DA report

The dependence of the gas price on electricity in the European Union

Authors:

- Jan Sawicki
- Mateusz Sołtys

Project repository:

github.com/JTSawicki/Projekt-DataAnalytics-2022

1. A short introduction to the issue

Electricity prices in most countries are directly dependent on the prices of fuels and raw materials used to produce energy. For example, in the European Union most of the electricity is produced using fossil fuels, in particular natural gas and crude oil imported from third countries such as Russia or the Persian Gulf countries. This results in a strong dependence of prices on imported raw materials.

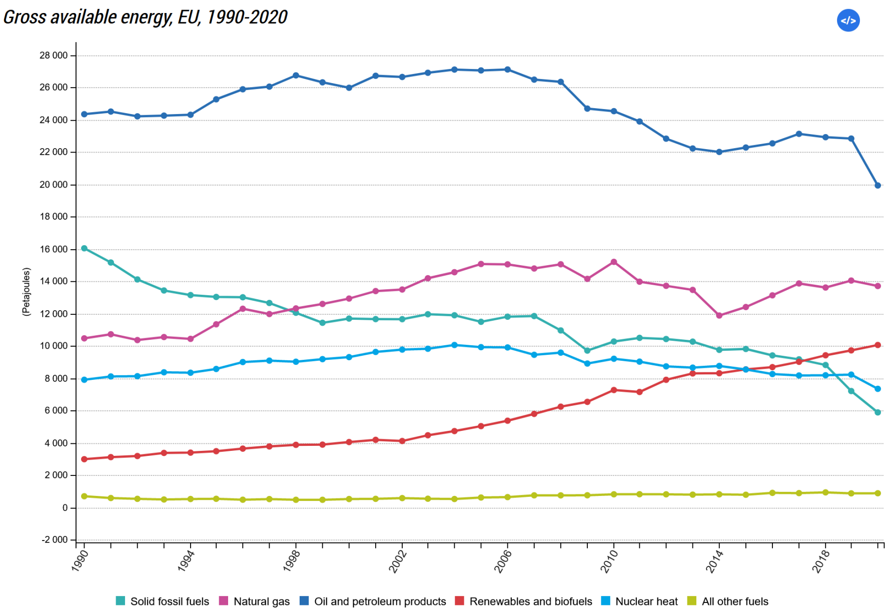


Fig. 01 Sources of energy in the European Union

An analysis of the dependence of these prices would allow for the creation of a model that would allow to predict the estimated increase in electricity prices in the event of an increase in fuel prices, e.g. as a result of embargoes and the war in Ukraine.

2. Data source

During the collection of data and their subsequent analysis, Europe was treated as a uniform entity in statistical terms. This can be done due to the strong interconnection between individual countries, which results in a de facto one electric network and the "global" nature of energy commodity prices. Therefore, the average energy prices obtained from Eurostat were used as a data source.

Only data on electricity and gas prices were available there, but it was found that the model can be based on it because the prices of other fuels are directly dependent on each other.
MJ for natural gas and KWh for electricity were selected as the data unit. However, all prices are in Euro (€).
Then the data available for the 25 EU countries that had the information required for the model were averaged. In addition, the data is broken down separately for private and commercial customers because in the event of price spikes, most EU governments grant aid packages (e.g. by lowering taxes or climate charges) to households.

