ØAMET4100 · Spring 2019
Worksheet 7

Instructor: Fenella Carpena

March 7, 2019

1 IV Regression

Exercise 1.1 (Adapted from Stock & Watson, Review the Concepts 12.1) Consider the problem of estimating the elasticity of supply for butter using the supply equation

$$ln(Q_i^{butter}) = \gamma_0 + \gamma_1 ln(P_i^{butter}) + v_i.$$

In this regression model, is $ln(P_i^{butter})$ positively or negatively correlated with the error v_i? If γ_1 is estimated by OLS, would you expect the estimated value to be larger or smaller than the true value of γ_1? Explain.

Exercise 1.2 Consider the problem of estimating the elasticity of demand for cigarettes using the demand equation

$$ln(Q_i^{cigs}) = \beta_0 + \beta_1 ln(P_i^{cigs}) + u_i.$$

Suppose that we used as an instrument for $ln(P_i^{cigs})$ the number of trees per capita in the state. Is this instrument relevant? Is it exogenous? Is it a valid instrument?

Exercise 1.3 (Adapted from Stock & Watson, Review the Concepts 12.3) Consider a study on the effect of incarceration (imprisonment) on crime rates. Specifically, we want to examine whether putting criminals in jail reduces crime.

(a) One strategy for estimating this effect is to regress *crime_rates* (crimes per 100,000 member of the general population against *incarceration_rate* (prisoners per 100,000) using annual data from U.S. states. Explain why this regression is subject to bias.

(b) Suppose that we use the number of lawyers per capita as an instrument for $incarceration\_rate$. Would this instrument be relevant? Would it be exogenous? Would it be a valid instrument?

Exercise 1.4 Consider the following regression of an individual i's wages on his/her years of education.

$$wage_i = \beta_0 + \beta_1 educ_i + u_i$$

Explain whether the following two candidate IVs are valid instruments for $educ_i$.

(a) The individual's IQ score

(b) The years of education of the individual's mother

Exercise 1.5 (Stock & Watson, Exercise 12.7) In an IV regression model with one regressor, X_i, and two instruments, Z_{1i} and Z_{2i}, the value of the J-statistic is $J = 18.2$. If you need to conduct a hypothesis test, use the 1% significance level.

(a) Does this suggest that $E(u_i|Z_{1i}, Z_{2i}) \neq 0$? Explain.

(b) Does this suggest that $E(u_i|Z_{1i}) \neq 0$? Explain.

Exercise 1.6 (Stock & Watson, Exercise 12.9) A researcher is interested in the effect of military service on human capital. He collects data from a random sample of 4000 workers aged 40 and runs the OLS regression $Y_i = \beta_0 + \beta_1 X_i + u_i$, where Y_i is the worker i's annual earnings and X_i is a binary variable that is equal to 1 if the person served in the military and 0 otherwise.

(a) Explain why the OLS estimates are likely to unreliable. (*Hint:* Which variables are omitted from the regression? Are they correlated with military service?)

(b) During the Vietnam War there was a draft, where priority for the draft was determined by a national lottery. (The days of the year were randomly reordered 1 through 365. Those with birthdates ordered first were drafted before those with birthdates ordered second, and so forth.) Explain how the lottery might be used as an instrument to estimate the effect of military service on earnings.

Exercise 1.7 As a credit risk analyst at a financial company, you have been tasked with researching whether having a credit card causes individuals to spend more money. You analyze records from 13,444 individuals, which contain the following variables:

- *lexpend*: natural log of average monthly expenditures over the last 12 months
- *age*: age in years
- *lincome* : natural log of annual income
- *ownrent*: a dummy variable equal to one if the individual is a home-owner and zero if he/she rents
- *cardhldr*: a dummy variable equal to one if the individuals has a credit card
- *gas*: number of gas stations in the individual's zip code
- *eatdrink*: number of eating and drinking establishments (e.g., restaurants) in the individual's zip code

The Stata results from the analysis can be found below on pages 4-5.

(a) Consider the OLS regression of *lexpend* on *age*, *lincome*, *ownrent*, and *cardhldr*. Interpret the coefficient on *cardhldr*.

(b) Give an example of an omitted variables in the regression in part (a) that would cause the coefficient on *cardhldr* to be biased. Would this omitted variable lead to upward or downward bias in *cardhldr*?

(c) Consider the variables *gas* and *eatdrink* as instruments for *cardhldr*. Are these instruments relevant? Explain.

