# The factors that determine retail fuel prices

Econometrics Empirical Groupe Assignment

Created by: Levente Attila Kiss, Marco Romanelli, Milán Szabolcs Sztojka

Introduction:

In this assignment we identify the factors that determine retail fuel prices at the station level on data containing fuel prices for all Hungarian stations for a randomly selected week. During the research, we first identify the relevant factors that influence fuel prices, then analyse the data and variables describing these factors, which, for the sake of easier utilization, are put trough several data cleaning processes and transformations. To make it easier to create a model that best describes the problem, we tried different models and select based on several critical factors. Finally, we presented the main results of the submission in the conclusion.

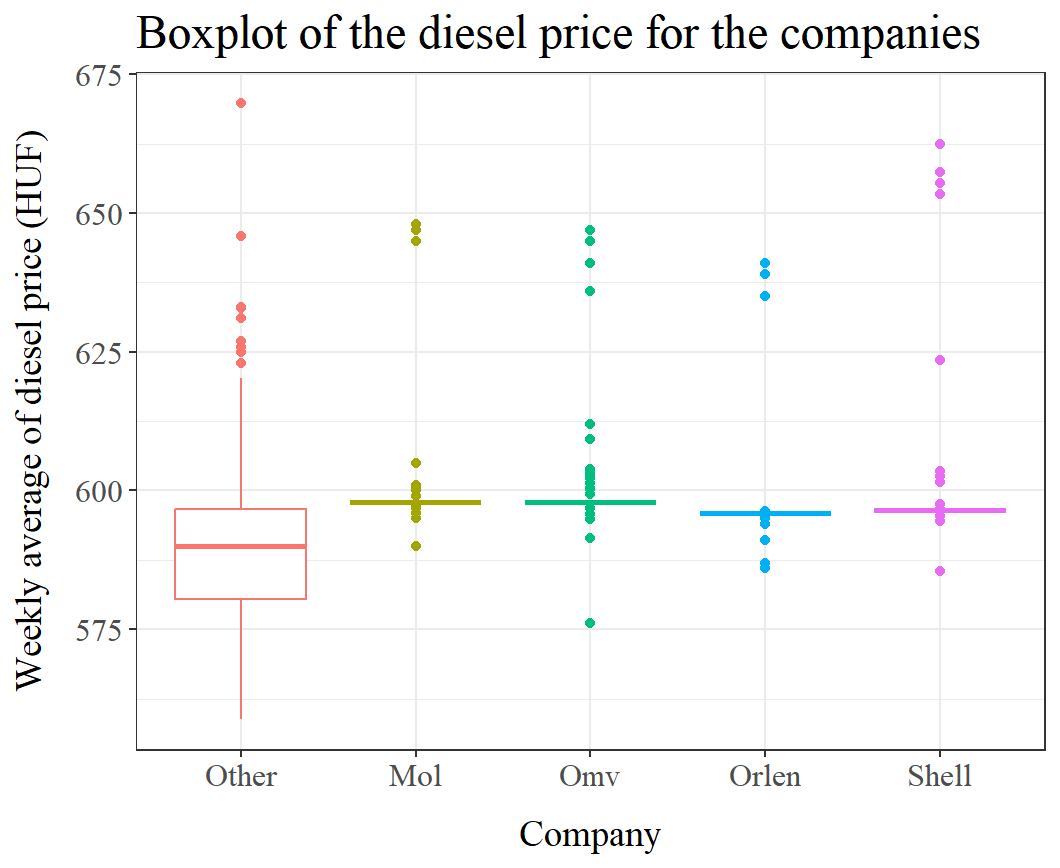
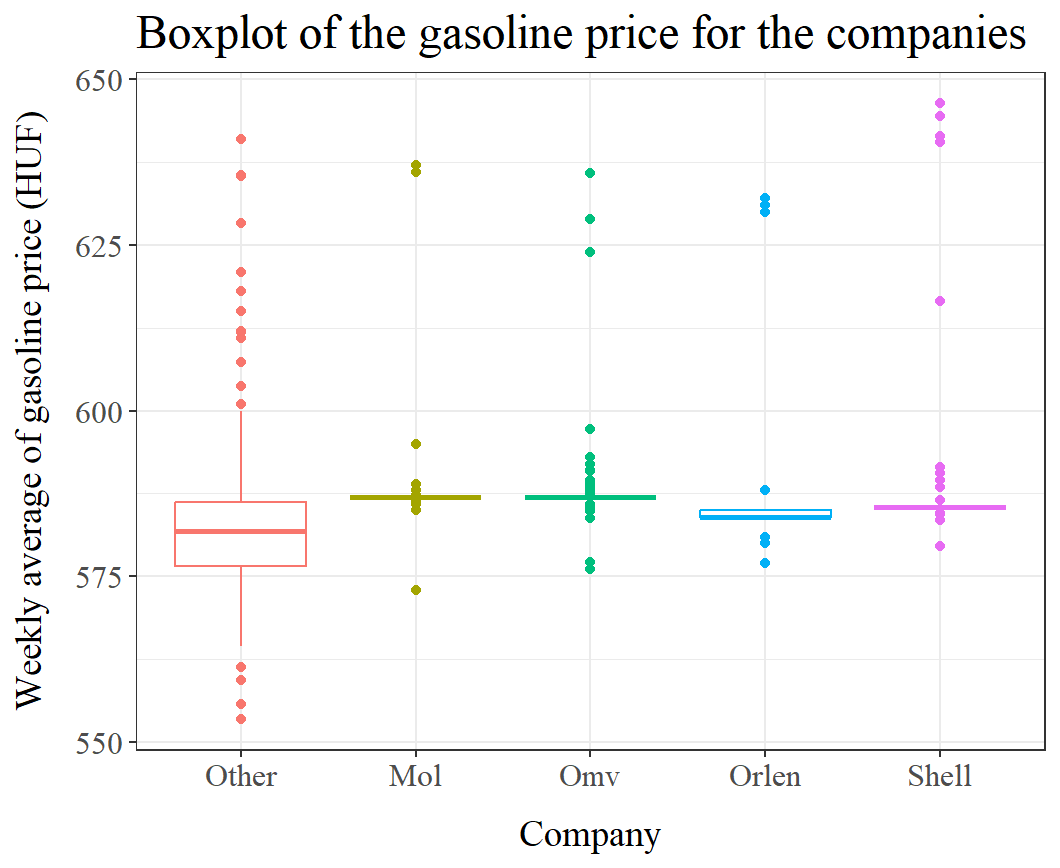
At the beginning of our research, we first examined the data for the week assigned to us (Dataset of 2024.04.11-04.17) to define the main factors behind gas and diesel prices all around Hungary. We identified two main components, the company and the location of the fuel station. The first one contains more than 100 fuel station owner companies, from multinational big firms like Mol or Shell, to small family businesses with only one station. Therefore, we assume that this huge different could lead to difference in prices. The second factor, the location is more complicated. We think this is the main component behind the fuel prices. On one hand because it is important what features a fuel station have, like based on our own experience a station on a highway or a motorway or in Budapest is more expensive. On the other hand, the location of the city also contains important economic factors like wage, what can impact prices, or car density, what can impact demand in the region.