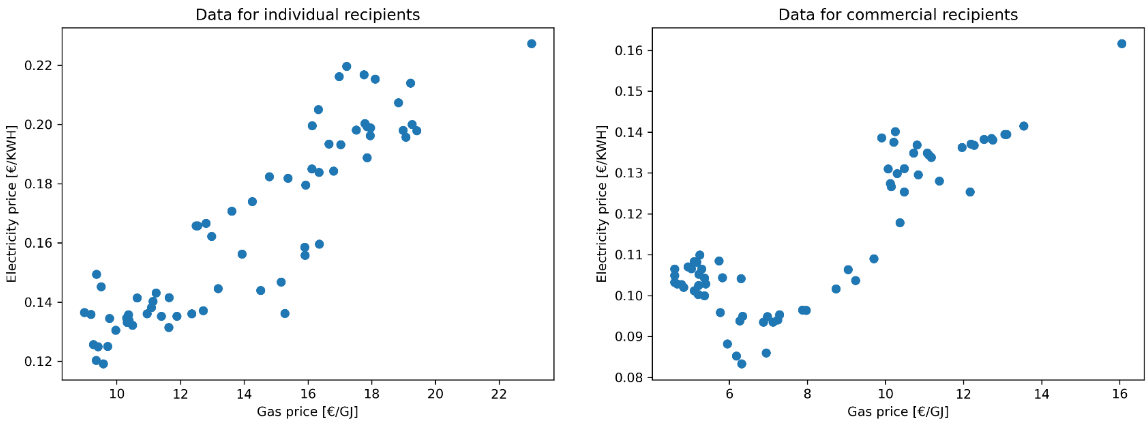


Fig. 02 Acquired data

Countries for which data were averaged: Austria, Belgium, Bulgaria, Czech Republic, Germany, Denmark, Estonia, Spain, France, Croatia, Hungary, Ireland, Italy, Lithuania, Luxembourg, Latvia, Netherlands, Poland, Portugal, Romania, Sweden, Slovakia, Turkey, Great Britain.

3. Model and priors

Two basic models were chosen. The first is the linear model and it describes the price resulting from the cost of energy production:

$$y = \alpha + \beta * x$$

Where in the model:

- y - energy price for 1 KWh
- x - gas price for 1 MJ
- α (alpha) - fixed costs
- β (beta) - conversion factor

As initial assumptions of the model (prior), the values of α and β were limited to positive values because, due to practical economic dependencies, a negative feedback from the raw material price in production is impossible. Additionally, the distribution value of normal beta was strongly reduced to normal (0.007,0.01). This is due to the linear dependence of the conversion of MJ gas to KWh of current:

$$Electricity[Kwh] = \frac{Gas[MJ]}{AverageEfficiency * MJtoKWhConversion}$$

$$1\ Electricity[Kwh] = \frac{1\ Gas[MJ]}{\frac{(0.45 + 0.57)}{2} * 277.78} \approx 0.0070 * 1\ Gas[MJ]$$

However, since the price of energy is also influenced by other factors, such as state subsidies for citizens at high prices, or the fact that some sectors of the economy will reduce consumption, and others will pay any price for the energy needed without limiting consumption, it was decided to create a second model describing the problem with the polynomial fourth degree.

$$y = \alpha + \beta x + \gamma x^2 + \delta x^3 + \varepsilon x^4$$

The same constraints as before were adopted for α and β, while the subsequent coefficients are also non-negative limited by the normal distribution (0.0.001). This is because there have been no dramatic energy disasters in Europe during the period considered (e.g. a complete lack of fuel supply) and normal additional market phenomena cannot have a greater impact on the energy price than production.

Additionally, in order to show the relationship better, calculations were also performed for additional second and third degree polynomial models.

4. Analysis of model results

It takes a long time to compile and match the data to the models. Using a platform with a six-core i5-8600K processor overclocked to 4.2 GHz, it took about 1.5 minutes each time.

After making the model adjustments, the following results were obtained. Sigma is the calculated standard deviation of the data. No sampling issues were encountered.

