Bendegúz István Biksi, Xia Mengzhen

BGSE

3. Econometrics Assigment

11/11/15

1. Tests of structural change

a. Estimate the model using the full sample, and then use two subperiods: 1960-1973 and 1974-1995. Perform a test of equality of the coefficients of the two equations using the sums of squares of these three regressions. The full sample model:

```
. reg l_gpop year l_y l_pg l_pnc l_puc
```

Source	SS	df	MS
Model	.780166195	5	.156033239
Residual	.025187735	30	.000839591
Total	.80535393	35	.023010112

Number of obs	=	36
F(5, 30)	=	185.84
Prob > F	=	0.0000
R-squared	=	0.9687
Adj R-squared	=	0.9635
Root MSE	=	.02898

l_gpop	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
year	-.0191181	.0059565	-3.21	0.003	-.031283	-.0069533
l_y	1.954627	.1928543	10.14	0.000	1.560766	2.348488
l_pg	-.1155303	.0334795	-3.45	0.002	-.1839046	-.0471561
l_pnc	.2052824	.1520194	1.35	0.187	-.1051825	.5157474
l_puc	-.1292739	.0714118	-1.81	0.080	-.2751162	.0165684
_cons	20.06664	10.11464	1.98	0.056	-.5902017	40.72349

Model on the subperiod before 1973:

```
. reg l_gpop year l_y l_pg l_pnc l_puc if year<=1973
```

Source	SS	df	MS
Model	.330965051	5	.06619301
Residual	.000652264	8	.000081533
Total	.331617315	13	.025509024

Number of obs	=	14
F(5, 8)	=	811.86
Prob > F	=	0.0000
R-squared	=	0.9980
Adj R-squared	=	0.9968
Root MSE	=	.00903

l_gpop	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
year	.0263662	.0106902	2.47	0.039	.0017144	.0510179
l_y	.4240038	.3517738	1.21	0.263	-.3871881	1.235196
l_pg	.0945454	.1505087	0.63	0.547	-.2525284	.4416192
l_pnc	.5838989	.1314914	4.44	0.002	.2806792	.8871187
l_puc	-.3346175	.0923259	-3.62	0.007	-.5475215	-.1217135
_cons	-55.78581	18.04974	-3.09	0.015	-97.40859	-14.16302

Model ont he subperiod after73:

```
. reg l_gpop year l_y l_pg l_pnc l_puc if year>1973

      Source |       SS       df       MS              Number of obs =      22
-------------+------------------------------           F(  5,    16) =   37.13
       Model |  .054101245     5  .010820249           Prob > F      =  0.0000
    Residual |  .004662195    16  .000291387           R-squared     =  0.9207
-------------+------------------------------           Adj R-squared =  0.8959
       Total |   .05876344    21  .002798259           Root MSE      =  .01707

------------------------------------------------------------------------------
      l_gpop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   -.012617   .0037736    -3.34   0.004    -.0206166   -.0046173
         l_y |   1.014073   .2856869     3.55   0.003     .4084437    1.619702
        l_pg |   -.242374   .0400297    -6.05   0.000    -.3272333   -.1575148
       l_pnc |   .3301687    .181118     1.82   0.087    -.0537843    .7141218
       l_puc |  -.0553741   .0507741    -1.09   0.292    -.1630103    .0522622
       _cons |   15.84117    6.83939     2.32   0.034     1.342313    30.34003
------------------------------------------------------------------------------
```

	Residual Sum of Squares
Full sample model	0.025187735
Before 1973 model	0.000652264
After 1973 model	0.004662195
r	6
k	6

The null hypothesis: the coefficients of the two subsample (before73, after73) regressions are the same. If one rejects the null hypothesis, s/he is stating that the coefficients are different in the two subsamples.

To test wheter there is a structural change in the data we performe a Chow test, by using the formula:

$$F(r, n-2k) = \frac{\frac{RSS_{all}-(RSS_{bef73}+RSS_{aft73})}{r}}{\frac{RSS_{bef73}+RSS_{aft73}}{n-2k}}$$

Where r is the number of restrictions, n is the sample size, k is the number of parameters and RSS_i is the residual sum of squares obtained from regression i.

In this case r=k=6, so F(6,24)= 14,95789, that is larger then the critical value of F(6,24)=2,51, so we can reject the null hypothesis that the coefficients of the two model ran on different subperiods are the same. Meaning they are different.

b. Now imagine that you know the intercepts to be different, so you want to test for a change in the other five coefficients beyond a simple shift in the constant term. How would you do this? (Hint: still you may want to run the two separate regressions for the two subperiods – "unrestricted" model-, but modify the full-sample regression in order to accommodate in that regression the differing intercepts). Do it.

