Empirical Industrial Organization (ECO 310)
Fall 2018. Victor Aguirregabiria

Solution to Problem Set #2
Due on Thursday, December 6th, 2018 [before 11:59pm]

INSTRUCTIONS. Please, follow the following instructions for the submission of your completed problem set.

1. Write your answers electronically in a word processor.

2. For the answers that involve coding in STATA, include in the document the code in STATA that you have used to obtain your empirical results.

3. Convert the document to PDF format.

4. Submit your problem set (in PDF) online via Quercus.

5. You should submit your completed problem set before 11:59pm of Monday, December 6th, 2018.

6. Problem sets should be written individually.

The total number of marks is 100.

QUESTION 1. [50 points]. Consider an industry with a differentiated product. There are two firms in this industry, firms 1 and 2. Each firm produces and sells only one brand of the differentiated product: brand 1 is produced by firm 1, and brand 2 by firm 2. The demand system is a logit demand model, where consumers choose between three different alternatives: $j = 0$, represents the consumer decision of no purchasing any product; and $j = 1$ and $j = 2$ represent the consumer purchase of product 1 and 2, respectively. The utility of no purchase ($j = 0$) is zero. The utility of purchasing product $j \in \{1, 2\}$ is $\beta\ x_j - \alpha\ p_j + \varepsilon_j$, where the variables and parameters have the interpretation that we have seen in class. Variable x_j is a measure of the quality of product j, e.g., the number of stars of the product according to consumer ratings. Therefore, we have that $\beta > 0$. The random variables ε_1 and ε_2 are independently and identically distributed over consumers with a type I extreme value distribution, i.e., Logit model of demand. Let H be the number of consumers in the market. Let s_0, s_1, and s_2 be the market shares of the three choice alternatives, such that s_j represents the proportion of consumers choosing alternative j and $s_0 + s_1 + s_2 = 1$.

Q1.1. (5 points) Based on this model, write the equation for the market share s_1 as a function of the prices and the qualities x's of all the products.

ANSWER. The logit model with two products and utilities $\beta\ x_j - \alpha\ p_j + \varepsilon_j$ implies that the market share of product 1 is:

$$s_1 = \frac{\exp\{\beta\ x_1 - \alpha\ p_1\}}{1 + \exp\{\beta\ x_1 - \alpha\ p_1\} + \exp\{\beta\ x_2 - \alpha\ p_2\}}$$

Q1.2. (10 points) Obtain the expression for the derivatives: (a) $\frac{\partial s_1}{\partial p_1}$; (b) $\frac{\partial s_1}{\partial p_2}$; (c) $\frac{\partial s_1}{\partial x_1}$; and (d) $\frac{\partial s_1}{\partial x_2}$. Write the expression for these derivatives in terms only of the market shares s_1 and s_2 and the parameters of the model.

ANSWER. We can write $s_1 = \frac{\exp\{\delta_1\}}{1+\exp\{\delta_1\}+\exp\{\delta_2\}}$ with $\delta_1 = \beta\ x_1 - \alpha\ p_1$ and $\delta_2 = \beta\ x_2 - \alpha\ p_2$. Note that:

$$\begin{aligned}
\frac{\partial s_1}{\partial \delta_1} &= \frac{\exp\{\delta_1\}\ \ [1 + \exp\{\delta_1\} + \exp\{\delta_2\}]}{[1 + \exp\{\delta_1\} + \exp\{\delta_2\}]^2} - \frac{\exp\{\delta_1\}\ \ \exp\{\delta_1\}}{[1 + \exp\{\delta_1\} + \exp\{\delta_2\}]^2} \\
&= s_1 - (s_1)^2 = s_1(1 - s_1)
\end{aligned}$$

And,

$$\begin{aligned}
\frac{\partial s_1}{\partial \delta_2} &= -\frac{\exp\{\delta_1\}\ \ \exp\{\delta_2\}}{[1 + \exp\{\delta_1\} + \exp\{\delta_2\}]^2} \\
&= -s_1\ s_2
\end{aligned}$$

(a) Using the chain rule of derivation, $\frac{\partial s_1}{\partial p_1} = \frac{\partial s_1}{\partial \delta_1}\ \frac{\partial \delta_1}{\partial p_1}$. Note that $\frac{\partial \delta_1}{\partial p_1} = -\alpha$. And as shown above $\frac{\partial s_1}{\partial \delta_1} = s_1(1 - s_1)$. Therefore:

$$\frac{\partial s_1}{\partial p_1} = -\alpha\ s_1(1 - s_1)$$

(b) Using the chain rule of derivation, $\frac{\partial s_1}{\partial p_2} = \frac{\partial s_1}{\partial \delta_2}\ \frac{\partial \delta_2}{\partial p_2}$. Note that $\frac{\partial \delta_2}{\partial p_2} = -\alpha$. And as shown above $\frac{\partial s_1}{\partial \delta_2} = -s_1\ s_2$. Therefore:

$$\frac{\partial s_1}{\partial p_2} = \alpha\ s_1\ s_2$$

(c) Using the chain rule of derivation, $\frac{\partial s_1}{\partial x_1} = \frac{\partial s_1}{\partial \delta_1}\frac{\partial \delta_1}{\partial x_1}$. Note that $\frac{\partial \delta_1}{\partial x_1} = \beta$. Therefore:

$$\frac{\partial s_1}{\partial x_1} = \beta\ s_1(1 - s_1)$$

(d) Using the chain rule of derivation, $\frac{\partial s_1}{\partial x_2} = \frac{\partial s_1}{\partial \delta_2}\frac{\partial \delta_2}{\partial x_2}$. Note that $\frac{\partial \delta_2}{\partial x_2} = \beta$. Therefore:

$$\frac{\partial s_1}{\partial x_2} = -\beta\ s_1\ s_2$$

—

The profit function of firm $j \in \{0, 1\}$ is $\pi_j = p_j\ q_j - \ c_j\ q_j - FC(x_j)$, where: q_j is the quantity sold by firm j (i.e., $q_j = H\ s_j$); c_j is firm $j's$ marginal cost, that is assumed constant, i.e., linear cost function; and $FC(x_j)$ is a fixed cost that depends on the level of quality of the firm.

