Instrumental Variables Estimation

Jing Bu

11/29/2018

目录

1 Instrumental Variables in Simple Regression Models

- We only analyze women with non-missing wage, so we extract a subset from our data, we want to estimate the return to education for these women. As an instrumental variable for education, we use the education of her father: fathereduc. First, we caculate the OLS and IV slope parameters, respectively. Remember that the *with* command defines that all variables names refer to our data frame oursample. Then the full OLS and IV estimates are caculated using the boxed routines lm and ivreg, respectively.

1.1 restrict to non-missing wage observations

```
#install.packages("AER")
library("foreign")
library("AER")
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

```
mroz<-read.dta("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assistant
oursample <- subset(mroz, !is.na(wage))
```

1.2 OLS slope parameter manually

```
with(oursample, cov(log(wage),educ) / var(educ) )
```

```
## [1] 0.1086487
```

1.3 IV slope parameter manually

```
with(oursample, cov(log(wage),fatheduc) / cov(educ,fatheduc) )
```

```
## [1] 0.05917348
```

1.4 OLS automatically

```
reg.ols <-lm(log(wage) ~ educ, data=oursample)
```

1.5 IV automatically

```
reg.iv <-ivreg(log(wage) ~ educ | fatheduc, data=oursample)
```

1.6 Pretty regression table

```
library("stargazer")
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics T
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(reg.ols,reg.iv, type="text")
```

```
##
## ===============================================================
##                                       Dependent variable:
##                                   -----------------------------
##                                           log(wage)
##                                     OLS           instrumental
##                                                     variable
##                                     (1)               (2)
## ---------------------------------------------------------------
## educ                              0.109***           0.059*
##                                   (0.014)           (0.035)
##
## Constant                          -0.185             0.441
##                                   (0.185)           (0.446)
##
## ---------------------------------------------------------------
## Observations                        428               428
## R2                                 0.118             0.093
## Adjusted R2                        0.116             0.091
## Residual Std. Error (df = 426)     0.680             0.689
## F Statistic                56.929*** (df = 1; 426)
## ===============================================================
## Note:                                *p<0.1; **p<0.05; ***p<0.01
```

2 More Exogenous Regressors

2.1 Using college proximity as an IV for education

- We use *CARD.dta* to estimate the return to education. Education is allowed to be endogenous and instrumented with the dummy variable *nearc4* which indicates whether the individual grew up close to a college. In addition, we control for experience, race, and regional

information. These variables are assumed to be exogenous and act as their own instruments. We first check for relevance by regressing the endogenous independent variable *educ* on all exogenous variables including the instrument *nearc4*. Its parameter is highly signficantly different from zero,so relevance is supported. We then estimate the log wage equation with OLS and IV. All results are displayed in one table with **stargazer**.

2.1.1 Checking for relevance: reduced form

```
library("AER")
library("foreign")
card<-read.dta("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assistant
```

```
redf<-lm(educ ~nearc4+exper+I(exper^2)+black+smsa+south+smsa66+reg662+reg663+reg664+reg
```

2.1.2 OLS

```
ols<-lm(log(wage)~educ+exper+I(exper^2)+black+smsa+south+smsa66+reg662+reg663+reg664+re
```

2.1.3 IV estimation

```
iv <-ivreg(log(wage)~educ+exper+I(exper^2)+black+smsa+south+smsa66+reg662+reg663+reg664
```

2.1.4 Pretty regression table of selected coefficients

```
library("stargazer")
stargazer(redf,ols,iv,type="text",keep=c("educ","nearc4","exper","black"),keep.stat=c("
```

```
##
## ===============================================
##                         Dependent variable:
##                     -----------------------------
```

```
##                 educ          log(wage)
##                  OLS       OLS    instrumental
##                                      variable
##                  (1)       (2)        (3)
## ------------------------------------------------
## nearc4       0.320***
##               (0.088)
##
## educ                    0.075***    0.132**
##                          (0.003)    (0.055)
##
## exper        -0.413*** 0.085***    0.108***
##               (0.034)   (0.007)    (0.024)
##
## I(exper2)      0.001   -0.002***  -0.002***
##               (0.002)  (0.0003)    (0.0003)
##
## black        -0.936*** -0.199***  -0.147***
##               (0.094)   (0.018)    (0.054)
##
## ------------------------------------------------
## Observations   3,010     3,010      3,010
## R2             0.477     0.300      0.238
## ================================================
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

3 Two Stage Least Squares

- Two stage least squares(2SLS) is a general approach for IV estimation when we have one or more endogenous regressors and at least as many additional instrumental variables.