The restricted or the whole sample model now is:

```
. reg l_gpop year l_y l_pg l_pnc l_puc after73

      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  6,    29) =  216.29
       Model |  .787750457     6  .131291743           Prob > F      =  0.0000
    Residual |  .017603473    29  .000607016           R-squared     =  0.9781
-------------+------------------------------           Adj R-squared =  0.9736
       Total |   .80535393    35  .023010112           Root MSE      =  .02464

------------------------------------------------------------------------------
      l_gpop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |  -.0168618   .0051048    -3.30   0.003    -.0273023   -.0064212
         l_y |   1.838168   .1672588    10.99   0.000     1.496086    2.180251
        l_pg |   -.178005   .0335078    -5.31   0.000    -.2465361   -.1094739
       l_pnc |   .2098421   .1292667     1.62   0.115    -.0545381    .4742222
       l_puc |  -.1281312   .0607215    -2.11   0.044    -.2523205   -.0039418
     after73 |   .0773106   .0218717     3.53   0.001      .032578    .1220433
       _cons |   16.65782   8.654259     1.92   0.064    -1.042128    34.35777
------------------------------------------------------------------------------
```

Where after73 takes one if the observation is taken after 1973.

	Residual Sum of Squares
Full sample model	0.017603473
Before 1973 model	0.000652264
After 1973 model	0.004662195
r	5
k	6

The value of the test statistic is F(5,24)=11.09939266, that is higher then F(5,24)=2.62, so we can reject the null hypothesis, that is the coefficients of year, log_income, log_price, log_usedcarprice and log_newcarprice are the same across the two subsample. Meaning they are different.

c. Now imagine that **you know** that the price elasticities on automobile prices and the coefficient of the time trend do not change, but the other three coefficients could change or not. How would you test for this change? Do it.

The restricted on the whole sample model now is:

```
. reg l_gpop year l_y l_pg l_pnc l_puc after73 l_yafter73 l_pgafter73

      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  8,    27) =  335.15
       Model |  .797324836     8  .099665605           Prob > F      =  0.0000
    Residual |  .008029094    27  .000297374           R-squared     =  0.9900
-------------+------------------------------           Adj R-squared =  0.9871
       Total |   .80535393    35  .023010112           Root MSE      =  .01724

------------------------------------------------------------------------------
      l_gpop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |  -.0129056   .0036746    -3.51   0.002    -.0204452    -.005366
         l_y |   1.517164   .1919942     7.90   0.000     1.123224    1.911103
        l_pg |   .0916647   .2283134     0.40   0.691    -.3767957     .560125
       l_pnc |   .5933324   .1350152     4.39   0.000     .3163041    .8703606
       l_puc |  -.1080754   .0447157    -2.42   0.023    -.1998245   -.0163263
     after73 |   7.866224   2.437508     3.23   0.003     2.864871    12.86758
  l_yafter73 |  -.8534086   .2699913    -3.16   0.004    -1.407385   -.2994323
 l_pgafter73 |  -.3841614   .2207667    -1.74   0.093    -.8371373    .0688145
       _cons |   11.71724   6.125905     1.91   0.066    -.8520753    24.28656
------------------------------------------------------------------------------
```

Where l_yafter73 is the interaction of the dummy variably after73 and the log of the income. Similarly l_pgafter73 is the interaction between after73 dummy and log of the price.

	Residual Sum of Squares
Full sample model	0,008029094
Before 1973 model	0.000652264
After 1973 model	0.004662195
r	3
k	6

Here the null hypothesis is that: price elasticities on automobile prices and the coefficient of the time trend are the same across the two subperiod. The value of the test-statistic: $F(3,24)= 4,08641$, meanwhile $F(3,24)= 3.0087866$. As the value of the test statistic is higher then the critical value of the F-distribution, we can reject the null hypothesis. So price elasticities on automobile prices and the coefficient of the time trend are not the same across the two subperiod. Note that the value of the test statisic is the lowest in this setting but we are still able to reject the null hypothesis.

2. On instrumental variables.

$$\ln(Q_i) = \beta_0 + \beta_1 \ln(P_i) + \beta_2 Ice_i + \sum_{j=1}^{12} \beta_{2+j} Seas_{j,i} + u_i$$

a) Why do we include the seasonal variables and why do we include only twelve of them?

i. We include seasonal dummy variables in the demand equation, because the demand for grain transportation can have seasonal fluctuation. Meaning, that the demand has a seasonal trend in it's time series data.

ii. We only include twelve of the thirteen seasonal dummy variables to avoid perfect collinearity and therefore to avoid biased estimation of the parameters.

b) Estimate this demand equation by OLS. What is the estimated value of the price elasticity of demand?

```
. reg lQ lP ice seas1 seas2 seas3 seas4 seas5 seas6 seas7 seas8 seas9 seas10 seas11 seas12
```

