect its PACF to cut off?

(b) (1 point) Do you expect its ACF to cut off or decay exponentially?

(c) (4 points) Simulate a series of length 1000 from this model. Obtain the sample PACF and SACF. Do these agree with your answers to (a) and (b)?

Question 6

Data on counts of wolves in Riding Mountain National Park (Manitoba, Canada) are contained in the file `wolfRMNP.R`. You can enter it into R using

```r
source("wolfRMNP.R")
```

Build an AR(p) model for the Riding Mountain National Park wolf count data.

(a) (5 points) Use the PACF and ACF to identify an appropriate order for your autoregressive model.

(b) (4 points) Use maximum likelihood estimation to fit the model.

(c) (5 points) Check the residual diagnostics. Are there any problems?

Question 7

Consider the following models.

$$x_t = .6x_{t-1} + z_t \tag{1}$$

($x_{100} = -0.5$, and z_t is NID(0, 1))

$$x_t = 3 - .3x_{t-1} + .2x_{t-2} + z_t \tag{2}$$

($x_{100} = 3.5$, $x_{99} = 2.9$, z_t is NID(0, 4))

(a) (2 points) Using each of the time series models, and the given data, make a point prediction for x_{101}.

(b) (2 points) In each case, specify the distribution of the margin of error for your prediction. (This is usually 1.96 $\times$ the standard error of the forecast.)

(c) (2 points) In each case, supply a 95 % prediction interval.

Question 8

In this question, you will see how you can improve forecasts for one time series by using information from a related time series for which future information is available. Suppose $\{z_t\}$ and $\{x_t\}$ are time series that satisfy the following relations. Suppose $\mu_X, \mu_Z, \alpha, \beta$ and ϕ are constants and

$$z_t = \mu_Z + \phi(z_{t-1} - \mu_Z) + \alpha(x_{t-1} - \mu_X) + \varepsilon_t \tag{3}$$

and

$$x_t = \beta(z_{t-1} - \mu_Z) + \mu_X + \eta_t \tag{4}$$

where η_t and ε_t are white noise processes with standard deviations σ_X and σ_Z, respectively. Suppose also that η_t is independent of z_t.

(a) (2 points) Show that z_t can be expressed in the form of an AR(2) process with mean μ_Z and parameters $\phi_1 = \phi$ and $\phi_2 = \alpha\beta$.

(b) (2 points) Show that the noise component for this model has variance $\alpha^2\sigma_X^2 + \sigma_Z^2$.

(c) (5 points) We have two annual time series. One is `z` and we have observations for it until 2019. The other is `x` and we have accurate predictions of it until 2039.

Fit an AR(2) model to the 100 observations on the time series in `z` which can be found in the `zx` list in *zx.R* using the code. Prior to doing the fitting, use the ACF and PACF functions to assure yourself that an AR(2) model is appropriate.

```
source("zx.R")
z <- zx$z
```

(d) (2 points) Use the following code to fit the simple regression model (4) relating x_t to $z_{t-1} - \mu_Z$, using the first 100 observations of the `x` time series in `zx`.

```
x <- zx$x
n <- length(z)  # n should be 100 in this case
muZhat <- coef(z.AR2)[3] # mean of z's
z0 <- z - muZhat # centered z's
z0 <- z0[-n] # z_{t-1}
x0 <- x[2:n]  # x_t
x0.lm <- lm(x0 ~ z0)  # this fits the simple linear regression model
```

From the output, identify an estimate for β.

(e) (2 points) Use the output from the fitted AR(2) model to identify an estimate of $\alpha\beta$. Use this and your previously obtained estimate of β to obtain an estimate of α.

(f) (2 points) You can also identify an estimate of μ_X from the simple regression output. What is it?

(g) (6 points) Now, we are going to use the fitted model (3) to make 20 years of forecasts on the `z` series.

(h) (4 points) Obtain the next 20 years of predictions on `z` using the AR(2) model and the `predict` function. Store the values in a vector called `zAR2pred`. Using a 1 by 3 layout, obtain trace plots of `z`, `znew` and `zAR2pred`, using `ylim = c(0, 20)` Comment on the results. Which forecasts do you think are more accurate and why?