Data Transformations

Data and Transformations

Annual data were collected for the period 1995–2022 at the NUTS III regional level, resulting in $N = 644$ observations. The dataset includes:

- Electricity consumption $\mathrm{EC}_{r,t,s}$ at region r, year t, and economic sector s.
- Population $\mathrm{POP}_{r,t}$.
- Cooling Degree Days $\mathrm{CDD}_{r,t}$ and Heating Degree Days $\mathrm{HDD}_{r,t}$ to control for weather effects.
- Gross Value Added per sector $\mathrm{GVA}_{r,t,s}$ and wages per employee $\mathrm{TW}_{r,t}$.
- Electricity prices $\mathrm{REP}_{r,t,s}$, $\mathrm{SEP}_{r,t,s}$, $\mathrm{AEP}_{r,t,s}$, $\mathrm{IEP}_{r,t,s}$ by consumption band and sector.
- Gas prices $\mathrm{D1}_{r,t}$, $\mathrm{D2}_{r,t}$, $\mathrm{I2}_{r,t}$, $\mathrm{I3}_{r,t}$ by consumption band.

The following transformations and normalizations were applied:

1. **Per capita normalization:** To adjust for population size, total and residential electricity consumption and total GVA were normalized per capita as

$$\mathrm{EC}^{pc}_{r,t,s} = \frac{\mathrm{EC}_{r,t,s}}{\mathrm{POP}_{r,t}}, \quad \mathrm{GVA}^{pc}_{r,t,s} = \frac{\mathrm{GVA}_{r,t,s}}{\mathrm{POP}_{r,t}}.$$

   This normalization facilitates comparison across regions of varying population.

2. **Weather variables:** Cooling Degree Days ($\mathrm{CDD}_{r,t}$) are calculated by

$$\mathrm{CDD}_{r,t} = \sum_i \max\left(0, T_{im} - 21^\circ\mathrm{C}\right),$$

   where T_{im} is the mean daily temperature for day i in region r, month m, and year t. If $T_{im} < 24^\circ$C, then $\mathrm{CDD}_{r,t} = 0$.

   Heating Degree Days ($\mathrm{HDD}_{r,t}$) are calculated as

$$\mathrm{HDD}_{r,t} = \sum_i \max\left(0, 18^\circ\mathrm{C} - T_{im}\right),$$

   with $\mathrm{HDD}_{r,t} = 0$ if $T_{im} \geq 15^\circ$C.

   These variables capture cooling and heating energy demands affecting electricity consumption.

3. **Price deflation:** All nominal prices, including electricity and gas prices, were deflated by the Consumer Price Index (CPI_t) from INE:

$$P_{r,t,s}^{\text{real}} = \frac{P_{r,t,s}^{\text{nominal}}}{\text{CPI}_t / \text{CPI}_{\text{base}}},$$

where CPI_{base} is the base year CPI.

4. **Gas price series construction:** To create a consistent gas price series over the period 1995–2022, natural gas prices are used directly from 2002 onwards. For the years between 1995 and 2002, when natural gas prices are not available, the prices are constructed by applying the relative variation of liquefied petroleum gas (LPG) prices to the natural gas price in 2002. Formally:

$$P_t^{\text{gas}} = \begin{cases} P_t^{\text{NG}}, & t \geq 2002, \\ P_{2002}^{\text{NG}} \times \dfrac{P_t^{\text{LPG}}}{P_{2002}^{\text{LPG}}}, & 1995 \leq t < 2002, \end{cases}$$

where P_t^{NG} and P_t^{LPG} denote the natural gas and LPG prices at time t, respectively. This approach ensures a continuous and comparable gas price series by using LPG price changes to estimate natural gas prices before 2002, based on the 2002 natural gas price as a reference.

5. **Logarithmic transformations:** Natural logarithms were applied to key variables to linearize relationships and interpret coefficients as elasticities:

$$\ln \text{EC}_{r,t,s}^{pc}, \quad \ln \text{GVA}_{r,t,s}^{pc}, \quad \ln \text{TW}_{r,t}, \quad \ln P_{r,t,s}^{\text{real}}.$$

6. **Growth rates and differencing:** When necessary, first differences of logarithms were computed to focus on relative changes and ensure stationarity:

$$\Delta \ln X_{r,t} = \ln X_{r,t} - \ln X_{r,t-1}.$$

Variables were chosen based on regional data availability and their proven relevance in the literature. Electricity and gas price bands were defined according to sector-specific consumption profiles from ERSE market studies. Data sources include DGEG, INE, Eurostat, and ERSE and are cited in the paper.