Source	SS	df	MS
Model	22.4696137	14	1.60497241
Residual	49.3989901	313	.157824249
Total	71.8686038	327	.219781663

Number of obs	=	328
F(14, 313)	=	10.17
Prob > F	=	0.0000
R-squared	=	0.3126
Adj R-squared	=	0.2819
Root MSE	=	.39727

lQ	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lP	-.6388847	.0823886	-7.75	0.000	-.8009902	-.4767793
ice	.4477537	.1196041	3.74	0.000	.2124239	.6830835
seas1	-.1328219	.1109586	-1.20	0.232	-.351141	.0854972
seas2	.0668882	.1112977	0.60	0.548	-.1520981	.2858745
seas3	.1114365	.1113083	1.00	0.318	-.1075707	.3304436
seas4	.1554219	.1107434	1.40	0.161	-.0624737	.3733175
seas5	.1096585	.1299182	0.84	0.399	-.1459648	.3652819
seas6	.0468325	.159596	0.29	0.769	-.267184	.3608491
seas7	.1225526	.1600407	0.77	0.444	-.1923389	.4374442
seas8	-.2350078	.1598562	-1.47	0.143	-.5495363	.0795207
seas9	.0035607	.160021	0.02	0.982	-.3112921	.3184135
seas10	.1692469	.1612946	1.05	0.295	-.1481118	.4866057
seas11	.2151845	.1600958	1.34	0.180	-.0998156	.5301846
seas12	.2196331	.1591364	1.38	0.169	-.0934792	.5327454
_cons	8.861233	.1713614	51.71	0.000	8.524067	9.198399

The estimated value of the price elasticity of demand is: -.6388847 (.0823886). So the demand is inelastic, since .64<1.

c) Interpret and explain the coefficient β2.

i. The coefficient of ice (β2) is: .4477537 with standard deviation: .1196041. Meaning that, if the Great Lakes are frozen (so you cannot use the alternative shipping method), the quantity that is demanded increases by approx. 56,48%.

ii. If the lakes are frozen the demand for grain transportation gets higher. Why? If the lakes are frozen, one can not use ships to transport the grain, therefore the only option is trains.

d) Explain why this OLS estimator is probably biased (think "supply and demand").

The OLS estimator is probably biased because of endogenity. We should estimate a system of equations of supply and demand at the same time. To scan demand and supply, one should have good exogenous supply and demand shifter variables, because price and quantity takes their values simultaneously, so one should estimate the two together.

e) You are thinking of using the Cartel variable as an IV for ln(P). Use your economic common sense to examine if this variable satisfies the conditions to be a valid instrument.

Cartel as an instrumental variable should fit into two conditions. The relevance condition and the exogenity condition. The verification of the relevance condition, namely $\mathrm{cov}(\mathrm{Cartel}, \ln(P)) \neq 0$ is relatively simple with common sense. If there is a cartel in the relevant market, it does have effects on the market price. Probably a cartel would sustain higher prices, compared to a non-cartel market price. The exogenity condition $\mathrm{cov}(\mathrm{Cartel}, \varepsilon) = 0$ is a little bit more difficult to verify, but one is able to think about, on avarage the demanded quantitiy has nothing to do with the market structure, except the price of the given commodity.

f) Estimate the first stage regression. Is Cartel a weak instrument? (Present a more or less formal test).

First stage regression:

```
. ivregress 2sls lQ ice seas1 seas2 seas3 seas4 seas5 seas6 seas7 seas8 seas9 seas10 seas11 seas12 (lP=cartel),
> first
```

First-stage regressions

Number of obs	=	328
F(14, 313)	=	21.32
Prob > F	=	0.0000
R-squared	=	0.4881
Adj R-squared	=	0.4652
Root MSE	=	0.2114

lP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ice	.035003	.0642519	0.54	0.586	-.0914171	.1614232
seas1	.0387253	.0590844	0.66	0.513	-.0775276	.1549782
seas2	.1362884	.0590844	2.31	0.022	.0200355	.2525413
seas3	.1890486	.0593185	3.19	0.002	.0723351	.3057621
seas4	.0895226	.059357	1.51	0.133	-.0272666	.2063119
seas5	.0178628	.0698693	0.26	0.798	-.11961	.1553356
seas6	-.025741	.0855293	-0.30	0.764	-.194026	.142544
seas7	-.0671265	.0855293	-0.78	0.433	-.2354115	.1011585
seas8	-.0358373	.0857092	-0.42	0.676	-.2044762	.1328017
seas9	-.0057758	.0863212	-0.07	0.947	-.175619	.1640673
seas10	-.1002111	.0863212	-1.16	0.247	-.2700542	.0696321
seas11	-.0867514	.0853616	-1.02	0.310	-.2547064	.0812036
seas12	.0116931	.0853616	0.14	0.891	-.1562619	.1796482
cartel	.3578984	.0248623	14.40	0.000	.30898	.4068168
_cons	-1.693741	.0783608	-21.61	0.000	-1.847922	-1.539561

