ECO310 - Tutorial 6

Demand Estimation

Francis Guiton

February 22, 2021

For this week's tutorial, we will look at Verboven's automobile dataset, which is available on Quercus under *verboven_cars.dta*. We will proceed in two steps: first, we will consider the demand model in the product space. Then, we will look at the model in the characteristics space, and see how to instrument the price with BLP instruments.

1 Set-Up

We load the dataset on STATA. The dataset contains the price, quantity sold, and characteristics of various cars sold in different markets across several years. The dataset is in panel format, according to three dimensions: year, market, and car model. Below, we generate summary statistics of our variables of interest in log-terms (*ln_p*, *ln_q*):

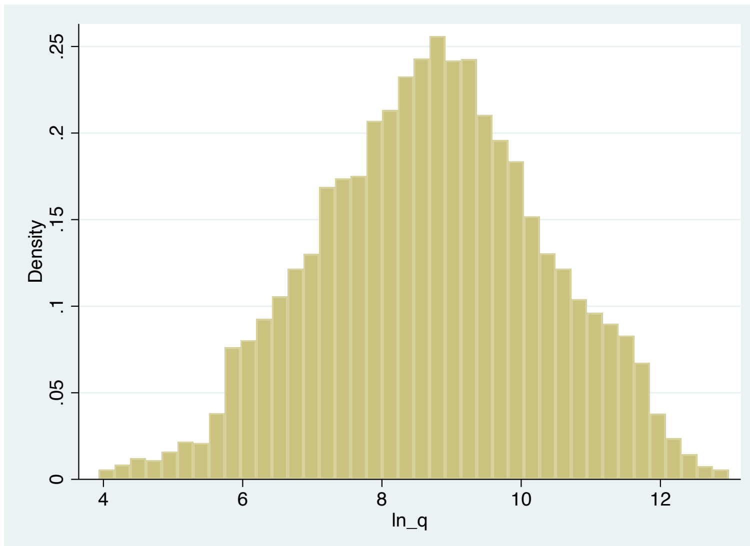
```
. sum ln_p

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        ln_p |     11,549    11.50098    2.650737     6.2106   18.57478
```

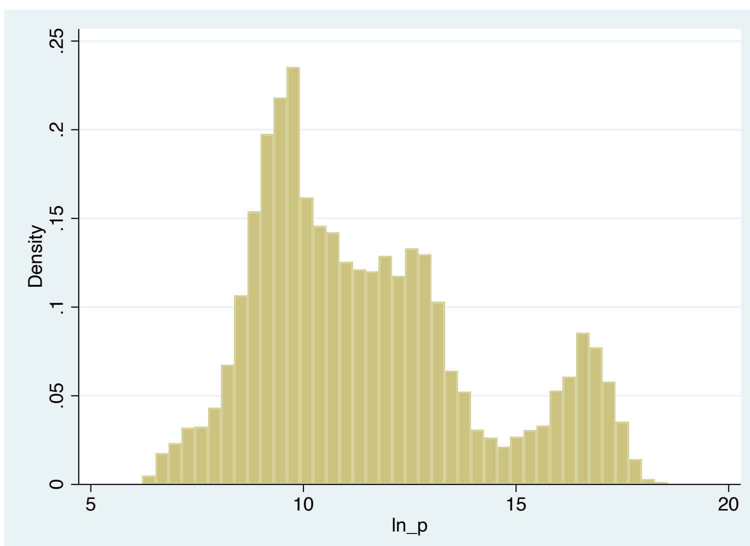
```
. sum ln_q

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        ln_q |     11,549    8.709449    1.626834   3.931826   12.98009
```

Additionally, we investigate the different markets in our dataset:

```
. tab ma

     market |
   (=second |
  dimension |
  of panel) |      Freq.     Percent        Cum.
------------+-----------------------------------
    Belgium |      2,673       23.14       23.14
     France |      2,265       19.61       42.76
    Germany |      2,283       19.77       62.52
      Italy |      2,027       17.55       80.08
         UK |      2,301       19.92      100.00
------------+-----------------------------------
      Total |     11,549      100.00
```

Finally, we explore the distributions of price and quantity in our dataset using histograms:





2 Model in Product Space

We will first consider demand in product space. That is, we assume that consumers have preferences over products, and we will estimate the elasticity of demand under various regression specifications.

