

Problem 1:

I have put three data set on the website.

ibm: daily IBM stock closing prices

internet: number of users log on to an internet server each minute

gasprices: average price (US dollars per gallons) for regular gasoline in the US. There are $n = 145$ weekly observations collected from 1/5/2009 to 10/10/2011.

Fit an ARMA or ARIMA(p, d, q) to each data. Perform all diagnostics check and identify model worthy of consideration. Write a brief summary of your findings.

Problem 2: The TSA library contains the dataset **CO2** which list the monthly carbon dioxide level in northern Canada from 1/1994 to 12/2004. We would like to fit the model

$$\text{CO2}_t = \beta_0 + \beta_1 t + \beta_2 \cos(2\pi f t) + \beta_3 \sin(2\pi f t) + \epsilon_t$$

where $E(\epsilon_t) = 0$. The deterministic part of the model is

$$\mu_t = \beta_0 + \beta_1 t + \beta_2 \cos(2\pi f t) + \beta_3 \sin(2\pi f t),$$

and contains both linear and trigonometric trend components. Note that there are 12 observations per year, so we take $f = 1$.

(a) Use the following R command to fit the model

```
har.=harmonic(co2, 1)
fit=lm(co2 har. + time(co2))
summary(fit)
```

Give the least square estimate. Interpret your fit.

(b) Plot the data and superimposed your model. How would you rate the fit overall?

(c) Get the studentized residuals **rstudent** from the summary fit. Use them perform all diagnostics check. That is: **hist**, **qqnorm**, **shapiro.test**, **acf**, and the **runs** test. **Comments** on your diagnostics.

Problem 3: Use the **tb** here. It pertains to the number of Tuberculosis cases (per month) in the US from 1/2000 to 12/2009.

(a) Plot the data and identify *candidate trend*, that either: **cosine trend**, **polynomial trend**, or **seasonal mean**. Each is of the form $TB_t = \mu_t + \epsilon_t$ as you know where $E(\epsilon_t) = 0$.

(b) Give the plot of the TB data with your model superimposed. Comment.

(c) Examine the studentized residuals **rstudent** from your model. Comment on your fit.