**Week 3 Lab Handout- Instrumental Variables and Two-Stage Lease Squares**

**PA 5033 – Multivariate Techniques**

March 26th, 2021

Wenchen Wang

**PART A: TWO-STAGE LEASE SQUARES ~25 min**

**PART B: STATA 2SLS SHORTCUT AND THE ORDER CONDITION~25 min**

Retrieve data (**Prtime.dta**) from class CANVAS site.

**PART A: TWO-STAGE LEASE SQUARES**

1. **Ordinary Least Squares (OLS) Regression and Simultaneity Bias:**

We are interested in examining the effect of wages on employment and employment on wages. However, wages and employment may be determined at the same time, causing bias in the coefficients and biasing downward the standard errors. This is called simultaneity bias. To correct for this, we will use two-stage least squares.

We will consider the following system of structural equations:

**Equation 1**: ***prepop = f(avewage, prgnpt-1, usgnp, yr)***

**Equation 2**: ***avewage = f(prepop, prgnp)***

Run each of these regressions. Note that ***prepop*** and ***avewage*** are simultaneously determined (endogenous).







How do you interpret the coefficients on the main independent variables in these two equations? Do they meet your assumption?

What might be wrong with this model?

1. **Two Stage Least Squares Regression, Stage 1: Get instruments to replace endogenous variables (*prepop*, *avewage*):**

Next, we will run regressions on the reduced-form equations for each of the endogenous variables that appear as explanatory variables in the structural equations in the system.

**Equation System**

**Structural Equation 1**: ***prepop = f(avewage, prgnpt-1, usgnp, yr)***

**Structural Equation 2**: ***avewage = f(prepop, prgnp)***

Remember that to run reduced form equations you must perform a regression with each endogenous variable regressed on all of the exogenous and lagged endogenous variables in the equation system. That is, all exogenous or lagged endogenous variables from every structural equation must be included in the regression for every reduced form regression.

**Reduced Form Regressions**

**Reduced Form Equation 1:**

**Reduced Form Equation 2:**

1. **Stage 1: Create instrument to replace *prepop*:**

Estimate an OLS equation using ***prepop***as the dependent variable and ***prgnp, prgnpt-1, usgnp,*** and ***yr*** (all predetermined/exogenous variables).



Save the predicted values after running the regression using the “predict” command shown below. The new variable created from this regression, ***pre\_1***, is an instrumental variable for ***prepop***.

Note: Predicting “yhat” calculates the value of the dependent variable based on the results of the most recent regression. In this case, the dependent variable is prepop. This new variable will, by default, be named “yhat”. Use the “rename” command below to rename the variable “pre\_1”. This is the instrumental variable we will use in the structural equation later.



1. **Stage 1: Create instrument to replace *avewage*:**

Estimate an OLS equation using ***avewage*** as the dependent variable and ***prgnp, prgnpt-1, usgnp,*** and ***yr*** as the independent variables. Again, save the predicted values. The new instrumental variable created will be **ave\_1**.



Using the same commands as above, save the predicted value of “avewage” and rename it “ave\_1”. This is the instrumental variable we will use in the structural equation later.



1. **Two Stage Least Squares Regression, Stage 2: Use the instruments in the original structural equations:**
2. To correct for the simultaneity bias in **Equation 1**, use the same model but replace the actual variable ***avewage*** with the **predicted values** of ***avewage*** (***ave\_1***). More specifically, run a regression with ***prepop*** as the dependent variable and ***ave\_1, prgnp\_1, usgnp,*** and ***yr*** as the dependent variables. What do your results show?

)

1. To correct for the simultaneity bias in **Equation 2***,* use the same model but replace the actual variable ***prepop***with the **predicted values** of ***prepop*** (***pre\_1***). More specifically, run a regressionwith ***avewage*** as the dependent variable and ***pre\_1,*** and ***prgnp*** as the independent variables. What do your results show?

)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| VARIABLES | OLS 1 | OLS 2 | 2SLS 1 | 2SLS 2 |
|  |  |  |  |  |
| avewage | -0.0412\*\*\* |  |  |  |
|  | (0.00681) |  |  |  |
| prgnp\_1 | 3.04e-05\*\* |  | 3.18e-05\*\* |  |
|  | (1.16e-05) |  | (1.23e-05) |  |
| usgnp | 0.000135\*\*\* |  | 0.000159\*\*\* |  |
|  | (2.79e-05) |  | (3.20e-05) |  |
| yr | -0.00969\*\*\* |  | -0.00984\*\*\* |  |
|  | (0.00190) |  | (0.00200) |  |
| prepop |  | -11.15\*\*\* |  |  |
|  |  | (2.544) |  |  |
| prgnp |  | 0.000935\*\*\* |  | 0.000864\*\*\* |
|  |  | (7.49e-05) |  | (7.90e-05) |
| avewage\_IV |  |  | -0.0541\*\*\* |  |
|  |  |  | (0.00984) |  |
| prepop\_IV |  |  |  | -15.91\*\*\* |
|  |  |  |  | (3.049) |
| Constant | 0.742\*\*\* | 3.994\*\*\* | 0.717\*\*\* | 6.006\*\*\* |
|  | (0.0749) | (1.150) | (0.0798) | (1.360) |
|  |  |  |  |  |
| Observations | 37 | 38 | 37 | 37 |
| R-squared | 0.862 | 0.945 | 0.848 | 0.952 |

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

How do the results change compared with OLS equations?

**PART B: STATA 2SLS SHORTCUT AND THE ORDER CONDITION**

1. **Two Stage Least Squares of the STATA “Shortcut”:**

Run a two-stage least squares regression of ***prepop*** as a function of ***avewage,*** and ***prgnpt-1, usgnp,*** and ***yr***. The command to run regressions using instrumental variables in STATA is “ivreg”. After the “ivreg” command, the syntax is basically the same as with the normal regress command. In addition, when instrumenting an endogenous variable with the “ivregress” command, you must tell STATA how the instrumental variable will be calculated. Effectively, add the Reduced Form equation in parentheses to the end of the command.

We will need to create an instrument for the endogenous variable ***avewage***. We will calculate the instrument as a function of ***prgnp, prgnpt-1, usgnp,*** and ***yr***.

Make sure, you do NOT write the instrumental variable twice in the regression.



Run a two-stage least squares regression of ***avewage*** as a function of ***prepop,*** and ***prgnp***. Use the “ivreg” command as we did above. We will need to create an instrument for the endogenous variable ***prepop***. We will calculate the instrument as a function of ***prgnp, prgnpt-1, usgnp,*** and ***yr***. Run that regression.



1. **Two Stage Least Squares and the Order Condition:**

Let’s try a slightly different regression but disregard the use of ***prgnp*.** Run a two-stage least squares regression of ***prepop*** as a function of ***avewage, prgnpt-1, usgnp,*** and ***yr***.

Use ***prepop*** as the dependent variable, and ***avewage, prgnpt-1, usgnp,*** and ***yr*** as the explanatory variables, with ***avewage*** being a function of ***prgnpt-1, usgnp,*** and ***yr*.**

**Structural Equation 1**: ***prepop = f(avewage, prgnpt-1, usgnp, yr)***

**Structural Equation 2**: ***avewage = f(prepop)***

In this case, our first-stage 2SLS equations will be as below:

**2SLS Equation 1**: ***prepop = f(prgnpt-1, usgnp, yr)***

**2SLS Equation 2**: ***avewage = f(prgnpt-1, usgnp, yr)***



Why won’t STATA run this regression?