Empirical assignment

Econometrics

The aim of this assignment is to identify the factors that determine retail fuel prices at the station level. Attached is a dataset (please check which one is assigned to your group) containing fuel prices for all Hungarian stations for a randomly selected week.[1] To minimise potential data problems, prepare the weekly average price per station and use this cross-sectional data for the exercise.

1. What factors do you think influence fuel prices?
2. Try to collect data on these factors. Some variables can be generated using the dataset at hand (e.g., the distance to the nearest competing station), but additional data can be obtained from the Hungarian Central Statistical Office:
    a. Regional data from the TIMEA app: https://map.ksh.hu/timea/?locale=en
    b. Settlement data: https://www.ksh.hu/apps/hntr.egyeb?p_lang=EN&p_sablon=LETOLTES

    (You do not need to use all the available variables; the focus should rather be on the interpretability of the relationships.)
3. Estimate different regression models and discuss the parameter estimates and the explanatory power of the model. Carefully interpret the results and draw conclusions about which factors are important in determining fuel prices and which are less important.
4. Illustrate your findings with appropriately chosen figures.
5. What differences can you observe between gasoline and diesel?

Please submit your solutions as a compact essay, with your R script in a separate file. Place special attention to the structure (e.g., introduction, conclusion) and formatting (e.g., well-designed tables and figures). Solutions that do not satisfy the basic requirements of structure and formatting will be sent back for revision. The essay should be complete and understandable on its own, without the R script.

See also the technical and reproducibility requirements below.

The submission deadline is 23.59 on 31st October 2025 via Moodle. If you would like to present your solution in the final practice class on 21st October 2025, you must submit it by 17.00 on 20th October 2025 at the latest.

Groups and dataset:

- Dataset of 2024.10.01-10.07
    - Martina Csikós, Márton Espán, Gergely Gajdócsi:

[1] Source: holtankoljak.hu; compiled by László Nagy-Czirok

- Edvárd Müllner, Samu Kaplony Tósoki, Sólyom András Vigh
- László Bendegúz Krihó, Ákos Perger

- Dataset of 2024.04.11-04.17
  - Chacon Sebastian Chinchilla, Malick Halle Cisse, Pham Thuy Trang Nguyen
  - Péter Csanád Fuchs, Andor Hajdu, Kristián Alex Tudisco
  - Levente Attila Kiss, Marco Romanelli, Milán Szabolcs Sztojka
- Dataset of 2025.09.02-09.08
  - Bátor Péter Birovecz, Soma Attila Horváth, Hunor Kelemen
  - Benedek Pótz-Nagy, Ágnes Vitálos, Attila Benjámin Vörös, Sharmila Mauro Cumbana

Technical and reproducibility requirements:

1. GitHub repository
   a. Each team must create a GitHub repository for the project, containing all codes and outputs.
   b. If the repository is private, add Martin Neubrandt (GitHub ID: martinneubrandt) as a collaborator; if it is public, simply share the repository link.
   c. Each team member must make at least one commit to the repository to demonstrate contribution.
2. R Script Requirements
   a. All code must be clearly commented to ensure readability.
   b. The workflow must be fully reproducible, starting from the raw dataset and ending with the generation of all outputs (tables, figures, and CSV files).
   c. Each output included in the essay must also be saved to an appropriate format (e.g., .csv, .xlsx, .png, .pdf).