Simple OLS

```
. reg ln_q ln_p, robust

Linear regression                               Number of obs     =     11,549
                                                F(1, 11547)       =     217.88
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0185
                                                Root MSE          =     1.6118

------------------------------------------------------------------------------
             |               Robust
        ln_q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.0834329   .0056524   -14.76   0.000    -.0945125   -.0723533
       _cons |   9.669009   .0650023   148.75   0.000     9.541594    9.796425
------------------------------------------------------------------------------
```

The above specification yields a negative estimate for price. However, given the format of our dataset, we must also control for endogeneity by including various fixed effects.

Below, we will consider the user-created command *reghdfe*. This command yields identical estimates to *xtreg, fe* or *reg* with dummy variables. However, it has several advantages. First, unlike including dummy variables, it allows us to ”absorb” the effects of these variables, and therefore does not require computing parameter estimates for each of the dummies. Second, it allows us to work with higher-dimension panels (i.e. more than two dimensions), as opposed to *xtreg*. In our case this will be useful, as our panel has three dimensions.

In terms of syntax, we simply write *reghdfe dep_var ind_vars*, a(*panel_vars*).

Car Model Fixed Effect

We first control for the car model below:

```
. reghdfe ln_q ln_p, vce(robust) a(co)
(dropped 15 singleton observations)
(MWFE estimator converged in 1 iterations)

HDFE Linear regression                            Number of obs   =     11,534
Absorbing 1 HDFE group                            F(   1,  11192) =     336.74
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.4556
                                                  Adj R-squared   =     0.4390
                                                  Within R-sq.    =     0.0323
                                                  Root MSE        =     1.2183

------------------------------------------------------------------------------
             |               Robust
        ln_q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.0873493   .0047601   -18.35   0.000    -.0966799   -.0780187
       _cons |   9.715253   .0543831   178.64   0.000     9.608653    9.821854
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
          co |       341           0         341     |
-----------------------------------------------------+
```

Car Model and Year Fixed Effects

Next, we include the time dimension of our panel:

```
. reghdfe ln_q ln_p, vce(robust) a(co ye)
(dropped 15 singleton observations)
(MWFE estimator converged in 10 iterations)

HDFE Linear regression                            Number of obs   =     11,534
Absorbing 2 HDFE groups                           F(   1,  11163) =     303.16
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.4591
                                                  Adj R-squared   =     0.4412
                                                  Within R-sq.    =     0.0293
                                                  Root MSE        =     1.2159

------------------------------------------------------------------------------
             |               Robust
        ln_q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.0840172   .0048254   -17.41   0.000    -.0934759   -.0745585
       _cons |   9.676925   .0551145   175.58   0.000     9.568891     9.78496
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
          co |       341           0         341     |
          ye |        30           1          29     |
-----------------------------------------------------+
```

Car Model, Year and Market Fixed Effects

Finally, we control for the market fixed effect:

```
. reghdfe ln_q ln_p, vce(robust) a(co ye ma)
(dropped 15 singleton observations)
(MWFE estimator converged in 11 iterations)

HDFE Linear regression                            Number of obs   =     11,534
Absorbing 3 HDFE groups                           F(   1,  11159) =      37.39
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.5691
                                                  Adj R-squared   =     0.5547
                                                  Within R-sq.    =     0.0044
                                                  Root MSE        =     1.0854

------------------------------------------------------------------------------
             |               Robust
        ln_q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.3278852    .053624    -6.11   0.000    -.4329976   -.2227727
       _cons |     12.482   .6174535    20.22   0.000     11.27168    13.69232
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
          co |       341           0         341     |
          ye |        30           1          29     |
          ma |         5           1           4    ?|
-----------------------------------------------------+
? = number of redundant parameters may be higher
```

Using the above command, we are therefore effectively running a three-dimension panel regression.

Including (log-) population as a covariate

Below, we add the (log-) population of each market as a covariate in our linear regression:

```
. reghdfe ln_q ln_p ln_pop, vce(robust) a(co ye ma)
(dropped 15 singleton observations)
(MWFE estimator converged in 11 iterations)

HDFE Linear regression                          Number of obs   =     11,534
Absorbing 3 HDFE groups                         F(   2,  11158) =      21.49
                                                Prob > F        =     0.0000
                                                R-squared       =     0.5693
                                                Adj R-squared   =     0.5548
                                                Within R-sq.    =     0.0047
                                                Root MSE        =     1.0853

------------------------------------------------------------------------------
             |               Robust
        ln_q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.3017443   .0556516    -5.42   0.000    -.4108314   -.1926573
      ln_pop |   .4440674   .2142268     2.07   0.038      .024145    .8639899
       _cons |   4.418463   3.967015     1.11   0.265    -3.357587    12.19451
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
          co |       341           0         341     |
          ye |        30           1          29     |
          ma |         5           1           4    ?|
-----------------------------------------------------+
? = number of redundant parameters may be higher
```

Although we obtain a reasonable (negative) coefficient for price in our regression, we should still be concerned with endogeneity. That is, we suspect that price is still correlated with demand unobservables in our model. We now move on to demand in the product characteristics space, where we will explore a useful method for instrumenting the price variable.

