MOOC Econometrics

Test Exercise 4

**Notes:**

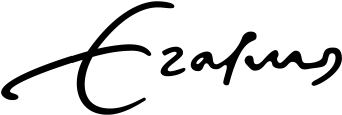
* See website for how to submit your answers and how feedback is organized

**Goals and skills being used:**

* Practice with identifying causes of endogeneity
* Practice with identifying valid instruments
* Obtain insight in the logic behind the 2SLS estimator

**Questions**

To run a study on the effect of a new diet a researcher runs a survey. The three most important questions in this survey are: (1) What was your weight one year ago? (2) What is your current weight? (3) Did you follow this diet in the past year?

We denote the anwers of individual *i* = 1,*...*,*n* to the first two questions with *yi*0 and *yi*1, the answer to the third question is denoted by *di* (where *di* = 1 if the diet was followed). Furthermore, some background characteristics of the respondents are collected. These characteristics are combined in the vector *xi*. Assume that all respondents are perfectly able to correctly answer the questions in the survey and do so truthfully.

(a) First of all the researcher uses OLS to estimate the parameters of the model

*yi*.

The OLS estimator for *β* is possibly not consistent as the variable *di* may be endogenous. Clearly explain why this may be the case. Indicate whether your reason would lead OLS to overestimate or underestimate the true effect of the diet.

The researcher finds out that in some regions of the country the diet was promoted via door-to-door advertising. The researcher manages to construct a variable *zi* that indicates whether individual *i* does (*zi* = 1) or does not (*zi* = 0) live in a region in which the diet was advertised.

1. In general there are two important conditions for variables *Z* to be useful as instruments. In formal terms these conditions are 0 and = 0 as the sample size *n* grows large. Rephrase these two conditions in words in the context of this application for the above mentioned advertising variable (no formulas!).
2. For both assumptions in (b), indicate whether it can be tested statistically given the available variables. If yes, indicate how. If no, why not?
3. Suppose that *zi* satisfies the conditions in (b) and suppose that *zi* is uncorrelated with *yi*0 and *xi*. In this case the 2SLS-estimator for *β* in the model *yi*1− *yi*0 = *α* + *βdi* + *ηi* is consistent when a constant and *zi* are used as instruments.

Show that we can write this 2SLS estimator for *β* in terms of simple sample averages. You can use the following averages:

* + Average weight change over all individuals: *∆*
  + Average weight change over individuals with *zi* = 1: *∆*1
  + Average weight change over individuals with *zi* = 0: *∆*0
  + Proportion of people taking the diet: *d¯*
  + Proportion of people with *zi* = 1 taking the diet: *d¯*1
  + Proportion of people with *zi* = 0 taking the diet: *d¯*0To further explain the notation, for example:

**Hint:** start with the formula: (*Z*’*X*)−1*Z*’*y*.  
=========================================================================================  
**Answers**.  
  
(a) The variable di may be endogenous since the individuals may decide to diet or not when they see the diet advertisements. In general, a person is more likely to diet when living in some regions of the country the diet was promoted via door-to-door advertising. Thus, the omitted variables, such as whether living a advertsing region, can cause endogenuity.  
  
Since the diet advertisement usually can help to lose more weight, it may lead OLS to under-estimate the true effect of the diet.  
  
(b) Two conditions for variable Z (whether in a region where the diet was advertised):  
 1) Whether living an advertised region(Z) affects (correlated or not) the annual weight change directly.  
 2) Whether living an advertised region(Z) affects the diet decision(di).  
  
(c) The first conditions in (b) can be tested with Sargan test, but need more instruments.  
 The second condition is tested by regressing the diet (di) against the advertisement (Z), , the current weight(y0), and the combined characteristics(xi). This is the 1st stage of Sargan Test, and its purpose is to test the significance of instrument variable(Z).  
  
To test the instrument validity,  
 1) 1st stage is to regress the diet (di) against constant, the advertisement(Z), the current weight(y0), and the combined characteristics(xi).  
 2) 2nd stage is to regress the annual weight change (yi1-yi0) agianst constant, the current weight(y0), the combined characteristics(xi), and diet-fit ().  
 3) Check the correclation between 2SLS residulas and instruments. If they are correlated, then the instruments are not valid.  
 Regress 2SLS residual against constant, current weight(yi0), combined characteristics(xi), and advertisement(Z).  
 Sargan test for H0: (valid instrument, or the coefficients of Z is 0)  
 For example, we can not reject H0 if sargan test statistic = nR2 < (χ2 critical value for a specific level)[[1]](#footnote-1).  
  
(d) Let’s start with the formula: (*Z*’*X*)−1*Z*’*y*.  
  
 …… (Refer to Training Exercize 4.3)  
  
 where and .  
  
 and   
  
 Thus,   
   
   
 Since and , we get

1. For χ2 test, the significance level is usually 5%, and the degree of freedom is (m-k), where m = # of instruments(Z), and k = # of explanatory variables. [↑](#footnote-ref-1)