Q1.3. (10 points) Suppose that firms take their qualities x_1 and x_2 as given and compete in prices ala Bertrand.
(a) Obtain the equation that describes the marginal condition of profit maximization of firm 1 in this Bertrand game. Write this equation taking into account the specific form of $\frac{\partial s_1}{\partial p_1}$ in the Logit model.
(b) Given this equation, write the expression for the equilibrium price-cost margin $p_1 - c_1$ as a function of s_1 and the demand parameter α.

ANSWER.
(a) The f.o.c. is: $q_1 + p_1\frac{\partial q_1}{\partial p_1} - \ c_1\frac{\partial q_1}{\partial p_1} = 0$. Since $q_1 = H\ s_1$, we have that $\frac{\partial q_1}{\partial p_1} = H\frac{\partial s_1}{\partial p_1}$. And taking into account the expression for $\frac{\partial s_1}{\partial p_1}$ in Q1.2, we have that $\frac{\partial q_1}{\partial p_1} = -H\ \alpha\ s_1(1-s_1) = -\alpha$ $q_1(1 - s_1)$. Therefore, the F.O.C. is:

$$q_1 - \alpha\ q_1\ (p_1 - c_1)\ (1 - s_1) = 0$$

Or equivalently,

$$1 - \alpha\ (p_1 - c_1)\ (1 - s_1) = 0$$

(b) Solving for the price-cost margin, we have that:

$$p_1 - c_1 = \frac{1}{\alpha\ (1 - s_1)}$$

—

Now, suppose that the researcher is not willing to impose the assumption of Bertrand competition and considers a conjectural variations model. Define the conjecture parameter CV_1 as the belief or conjecture that firm 1 has about how firm 2 will change its price when firm 1 changes marginally its price. That is, CV_1 represents the belief or conjecture of firm 1 about $\frac{\partial p_2}{\partial p_1}$. Similarly, CV_2 represents the belief or conjecture of firm 2 about $\frac{\partial p_2}{\partial p_1}$.

Q1.4. (10 points) Suppose that firm 1 has a conjectural variation CV_1.
(a) Obtain the equation that describes the marginal condition of profit maximization of firm 1 under this conjectural variation. Write this equation taking into account the specific form of $\frac{\partial s_1}{\partial p_1}$ in the Logit model. [Hint: Now, we have that: $\frac{dq_1}{dp_1} = \frac{\partial q_1}{\partial p_1} + \frac{\partial q_1}{\partial p_2}\frac{\partial p_2}{\partial p_1}$, where $\frac{\partial q_1}{\partial p_1}$ and $\frac{\partial q_1}{\partial p_2}$ are the expressions you have derived in Q1.2].
(b) Given this equation, write the expression for the equilibrium price-cost margin $p_1 - c_1$ as a function of the market shares s_1 and s_2, and the parameters α and CV_1.

ANSWER.
(a) The f.o.c. is: $q_1 + (p_1 - c_1)\frac{dq_1}{dp_1} = 0$. But now we have that $\frac{dq_1}{dp_1} = \frac{\partial q_1}{\partial p_1} + \frac{\partial q_1}{\partial p_2}\frac{\partial p_2}{\partial p_1} = \frac{\partial q_1}{\partial p_1} + \frac{\partial q_1}{\partial p_2}CV_1$. Therefore, the F.O.C. is:

$$q_1 + (p_1 - c_1)\left[\frac{\partial q_1}{\partial p_1} + \frac{\partial q_1}{\partial p_2}CV_1\right] = 0$$

Now, we particularize this equation for the Logit model. Since $q_1 = H\ s_1$, we have that $\frac{\partial q_1}{\partial p_1} = H\frac{\partial s_1}{\partial p_1}$, and taking into account the expression for $\frac{\partial s_1}{\partial p_1}$ in Q1.2, we have that $\frac{\partial q_1}{\partial p_1} = -H\ \alpha\ s_1(1 - s_1) = -\alpha\ q_1(1 - s_1)$. Similarly, $\frac{\partial q_1}{\partial p_2} = H\frac{\partial s_1}{\partial p_2}$, and taking into account the expression for $\frac{\partial s_1}{\partial p_2}$ in Q1.2, we have that $\frac{\partial q_1}{\partial p_2} = H\ \alpha\ s_1\ s_2 = \alpha\ q_1\ s_2$. Therefore, $\frac{\partial q_1}{\partial p_1} + \frac{\partial q_1}{\partial p_2}CV_1$ is equal to $-\alpha\ q_1(1 - s_1) + \alpha\ q_1\ s_2\ CV_1$. The F.O.C. becomes:

$$q_1 + (p_1 - c_1)\left[-\alpha q_1(1 - s_1) + \alpha q_1 s_2 CV_1\right] = 0$$

Or equivalently,

$$1 + (p_1 - c_1)\left[-\alpha(1 - s_1) + \alpha s_2 CV_1\right] = 0$$

(b) Solving for the price-cost margin, we have that:

$$p_1 - c_1 = \frac{1}{\alpha\ (1 - s_1 - s_2 CV_1)}$$

Q1.5. (15 points) Suppose that the researcher does not know the magnitude of the marginal costs c_1 and c_2, but she knows that the two firms use the same production technology, they use the same type of variable inputs, and they purchase these inputs in the same markets where they are price takers. Under these conditions, the researcher knows that $c_1 = c_2 = c$, though she does not know the magnitude of the marginal cost c.