3 Model in Characteristics Space

We now consider the model in the product characteristics space. That is, consumers now have preferences over product characteristics.

First, we construct the market shares s_{jmt}. For each car model j, market m and year t, the market share is computed as the quantity sold of the car divided by the market size M (here, we will use population as the market size). Then, given these market shares, we construct the outside option s_{0mt} as:

$$s_{0mt} = 1 - \sum_j s_{jmt}$$

Given this outside option, we can now construct the logarithm of the odds-ratio which will serve as our dependent variable in our logit regression model: $\log(\frac{s_{jmt}}{s_{0mt}})$.

Simple OLS

```
. reg sj_s0 ln_p, vce(robust)

Linear regression                               Number of obs     =     11,549
                                                F(1, 11547)       =       6.63
                                                Prob > F          =     0.0100
                                                R-squared         =     0.0006
                                                Root MSE          =     1.4994

------------------------------------------------------------------------------
             |               Robust
       sj_s0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.0136231   .0052896    -2.58   0.010    -.0239917   -.0032545
       _cons |  -8.582377   .0620086  -138.41   0.000    -8.703925    -8.46083
------------------------------------------------------------------------------
```

Again, we would like to control for the car model, market and time fixed effects of our panel dataset. Similarly to the previous section, we will now use the *reghdfe* command in order to run a three-dimensional panel regression.

Car Model Fixed Effect

We first control for the car model below:

```
. reghdfe sj_s0 ln_p, vce(robust) a(co)
(dropped 15 singleton observations)
(MWFE estimator converged in 1 iterations)

HDFE Linear regression                            Number of obs   =     11,534
Absorbing 1 HDFE group                            F(   1,  11192) =       1.01
                                                  Prob > F        =     0.3141
                                                  R-squared       =     0.4626
                                                  Adj R-squared   =     0.4462
                                                  Within R-sq.    =     0.0001
                                                  Root MSE        =     1.1161

------------------------------------------------------------------------------
             |               Robust
       sj_s0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.0046363   .0046057    -1.01   0.314    -.0136644    .0043917
       _cons |  -8.684781   .0529518  -164.01   0.000    -8.788576   -8.580987
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
          co |       341           0         341     |
-----------------------------------------------------+
```

Car Model and Year Fixed Effect

Next, we include the time dimension of our panel:

```
. reghdfe sj_s0 ln_p, vce(robust) a(co ye)
(dropped 15 singleton observations)
(MWFE estimator converged in 10 iterations)

HDFE Linear regression                            Number of obs   =     11,534
Absorbing 2 HDFE groups                           F(   1,  11163) =       0.23
                                                  Prob > F        =     0.6329
                                                  R-squared       =     0.4706
                                                  Adj R-squared   =     0.4530
                                                  Within R-sq.    =     0.0000
                                                  Root MSE        =     1.1093

------------------------------------------------------------------------------
             |               Robust
       sj_s0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |   .0022322   .0046737     0.48   0.633     -.006929    .0113934
       _cons |  -8.763786   .0536031  -163.49   0.000    -8.868858   -8.658715
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
          co |       341           0         341     |
          ye |        30           1          29     |
-----------------------------------------------------+
```

Car Model, Year and Market Fixed Effect

Finally, we control for the market fixed effect:

```
. reghdfe sj_s0 ln_p, vce(robust) a(co ye ma)
(dropped 15 singleton observations)
(MWFE estimator converged in 11 iterations)

HDFE Linear regression                            Number of obs   =     11,534
Absorbing 3 HDFE groups                           F(   1,  11159) =      24.82
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.4929
                                                  Adj R-squared   =     0.4759
                                                  Within R-sq.    =     0.0029
                                                  Root MSE        =     1.0858

------------------------------------------------------------------------------
             |               Robust
       sj_s0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.2671712   .0536294    -4.98   0.000    -.3722942   -.1620482
       _cons |  -5.664992   .6175079    -9.17   0.000    -6.875416   -4.454567
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
          co |       341           0         341     |
          ye |        30           1          29     |
          ma |         5           1           4    ?|
-----------------------------------------------------+
? = number of redundant parameters may be higher
```