(d) Are the two candidate instruments in part (c) jointly exogenous? Explain.

```
. regress lexpend age lincome i.ownrent i.cardhldr , robust

Linear regression                                      Number of obs =   13444
                                                       F(  4, 13439) =28193.51
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.7480
                                                       Root MSE      =  1.1907

------------------------------------------------------------------------------
             |               Robust
     lexpend |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0123622   .0012148   -10.18   0.000    -.0147434   -.0099809
     lincome |   .8992942   .0277035    32.46   0.000     .8449915    .9535969
   1.ownrent |  -.1704012   .0232584    -7.33   0.000     -.215991   -.1248115
  1.cardhldr |   4.736288   .0157039   301.60   0.000     4.705506     4.76707
       _cons |   -6.24207   .2709356   -23.04   0.000    -6.773142   -5.710998
------------------------------------------------------------------------------
. regress cardhldr gas eatdrink , r

Linear regression                                      Number of obs =   13444
                                                       F(  2, 13441) =   11.07
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0022
                                                       Root MSE      =  .41319

------------------------------------------------------------------------------

             |               Robust
    cardhldr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gas |  -.0066057   .0024981    -2.64   0.008    -.0115024   -.0017091
    eatdrink |  -.0037578   .0010356    -3.63   0.000    -.0057877   -.0017279
       _cons |   .8176533   .0084165    97.15   0.000     .8011558    .8341509
------------------------------------------------------------------------------
```

```
. ivregress 2sls lexpend age lincome i.ownrent (cardhldr = gas eatdrink) , robust

Instrumental variables (2SLS) regression          Number of obs =   13444
                                                  Wald chi2(4)  = 2093.17
                                                  Prob > chi2   =  0.0000
                                                  R-squared     =  0.5131
                                                  Root MSE      =  1.6547

------------------------------------------------------------------------------
             |               Robust
     lexpend |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cardhldr |   7.559458   1.027133     7.36   0.000     5.546315    9.572601
         age |  -.0089305   .0020862    -4.28   0.000    -.0130193   -.0048416
     lincome |   .4721932   .1590177     2.97   0.003     .1605243    .7838621
   1.ownrent |   -.285823   .0535069    -5.34   0.000    -.3906946   -.1809513
       _cons |  -4.148696   .8400043    -4.94   0.000    -5.795075   -2.502318
------------------------------------------------------------------------------
Instrumented:  cardhldr
Instruments:   age lincome 1.ownrent gas eatdrink

. predict u2slshat , resid

. regress u2slshat age lincome ownrent gas eatdrink

      Source |       SS       df       MS              Number of obs =   13444
-------------+------------------------------           F(  5, 13438) =    0.73
       Model |  10.0298087     5  2.00596175           Prob > F      =  0.5989
    Residual |  36798.2906 13438  2.73837555           R-squared     =  0.0003
-------------+------------------------------           Adj R-squared = -0.0001
       Total |  36808.3204 13443  2.73810313           Root MSE      =  1.6548

------------------------------------------------------------------------------
    u2slshat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   2.62e-06   .0015645     0.00   0.999     -.003064    .0030693
     lincome |   -.000407   .0340592    -0.01   0.990    -.0671677    .0663538
     ownrent |  -.0042808   .0318055    -0.13   0.893    -.0666241    .0580625
         gas |   .0104032   .0079875     1.30   0.193    -.0052534    .0260598
    eatdrink |  -.0053187   .0036228    -1.47   0.142    -.0124198    .0017825
       _cons |   .0231057   .3410017     0.07   0.946    -.6453055    .6915168
------------------------------------------------------------------------------

. test gas eatdrink

 ( 1)  gas = 0
 ( 2)  eatdrink = 0

       F(  2, 13438) =    1.83
            Prob > F =    0.1602
```

TABLE 3 Critical Values for the χ^2 Distribution

Degrees of Freedom	Significance Level		
	10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21
11	17.28	19.68	24.72
12	18.55	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.59	33.41
18	25.99	28.87	34.81
19	27.20	30.14	36.19
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.41	42.98
25	34.38	37.65	44.31
26	35.56	38.89	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59
30	40.26	43.77	50.89

This table contains the 90^{th}, 95^{th}, and 99^{th} percentiles of the χ^2 distribution. These serve as critical values for tests with significance levels of 10%, 5%. and 1%.