Empirical Industrial Organization (ECO 310)

Winter 2021. Victor Aguirregabiria

Problem Set #2

Constructing BLP Instruments for Question 2

This document describes how to construct BLP instrumental variables using only the commands `egen` and `gen` in STATA. More specifically, it describes the construction of the instrument $IV1$ and an instrument $IV3$ that is closely related to $IV2$.

Since the construction of instrument $IV2$ requires using loops in STATA, students can choose between using $IV1$ and $IV3$ instead of $IV1$ and $IV2$. Of course, students can still choose using $IV2$ and construct this variable using loops.

[1] The definition of the instruments.

Let j index car model (variable `co`), let m index country (variable `country`), and let t index year (variable `year`). In this dataset we have 6 product characteristics: `horsepower`, `fuel`, `width`, `height`, `weight`, and `domestic`. For the sake of concreteness, I focus here on the construction of the instrumental variables associated with one of these product characteristics, say `fuel`. Note that $\texttt{fuel}_{jmt}$ represents observation (j, m, t) of variable `fuel`. Let $IV1\_fuel$, $IV2\_fuel$, and $IV3\_fuel$ be the three instrumental variables associated with `fuel`. $IV1\_fuel_{jmt}$ represents observation (j, m, t) of variable $IV1\_fuel$. Similarly, $IV2\_fuel_{jmt}$ and $IV3\_fuel_{jmt}$ represent observation (j, m, t) of variables $IV2\_fuel$ and $IV3\_fuel$, respectively. These instruments are defined as follows:

$$IV1\_fuel_{jmt} = \sum_{i=1}^{J_{mt}} (\texttt{fuel}_{jmt} - \texttt{fuel}_{imt})$$

$$IV2\_fuel_{jmt} = \sum_{i=1}^{J_{mt}} |\texttt{fuel}_{jmt} - \texttt{fuel}_{imt}|$$

$$IV3\_fuel_{jmt} = \sum_{i=1}^{J_{mt}} (\texttt{fuel}_{jmt} - \texttt{fuel}_{imt})^2$$

where J_{mt} is the number of car models in the dataset in country m at year t.

You might find some discrepancy between the definitions of $IV1\_fuel_{jmt}$ and $IV2\_fuel_{jmt}$ above and those in the enunciate of Problem Set #2. Note that they are exactly equivalent because for $j = i$, $\texttt{fuel}_{jmt} - \texttt{fuel}_{imt} = 0$ and $|\texttt{fuel}_{jmt} - \texttt{fuel}_{imt}| = 0$.

[2] STATA Code to construct IV1_fuel

Note that:

$$\begin{aligned} IV1\_fuel_{jmt} &= \sum_{i=1}^{J_{mt}} (\texttt{fuel}_{jmt} - \texttt{fuel}_{imt}) \\ &= J_{mt} * \texttt{fuel}_{jmt} - \sum_{i=1}^{J_{mt}} \texttt{fuel}_{imt} \\ &= J_{mt} * \texttt{fuel}_{jmt} - SUMfuel_{mt} \end{aligned}$$

where $SUMfuel_{mt}$ is $\sum_{i=1}^{J_{mt}} \texttt{fuel}_{imt}$. Based on this formula, we can construct variable $IV1\_fuel$ using the following three lines of code in STATA.

```
egen sumfuel = sum(fuel), by(country year)
egen numJ = sum(1), by(country year)
gen IV1_fuel = numJ * fuel - sumfuel
```

[3] STATA Code to construct IV3_fuel

Define $\texttt{meanfuel}_{mt}$ as the mean value $\frac{1}{J_{mt}} \sum_{i=1}^{J_{mt}} \texttt{fuel}_{imt}$, where J_{mt} is the number of car models in the dataset in country m at year t. It is simple to verify that $\sum_{j=1}^{J_{mt}} (\texttt{fuel}_{jmt} - \texttt{meanfuel}_{mt}) = 0$, and we will use this result below. By adding and substracting $\texttt{meanfuel}_{mt}$ in the expression for $IV3\_fuel_{jmt}$, we can get the following result:

$$\begin{aligned} IV3\_fuel_{jmt} &= \sum_{i=1}^{J_{mt}} \left([\texttt{fuel}_{jmt} - \texttt{meanfuel}_{mt}] - [\texttt{fuel}_{imt} - \texttt{meanfuel}_{mt}] \right)^2 \\ &= \sum_{i=1}^{J_{mt}} [\texttt{fuel}_{jmt} - \texttt{meanfuel}_{mt}]^2 \\ &\quad -2 \sum_{i=1}^{J_{mt}} [\texttt{fuel}_{jmt} - \texttt{meanfuel}_{mt}][\texttt{fuel}_{imt} - \texttt{meanfuel}_{mt}] \\ &\quad + \sum_{i=1}^{J_{mt}} [\texttt{fuel}_{imt} - \texttt{meanfuel}_{mt}]^2 \\ &= J_{mt} * [\texttt{fuel}_{jmt} - \texttt{meanfuel}_{mt}]^2 \\ &\quad -2[\texttt{fuel}_{jmt} - \texttt{meanfuel}_{mt}] \sum_{i=1}^{J_{mt}} [\texttt{fuel}_{imt} - \texttt{meanfuel}_{mt}] \\ &\quad + J_{mt} * \texttt{varfuel}_{mt} \\ &= J_{mt} * [\texttt{fuel}_{jmt} - \texttt{meanfuel}_{mt}]^2 + J_{mt} * \texttt{varfuel}_{mt} \end{aligned}$$

where $\texttt{varfuel}_{mt}$ is the variance $\frac{1}{J_{mt}} \sum_{i=1}^{J_{mt}} [\texttt{fuel}_{imt} - \texttt{meanfuel}_{mt}]^2$, and we have used the result $\sum_{j=1}^{J_{mt}} (\texttt{fuel}_{jmt} - \texttt{meanfuel}_{mt}) = 0$. Using this formula, we can construct variable $IV3\_fuel$ using the following lines of code in STATA.

```
egen meanfuel = mean(fuel), by(country year)
egen varfuel = var(fuel), by(country year)
egen numJ = sum(1), by(country year)
gen IV3_fuel = numJ * (fuel - meanfuel)^2 + numJ * varfuel
```