(a) The marginal conditions for profit maximization in Q1.4(b), for the two firms, together with the condition $c_1 = c_2 = c$, imply that price difference between these two firms, $p_1 - p_2$, is a particular function of their markets shares and their conjectural variations. Derive the equation that represents this condition.

(b) The researcher observes prices $p_1 = \$200$ and $p_2 = \$195$ and market shares $s_1 = 0.5$ and $s_2 = 0.2$. Firm 1 has both a larger price and a larger market share because its product has better quality, i.e., $x_1 > x_2$. The researcher has estimated the demand system and knows that $\alpha = 0.01$. Plug in these data into the equation in Q1.5(a) to obtain a condition that the parameters CV_1 and CV_2 should satisfy in this market.

(c) Using the equation in Q1.5(b), show that the hypothesis of Nash-Bertrand competition (that requires $CV_1 = CV_2 = 0$) implies a prediction about the price difference $p_1 - p_2$ that is substantially larger than the price difference that we observe in the data.

(d) Using the equation in Q1.5(b), show that the hypothesis of Collusion (that requires $CV_1 = CV_2 = 1$) implies a prediction about the price difference $p_1 - p_2$ that is much closer to the price difference that we observe in the data.

ANSWER.

(a) The marginal conditions for firms 1 and 2 are $p_1 - c = \frac{1}{\alpha\ (1 - s_1 - s_2 CV_1)}$ and $p_2 - c = \frac{1}{\alpha\ (1 - s_2 - s_1 CV_1)}$, respectively. The difference between these two equations implies:

$$p_1 - p_2 = \frac{1}{\alpha\ (1 - s_1 - s_2 CV_1)} - \frac{1}{\alpha\ (1 - s_2 - s_1 CV_2)}$$

This is the condition we are looking for.

(b) Plugging our data on prices and market share into the previous equation, we have:

$$\$200 - \$195 = \frac{100}{1 - 0.5 - 0.2\ CV_1} - \frac{100}{1 - 0.2 - 0.5\ CV_2}$$

Or equivalently,

$$\$5 = \frac{100}{0.5 - 0.2\ CV_1} - \frac{100}{0.8 - 0.5\ CV_2}$$

(c) The hypothesis of Nash-Bertrand competition, $CV_1 = CV_2 = 0$, implies that the right hand side of the equation in Q5(b) is:

$$\frac{100}{0.5} - \frac{100}{0.8} = 200 - 125 = \$75$$

That is, the assumption of Nash-Bertrand competition implies a price difference of \$75. However, the actual price difference in the data is only \$5. The assumption of Nash-Bertrand competition over-estimates the observed price difference by 1400%.

(d) The hypothesis of Collusion, $CV_1 = CV_2 = 1$, implies that the right hand side of the equation in Q5(b) is:

$$\frac{100}{0.5 - 0.2} - \frac{100}{0.8 - 0.5} = \$0$$

That is, Collusion implies a price difference of \$0, which is closer to the actual price difference of \$5 that we observe in the data.

QUESTION 2. [50 points]. To answer the questions in this part of the problem set you need to use the dataset verboven_cars.dta Use this dataset to implement the estimations describe below. Please, provide the STATA code that you use to obtain the results. For all the models that you estimate below, impose the following conditions:

- For market size (number of consumers), use Population/4, i.e., pop/4

- Use prices measured in euros (eurpr).

- For the product characteristics in the demand system, include the characteristics: hp, li, wi, cy, le, and he.

- Include also as explanatory variables the market characteristics: ln(pop) and log(gdp).

- In all the OLS estimations include fixed effects for market (ma), year (ye), and brand (brd).

- Include the price in logarithms, i.e., ln(eurpr).

- Allow the coefficient for log-price to be different for different markets (countries). That is, include as explanatory variables the log price, but also the log price interacting (multiplying) each of the market (country) dummies except one country dummy (say the dummy for Germany) that you use as a benchmark.

Q2.1. (15 points)

(a) Obtain the OLS-Fixed effects estimator of the Standard logit model. Interpret the results.

(b) Test the null hypothesis that all countries have the same price coefficient.

(c) Based on the estimated model, obtain the average price elasticity of demand for each country evaluated at the mean values of prices and market shares for that country.

ANSWER. The complete code is provided at the end of this document. Here I include some parts of the code. First, we read the dataset and construct some new variables.

```
use "C:\verboven_cars.dta", clear
gen logq = ln(qu)
gen logp = ln(eurpr)
gen logpop = ln(pop)
gen loggdp = ln(ngdp)
gen msize = pop/4
```

We need to construct the variable with the market shares (s_{jmt}), the market share of the outside alternative (s_{0mt}), and the log-odds ratio ($\ln(s_{jmt}/s_{0mt})$).

```
//construct market share s_j
gen share = qu/msize
//construct outside good's market share s_0
egen sum_share = sum(share), by(ma ye)
gen share0 = 1 - sum_share
//generate log odd ratio
gen lsj_ls0 = ln(share/share0)
```

And we need to generate the dummy variables for country (market) and the product of each of these dummies with log-price.

```
//generate country dummies
tab ma, gen(countrydum_)
gen dumBel_logp = countrydum_1 * logp
gen dumFra_logp = countrydum_2 * logp
gen dumGer_logp = countrydum_3 * logp
gen dumIta_logp = countrydum_4 * logp
```

```
gen dumUK_logp = countrydum_5 * logp
```

