02424 Week 3

Exercise 1

Question 1

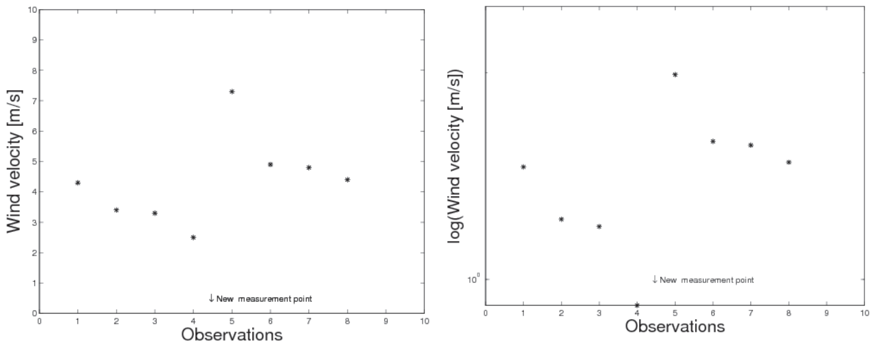

Figure 1: Left: Plot of wind velocity vs. observations. Right: Plot of log wind velocity vs. observations.

In figure 1 the velocity is plotted as a function of observations. By inspecting the graph it is assumed that $\log Y_t$ depends linear on t in each area (due to the few observations it is impossible to determine if the dependency is more linear in the logarithmic case). Furthermore the slope is believed to be the same in the two areas. It is obvious that having measurement from a longer period of time the model will not could be described by a linear model, but for the restricted area, which is considered, it is reasonable to use a linear model constant parameters.

Summarized models

$$\log Y_t = \theta_1 + \theta_2 t + \theta_3 \rho_t + \varepsilon_t$$

where

$$\rho_t = \begin{cases} 0 & \text{for} \quad t \leq 4 \\ 1 & \text{for} \quad t \geq 5 \end{cases}$$

It is assumed that $V[\varepsilon] = \sigma^2$ (i.e. constant).

Question 2

Using the model from Question 1 the observations can be written as

$$\begin{bmatrix} \log Y_1 \\ \vdots \\ \log Y_4 \\ \log Y_5 \\ \vdots \\ \log Y_8 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 0 \\ 1 & 5 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 8 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_4 \\ \varepsilon_5 \\ \vdots \\ \varepsilon_8 \end{bmatrix}$$

or

$$\mathbf{Y}^{\star} = \mathbf{x}\theta + \varepsilon$$

We find

$$\mathbf{x}^{\top}\mathbf{x} = \begin{bmatrix} 8 & 36 & 4 \\ 36 & 204 & 26 \\ 4 & 26 & 4 \end{bmatrix}, \ \mathbf{x}^{\top}\mathbf{Y}^{\star} = \begin{bmatrix} 4.9696 \\ 23.2280 \\ 2.8782 \end{bmatrix}$$

which leads to

$$\hat{\theta} = (\mathbf{x}^{\top}\mathbf{x})^{-1}\mathbf{x}^{\top}\mathbf{Y}^{\star} = \begin{bmatrix} 0.875 & -0.25 & 0.75 \\ -0.25 & 0.1 & -0.4 \\ 0.75 & -0.4 & 2.1 \end{bmatrix} \begin{bmatrix} 4.9696 \\ 23.228 \\ 2.8782 \end{bmatrix} = \begin{bmatrix} 0.700 \\ -0.0709 \\ 0.480 \end{bmatrix}$$

An unbiased estimate of σ^2

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{8}(\log Y_t - \mathbf{x}_t\hat{\theta})^2}{8-3} = 0.0419^2$$

The difference in wind speed at the two measurement points is determined by θ_3. Let $Y_t^{\downarrow}$ and $Y_t^{\uparrow}$ be the wind speed at the old and new measurement point respectively.

$$\log Y_t^{\uparrow} - \log Y_t^{\downarrow} = \widehat{\theta}_3$$

or

$$\frac{Y_t^{\uparrow}}{Y_t^{\downarrow}} = 10^{\hat{\theta}_3} = 3.02$$

The estimated model indicates that the wind speed at the top of the building is approx. 3 times higher than the wind speed at the normal measurement point (2 m. above ground level).

Question 3

The predicted wind speed in one hour ($t = 9$) at the normal measurement point is given as

$$\begin{aligned}\log \hat{Y}_9 &= \mathbf{x}_9 \hat{\theta} = \begin{bmatrix} 1 & 9 & 0 \end{bmatrix} \begin{bmatrix} 0.700 \\ -0.0709 \\ 0.480 \end{bmatrix} = 0.0619 \quad \Rightarrow \\ \hat{Y}_9 &= 1.15 \text{ m/s}\end{aligned}$$

Exercise 2

Question 1 - 2

See the R code in `so2sol.R`

Question 3

The correlation between cities can be taken into account by specifying the correlation between the observations in the statistical model, ie. the model is of the form

$$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Sigma})$$

where the correlation between cities is described by the matrix $\boldsymbol{\Sigma}$. A first (and reasonable) assumption would be to assume that the correlation coefficient is exponentially decaying with the distance between cities, ie.

$$\boldsymbol{\Sigma}_{ij} = \rho^{d_{ij}}$$

where d_{ij} is the distance between city i and city j, and ρ is the correlation for unit distance.