Take Home Assignment

Big Data Econometrics

After importing dataset “Gasoline.csv” in Gretl and removing rows which contains NA values we can start to build our model.

First of all, in order to reach the assignment’s second goal (elasticity), we want to use the log of cons (lc), price (lp) and gdp (ly).

The first ADL model was builded using 8 lags on lc,lp,ly, the model was quite good it scores 0.947 on Adjusted R-squared and the p-value of the Godfrey test was 0.134 but we knew that the model could be better, presumably the quantity of lags was too large therefore it contains the “true” “p” and “qs” (there’s not a TRUE number of lags for the model).

After some shrinking and tests the final model, as can be seen in FIG.1, has: 3 lags on lc, 2 lags on lp and 6 lags on ly, plus the three dummy (to prevent dummy trap) for the quarters (it seems that gasoline consumotion grows from April to September, which is ok, it’s hotter and people go out) and the covid features (this one’s coefficent is -0.23 and it is in line with what we saw, people in lockdown means less gasoline consumption).

```
Model 1: OLS, using observations 1993:4-2021:4 (T = 113)
Dependent variable: lc

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       0.630608     0.294025      2.145    0.0345    **
  lc_1        0.682670     0.0829470     8.230    9.83e-013 ***
  lc_2       -0.0520292    0.107372     -0.4846   0.6291
  lc_3        0.223734     0.0799096     2.800    0.0062    ***
  lp         -0.0190844    0.0180659    -1.056    0.2935
  lp_1       -0.0186377    0.0264564    -0.7045   0.4829
  lp_2        0.0241374    0.0180346     1.338    0.1840
  ly          0.835847     0.242845      3.442    0.0009    ***
  ly_1       -1.26407      0.322772     -3.916    0.0002    ***
  ly_2        0.997032     0.339057      2.941    0.0041    ***
  ly_3       -0.393689     0.266314     -1.478    0.1426
  ly_4       -0.237846     0.148737     -1.599    0.1131
  ly_5       -0.0928693    0.141406     -0.6568   0.5129
  ly_6        0.227738     0.112039      2.033    0.0449    **
  dq1        -0.0263276    0.00606372   -4.342    3.53e-05  ***
  dq2         0.0370658    0.00855078    4.335    3.63e-05  ***
  dq3         0.0269369    0.00736524    3.657    0.0004    ***
  covid      -0.232152     0.0288383    -8.050    2.36e-012 ***

Mean dependent var   9.074395   S.D. dependent var   0.069755
Sum squared resid    0.018636   S.E. of regression   0.014006
R-squared            0.965804   Adjusted R-squared   0.959684
F(17, 95)            157.8282   P-value(F)           9.07e-62
Log-likelihood       331.7789   Akaike criterion    -627.5578
Schwarz criterion   -578.4648   Hannan-Quinn        -607.6364
rho                  0.057925   Durbin-Watson        1.872362
```

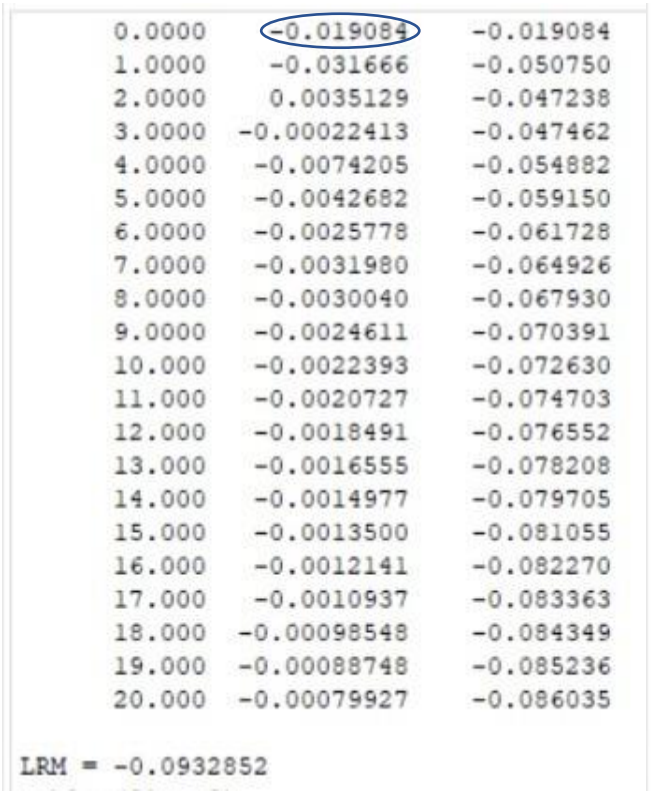
FIG.1

The model has a nice adjusted R-squared and the Godfrey test (FIG.2) let us assume that there is no auto-correlation between errors, so they are MDS and we can compute elasticities.

```
Alternative statistic: TR^2 = 8.507686,
with p-value = P(Chi-square(4) > 8.50769) = 0.0747
```