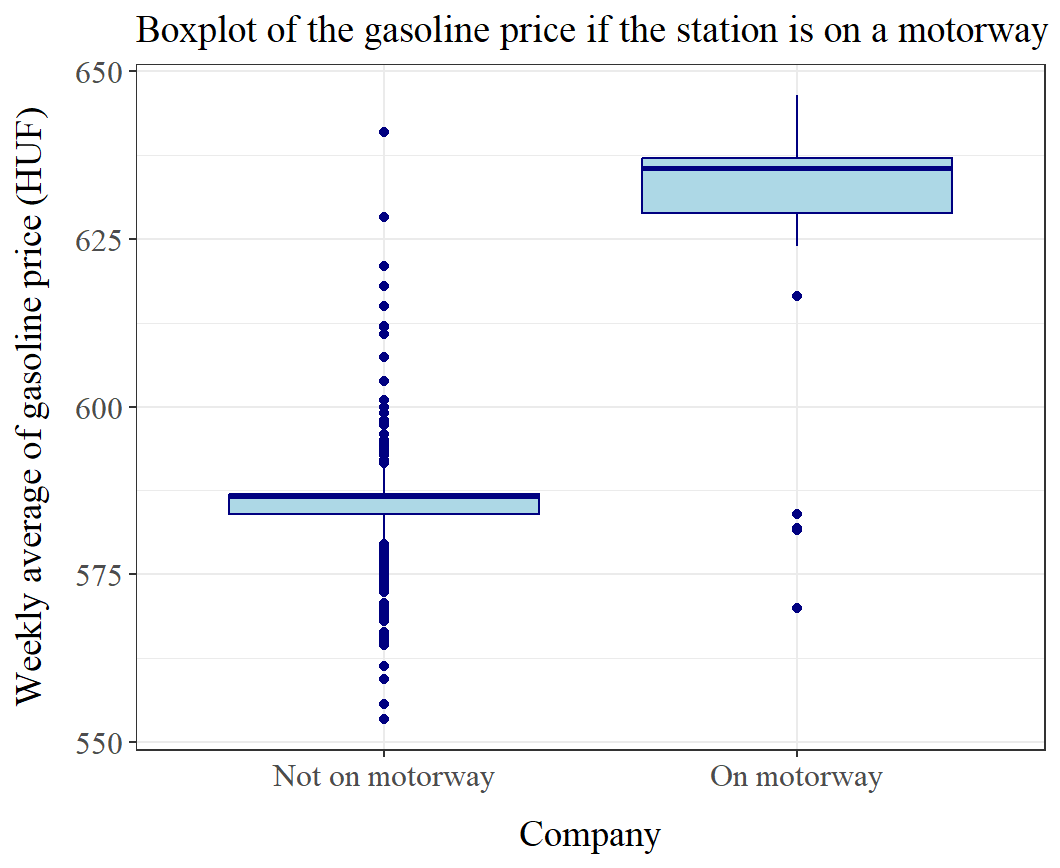
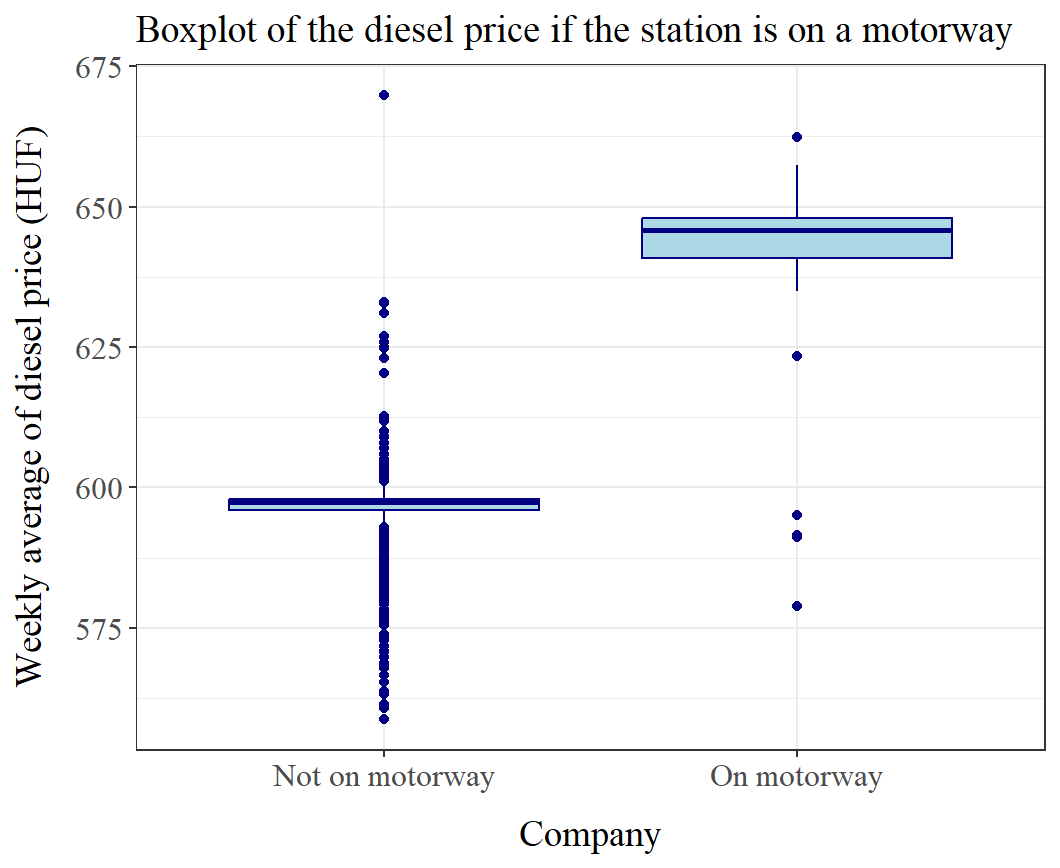
Data:

The attached dataset contains cross-sectional data, therefore in the external data collection we focused on searching the closest time period to the original data, which is about fuel prices in the period of 11.-17. of April 2025. As additional data we obtained three datasets from the Hungarian Central Statistical Office. The first data is about passenger cars per thousand capita, so the car density, for every Hungarian county from 2024. It is from the TIMEA app. We think that areas, that have a higher average of cars per 1000 people might operate with higher fuel prices, as the demand should be higher as well. We also added the average wage of the county, that the petrol station is located in. We try to measure income effects by adding this variable. For this we used a dataset that contained the wages, and another that we only used, to connect the cities with the counties.

The original dataset contains variables about the name of the company, the city and the diesel and gasoline prices of the day in HUF/liter. As it is asked, we prepare the weekly average price per station, to minimised potential data problems, in both kind of fuels. After filtering for not available data the final dataset contains 1177 observations. This will be used as the dependent variable for our models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Descriptive statistics for the dependent variables | | | | | |
| Price | Mean | SD | Min | Max | IQR |
| Diesel | 598.13 | 14.46 | 558.76 | 669.90 | 2.00 |
| Gasoline | 588.07 | 13.31 | 553.43 | 646.47 | 2.53 |

From the company variable we created categories as a factor variable. After analysing the distribution of the fuel station owner firms we created five categories, four for the biggest companies with more than 100 fuel stations, they are Mol, Shell, Orlen and OMV. The fifth category contains every other smaller firm, as control category. We believe that fuel prices are not very hectic across stations of the same company.

Finally, from the address variable we created two dummy variables, as important features of the station, one if it is on a motorway and one if it is in the capital city. According to our expectations, the location of the stations highly influence their prices, as the demand for fuel depends on how the traffic is in the area.

Modelling:

We built the model step by step, by adding the variables. After checking AIC and BIC scores and analysing the variance of the different models we conclude that the best model is the following:

The variable of the average income proved not to improve the model, so it has been left out. We use the same model to estimate the gasoline and diesel prices. Any difference between the effects that determine the prices shall be present in the models’ results.

We also tested if the error term is heteroscedastic and based on a Breusch-Pagan test we consider it to be, so we are using heteroscedasticity-robust standard errors.

==================================================

Dependent variable:

------------------------------------------------------------------------------------

Diesel price Gasoline price

-------------------------------------------------------------------------------------

Mol 10.204\*\*\* 6.528\*\*\*

(0.892) (0.746)

Omv 10.015\*\*\* 6.230\*\*\*

(0.927) (0.780)

Orlen 6.885\*\*\* 3.148\*\*\*

(0.946) (0.798)

Shell 8.401\*\*\* 4.756\*\*\*

(0.948) (0.790)

On motorway 43.773\*\*\* 43.779\*\*\*

(1.661) (1.613)

In Budapest -1.158\*\* -1.089\*\*\*

(0.506) (0.423)

Car/1000 people 0.010\* 0.005

(0.006) (0.004)

Constant 583.927\*\*\* 578.885\*\*\*

(2.950) (1.951)

--------------------------------------------------------------------------------------

Observations 1,177 1,177

Adjusted R2 0.699 0.753

===================================================

Note: \* p < 0.1; \*\* p < 0.05; \*\*\* p < 0.01

If our results are considered on a 5% significancy level, most of our parameters are considered to be significant. We see, that on average, the biggest companies tend to price diesel higher between 7-10 Forints/l, while everything is held constant. This effect is a little bit less, if we consider gasoline prices, it is around 3-6.5 Forints per liter. We find that being on the motorway has the biggest effect. In both the gasoline and diesel prices, on average a petrol station uses 43.8 Forints/l higher prizes if it is located on a motorway (ceteris paribus). This is a huge difference, as it is almost 10% of the means of the prices. The parameter of the dummy Variable of Budapest is also significant on the 5% level for both dependent variables, however the direction of the effect may be counterintuitive. We would imagine, that in Budapest prices are higher, as the population is relatively big, that creates more demand, but the number of stations in the city is also more, so the competition could still be high, that pushes prices down. The effect isn’t too high though, petrol stations on average have 1 Forint/l lower prices in the capital, than stations elsewhere (ceteris paribus). The parameter of the car density variable is insignificant, so it seems we can’t find a relation with prices. The adjusted R2 for both models are relatively high, so we consider we could describe a big chunk of the variance of the prices. We don’t see differences in the direction of the parameters, so we think that the same effects determine the prices for both diesel and gasoline, only the magnitude changes some.

Conclusion:

We modelled fuel prices based on their companies and other geographical factors. After we collected and structured our data, we built our models and selected the one that we thought is best for our purposes. We have found that the biggest factor is if a petrol station is located at a motorway, which boosts both diesel and gasoline prices by more than 40 Forints/l across stations with similar features. Beside this, the company of the station also proves to be a factor, with bigger companies on average pricing higher, than small companies. As our R2 is high in both models, we are satisfied with our models’ performance, however it doesn’t mean they don’t have any limitations. As we could only gather data from limited sources, there could be other variables that we missed, based on the local competition of the petrol stations and other factors regarding the wealth and preferences of the stations’ customers. This could be a baseline to a more detailed and profound study.

GitHub repository:

Link: <https://github.com/Lecsak014/Econ_empirical_assignment.git>

Name: Econ\_empirical\_assignment (Public) created by Lecsak014 (Levente Kiss)