(a) Obtain the OLS-Fixed effects estimator of the Standard logit model. Interpret the results. The code is (including logp and omitting Germany * logp)

```
// With logp and Germany as excluding dummy * logp
reghdfe lsj_ls0 logp dumBel_logp dumFra_logp dumIta_logp dumUK_logp hp li wi
cy le he logpop loggdp, vce(robust) a(ma ye brd)
```

This is the table with the estimation results.

```
. // With logp and Germany as excluding dummy * logp
. reghdfe lsj_ls0 logp dumBel_logp dumFra_logp dumIta_logp dumUK_logp hp li wi cy le he logpop loggdp
(converged in 8 iterations)

HDFE Linear regression                         Number of obs   =     11,549
Absorbing 3 HDFE groups                        F(  13,  11463) =     243.25
Statistics robust to heteroskedasticity        Prob > F        =     0.0000
                                               R-squared       =     0.4099
                                               Adj R-squared   =     0.4056
                                               Within R-sq.    =     0.2310
                                               Root MSE        =     1.1578
```

lsj_ls0	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
logp	-.6834595	.1202965	-5.68	0.000	-.9192613	-.4476577
dumBel_logp	-.3969749	.0573774	-6.92	0.000	-.5094444	-.2845054
dumFra_logp	-.7097122	.0636482	-11.15	0.000	-.8344736	-.5849507
dumIta_logp	-.8742421	.0857869	-10.19	0.000	-1.042399	-.7060851
dumUK_logp	-.5035495	.0713707	-7.06	0.000	-.6434483	-.3636508
hp	-.013324	.0017628	-7.56	0.000	-.0167794	-.0098685
li	-.0408816	.0136327	-3.00	0.003	-.067604	-.0141592
wi	.0648839	.0033021	19.65	0.000	.0584111	.0713567
cy	-.0007224	.0000852	-8.48	0.000	-.0008893	-.0005554
le	-.0001112	.0007775	-0.14	0.886	-.0016352	.0014128
he	-.0173409	.0029955	-5.79	0.000	-.0232125	-.0114693
logpop	-.7868595	.2654413	-2.96	0.003	-1.30717	-.2665493
loggdp	.8262299	.0877512	9.42	0.000	.6542226	.9982372

Absorbed degrees of freedom:

Absorbed FE	Num. Coefs. =	Categories -	Redundant
ma	5	5	0
ye	29	30	1
brd	39	40	1 ?

? = number of redundant parameters may be higher

Alternatively, but with identical results, we can implement this estimator without including logp and including all the country dummies interacted with logp. The code is:

```
. // without logp and including all country dummies * logp
. reghdfe lsj_ls0 dumGer_logp dumBel_logp dumFra_logp dumIta_logp dumUK_logp hp li wi cy le he logpop
(converged in 8 iterations)

HDFE Linear regression                            Number of obs   =     11,549
Absorbing 3 HDFE groups                           F(  13,  11463) =     243.25
Statistics robust to heteroskedasticity           Prob > F        =     0.0000
                                                  R-squared       =     0.4099
                                                  Adj R-squared   =     0.4056
                                                  Within R-sq.    =     0.2310
                                                  Root MSE        =     1.1578

------------------------------------------------------------------------------
             |               Robust
     lsj_ls0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 dumGer_logp |  -.6834595   .1202965    -5.68   0.000    -.9192613   -.4476577
 dumBel_logp |  -1.080434   .1088015    -9.93   0.000    -1.293704   -.8671648
 dumFra_logp |  -1.393172   .1140136   -12.22   0.000    -1.616658   -1.169686
 dumIta_logp |  -1.557702   .1150547   -13.54   0.000    -1.783228   -1.332175
  dumUK_logp |  -1.187009   .1075541   -11.04   0.000    -1.397833   -.9761846
          hp |   -.013324   .0017628    -7.56   0.000    -.0167794   -.0098685
          li |  -.0408816   .0136327    -3.00   0.003     -.067604   -.0141592
          wi |   .0648839   .0033021    19.65   0.000     .0584111    .0713567
          cy |  -.0007224   .0000852    -8.48   0.000    -.0008893   -.0005554
          le |  -.0001112   .0007775    -0.14   0.886    -.0016352    .0014128
          he |  -.0173409   .0029955    -5.79   0.000    -.0232125   -.0114693
      logpop |  -.7868595   .2654413    -2.96   0.003     -1.30717   -.2665493
      loggdp |   .8262299   .0877512     9.42   0.000     .6542226    .9982372
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Num. Coefs.  =   Categories  -   Redundant  |
-------------+---------------------------------------|
          ma |           5               5              0    |
          ye |          29              30              1    |
         brd |          39              40              1 ?  |
-----------------------------------------------------+
? = number of redundant parameters may be higher
```

Interpretation:

- In the estimation without including logp and including all the country dummies interacted with logp, the parameter estimates are estimates of $\alpha_{Germany}$, $\alpha_{Belgium}$, α_{France}, α_{Italy}, and α_{UK}. In the estimation including logp and omitting Germany*logp, the parameter estimate associated to regressor logp is $\alpha_{Germany}$, and the parameter estimates for the interactions of country dummies with logp are $(\alpha_{Belgium} - \alpha_{Germany})$, $(\alpha_{France} - \alpha_{Germany})$, $(\alpha_{Italy} - \alpha_{Germany})$, and $(\alpha_{UK} - \alpha_{Germany})$.

- The parameters α that measure price sensitivity are all significantly smaller than zero as we expect with a downward sloping demand curve. There are differences between countries in these parameters. Below we present a formal test of the null hypothesis that these parameters are the same. But it is already clear that the parameter α for Germany is substantially smaller in absolute value than the corresponding parameters for other countries. Italyis the market with the largest price sensitivity.

