4
Empirical Model

As mentioned previously, our objective is to estimate the competitive effect on the ethanol and gasoline retail markets generated by the raise of percentage of flex fuel cars in relation to percentage of cars that use only one type of fuel. To analyze this issue, we will use the unbalanced panel data described in section 3.1.

We will estimate the model given by the following equations:

$$\text{Pgas}_{it} = \delta^{g}.\%\text{Flex}_{\text{cm}} + \phi^{g}.\text{totgas}_{\text{cm}} + \lambda^{g}.\text{Cgas}_{\text{it}} + \gamma_1^{g}.\text{Peth}_{\text{it}} + \gamma_2^{g}.\%\text{Flex.Peth}_{\text{it}} + \beta^{g}{}'.X_{cm} + \varepsilon^{g}{}_{it}$$

$$\text{Peth}_{it} = \delta^{a}.\%\text{Flex}_{\text{cm}} + \phi^{a}.\text{toteth}_{\text{cm}} + \lambda^{a}.\text{Ceth}_{\text{it}} + \gamma_1^{a}.\text{Pgas}_{\text{it}} + \gamma_2^{a}.\%\text{Flex.Pgas}_{\text{it}} + \beta^{a}{}'.X_{cm} + \varepsilon^{a}{}_{it}$$

Where:

i indexes a gas station, t indexes a week, m indexes a month and c indexes a city[5].

Pgas_{it} and Peth_{it} are gasoline and ethanol prices (in R$).

Cgas_{it} and Ceth_{it} are gasoline and ethanol costs (in R$).

toteth_{cm} and totgas_{cm} are the stock of cars that can run on ethanol and gasoline, respectively (in thousands of cars) – each variable includes the total number of flex cars.

$\%\text{Flex}_{cm}$ is the percentage of flex cars in relation to the total number of vehicles.

$\%\text{Flex.Pgas}_{it}$ is the interaction between percentage of flex cars and gasoline price.

$\%\text{Flex.Peth}_{it}$ is the interaction between percentage of flex cars and ethanol price.

---

[5] The variables indexed by *c* do not vary within stations in a given city and variables indexed by *m* do not vary within weeks in a given month.

X_{cm} is a vector containing the following variables: municipal GDP per capita (in R$), an approximation of the number of fuel stations per vehicles and an approximation of the number of hotels per km².

Additionally, we will include time dummy variables, city dummy variables and/or station dummy variables, according to specifications described in section 5.

With regard to the coefficients of interest, we expect that, controlling for other variables, an increase in percentage of flex cars will reduce fuel prices. That is, we conjecture that the coefficient of $\%Flex_{it}$ will be negative in both equations:

$$\delta^a < 0 \qquad \delta^g < 0$$

Other important coefficients to be analyzed are the ones of the interactions between the percentage of flex cars and fuel prices. As previously mentioned, we believe that an increase in the percentage of flex cars will raise the degree of substitution between the two products. That is to say we expect that an increase in the percentage of flex cars will augment the sensitivity of one fuel price to the other:

$$\gamma_2^a > 0 \qquad \gamma_2^g > 0$$

To test our main hypothesis we are using the variable percentage of flex cars and not the total stock of flex cars in each city. We cannot assume that by simply raising the number of flex cars in the market the fuel prices will get lower. That is owing to the increase in the numbers of flex cars having an ambiguous effect on prices: it narrows the difference between ethanol and gasoline (which forces the prices to go down) but at the same time it increases the demand size for both fuels (which might makes prices go up). So, we are considering that the drop in ethanol and gasoline prices is due to the substitution of mono-fuel cars for bi-fuel cars, not simply to an increase in the number of flex cars.

As we are estimating a reduced form model (that is to say we are estimating neither a supply nor a demand equation), we do not consider consumed fuel quantity a fundamental control variable to test the hypothesis mentioned above. Besides, the number of cars by fuel type is a good proxy for consumed fuel quantity in each city.