FIG. 2

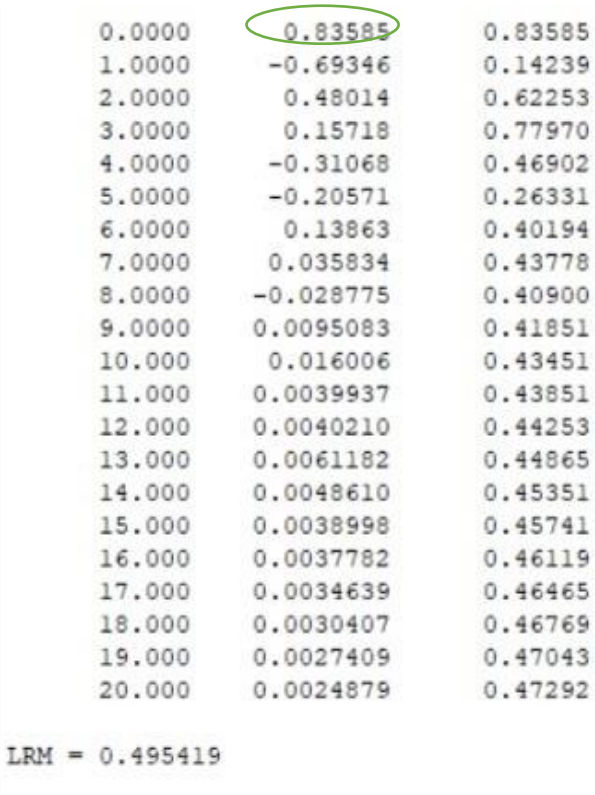
Given that our dependent variable has A(0)=1, the impact multiplier d0 is easy to find both for cons to price and cons to gdp, it's just the coefficent of lp and ly, since d0= B(0)/A(0); -0.0190844 for cons to price and 0.835847 for cons to gdp, but to be sure we can computer the long run multipliers (FIG.3-FIG.4) and take the first values that are the d0s.



0.0000	-0.019084	-0.019084
1.0000	-0.031666	-0.050750
2.0000	0.0035129	-0.047238
3.0000	-0.00022413	-0.047462
4.0000	-0.0074205	-0.054882
5.0000	-0.0042682	-0.059150
6.0000	-0.0025778	-0.061728
7.0000	-0.0031980	-0.064926
8.0000	-0.0030040	-0.067930
9.0000	-0.0024611	-0.070391
10.000	-0.0022393	-0.072630
11.000	-0.0020727	-0.074703
12.000	-0.0018491	-0.076552
13.000	-0.0016555	-0.078208
14.000	-0.0014977	-0.079705
15.000	-0.0013500	-0.081055
16.000	-0.0012141	-0.082270
17.000	-0.0010937	-0.083363
18.000	-0.00098548	-0.084349
19.000	-0.00088748	-0.085236
20.000	-0.00079927	-0.086035

LRM = -0.0932852

FIG.3



0.0000	0.83585	0.83585
1.0000	-0.69346	0.14239
2.0000	0.48014	0.62253
3.0000	0.15718	0.77970
4.0000	-0.31068	0.46902
5.0000	-0.20571	0.26331
6.0000	0.13863	0.40194
7.0000	0.035834	0.43778
8.0000	-0.028775	0.40900
9.0000	0.0095083	0.41851
10.000	0.016006	0.43451
11.000	0.0039937	0.43851
12.000	0.0040210	0.44253
13.000	0.0061182	0.44865
14.000	0.0048610	0.45351
15.000	0.0038998	0.45741
16.000	0.0037782	0.46119
17.000	0.0034639	0.46465
18.000	0.0030407	0.46769
19.000	0.0027409	0.47043
20.000	0.0024879	0.47292

LRM = 0.495419

FIG.4

The results make sense, infact they says that:

- If Price goes up by 100% the consumption goes down by ~ 2%, as we know gasoline is primary for our lives so there is almost anaelasticity, so we can't consume THAT less if we want to keep our jobs, goods in supermarkets etc.., BUT, in the long term, we could find other ways such as more public transports, bicycles perhaps or smart working and so it would decrease by ~ 9%, that, considering the good, is quite a lot in my opinion.
- if GDP goes up by 100% the consumption goes up by ~ 84% which is quite understandable because people will spend more maybe for some trip out of town or whatever. The long term multiplier instead is 49% so this trend should continue but people, maybe, will save more money because of marginal propensity to consume that becomes lower at higher incomes.

PS:

In the script i sent there is anther model, more parsiomonious, with "time" feature in order to modelling the trend, said that, the gdp long run multiplier is weird so i preferred the first model even if the are a lot of non-significant lags.