Pricing Analytics Project 1: Estimating Demand for Cars

Team members:

Moldir Sailike 32928624
Xiaoyan(Nora) Hu 32807851
Xueyuan(Serena) Hu 32909760
Zhian Zhang 32910216
Zhiying Liu 32917031

4. Interpreting a log-log regression:

Answer to question 4-1:
The coefficient is -0.2925, which represents the price elasticity of demand, it means one percent change in the price of a car would result in a 0.2925% decrease in quantity demanded. Since the price elasticity is less than 1, car demand in Europe is inelastic, suggesting that we can raise the price.
The intercept is 11.2904 is the expected value of log(qujmt) when log(eurprjmt) equals to 0, which corresponds to a price of 1 Euro. Therefore, the market size is approximately 195,164,130,028.58 units.

Answer to question 4-2:
The error term ϵjmt is unobservable and correlated with the price variable log(eurprjmt). Without adding other demand shifters, this leads to omitted variable bias and other issues, that's why β1 estimate from the regression is likely not causal.

5. Adding control variables:

Answer to question 5-1:

```
reg3_5= felm( log(qu) ~ log(eurpr) + ac:hp + cy:li + ergdpc + tax + avgurprrival
  + nco | factor(ye):factor(co) + factor(brand):factor(ma) + factor(cla),
  data = cardata)
summary(reg3_5)


Residuals:
    Min      1Q  Median      3Q     Max
-2.6003 -0.2648  0.0000  0.2837  3.1870

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
log(eurpr)   -1.844e+00  9.985e-02 -18.472  < 2e-16 ***
ergdpc       -3.869e-05  3.222e-06 -12.008  < 2e-16 ***
tax           4.636e-01  2.878e-01   1.611 0.107303
avgurprrival  2.351e-04  1.366e-05  17.206  < 2e-16 ***
nco           5.619e-03  1.593e-03   3.527 0.000423 ***
ac:hp        -6.470e-05  1.798e-04  -0.360 0.718929
cy:li         6.475e-06  7.629e-06   0.849 0.396027
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5696 on 6249 degrees of freedom
  (2802 observations deleted due to missingness)
Multiple R-squared(full model): 0.9134   Adjusted R-squared: 0.8804
```

Multiple R-squared(proj model): 0.08975   Adjusted R-squared: -0.2568
F-statistic(full model):27.69 on 2379 and 6249 DF, p-value: < 2.2e-16
F-statistic(proj model): 88.02 on 7 and 6249 DF, p-value: < 2.2e-16
*** Standard errors may be too high due to more than 2 groups and exactDOF=FALSE

Answer to question 5-2:
Fixed effect:
(ye:co): each year-car type combination represents the different car product features in different years, and that is the key measurement of the product features.
(brand:ma): each brand-market combination epresents different car brands' popularity in different markets, which is associated with the brand value or perception.
(cla): car category is represents category-sepcific consumer preferences, as the demand for different types of cars

Control variable:
The variable ergdpc**:** it represents the real GDP per capita in Euro measuring the economy where the car was sold, the price is also related to different economies. Tax is also included because the percentage of VAT is an important variable that influences the after-tax price.
We use avgurprrival and nco to measure the rival's average price and numbers in the same year and market, we believe the competition is an important variable to take into account when considering our price.
We use the interaction term between time to acceleration and horsepower(ac:hp) to represent the car's attribute of power and how fast it could be, which is a key variable influence the demand. We also include the interaction term between cylinder and fuel efficiency(cy:li) to measure the fuel ability as a key attribute for the car. In order to analyze how various car-specific size attributes combine to influence demand rather the individual car attributes.

6. Instrumental variables:

Answer to question 6-1:
```
reg6=felm(formula = log(qu) ~ac:hp +cy:li + ergdpc+avgurprrival+ nco
        | factor(year):factor(co) +
          factor(brand):factor(ma)+ factor(cla)|
          (log(eurpr) ~ tax + we*unit_value_98), data = fulldata)
summary(reg6)
```

```
Call:
  felm(formula = log(qu) ~ ac:hp + cy:li + ergdpc + avgurprrival +      nco | factor(year):factor(co)
+ factor(brand):factor(ma) +      factor(cla) | (log(eurpr) ~ tax + we * unit_value_98), data =
fulldata)

Residuals:
```

```
   Min      1Q  Median      3Q     Max
-2.5929 -0.2624  0.0000  0.2838  3.1894

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
ergdpc           -3.870e-05  3.225e-06 -12.002  < 2e-16 ***
avgurprrival      2.171e-04  4.359e-05   4.982 6.48e-07 ***
nco               6.691e-03  1.482e-03   4.515 6.45e-06 ***
ac:hp            -1.188e-04  2.039e-04  -0.583  0.56004
cy:li             1.892e-06  1.290e-05   0.147  0.88340
`log(eurpr)(fit)` -1.617e+00  5.119e-01  -3.159  0.00159 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5699 on 6250 degrees of freedom
  (2888 observations deleted due to missingness)
Multiple R-squared(full model): 0.9133   Adjusted R-squared: 0.8803
Multiple R-squared(proj model): 0.08873   Adjusted R-squared: -0.258
F-statistic(full model):27.54 on 2378 and 6250 DF, p-value: < 2.2e-16
F-statistic(proj model): 47.36 on 6 and 6250 DF, p-value: < 2.2e-16
F-statistic(endog. vars):9.979 on 1 and 6250 DF, p-value: 0.001591
*** Standard errors may be too high due to more than 2 groups and exactDOF=FALSE
```