Including (log-) population as a covariate

Below, we add the (log-) population of each market as a covariate in our regression:

```
. reghdfe sj_s0 ln_p ln_pop, vce(robust) a(co ye ma)
(dropped 15 singleton observations)
(MWFE estimator converged in 11 iterations)

HDFE Linear regression                            Number of obs   =     11,534
Absorbing 3 HDFE groups                           F(   2,  11158) =      15.27
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.4932
                                                  Adj R-squared   =     0.4761
                                                  Within R-sq.    =     0.0035
                                                  Root MSE        =     1.0856

------------------------------------------------------------------------------
             |               Robust
       sj_s0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.3002158    .0556574    -5.39   0.000    -.4093141   -.1911176
      ln_pop |  -.5613457    .2142543    -2.62   0.009     -.981322   -.1413694
       _cons |   4.528126    3.967477     1.14   0.254    -3.248829    12.30508
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
          co |       341           0         341     |
          ye |        30           1          29     |
          ma |         5           1           4    ?|
-----------------------------------------------------+
? = number of redundant parameters may be higher
```

Including product characteristics as covariates

```
. reghdfe sj_s0 ln_p li wi cy we pl do le ln_pop, a(co ye ma)
(dropped 15 singleton observations)
(MWFE estimator converged in 11 iterations)

HDFE Linear regression                            Number of obs   =     11,532
Absorbing 3 HDFE groups                           F(   9,  11149) =      56.74
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.5136
                                                  Adj R-squared   =     0.4969
                                                  Within R-sq.    =     0.0438
                                                  Root MSE        =     1.0638

------------------------------------------------------------------------------
       sj_s0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.2146746   .0486984    -4.41   0.000    -.3101322   -.1192171
          li |  -.1026867   .0157396    -6.52   0.000     -.133539   -.0718344
          wi |   .0513916   .0052541     9.78   0.000     .0410926    .0616906
          cy |  -.0015656   .0000976   -16.03   0.000     -.001757   -.0013742
          we |   .0007817   .0002556     3.06   0.002     .0002806    .0012827
          pl |   .0401136   .0494747     0.81   0.418    -.0568655    .1370926
          do |  -.0313473   .0232265    -1.35   0.177    -.0768754    .0141807
          le |   .0024706   .0018671     1.32   0.186    -.0011892    .0061303
      ln_pop |  -.4668392   .2197867    -2.12   0.034    -.8976599   -.0360184
       _cons |  -5.300284     4.1156    -1.29   0.198    -13.36759    2.767019
------------------------------------------------------------------------------
```

Again, we should be concerned with the endogeneity of the price variable. We suspect that price is correlated with the demand unobservables. Fortunately, expressing our model in the characteristics space allows us to include BLP instruments in order to control for this endogeneity.

Specifically, we will instrument the price with the average characteristics of other products. Under the assumption that the price of a product depends not only on its own characteristics but also the characteristics of its competitors, these instruments are valid.

Using BLP Instruments

Below, we instrument the price using three product characteristics: horse power, time to acceleration, and maximum speed. We use the IV-counterpart to the *reghdfe* command, which is simply called *ivreghdfe*. The syntax is identical to the usual *ivreg2* command on STATA, but includes the *a()* option in order to absorb the fixed effects.

```
. ivreghdfe sj_s0 li wi cy we pl do le ln_pop (ln_p= av_ac av_hp av_sp), a(co ye ma) robust
(dropped 15 singleton observations)
(MWFE estimator converged in 11 iterations)

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                      Number of obs =     11532
                                                      F(  9, 11149) =     46.23
                                                      Prob > F      =    0.0000
Total (centered) SS     =   13194.6857                Centered R2   =   -0.1023
Total (uncentered) SS   =   13194.6857                Uncentered R2 =   -0.1023
Residual SS             =  14544.57103                Root MSE      =     1.142

------------------------------------------------------------------------------
             |               Robust
       sj_s0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -2.224643   .3302294    -6.74   0.000    -2.871951   -1.577336
          li |  -.0909261   .0170802    -5.32   0.000    -.1244063    -.057446
          wi |   .0573559   .0074319     7.72   0.000     .0427881    .0719237
          cy |  -.0011164   .0001302    -8.57   0.000    -.0013717   -.0008612
          we |   .0016828   .0003393     4.96   0.000     .0010176    .0023479
          pl |  -.0633531   .0585873    -1.08   0.280    -.1781945    .0514883
          do |  -.0092754   .0272436    -0.34   0.734    -.0626776    .0441268
          le |   .0029451   .0019652     1.50   0.134     -.000907    .0067971
      ln_pop |  -2.971021   .4609101    -6.45   0.000    -3.874486   -2.067556
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):            265.971
                                                   Chi-sq(3) P-val =    0.0000
------------------------------------------------------------------------------
```