计量经济学：作业三

刘泓尊　2018011446　计 84

liu-hz18@mails.tsinghua.edu.cn

2020.12

1. (a)

$$\beta_{IV} = \frac{\sum_{i=0}^{n}(y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=0}^{n}(x_i - \bar{x})(z_i - \bar{z})}$$

设 $z_i = 1$ 的那部分样本有 k 个, 则 $z_i = 0$ 的那部分样本有 $n - k$ 个。

所以分子:

$$\begin{aligned}\sum_{i=0}^{n}(y_i - \bar{y})(z_i - \bar{z}) &= k(y_1 - \bar{y})(1 - \bar{z}) + (n-k)(y_0 - \bar{y})(-\bar{z}) \\ &= k(1 - \frac{k}{n})(y_1 - \frac{k\bar{y_1} + (n-k)\bar{y_0}}{n}) - (n-k)\frac{k}{n}(\bar{y_0} - \frac{k\bar{y_1} + (n-k)\bar{y_0}}{n}) \\ &= k(1 - \frac{k}{n})^2(\bar{y_1} - \bar{y_0}) + (n-k)(\frac{k}{n})^2(\bar{y_1} - \bar{y_0}) \\ &= (\bar{y_1} - \bar{y_0})[k(1 - \frac{k}{n})^2 + (n-k)(\frac{k}{n})^2]\end{aligned} \tag{1}$$

同理得到分母：

$$\sum_{i=0}^{n}(x_i - \bar{x})(z_i - \bar{z}) = (\bar{x_1} - \bar{x_0})[k(1 - \frac{k}{n})^2 + (n-k)(\frac{k}{n})^2]$$

所以 IV 估计量可以写成:

$$\beta_{IV} = \frac{\hat{y_1} - \hat{y_0}}{\hat{x_1} - \hat{x_0}}$$

(b) 是一个合理的 IV

设第一胎是否是双胞胎即为 z. OLS 得到误差为 u，儿女数目为 x，那么在满足

$$\begin{aligned}cov(u, z) &= 0 \\ cov(x, z) &\neq 0\end{aligned} \tag{2}$$

的时候，双胞胎可以作为儿女数目的 IV。

从第一问可以得出来，只需要统计第一胎是双胞胎 ($z_i = 1$) 的样本，得到其儿女数目均值 $\bar{x_1}$ 和女性就业率的均值 $\bar{y_1}$，以及第一胎不是双胞胎 ($z_i = 0$) 的样本的儿女数目均值 $\bar{x_0}$ 和女性就业率的均值 $\bar{y_0}$。进一步构造 IV 估计量：

$$\beta_{IV} = \frac{\hat{y_1} - \hat{y_0}}{\hat{x_1} - \hat{x_0}}$$

即可。

2. (a) 回归结果为

```
. * Estimate original model, get residuals
. reg colGPA hsGPA ACT skipped PC

      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(4, 136)       =     11.90
       Model |  5.03143078         4  1.2578577   Prob > F        =    0.0000
    Residual |  14.3746687       136  .105696093   R-squared       =    0.2593
-------------+----------------------------------   Adj R-squared   =    0.2375
       Total |  19.4060994       140  .138614996   Root MSE        =    .32511

------------------------------------------------------------------------------
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsGPA |   .4129522   .0924311     4.47   0.000     .2301642    .5957403
         ACT |    .013344   .0104437     1.28   0.204    -.0073091    .0339972
     skipped |  -.0710336   .0262494    -2.71   0.008    -.1229435   -.0191238
          PC |   .1244391   .0573115     2.17   0.032     .0111021    .2377762
       _cons |   1.356509   .3275021     4.14   0.000     .7088537    2.004164
------------------------------------------------------------------------------
```