After we included the IVs, the absolute value of the coefficient for log(eurpr) dropped from 1.844e+00 to 1.617e+00, but still greater than 1, so we still believe this model is more causal than than the model my college did.

Answer to question 6-2:

We believe that tax and weight*iron's unit value are good instrumental variables. Heavier cars consume more raw materials, so when iron ore prices rise, heavier cars areaffected more than light cars. Furthermore, weight*iron strengthening the correlation with price, making it a strong instrument.

Overall the IV regression shows a less biased estimate compared to the previous regression and is more consistent of the true price elasticity of demand.Moreover, the price elasticity estimates from the IV regression are more reliable since the instrument affects price but does not directly influence demand.

7. Cross-elasticities and competitive effects

Answer to 7-1:

In our regression model, log(avgeurprrival) represents the average price of competitors' cars, and its coefficient β_2 reflects how changes in competitors' prices affect our sales.

From our regression results:

$$\beta_2 = 0.0002171$$

This coefficient is positive, meaning that when competitors' prices increase, our sales slightly increase. This confirms the substitution effect—some consumers may switch from competitors' cars to ours when competitors raise their prices. However, β_2 is very small, indicating that competitors' prices have little impact on our sales.

Answer to 7-2:
In a competitive market, β_2 is expected to be positive ($\beta_2 > 0$), as cars are substitutes—when competitors raise their prices, some consumers switch to our brand (substitution effect). Conversely, a negative β_2 would suggest that an increase in competitor prices leads to an overall decline in demand, rather than brand switching, which is unlikely in this context.
Our estimate, $\beta_2 = 0.0002171$, is positive, which aligns with theoretical expectations. This confirms that higher competitor prices can slightly increase our sales. However, the small magnitude suggests that while substitution occurs, its effect is limited—likely due to strong brand loyalty or intense market competition.

Answer to 7-3:
Since β_2 is positive, there is evidence of some competitive pressure in the market, as higher competitor prices slightly increase demand for our brand. This supports the substitution effect—some consumers switch to our brand when rival prices rise.

However, β_2 is extremely small (0.0002171), suggesting that competitive pressure is weak. This could indicate high brand loyalty, meaning consumers are not highly responsive to competitor price changes. It may also have market saturation, where most customers have already chosen a preferred brand. Further, data limitations could contribute to this result, as rival prices might not fully capture all competitive market dynamics (e.g., marketing expenditures, model releases).

In summary, the market is somewhat competitive, but brand loyalty or differentiation limits price-driven switching.

If we suspect that β_2 is underestimated or biased, we can refine the regression model by including the number of competitors (nco) to account for market competition. Additionally, an Instrumental Variable (IV) approach could help address potential endogeneity in rival prices.

8. Recovering costs

Answer to question 8-1:
As $\beta_1 = -0.2925$, Cost = Price × (1 + β_1) / β_1 and the price is given by eurpr, the results can be summarized as follows:

	eurpr	Estimated_cost
count	11,431	11,431
mean	8,401.96	-20,322.70
std.	5,542.15	13,405.37
min	830.67	-119,925.84
25%	4,435.54	-26,187.36

1) The negative cost estimates indicate that the colleague's $\beta_1 = -0.2925$ leads to incorrect cost calculations.

2) The mean estimated cost is -20,322.70, which is more than twice the mean price (8,401.96), an economically unrealistic result. This suggests that the demand function is misspecified, likely due to omitted variable bias or endogeneity in price determination.

Since the colleague's β_1 is likely biased, an Instrumental Variable (IV) approach is necessary to correct for endogeneity (i.e., price being correlated with unobserved demand factors).

Answer to question 8-2:
From our instrumental variable (IV) regression in Section 6, we obtained $\beta1=-1.617$
According to the formula, cost = $\frac{1-1.617}{-1.617} \approx 0.38$

This means the cost is about 38% of the price. Compared with the result when $\beta 1$ is -0.2925, our result is more reasonable. So, our result validates our improved demand estimate using IV regression.

When we compare these two analyses, the initial OLS estimate of β1 used by our colleagues was biased, likely due to some problems in pricing decisions, for example they may omit variable bias like brand value and marketing influenced both price and demand. However, the IV approach in our case, by isolating exogenous price variation, provides a more reliable demand elasticity estimate, which in turn produces a more credible cost estimate.