For individual clients

Linear model:

alpha = 5.17222 * 10 $^{-2}$

beta = 8.04387 * $^{10-3}$

sigma = 0.0135775 * 10 $^{-}$

Total difference in probability distribution between input and output data = 459,789

Stage 4 model:

alpha = 8.32107 * 10 $^{-2}$

beta = 4.06172 * 10 $^{-3}$

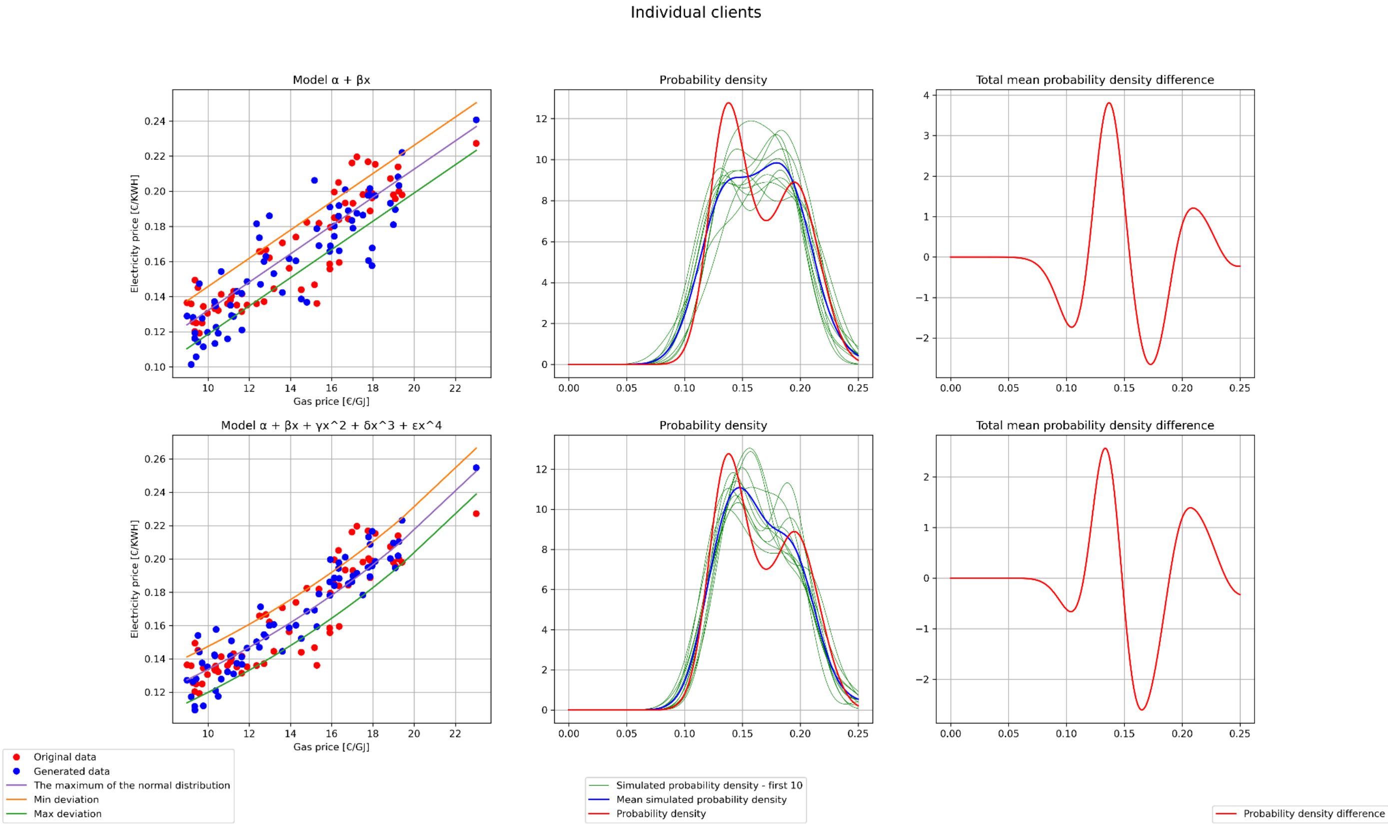
gamma = 7.56164 * 10 $^{-5}$

delta = 1.77194 * 10 $^{-6}$

epsilon = 5.12730 * 10 $^{-8}$

sigma = 0.0137871 * 10 $^{-}$

Total difference in probability distribution between input and output data = 355,638