图 1: OLS 回归结果

同方差下标准误:

hsGPA	0.0924311
ACT	0.0104437
skipped	0.0262494
PC	0.0573115
_cons	0.3275021

(b) 残差结果如下：

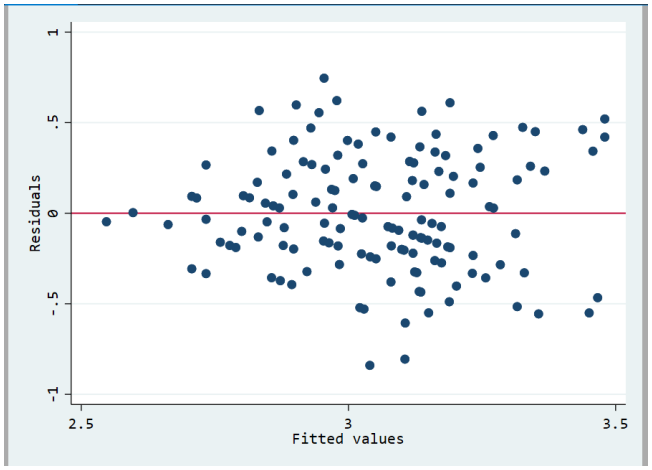

图 2: 残差图

从数据上看，可能存在异方差性。因为在拟合值较小的区域残差的波动较小，但是在拟合值较大的区域残差波动较大。

(c) 进行 BP 检验得到的结果如下：

```
. * BP test
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of colGPA

         chi2(1)      =     5.27
         Prob > chi2  =   0.0217
```

图 3:

从图中可以看到 $p = 0.0217 < \alpha = 0.05$. 所以拒绝原假设 H_0.

(d) 方程的拟合结果如下：

```
. reg u2 colGPA colGPA2

      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(2, 138)       =     44.84
       Model |  .874631551         2  .437315775   Prob > F        =    0.0000
    Residual |  1.34599908       138  .009753617   R-squared       =    0.3939
-------------+----------------------------------   Adj R-squared   =    0.3851
       Total |  2.22063063       140  .015861647   Root MSE        =    .09876

------------------------------------------------------------------------------
          u2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      colGPA |  -2.856173   .308956    -9.24   0.000    -3.467072   -2.245273
     colGPA2 |   .4631145   .0494299    9.37   0.000     .3653765    .5608525
       _cons |   4.441601   .4782571    9.29   0.000     3.495942    5.387261
------------------------------------------------------------------------------
```

图 4: WLS 权重

求出拟合值之后，通过 $countifhhat <= 0$ 来检验是否存在负的权重。得到的结果是 0，也就是 $\hat{h}_i$ 都是正数。

(e) 用 WLS 估计的回归结果为:

```
. * Reestimate the model using WLS
. reg colGPA hsGPA ACT skipped PC [weight = 1/hhat]
(analytic weights assumed)
(sum of wgt is 2,055.27277870609)

      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(4, 136)       =      3.55
       Model |  .864569374         4  .216142343   Prob > F        =    0.0086
    Residual |  8.27035658       136  .060811445   R-squared       =    0.0946
-------------+----------------------------------   Adj R-squared   =    0.0680
       Total |  9.13492595       140  .065249471   Root MSE        =    .2466

------------------------------------------------------------------------------
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsGPA |   .1673402   .0757196     2.21   0.029     .0176002    .3170803
         ACT |   .0085697   .0077582     1.10   0.271    -.0067727     .023912
     skipped |  -.0438268   .0216791    -2.02   0.045    -.0866986   -.0009551
          PC |   .0469237    .043948     1.07   0.288    -.0399862    .1338336
       _cons |   2.294579   .2702302     8.49   0.000     1.760183    2.828976
------------------------------------------------------------------------------
```

图 5:

同方差下标准误:

hsGPA	0.0757196
ACT	0.0077582
skipped	0.0216791
PC	0.0439480
_cons	0.2702302

(f) skipped 和 PC 的估计值变化如下：

	skipped	PC	p-value
OLS	-0.0710336	0.1244391	0.0
WLS	-0.0438268	0.049237	0.0086

从表中可以看到变化比较大。

p 值变大，所以显著性水平降低了。

(g) 用 WLS 估计的回归结果为:

```
. * Report WLS with heteroskedasticity-robust standard errors
. reg colGPA hsGPA ACT skipped PC [weight = 1/hhat], vce(robust)
(analytic weights assumed)
(sum of wgt is 2,055.27277870609)

Linear regression                               Number of obs     =        141
                                                F(4, 136)         =       4.18
                                                Prob > F          =     0.0032
                                                R-squared         =     0.0946
                                                Root MSE          =      .2466

------------------------------------------------------------------------------
             |               Robust
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsGPA |   .1673402   .0698415     2.40   0.018     .0292244     .305456
         ACT |   .0085697   .0070603     1.21   0.227    -.0053924    .0225318
     skipped |  -.0438268   .0208148    -2.11   0.037    -.0849895   -.0026642
          PC |   .0469237   .0394494     1.19   0.236    -.0310899    .1249373
       _cons |   2.294579   .2559692     8.96   0.000     1.788384    2.800774
------------------------------------------------------------------------------
```