Testing β_Cartel=0

```
. test cartel

 ( 1)  cartel = 0

       F(  1,   313) =  207.22
            Prob > F =    0.0000
```

According to the test as the value of the F-statistic is higher, then 10, we can conclude, that cartel is a strong IV, because we reject the null hypothesis, namely the impact of cartel on price is zero.

g) Estimate the demand equation using instrumental variables. What is the estimated value of demand elasticity now?

```
. ivregress 2sls lQ ice seas1 seas2 seas3 seas4 seas5 seas6 seas7 seas8 seas9 seas10 seas11 seas12 (lP=cartel)

Instrumental variables (2SLS) regression               Number of obs =     328
                                                       Wald chi2(14) =  129.21
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.2959
                                                       Root MSE      =  .39279
```

lQ	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lP	-.8665865	.1290666	-6.71	0.000	-1.119552	-.6136206
ice	.422934	.1187568	3.56	0.000	.190175	.6556929
seas1	-.1309732	.1097094	-1.19	0.233	-.3459997	.0840532
seas2	.0909521	.1105491	0.82	0.411	-.1257201	.3076244
seas3	.135872	.1105753	1.23	0.219	-.0808516	.3525956
seas4	.1525109	.1095011	1.39	0.164	-.0621072	.367129
seas5	.0735618	.1294287	0.57	0.570	-.1801137	.3272373
seas6	-.0060642	.1594995	-0.04	0.970	-.3186775	.3065491
seas7	.0602324	.1605893	0.38	0.708	-.2545169	.3749816
seas8	-.2935991	.1601377	-1.83	0.067	-.6074631	.020265
seas9	-.0583723	.1605412	-0.36	0.716	-.3730272	.2562827
seas10	.0858109	.1636392	0.52	0.600	-.234916	.4065378
seas11	.1517912	.1607242	0.94	0.345	-.1632225	.4668048
seas12	.1786558	.1583685	1.13	0.259	-.1317408	.4890524
_cons	8.573535	.2114377	40.55	0.000	8.159124	8.987945

```
Instrumented:  lP
Instruments:   ice seas1 seas2 seas3 seas4 seas5 seas6 seas7 seas8 seas9
               seas10 seas11 seas12 cartel
```

The estimated value of the demand elasticity is: -.8665865 (.1290666). The value is higher as it's standard deviation /compared to the results that were obtained in b)/, although the demand is still inelastic.

h) Carry out and comment on the Hausman test for comparison of the two sets of estimates. Is there evidence of a problem of endogeneity?

```
. hausman iv

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      iv            .           Difference          S.E.
-------------+----------------------------------------------------------------
          lP |   -.8665865    -.6388847       -.2277018        .0993495
         ice |     .422934     .4477537       -.0248198               .
       seas1 |   -.1309732    -.1328219        .0018486               .
       seas2 |    .0909521     .0668882        .0240639               .
       seas3 |     .135872     .1114365        .0244355               .
       seas4 |    .1525109     .1554219        -.002911               .
       seas5 |    .0735618     .1096585       -.0360967               .
       seas6 |   -.0060642     .0468325       -.0528967               .
       seas7 |    .0602324     .1225526       -.0623203        .0132632
       seas8 |   -.2935991    -.2350078       -.0585912         .009491
       seas9 |   -.0583723     .0035607        -.061933        .0129132
      seas10 |    .0858109     .1692469        -.083436        .0276012
      seas11 |    .1517912     .2151845       -.0633933        .0141983
      seas12 |    .1786558     .2196331       -.0409773               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from ivregress
           B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic

                  chi2(14) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                           =        5.25
                Prob>chi2 =      0.9820
                (V_b-V_B is not positive definite)
```

According to the Hausman-test we can not reject the null, because the p-value is higher then, .05, so ln(P) looks like an exogenous variable, but from economic theory we should know, that it is not exogenous. This is usually the case with the Hausman-test, because of the high variance of the coefficient of the 2sls estimation, the test usually can not be rejected. In order to work this test and get reasonable results one should increase the sample size.

i) Do your estimates suggest that the cartel was charging a price that maximized profits (that is, the price that a monopolist would charge? (Hint: what should a monopolist do if price elasticity is lower than one?)

If the price elasticity is lower than one, then a monopolist would increase price, up to a point where the own price elasticity of the demand is at least one. According to this I obtained from 2sls the following price elasticity: -.8665865, that is smaller then one in absolute terms, so a monopolist or a cartel that acts like a monopolist would increase the price. An interesting question could be that, comparing cartel prices and competitive prices, or price margins, althought that would need the values of marginal costs.