For commercial customers

Linear model:

alpha = 7.06123 * 10^{-2}

beta = 5.25652 * 10^{-3}

sigma = 0.00996989 * 10^{-}

Total difference in the probability distribution of the input and output data = 449,915

Stage 4 model:

alpha = 8.84904 * 10^{-2}

beta = 1.37348 * 10^{-3}

gamma = 1.08631 * 10^{-4}

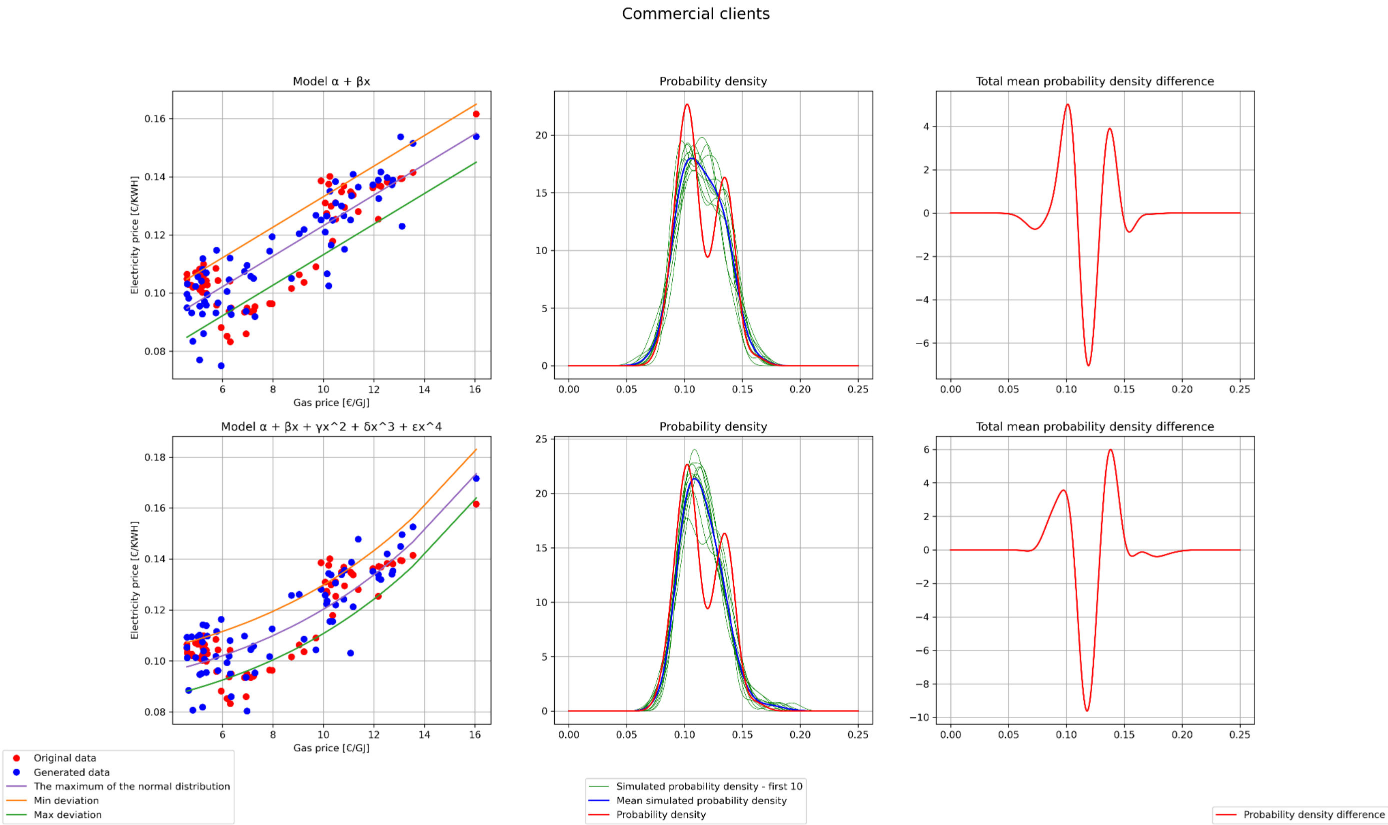
delta = 5.12428 * 10^{-6}

epsilon = 2.07216 * 10^{-7}

sigma = 0.00952928 * 10^{-}

