Problem 1: Analysing CO_2 emissions of European c1itizens

The file `co2.txt` contains data on the estimated (direct or not) CO_2 emissions of 3000 individuals participating in an app-based tracking program across 30 different European cities, identified by the variable `IDCity`. The app monitors various lifestyle behaviors, including purchasing habits, travel patterns, and more.

Consider the following linear mixed-effects model:

$$\texttt{co2}_i = \beta_0 \mathbf{1}_{n_i} + \beta_1 \texttt{purchases}_i + \beta_2 \texttt{heating}_i + \beta_3 \texttt{flights}_i + b_i \mathbf{1}_{n_i} + \boldsymbol{\epsilon}_i \qquad \textbf{(M1)}$$

for $i \in \texttt{IDCity}$ with $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_{n_i})$, $b_i \sim \mathcal{N}(0, \sigma_b^2)$ and n_i the number of individuals in city i.

- $\texttt{co2}_i$ is the n_i-dimensional vector of the normalised CO_2 emitted by individuals in city i;

- $\texttt{purchases}_i$ is the n_i-dimensional vector representing the standardized quantity of new purchases (such as clothing, electronics, and furniture) made per month by individuals in city i;

- $\texttt{heating}_i$ is the n_i-dimensional vector of binary variables indicating the energy sources used for home heating by individuals in city i (0 = electric, 1 = gas);

- $\texttt{flights}_i$ is the n_i-dimensional vector of binary variables indicating whether individuals in city i took more than three flights per year (0 = no, 1 = yes).

a) Fit the model **M1**, estimate β_1 and the PVRE.

b) On top of **M1**, fit now a model **M2**, introducing heteroscedastic residuals $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)$ with

$$\sigma_{ij} = \sigma \cdot |\texttt{purchases}_{ij}|^\delta$$

for individual $j \in \{1, \ldots, n_i\}$, for city $i \in \texttt{IDCity}$.

Estimate δ and the PVRE for **M2**.

c) Perform a likelihood ratio test to compare **M1** and **M2**. Which model would you choose?

d) On top of the selected model, net of the impact of fixed effect covariates, which are the `IDCity` associated with the highest CO_2 emissions?

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