3.1 Return to education for working women

- We still want to estimate the return to education for women using the data in *MROZ.dta*. Now,we use both mother's and father's education as instruments for the own education.In the following script,we obtain 2SLS estimates in two ways:First,we do both stages manually,including fitted education as **fitted(stage1)** as a regressors in the second stage. **ivreg**does this automatically and delivers the same parameter estimates as the output table reveals. AABut the standard errors differ slightly because the manual two stage version did not correct them.

3.1.1 restrict to non-missing wage observations

```
library("AER")
library("foreign")
oursample <- subset(mroz, !is.na(wage))
```

3.1.2 1st stage: reduced form

```
stage1 <- lm(educ~exper+I(exper^2)+motheduc+fatheduc, data=oursample)
```

3.1.3 2nd stage

```
man.2SLS<-lm(log(wage)~fitted(stage1)+exper+I(exper^2), data=oursample)
```

3.1.4 Automatic 2SLS estimation

```
aut.2SLS<-ivreg(log(wage)~educ+exper+I(exper^2) | motheduc+fatheduc+exper+I(exper^2),da
```

3.1.5 Pretty regression table

```
library("stargazer")
stargazer(stage1,man.2SLS,aut.2SLS,type="text",keep.stat=c("n","rsq"))
```

```
##
## ===============================================
##                       Dependent variable:
##                   ----------------------------
##                     educ          log(wage)
##                     OLS       OLS    instrumental
##                                          variable
##                     (1)       (2)         (3)
## -----------------------------------------------
## fitted(stage1)             0.061*
##                           (0.033)
##
## educ                                     0.061*
##                                         (0.031)
##
## exper              0.045   0.044***    0.044***
##                   (0.040)  (0.014)     (0.013)
##
## I(exper2)         -0.001   -0.001**    -0.001**
##                   (0.001)  (0.0004)    (0.0004)
##
## motheduc          0.158***
##                   (0.036)
##
## fatheduc          0.190***
##                   (0.034)
##
## Constant          9.103***  0.048       0.048
##                   (0.427)  (0.420)     (0.400)
##
```

```
## ------------------------------------------
## Observations     428      428       428
## R2              0.211    0.050     0.136
## ==========================================
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

4 Testing for Exogeneity of the Regressor

- There is another way to get the same IV parameter estimates as with 2SLS.In the same setup as above,this "control function approach"also consist of two stages.
- We use both mother's and father's education as instruments.The first stage regression is identical as in last section.The second stage adds the first stage residuals to the original list of regressors.The parameter estimates are identical to both the manual 2SLS and the automatic **ivreg**results.We can directly interpret the t test from the regression table as a test for exogeneity. Here,$t = 1.6711$ with a two-sided p value of $p = 0.095$, indicating a marginally significant evidence for endogeneity.

```
library("AER")
library("lmtest")
```

4.1 1st stage: reduced form

```
stage1<-lm(educ~exper+I(exper^2)+motheduc+fatheduc, data=oursample)
```

4.2 2nd stage

```
stage2<-lm(log(wage)~educ+exper+I(exper^2)+resid(stage1),data=oursample)
```

4.3 results including t tests

```
coeftest(stage2)
```

```
##
## t test of coefficients:
##
##                   Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)    0.04810030  0.39457526  0.1219 0.9030329
## educ           0.06139663  0.03098494  1.9815 0.0481824 *
## exper          0.04417039  0.01323945  3.3363 0.0009241 ***
## I(exper^2)    -0.00089897  0.00039591 -2.2706 0.0236719 *
## resid(stage1)  0.05816661  0.03480728  1.6711 0.0954406 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```