图 6:

异方差稳健下的标准误:

hsGPA	0.0698415
ACT	0.0070603
skipped	0.0208148
PC	0.0394494
_cons	0.2559692

和 (5) 中标准误的变化不大。

3. (a) 回归结果如下：

```
. * Estimate OLS
. reg lwage educ smsa exper expersq motheduc
```

Source	SS	df	MS	Number of obs	=	2,657
				F(5, 2651)	=	163.25
Model	123.302306	5	24.6604611	Prob > F	=	0.0000
Residual	400.454412	2,651	.15105787	R-squared	=	0.2354
				Adj R-squared	=	0.2340
Total	523.756718	2,656	.197197559	Root MSE	=	.38866

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0813949	.0039528	20.59	0.000	.073644	.0891459
smsa	.1724038	.0171025	10.08	0.000	.1388682	.2059395
exper	.0879387	.0073632	11.94	0.000	.0735005	.1023769
expersq	-.0022682	.0003602	-6.30	0.000	-.0029745	-.0015619
motheduc	.0143374	.0026633	5.38	0.000	.009115	.0195597
_cons	4.352978	.0720664	60.40	0.000	4.211666	4.49429

图 7: OLS 回归结果

(b) nearc4 对 educ 的回归结果如下：

```
. * Verify first stage
. reg educ nearc4
```

Source	SS	df	MS	Number of obs	=	3,010
				F(1, 3008)	=	63.91
Model	448.604204	1	448.604204	Prob > F	=	0.0000
Residual	21113.4759	3,008	7.01910767	R-squared	=	0.0208
				Adj R-squared	=	0.0205
Total	21562.0801	3,009	7.16586243	Root MSE	=	2.6494

educ	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nearc4	.829019	.1036988	7.99	0.000	.6256912	1.032347
_cons	12.69801	.0856416	148.27	0.000	12.53009	12.86594

图 8:

从图中可以看到，nearc4 的系数不为 0 且 $F = 63.91 > 10$，所以满足工具变量的相关性。

(c) 我认为不满足。居住地附近有大学可能会通过影响区域经济等来影响收入。所以可能这个变量和 u 相关。因此这个变量可能是内生的。

(d) 用 2SLS 来估计上述模型的回归结果：

```
. * Estimate 2SLS
. ivregress 2sls lwage smsa exper expersq motheduc (educ = nearc4)

Instrumental variables (2SLS) regression          Number of obs   =      2,657
                                                  Wald chi2(5)    =     341.02
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0895
                                                  Root MSE        =     .42365

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1703064   .0516094     3.30   0.001     .0691538     .271459
        smsa |   .1355772   .0283071     4.79   0.000     .0800963    .1910582
       exper |   .1212062   .0208497     5.81   0.000     .0803415    .1620708
     expersq |  -.0023351   .0003945    -5.92   0.000    -.0031083   -.0015618
    motheduc |  -.0050222   .0115684    -0.43   0.664    -.0276959    .0176514
       _cons |    3.10564    .725764     4.28   0.000     1.683169    4.528111
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   smsa exper expersq motheduc nearc4
```

图 9:

(e) 2SLS 的估计值中，$educ, exper, expersq$ 的估计值比 OLS 大，其他变量比 OLS 估计值小。

2SLS 估计值的方差都大于 OLS 估计值的方差。

2SLS 和 OLS 估计的 p 值均为 0.0，所以显著性水平一致。

2SLS 的拟合优度降低了。

(f) 进行内生性检验的结果如下:

```
. * Endogenous test
. estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Durbin (score) chi2(1)          =   3.5592  (p = 0.0592)
  Wu-Hausman F(1,2650)            =  3.55458  (p = 0.0595)
```

图 10:

在 10% 的显著性水平下，因为 $p = 0.0595 < \alpha$, 所以拒绝原假设 H_0。