Total difference in the probability distribution of the input and output data = 591,679

Energy statistics - an overview



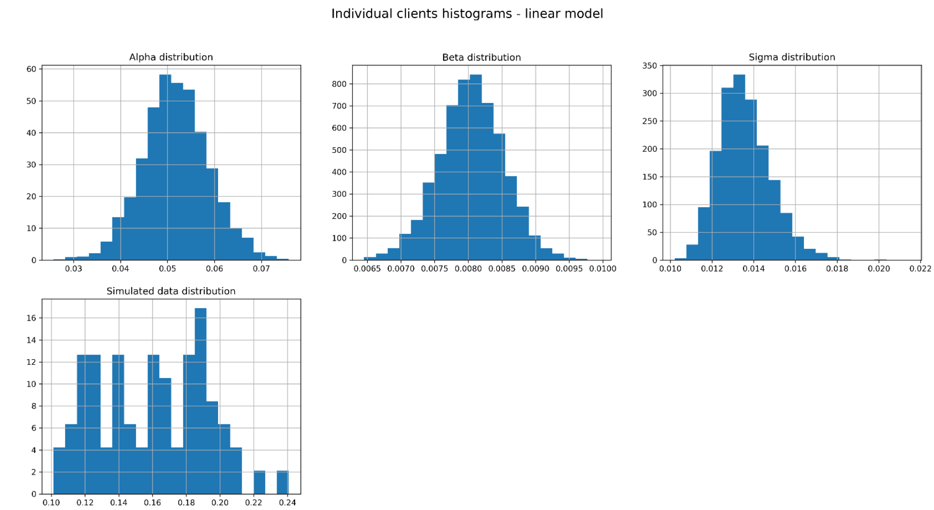
For all results (also for the 2nd and 3rd stage models), the status fit diagnostic function returned information that there were no problems.

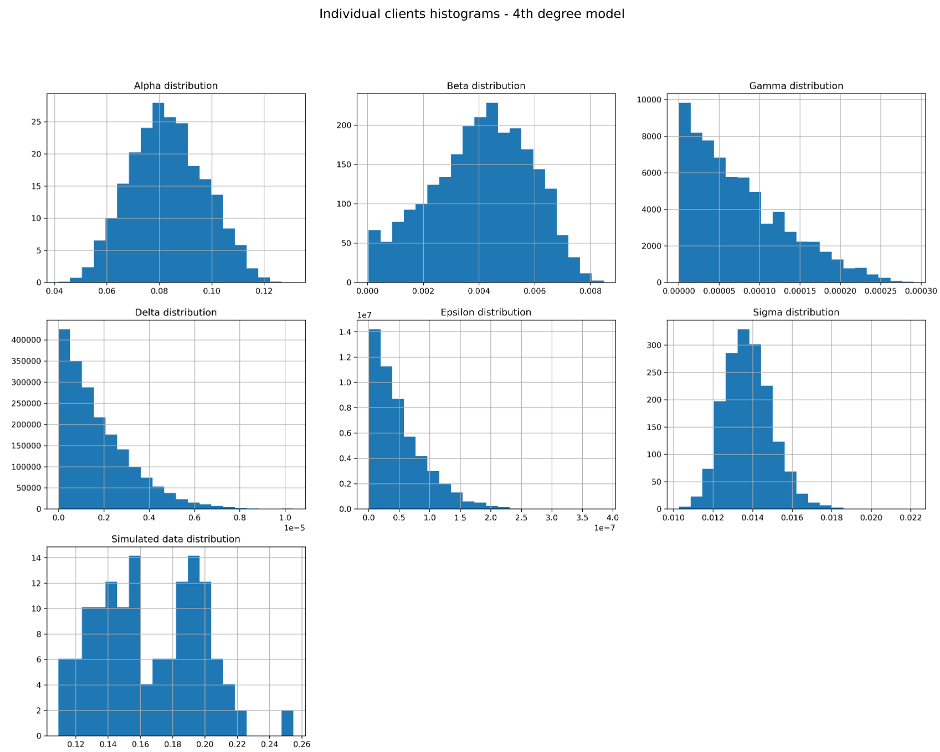
As can be seen from the above examples, the linear relationship is a good estimate, and in the case of commercial customers, even the best.