- The parameter estimate for the characteristic "width" (wi) is positive and statistically significant. However, the parameter estimates of the other product characteristics are all negative and most of them significant. This result is not reasonable, especially for characteristics such as cylinder volume (cy) and horsepower (hp) which are measures of quality. There are several possible interpretation for this implausible estimates of these parameters. Let me include here two possible explanations. First, consumers may have heterogeneous taste for these characteristics. Some consumers like then and others dislike them. We are estimating just the average taste. Though this is possible, it is hard to believe that the average consumer valuation of cylinder volume (cy) or horsepower (hp) is negative. A second and more plausible explanation is that some of these observable characteristics are negatively correlated with the unobserved quality of the product, ξ_j. The estimated parameters are capturing both the direct (ceteris paribus) effect of these characteristics on demand, but also the indirect (non causal) effect because their correlation with the error term.

- The effect of loggdp is positive on the demand of cars. Markets and time periods with more loggdp have a larger demand for all the products. This positive income effect makes economic sense.

- Instead, logpop has a negative effect on the demand for cars. This could be interpreted as a causal effect: markets with more population tend to have consumers with a lower taste for cars. But it can be also interpreted as a correction for our measure of market size, pop/4.

(b) Test the null hypothesis that all countries have the same price coefficient. The implementation of the test depends on how we have implemented the estimation: including logp and omitting one country dummy; or omitting logp and including all the country dummies.

If we include logp and omit one Germany*logp, this is the code for the test.

```
// Test of null hypothesis same alphas across countries
test dumBel_logp = dumFra_logp = dumIta_logp = dumUK_logp = 0
```

And this is the result or output from STATA:

```
. test dumBel_logp = dumFra_logp = dumIta_logp = dumUK_logp = 0

 ( 1)  dumBel_logp - dumFra_logp = 0
 ( 2)  dumBel_logp - dumIta_logp = 0
 ( 3)  dumBel_logp - dumUK_logp = 0
 ( 4)  dumBel_logp = 0

       F(  4, 11463) =    37.08
            Prob > F =    0.0000
```

If we do not include logp and include all the country dummies interacted with logp, this is the code for the test.

```
// Test of null hypothesis same alphas across countries
test dumGer_logp = dumBel_logp = dumFra_logp = dumIta_logp = dumUK_logp
```

And this is the result or output from STATA:

```
. test dumGer_logp = dumBel_logp = dumFra_logp = dumIta_logp = dumUK_logp

 ( 1)  dumGer_logp - dumBel_logp = 0
 ( 2)  dumGer_logp - dumFra_logp = 0
 ( 3)  dumGer_logp - dumIta_logp = 0
 ( 4)  dumGer_logp - dumUK_logp = 0

       F(  4, 11463) =    37.08
            Prob > F =    0.0000
```

Interpretation. The two approaches give us exactly the same result for the F-test. This should be necessarily the case by construction. The p-value of this null hypothesis is practically zero. Therefore, under practically any significance level, we reject the null hypothesis that the α parameters are the same across countries.

(c) Based on the estimated model, obtain the average price elasticity of demand for each country evaluated at the mean values of prices and market shares for that country.

ANSWER. The demand elasticity for product-country-year (j, m, t) is:

$$\eta_{jmt} = \frac{\partial s_{jmt}}{\partial p_{jmt}} \frac{p_{jmt}}{s_{jmt}}$$

Now, we take into account the form of $\partial s_{jmt}/\partial p_{jmt}$ in our Logit model.

$$\frac{\partial s_{jmt}}{\partial p_{jmt}} = \frac{\partial s_{jmt}}{\partial \delta_{jmt}} \frac{\partial \delta_{jmt}}{\partial p_{jmt}}$$

In the Logit model, $\dfrac{\partial s_{jmt}}{\partial \delta_{jmt}} = s_{jmt}(1 - s_{jmt})$. And given that $\delta_{jmt} = -\alpha_m \ln(p_{jmt}) + ...$, we have that $\dfrac{\partial \delta_{jmt}}{\partial p_{jmt}} = \dfrac{-\alpha_m}{p_{jmt}}$. Therefore,

$$\eta_{jmt} = \left[s_{jmt}(1 - s_{jmt}) \frac{-\alpha_m}{p_{jmt}} \right] \frac{p_{jmt}}{s_{jmt}} = -\alpha_m \ (1 - s_{jmt})$$

As for the code, we start by creating a variable that contains the information about the countries' alpha parameters. This is the code.

```
// Creating variable with alpha parameters
gen alpha = .
replace alpha = -_b[dumBel_logp] if ma==1
replace alpha = -_b[dumFra_logp] if ma==2
```

```
replace alpha = -_b[dumGer_logp] if ma==3
replace alpha = -_b[dumIta_logp] if ma==4
replace alpha = -_b[dumUK_logp] if ma==5
```

Variable **alpha** contains the estimated parameter $\alpha_{Belgium}$ for all the observations that belong to Belgium, the estimated parameter α_{France} for all the observations that belong to France, and so on for each country. Next, we create a variable that contains the values of the elasticities $\eta_{jmt} = -\alpha_m (1 - s_{jmt})$ for every product-country-year observation in the data. This is the code.

```
// Creating variable with the elasticities
gen elasticity = -alpha * (1-share)
```

To obtain the mean value of this variable by country, we could do it in different ways. For instance,

```
// Mean elasticities by country
sum elasticity if ma==1
sum elasticity if ma==2
sum elasticity if ma==3
sum elasticity if ma==4
sum elasticity if ma==5
```

Or in a more compact form

```
// Mean elasticities by country
tab ma, sum(leasticity)
```

These are the results

```
. tab ma, sum(elasticity)
```

market (=second dimension of panel)	Summary of elasticity Mean	Std. Dev.	Freq.
Belgium	-1.0787248	.00193667	2,673
France	-1.3908184	.00384684	2,265
Germany	-.68220874	.00203348	2,283
Italy	-1.5550288	.0050353	2,027
UK	-1.1853625	.00248949	2,301
Total	-1.1663937	.29085748	11,549

Interpretation of the results. In principle, the differences in the mean demand elasticities across countries come from two sources: (1) differences between countries in the α parameters; and (2) differences between countries in the mean market shares. However, mean market shares are very similar across countries. Therefore, mean demand elasticities are practically identical to the α parameters.

