Team Project 1
Pricing Analytics (MKT 440/31)
by Prof. Takeaki Sunada

Submitted by
Qianyu Dong (Qianyu.Dong@simon.rochester.com)
Yaqing Sun (Yaqing.Sun@simon.rochester.com)
Yilan Zhang (Yilan.Zhang@simon.rochester.com)

Jan. 29, 2020

4 Control variables

4.1 Interpreting a log-log regression

```
Call:
lm(formula = log(qu) ~ log(eurpr), data = data)

Coefficients:
(Intercept)   log(eurpr)
     11.322       -0.296
```

The **intercept** means that when log(eurpr) = 0, namely when price = 1 euro, demand of vehicle is e^11.322.
The **coefficient** indicates that the price elasticity of vehicle is -0.296, which is inelastic. The number doesn't look reasonable, since common sense tells us that vehicle as a type of luxury goods, should have a high elasticity of demand, at least >1 (absolute value).

4.2 Adding control variables

The log(qu)~log(eurpr) model seems unreasonable, because it failed to take some intrinsic discrepancies between different years, markets and car models into account. To be more specific, the pricing strategies of automobile companies definitely differentiate from markets to markets and customers' demand for cars also differ among markets; cars' price also vary from different years and types. Hence, we first incorporate **"ye", "co" and "ma" as fixed effects** into the regression model.

```
          cy        hp        we        le
cy 1.0000000 0.9055687 0.8905680 0.8249275
hp 0.9055687 1.0000000 0.8966029 0.7924450
we 0.8905680 0.8966029 1.0000000 0.8862262
le 0.8249275 0.7924450 0.8862262 1.0000000
```

As for the control variable part, we check the correlation between all attributes and price, all attributes and quantity, respectively. The result shows that "hp"(horsepower), "engdp"(nominal GDP in Euro), "cy"(cylinder volume or displacement), "we"(weight), "le"(length) have strongest correlations with both price and quantity. However, since "hp", "cy", "we" and "le" are all car attributes and may have some correlation with each other, we run "cor" to check the colinearity. Based on the result, the four attributes all have strong correlations with each other. To avoid bias of our result, we pick "hp" as representative for the following two reasons: 1) "hp" has the strongest correlation with both price and quantity compared with other three attributes thus should definitely be included as an omitted variable 2) "hp" has the strongest correlation with other three attributes which means it can serve as the best representative. From the perspective of intuition, "hp" and "engdp" are also reasonable to be added as control variables. Firstly, as the main measure of vehicle performance, horsepower is highly valued by consumers, thus having great influence on both price and quantity sold. "Engdp" is a major measurement of economic level in a specific area during a specific time period, which is both related to overall sales volume of products (especially luxury goods) and price setting. So we finally decided to add these two attributes as control variables to our regression model in this part.

```
Call:
   felm(formula = log(qu) ~ log(eurpr) + hp + log(engdp) | factor(ma) +      factor(co) + factor(ye), data = cardata)

Residuals:
    Min      1Q  Median      3Q     Max
-4.4797 -0.6579 -0.0173  0.6413  4.1219

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
log(eurpr) -1.604695   0.113427 -14.147 < 0.0000000000000002 ***
hp         -0.014669   0.001812  -8.097  0.00000000000000062 ***
log(engdp)  1.607031   0.136921  11.737 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

As the picture above shown, the coefficient of log price changes from -0.296 to -1.605. Given that automobile should be categorized into luxury goods, the elasticity of -1.605, which is smaller than -1, is apparently more reasonable than the log(qu)~log(eurpr) model.

5 Instrumental variables

```
Call:
   felm(formula = log(qu) ~ hp + log(engdp) | factor(co) + factor(ma) |      (log(eurpr) ~ unit_value), data = cardata)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6825 -0.6682 -0.0151  0.6525  4.3052

Coefficients:
                  Estimate Std. Error t value    Pr(>|t|)
hp               -0.006055   0.003993  -1.517       0.129
log(engdp)        2.490977   0.468205   5.320 0.000000106 ***
`log(eurpr)(fit)` -2.967028   0.570239  -5.203 0.000000199 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We chose unit value of iron ore as IV, which is an important input product of motor vehicle. This IV works, as the regression above shows. It changes the coefficient of price to -2.97. Unit value of iron ore is a good instrumental variable, since it satisfies the two conditions for IV: It is correlated with price, yet independent from demand shifts. Price changes in iron ore are passed on to final products (cars), but they might not affect consumers' demand.
As the result of adding unit value of iron ore, we remove factor(year) from the fixed effects. Notice that the unit value of iron ore remains the same in a specific year which will invalidate the regression between unit value and price if we keep factor(year).

6 Recovering costs

According to log-log demand model, we have:

$$log(Q) = \beta 1 * log(P) + X$$

So $Q = e^{[\beta 1 * log(P) + X]}$

$$= P^{\beta 1} * e^{X}$$

Profit $\Pi = Q*(P - C)$

$$= P^{(\beta 1+1)} * e^{X} - P^{\beta 1} * e^{X} * C$$

To maximize profit and find the profit maximization price, we take the derivative of Π and let it equals to 0, then get:

$$C = (\beta 1 + 1) / \beta 1 * P$$

According to our demand model in Part 5, β1 = -2.967. So for a car in 1970, market 1 and model code 15 whose price in Euros is €3018.003, the cost should be (-2.967 + 1)/-2.967*3018.003 = 2000.813.

The calculation makes sense, since the cost takes up roughly ⅔ of its selling price, leaving profit margin of about 33%.

7 Cross-elasticities and competitive effects

```
Call:
   felm(formula = log(qu) ~ hp + log(engdp) + log(avgurprrival) |      factor(ma) + factor(co) | (log(eurpr) ~ unit_value), data = cardata_new)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6238 -0.6565 -0.0191  0.6445  4.1866

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
hp                -0.009701   0.003316  -2.925 0.003447 **
log(engdp)         0.754168   0.195472   3.858 0.000115 ***
log(avgurprrival)  1.644551   0.357545   4.600 4.28e-06 ***
`log(eurpr)(fit)` -2.555004   0.484888  -5.269 1.40e-07 ***
```

The coefficient of log(average rival price) means that when the average rival price increase by 1%, the demand for a certain car will increase by 1.645% correspondingly.

Given that European automobile industry is a well-developed market with fierce competition, we expect an coefficient(cross elasticity) at least bigger than 1, which indicates that the change of rivals' price will exert strong pressure to an individual product. Thus, 1.645 just satisfies this criteria and it proves that European region is definitely a market with intense competition.