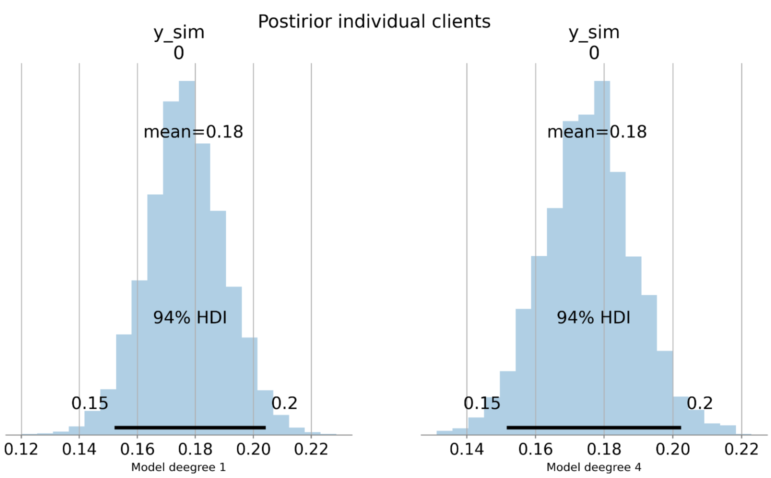
In the case of matching to given households, slightly better results were achieved in the case of using higher-order models, however, a question can be raised whether increasing the complexity of the model is justified in view of a slight improvement.

5. Parameter histograms & posterior

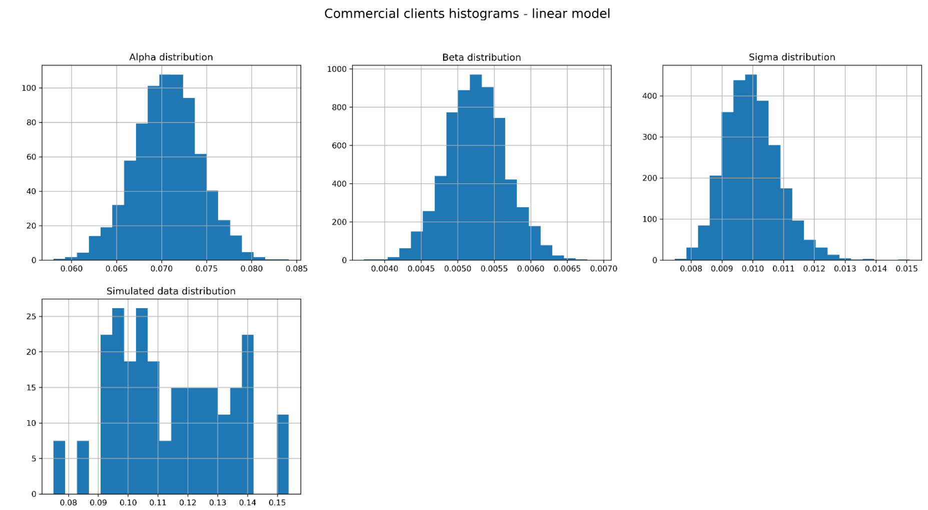
Individual clients:

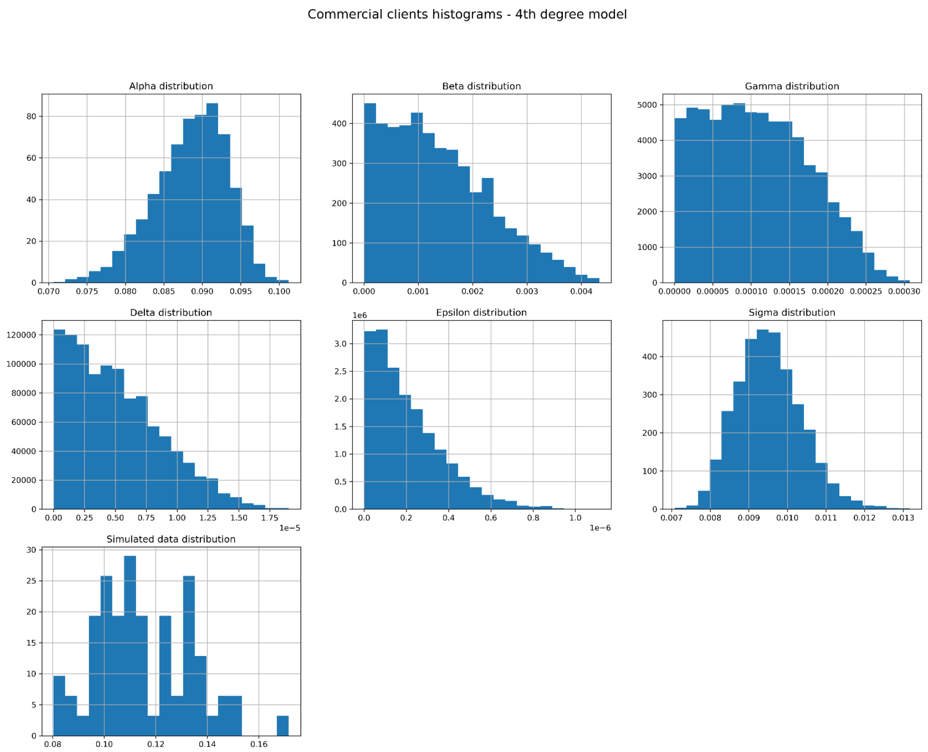


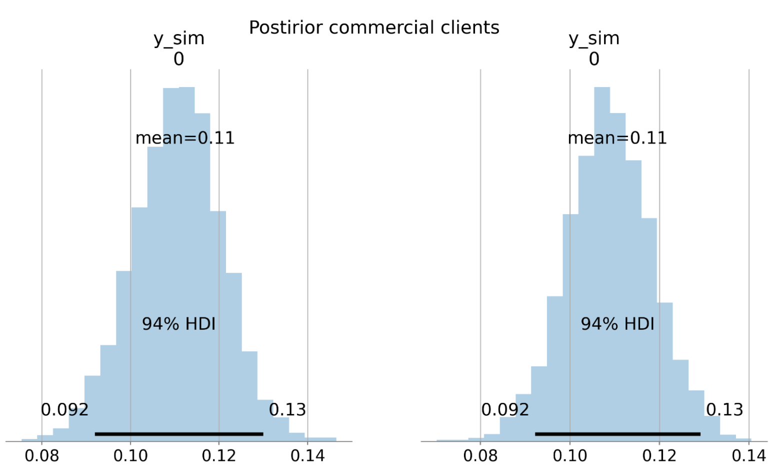




Commercial clients:







6. Bibliography

- Eurostat energy statistics - an overview
- Eurostat Database
- ScienceDirect