Q2.2. (20 points) Consider the equilibrium condition (first order conditions of profit maximization) under the assumption that each product is produced by only one firm.

(a) Write the equation for this equilibrium condition. Write this equilibrium condition as an equation for the Lerner Index, $\frac{p_j - MC_j}{p_j}$.

ANSWER. The profit function of firm j is $\pi_j = p_j \; q_j - C_j(q_j)$. The first order condition with respect to price is:

$$q_j + p_j \; \frac{\partial q_j}{\partial p_j} - MC_j \; \frac{\partial q_j}{\partial p_j} = 0$$

Solving for $p_j - MC_j$, we have that:

$$p_j - MC_j = \frac{-q_j}{\partial q_j / \partial p_j}$$

And the Lerner index is:

$$\frac{p_j - MC_j}{p_j} = \frac{-q_j}{p_j \; [\partial q_j / \partial p_j]}$$

Now, we take into account the form of $\partial q_j / \partial p_j$ in our Logit model.

$$\frac{\partial q_j}{\partial p_j} = H \; \frac{\partial s_j}{\partial p_j} = H \; \frac{\partial s_j}{\partial \delta_j} \; \frac{\partial \delta_j}{\partial p_j}$$

In the Logit model, $\frac{\partial s_j}{\partial \delta_j} = s_j(1 - s_j)$. And given that $\delta_j = -\alpha \; \ln(p_j) + ...$, we have that $\frac{\partial \delta_j}{\partial p_j} = \frac{-\alpha}{p_j}$. Therefore,

$$\frac{\partial q_j}{\partial p_j} = q_j \; (1 - s_j) \; \frac{-\alpha}{p_j}$$

And plugging this expression in the Lerner Index, we have:

$$\frac{p_j - MC_j}{p_j} = \frac{-q_j}{p_j \; [q_j \; (1 - s_j) \; \frac{-\alpha}{p_j}]} = \frac{1}{\alpha(1 - s_j)}$$

(b) Using the previous equation in Q2.2(a) and the estimated demand in Q2.1, calculate the Lerner index for every car-market-year observation in the data.

ANSWER. Note that, according to the estimated models, different countries have different value for the parameter α. Furthermore, every product-market-year observation (j, m, t) has a different market share, s_{jmt}. Therefore, the Lerner index variable is:

$$\frac{p_{jmt} - MC_{jmt}}{p_{jmt}} = \frac{1}{\alpha_m \left(1 - s_{jmt}\right)}$$

Note that, in this model, the Lerner index is exactly equal to the inverse of absolute value of the demand elasticity:

$$\frac{p_{jmt} - MC_{jmt}}{p_{jmt}} = \frac{1}{|\eta_{jmt}|}$$

This is the code to obtain the Lerner index variable.

```
// Lerner indexes
gen lerner = 1/abs(elasticity)
```

(c) Report the mean values of the Lerner Index for each of the counties/markets. Comment the results.

Again there are different way of computing the mean values of the Lerner index by country. We can obtain them as follows:

```
// Mean lerner indexes by country
// Note that Belgium (ma==1), France (ma==2),
// Germany (ma==3), Italy (ma==4), and UK (ma==5)
sum lerner if ma==1
sum lerner if ma==2
sum lerner if ma==3
sum lerner if ma==4
sum lerner if ma==5
```

Or in a more compact form:

```
// Mean lerner indexes by country
tab ma, sum(lerner)
```

These are the results:

```
. tab ma, sum(lerner)

     market |
   (=second |
  dimension |        Summary of lerner
  of panel) |        Mean   Std. Dev.       Freq.
------------+------------------------------------
    Belgium |   .92702347   .00167203       2,673
     France |   .71900666   .00200461       2,265
    Germany |   1.4658401   .00441103       2,283
      Italy |   .64308173   .00211327       2,027
         UK |   .84362754   .00178202       2,301
------------+------------------------------------
      Total |    .9262889   .28535342      11,549

.
```

The estimated demand model implies Lerner indexes that are very high, especially for Germany where the index implies a markup of 146%. This is because the elasticities are small, especially in Germany.

(d) Report the mean values of the Lerner Index for each of the top five car manufacturers (i.e., the five car manufacturers with largest total aggregate sales over these markets and sample period). Comment the results.

