Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
June 13th, 2023

Problem n.1

The file `StorageCentres.txt` contains data regarding 40 storage centres (`id_storage_centre` $\in \{1, ..., 40\}$) located in Regione Lombardia. For each storage centre, information about whether it is located within a radius of 15 km from a city (`rad_less_15_city` $\in \{0, 1\}$) and its size measured in m^2 (`size` $\in \mathbb{R}$) are available. Moreover, the semestral costs [k€] for manteinance (`costs` $\in \mathbb{R}$) are provided at 5 different time points (`time` $\in \{1, ..., 5\}$), starting to the 1st semester of 2021 up to the 1st semester of 2023.

At *baseline* (2nd semester of 2020) the information related to the costs in [k€] (`costs0` $\in \mathbb{R}$) and the economic wellbeing of the storage centre (`growth` $\in \{+1, 0, -1\}$) are also provided.

a) Implement the following linear regression model **M0**:

$$\texttt{costs}_{it} = \beta_0 + \beta_{1t} + \beta_2\, \texttt{costs0}_i + \beta_{3t}\, \texttt{growth}_i + \beta_4\, \texttt{rad\_less\_15\_city}_i + \beta_5\, \texttt{size}_i + \epsilon_{it} \tag{1}$$

$$\forall i \in \texttt{id\_storage\_centre} \quad \text{and} \quad \forall t \in \texttt{time} \tag{2}$$

with $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$. Report the estimates of the *parameters* of the model, the standard deviation σ of the error term and the AIC.

b) Provide the plot of the standardized residuals and comment on it. Do you believe that the hypothesis of homoscedastic residuals is satisfied?

In your opinion, what factors or aspects does **M0** fail to consider, considering the potential impact of Covid-19 on prices? Support your reasoning with appropriate plots (e.g. boxplots).

c) Implement a model **M1** in which the error term assumes the expression $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2_{it})$ with

$$\sigma_{it} = \begin{cases} \sigma \cdot |\texttt{time}_{it}|^x & \text{if } \texttt{growth} = 1 \\ \sigma \cdot |\texttt{time}_{it}|^y & \text{if } \texttt{growth} = 0 \\ \sigma \cdot |\texttt{time}_{it}|^z & \text{if } \texttt{growth} = \text{-1} \end{cases}$$

Report the estimates of x, y, z, σ and the AIC of **M1**.

Perform a test assessing whether it's better to consider $x = y = z$, report the p-value of the test and draw your conclusions.

d) Implement a model **M2** with the same within-group heteroscedasticity structure of **M1**, but with Heteroscedastic Autoregressive Residual Errors. Estimate the parameters of the model, the standard deviation σ of the error term, the parameter ρ in the matrix of the correlation structure and the AIC.

Comment on whether **M1** or **M2** is better, both performing a test and according to the best AIC.

Upload your solution here