Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
January 17th, 2024

Problem 3: Maintenance costs of storage centers

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 maintenance (`costs` $\in \mathbb{R}$) are provided at 5 different time points (`time` $\in \{1, ..., 5\}$), starting 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 $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$ and independent. 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? Support your reasoning with appropriate plots (e.g. boxplots).

c) Implement a model **M1** such that the independent error terms have heterogeneous variances; in particular, assume $\epsilon_{it} \sim \mathcal{N}(0, \sigma_{it}^2)$ with

$$\sigma_{it} = \sigma \cdot |\texttt{time}_{it}|^{\delta}$$

Report the estimates of δ and the AIC.

d) Implement a model **M2** with the same within-group heteroscedasticity structure of **M1**, but with Heteroscedastic Autoregressive (AR1) Residual Errors. Provide a 95% confidence interval for ρ in the matrix of the correlation structure.

Comment on whether **M1** or **M2** is better, supporting your answer with a test.

Upload your results here: `https://forms.office.com/e/zHKgAudp2z`