First, we need to figure out which are the five car manufacturers with the largest aggregate sales during the years of the sample. We calculate aggregate sales by firm.

```
// Aggregate sales by car manufacturer
egen totq_frm = sum(qu), by(frm)
```

We can figure out the top-5 manufacturers in different way. A very simple way is to sort the data by the variable `totq_frm` but in inverse order (using the command `gsort -`), then list the variables `frm` and `totq_frm` avoiding repetition of firms, and finally take the first 5 firms in this list. This is the code:

```
// Top 5 manufacturers
gsort - totq_frm
list frm totq_frm if frm~=frm[_n-1]
```

And we get (sorted in inverse order)

```
. gsort - totq_frm

. list frm totq_frm if frm~=frm[_n-1]
```

	frm	totq_frm
1.	Fiat	3.70e+07
1692.	VW	3.42e+07
2972.	Peugeot	2.80e+07
4342.	Renault	2.75e+07
5232.	Ford	2.68e+07

To obtain the mean Lerner indexes for these manufacturers, we can use the command `summarize` with the "if" option. Or we can use the following more compact way.

```
// Mean lerner indexes for the top-5 firms
// Note that Fiat (frm==4), VW (frm==26),
// Peugeot (frm==16), Renault (frm==18), and Ford (frm==5)

tab frm if (frm==4) | (frm==26) | (frm==16) | (frm==18) | (frm==5), sum(lerner)
```

These are the results.

```
. tab frm if (frm==4) | (frm==26) | (frm==16) | (frm==18) | (frm==5), sum(lerner)
```

firm code	Summary of lerner Mean	Std. Dev.	Freq.
Fiat	.90881441	.29116801	1,691
Ford	.92270469	.29024646	646
Peugeot	.91606159	.2855043	1,370
Renault	.91468186	.28602844	890
VW	.92550699	.2955836	1,280
Total	.9165548	.28992397	5,877

Interpretation. The differences in the mean Lerner indexes of these firms come from two sources: (1) firms can have different presence in different markets/counties, and countries have different values of α that generates important differences in the lerner indexes; and (2) firms have different market shares, and this affects their Lerner indexes. Given that these top-5 firms have similar market shares, most of the difference in their Lerner differences comes from their different presence across countries, and in particular their presence in Germany that is the country with smaller demand elasticity. Thus, we see that the German company

VW is the firm with the larger Lerner index because it is also the top-5 firm with stronger presence in Germany.

Q2.3. (15 points)

(a) Using the equilibrium condition and the estimated demand, obtain an estimate of the marginal cost for every car-market-year observation in the data.

ANSWER. The equilibrium condition of the model implies that:

$$p_{jmt} - MC_{jmt} = \frac{p_{jmt}}{\alpha_m \left(1 - s_{jmt}\right)}$$

And solving for the marginal cost:

$$MC_{jmt} = p_{jmt} - \frac{p_{jmt}}{\alpha_m \left(1 - s_{jmt}\right)}$$

By looking at this expression, we can already see that the small estimate of the parameter $\alpha_{Germany}$ (smaller than one) implies that the estimated marginal costs for Germany are negative, i.e., because with $\alpha < 1$, $p/\alpha(1-s) > p$. Of course, this is a very implausible prediction. There are two main aspects where the model could be failing and generating this implausible prediction: (1) ignoring that firms are multi-product; and (2) the logit model is very restrictive. A nested logit that accounts for the multiproduct nature of these firms does not generate these negative estimates for the marginal costs.

This is the code for calculating the marginal cost.

```
// Estimating the values of Marginal Costs
gen mc = eurpr - eurpr/(alpha * (1-share))
```

(b) Run an OLS-Fixed effects regression where the dependent variable is the estimated value of the marginal cost, and the explanatory variables (regressors) are the product characteristics hp, li, wi, cy, le, and he. Interpret the results.

ANSWER. This is the code for the estimation of the regression for the marginal cost function.

```
// Estimating marginal cost function
reghdfe mc hp li wi cy le he, vce(robust) a(ma ye brd)
```

And this is the output from this regression:

```
. // Estimating marginal cost function
. reghdfe mc hp li wi cy le he, vce(robust) a(ma ye brd)
(converged in 8 iterations)

HDFE Linear regression                          Number of obs   =     11,549
Absorbing 3 HDFE groups                         F(      6,  11470) =   27.08
Statistics robust to heteroskedasticity         Prob > F        =     0.0000
                                                R-squared       =     0.7028
                                                Adj R-squared   =     0.7008
                                                Within R-sq.    =     0.0201
                                                Root MSE        =  1560.4886

------------------------------------------------------------------------------
             |               Robust
          mc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          hp |   15.13636   3.143366     4.82   0.000     8.974829    21.2979
          li |    6.80395   19.81002     0.34   0.731    -32.02708   45.63498
          wi |    1.46403   3.918408     0.37   0.709    -6.216719   9.144779
          cy |  -.4246081    .110773    -3.83   0.000     -.641742  -.2074742
          le |    2.03699   1.020322     2.00   0.046     .0369856   4.036995
          he |   2.574545   3.428862     0.75   0.453     -4.14661     9.2957
------------------------------------------------------------------------------
```

Interpretation. The product characteristic with the strongest effect on the marginal cost is horse power (`hp`). Since marginal cost (the dependent variable) is measured in Euros and horse power is measured in kW, we have that an increase in 1kW of horse power implies a 15 euros increase in the marginal cost of a car. The empirical distribution of horse power (`sum hp, detail`) has 10% percentile of 31 kW, a median of 55 kW, and a 90% percentile of 90 kW. Therefore, ceteris paribus, the difference in marginal cost between a car at the 90-percentile and a car at the 10-percentile of horse power is equal to 15*(90-31) = 885 euros. Length (`le`) has also a significant and positive effect on marginal cost. Since length (`le`) is measured in centimeters, we have that a 1 meter increase in length (which is approximately the difference between the 90-percentile and the 10-percentile in the empirical distribution of length) implies an increase in marginal cost of 203 euros. The characteristic cylinder (`cy`) has a negative effect on marginal cost, which is not a plausible prediction. A possible explanation is that this characteristic is negatively correlated with the error term, i.e., with omitted / unobserved product characteristics affecting marginal cost. The other characteristics have a positive effect but not statistically significant to zero.

