Prof. Javier Gómez Biscarri
MSc in Finance/EPP/CMR
ECONOMETRICS

Assignment 3: The third on multiple regression - tests of structural change, Instrumental variables

Submission deadline: November 23, 17:30. This is an individual or group assignment: submit your answers to daniel.dejuan@barcelonagse.eu before the beginning of Monday's class.

1) **Tests of structural change**
The Stata file **gasoline.dta** has data on the US gasoline market (36 yearly observations, 1960 to 1995). The variables are:
G – total US gasoline consumption (total expenditure divided by price index)
Pg – price index for gasoline
Y – per capita disposable income
Pnc – price index for new cars
Puc – price index for used cars
Ppt – price index for public transportation
Pd – aggregate price index for consumer durables
Pn – aggregate price index for consumer nondurables
Ps – aggregate price index for consumer services
Pop – US total population in millions
Year – year of the data
Take the log(*G/pop*) that you used in Assignment 3, and regress it on Year (or Year-1960), log(*Y*), log(P_g), log(P_{nc}) and log(P_{uc}). Whenever we analyze oil, gasoline or the like we tend to be concerned with possible structural changes due to the oil crises. We will use this (limited) regression to test whether something really changed in 1973 that made the consumption of gasoline behave differently. We'll go step by step. *For this assignment, assume that homoskedasticity holds, so we can use the simple versions of the tests which use sums of squares.*

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.

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.

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.

2) **On instrumental variables.**
During the 1880's, a cartel known as the Joint Executive Committee (JEC) controlled the railroad transportation of grain from the Midwest to the East Coast. This cartel could affect the market price of grain so that it would be above competitive prices although from time to time, some member of the cartel would cheat and prices would drop. We will use these "cheating" moments to try to estimate the price elasticity of the demand for grain. The dataset **JEC.dta** has weekly data (1880-1886) on several variables which are described at the end.

Assume that the demand for grain transportation is

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

where Q_i is total tons transported in week i, P_i is the price of transportation per ton, Ice_i is a binary variable which takes value "1" if the Great Lakes are frozen (so you cannot use the alternative shipping method), and $Seas_{j,i}$ is a binary variable that captures seasonal demand factors in each month of the year (see below: there are actually 13 of these variables!).

a) Why do we include the seasonal variables and why do we include only twelve of them?
b) Estimate this demand equation by OLS. What is the estimated value of the price elasticity of demand?
c) Interpret and explain the coefficient β_2.
d) Explain why this OLS estimator is probably biased (think "supply and demand").
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.
f) Estimate the first stage regression. Is *Cartel* a weak instrument? (Present a more or less formal test).
g) Estimate the demand equation using instrumental variables. What is the estimated value of demand elasticity now?
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?
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?)

Description of variables in JEC.dta

week = week counter: = 1 for the first week (1/1/1880-1/7/1880), = 2 for the second week (1/8/1880-1/14/1880)=,…, = 328 for the last week

price = Weekly index of the price of transportation (by railroad) of a ton of grain

ice = 1 if the Great Lakes are frozen (so transportation by ship is not possible), =0 otherwise

cartel = 1 if the cartel is "operational", = 0 otherwise

quantity = total quantity of grain (in tons) transported in that specific week

*seas*1 – *seas*13 = thirteen "monthly" binary variables. In order to adjust to the weekly frequency, the year has been divided in 13 periods of 4 weeks. Thus ,
seas1 = 1 if the date is between January 1 and January 28, =0 otherwise
seas2 = 1 if the date is between January 29 and February 25, =0 otherwise
…
seas13 = 1 if the date is between December 4 and December 31, =0 otherwise

Stata commands

Regression of a specific subperiod can be easily done using a command such as:

regress y x1 x2… in 1/14

where 1/14 says that the first 14 observations should be used. Alternatively,

regress y x1 x2… if year<=1973

which is self-explanatory.

Instrumental variables regressions can be easily done using:

ivregress 2sls depvar *list of exogenous variables* (*x1 = list of instruments*), first

where the "first" option gives you the first stage equation and the test for weak instruments. Instead of **2sls** you can choose the **gmm** or **liml** options (but we haven't covered those, so don't worry about them).

A simple procedure to carry out the Hausman test in Stata is:

ivregress 2sls depvar *list of exogenous variables* (*x1=list of instruments*)
estimates store iv

so you obtain the IV estimates, which are always consistent, but inefficient if there is no endogeneity. Then you obtain the regular OLS estimates:

regress depvar *list of exogenous variables including x1*
estimates store ols

which are inconsistent under homogeneity but efficient if there is no homogeneity. Then you do the test:

hausman iv ols

Or you can do the regression-based version by hand, by getting the residuals from the first-stage regression and including them in the ivregress function:

ivregress 2sls depvar *list of exogenous variables first_stage_residuals* (*x1=list of instruments*)
test *first_stage_residuals*

The tests for overidentifying instruments (exogeneity of Z) can be easily done after estimation by using the command

estat overid

Alternatively, you can carry out the three tests by hand by constructing the appropriate regressions.

R commands

Regression of a specific subperiod can be easily done using a command such as:

lm(y~x1+x2,… , subset=(year<=1973))

which is self-explanatory.

Instrumental variables regressions can be done using:

library(AER)
ivreg(depvar ~x1+x2+w1+w2+... | z1+z2+w1+w2+..., data=df)

which, nicely, will also do the Hausman, first-stage and overid tests for you by typing:
summary(iv_model, diagnostics=TRUE)

Alternatively, you can carry out the three tests by hand by constructing the appropriate regressions.