We can obtain the estimates of country fixed effects, year fixed effects, and brand fixed effects, by using the option `savefe` in the estimation command (`reghdfe mc hp li wi cy le he, vce(robust) a(ma ye brd, savefe)`. Then, we can tabulate the values of these variables:

```
tab ma, sum(maFE)
tab ye, sum(yeFE)
tab brd, sum(brdFE)
```

This shows that there is substantial heterogeneity in marginal costs (keeping constant the observable characteristics) across countries, and especially over brands. There is also a clear increasing time trend in the evolution of marginal costs over time.

STATA CODE

```
// ----------------------------------------------
// ECO310 - EMPIRICAL INDUSTRIAL ORGANIZATION
//
// PROBLEM SET 2
//    using Verboven's data on automobiles in Europe
// Due of December 6th, 2018
//
// by Victor Aguirregabiria
//
//-----------------------------------------------
//-------------------------------------------------
//    1.  Reading dataset & generating new variables
//-------------------------------------------------
use "C:\Dropbox\problem_set_02\verboven_cars.dta", clear
gen logq = ln(qu)
gen logp = ln(eurpr)
gen logpop = ln(pop)
gen loggdp = ln(ngdp)
gen msize = pop/4
//construct market share s_j
gen share = qu/msize
//construct outside good's market share s_0
egen sum_share = sum(share), by(ma ye)
gen share0 = 1 - sum_share
//generate log odd ratio
gen lsj_ls0 = ln(share/share0)
//generate country dummies
tab ma, gen(countrydum_)
// countrydum_1 ma==Belgium
// countrydum_2 ma==France
// countrydum_3 ma==Germany
// countrydum_4 ma==Italy
// countrydum_5 ma==UK
gen dumBel_logp = countrydum_1 * logp
```

```
gen dumFra_logp = countrydum_2 * logp
gen dumGer_logp = countrydum_3 * logp
gen dumIta_logp = countrydum_4 * logp
gen dumUK_logp = countrydum_5 * logp
//---------------------------------------------------------------
//    3.  Estimation of the Logit model and test of null hypothesis
//        Questions 2.1(a) and 2.1(b)
//---------------------------------------------------------------
// With logp and Germany as excluding dummy * logp
reghdfe lsj_ls0 logp dumBel_logp dumFra_logp dumIta_logp dumUK_logp hp li wi
cy le he logpop loggdp, vce(robust) a(ma ye brd)
// Test of null hypothesis same alphas across countries
test dumBel_logp = dumFra_logp = dumIta_logp = dumUK_logp = 0
// Without logp and including all country dummies* logp
reghdfe lsj_ls0 dumGer_logp dumBel_logp dumFra_logp dumIta_logp dumUK_logp
hp li wi cy le he logpop loggdp, vce(robust) a(ma ye brd)
// Test of null hypothesis same alphas across countries
test dumGer_logp = dumBel_logp = dumFra_logp = dumIta_logp = dumUK_logp
//---------------------------------------------------------------
//    4.  Calculating demand elasticities
//        Elasticity = -alpha * (1- share)
//        Question 2.1(c)
//---------------------------------------------------------------
tab ma, sum(ma)
// Creating variable with alpha parameters
gen alpha = .
replace alpha = -_b[dumBel_logp] if ma==1
replace alpha = -_b[dumFra_logp] if ma==2
replace alpha = -_b[dumGer_logp] if ma==3
replace alpha = -_b[dumIta_logp] if ma==4
replace alpha = -_b[dumUK_logp] if ma==5

// Creating variable with the elasticities
gen elasticity = -alpha * (1-share)
// Mean elasticities by country
sum elasticity if ma==1
```

```
sum elasticity if ma==2
sum elasticity if ma==3
sum elasticity if ma==4
sum elasticity if ma==5
// ...  or in a more compact form
tab ma, sum(elasticity)
//------------------------------------------------------------------
//    4.  Lerner indexes and their means by country and car manufacturer
//        Question 2.2(b), 2.2(c), and 2.2(d)
//------------------------------------------------------------------
gen lerner = 1/abs(elasticity)
// Mean lerner indexes by country
sum lerner if ma==1
sum lerner if ma==2
sum lerner if ma==3
sum lerner if ma==4
sum lerner if ma==5
// ...  or in a more compact form
tab ma, sum(lerner)
// Aggregate sales by car manufacturer
egen totq_frm = sum(qu), by(frm)
// Top 5 manufacturers
gsort - totq_frm
list frm totq_frm if frm~=frm[_n-1]
// Mean lerner indexes for the top-5 firms
// Note that Fiat (frm==4), VW (frm==26),
// Peugeot (frm==16), Renault (frm==18), and Ford (frm==5)
tab frm, sum(frm)

tab frm if (frm==4) | (frm==26) | (frm==16) | (frm==18) | (frm==5), sum(lerner)
//------------------------------------------------------------------
//    5.  Estimating Marginal Costs and MArginal Cost function
//        Question 2.3(a) and 2.3(b)
//------------------------------------------------------------------
// Estimating the values of marginal costs
gen mc = eurpr - eurpr/(alpha * (1-share))
```

```
// Estimating marginal cost function
reghdfe mc hp li wi cy le he, vce(robust) a(ma ye brd)
// Estimating marginal cost function
reghdfe mc hp li wi cy le he, vce(robust) a(ma ye brd, savefe)
// Estimating marginal cost function without Germany
reghdfe mc hp li wi cy le he if ma~=3, vce(robust) a(ma ye brd, savefe)
rename __hdfe1__ maFE
rename __hdfe2__ yeFE
rename __hdfe3__ brdFE
tab ma, sum(maFE)
tab ye, sum(yeFE)
tab brd, sum(brdFE)
```