ECO 394M Homework 3

Steven Kim

Question 1

The -6.38 Wald estimate for the weeks worked variable means the causal effect of *More than 2 children* on $\mathbb{E}($*Weeks worked*$|$*More than 2 children*$)$ identified by the subpopulation for which by variation in *More than 2 children* is induced by a shift in sex of first two children in families with two or more children. In other words, for individuals whose fertility has been affected by their children's sex mix, *Weeks worked* is expected to decrease by 6.38 if they have more than two children. The variation of the *Weeks worked* from when the first two children's genders are the same to different divided by the variation of *More than 2 children* from when the first two children's genders are the same to different. The Z-statistic is -5.45, which means it is statistically significant at any conventional confidence level.

Question 2

(a) There might be other biases such as more smart students are more likely to choose to attend a choice school.

(b) Yes, as *grant* is not correlated with u_1.

(c) $choice = \delta_1 + \delta_2 faminc + \pi_1 grant + v_2$ where $E(v_2) = Cov(faminc, v_2) = Cov(grant, v_2) = 0$. $\pi_1 \neq 0$ is needed for *grant* to be partially correlated with *choice*. In words, *grant* has to have an effect on *choice*.

(d) $score = \beta_0 + \beta_1(\delta_1 + \delta_2 faminc + \pi_1 grant + v_2) + \beta_2 faminc + u_1$
$= \beta_0 + \beta_1\delta_1 + \beta_1\pi_1 grant + (\beta_1\delta_2 + \beta_2) faminc + u_1 + \beta_1 v_2$
We can see the (indirect) effect of *grant* on *score* (via *choice*).

Question 3

$$\begin{aligned}
\widehat{\beta} &= (X'P_zX)^{-1}X'P_zy \\
&= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y \\
&= (Z'X)^{-1}(Z'Z)(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'y \\
&= (Z'X)^{-1}Z'y
\end{aligned}$$

Question 4

(a) Since $E(u|x) = 0$, $E(xu) = E(x^2u) = E(x^3u) = 0$. Also, $Corr(x, x) \neq 0$, $Corr(x, x^2) \neq 0$, and $Corr(x, x^3) \neq 0$. Therefore, any of them each can be used as an instrument variable.

(b) First, regress x on x, x^2, and x^3. Since there is no correlation between u and any of the instrument variables, the fitted values would be given as

$$\widehat{x} = \delta_1 + \pi_1 x + \pi_2 x^2 + \pi_3 x^3$$

In the second stage, regress y on $\widehat{x}$:

$$\begin{aligned}
y &= \beta_1 + \beta_2\widehat{x} + u \\
&= \beta_1 + \beta_2(\delta_1 + \pi_1 x + \pi_2 x^2 + \pi_3 x^3) + u
\end{aligned}$$

Compare this to the OLS regression of y on x, x^2, x^3:

$$y = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3 + u'$$

They are essentially same and we would get the same results in OLS and 2SLS.

(c) OLS (which is the same as 2SLS) minimizes the objective function

$$S(b) = \sum_i (y_i - x_i b)^2 \text{ where } x_i = \begin{bmatrix} 1 & x_i & x_i^2 & x_i^3 \end{bmatrix}$$

GMM minimizes the objective function

$$S(b) = \left(\frac{1}{n}\sum_i g_i(b)\right)' \widehat{W} \left(\frac{1}{n}\sum_i g_i(b)\right) \text{ where } g_i(b) = \begin{bmatrix} (y_i - x_i b) \\ x_i(y_i - x_i b) \\ x_i^2(y_i - x_i b) \\ x_i^3(y_i - x_i b) \end{bmatrix} \text{ and } \widehat{W} = \left(\frac{1}{n}\sum_i g_i(b) g_i(b)'\right)^{-1}$$

Since they minimize different objective functions, they would be different estimators. In this case, since this IV model is over-identified, GMM estimator is more efficient than the 2SLS estimator.

Question 5

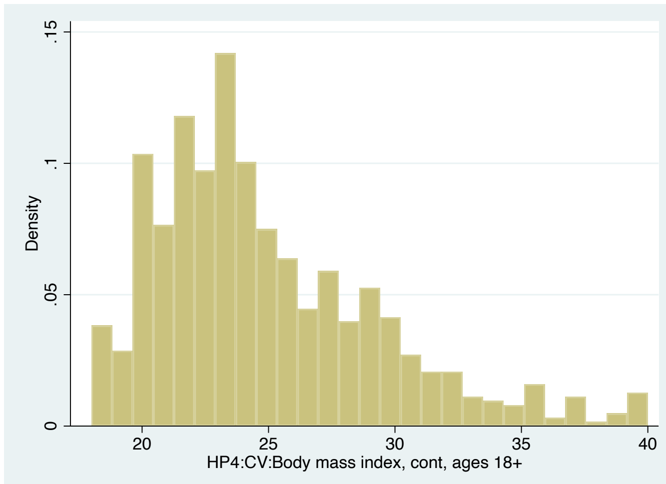
(a)

```
. use "children_sample.dta", clear

. keep if white & male
(1,023 observations deleted)

. tabstat bmi, statistics(p10 p25 p50 mean p75 p90)
```

Variable	p10	p25	p50	Mean	p75	p90
bmi	20.2	21.7	24	24.91896	27.4	30.95

(b)

```
. hist bmi
(bin=27, start=18, width=.81481481)
```



Yes, it makes sense. Mean is higher than median because of the righthand values away from the average. Other quantiles also seems right.

(c)

```
. regr bmi educ age mombmi dadbmi, r

Linear regression                               Number of obs     =        770
                                                F(4, 765)         =      28.24
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1393
                                                Root MSE          =     4.1536
```

bmi	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
educ	.003787	.1008632	0.04	0.970	-.1942145	.2017884
age	.2799552	.069457	4.03	0.000	.1436063	.4163041
mombmi	.18911	.0310735	6.09	0.000	.1281105	.2501095
dadbmi	.1728275	.0402927	4.29	0.000	.0937301	.2519249
_cons	9.045149	1.718142	5.26	0.000	5.672316	12.41798

The slope estimates tells us how a unit increase of each covariate would affect **bmi**, holding all others fixed.

(d)

```
. sqreg bmi educ age mombmi dadbmi, reps(500)
(fitting base model)

Bootstrap replications (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
```

```
..................................................  400
..................................................  450
..................................................  500
Simultaneous quantile regression                    Number of obs =        770
  bootstrap(500) SEs                                .50 Pseudo R2 =     0.0796

------------------------------------------------------------------------------
             |              Bootstrap
         bmi | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
q50          |
        educ |    .053927   .0916563     0.59   0.556    -.1260008    .2338548
         age |   .3358445    .068398     4.91   0.000     .2015744    .4701145
      mombmi |   .1315307   .0309848     4.25   0.000     .0707054    .1923561
      dadbmi |   .1777306   .0440507     4.03   0.000     .0912561    .2642051
       _cons |   7.681324   1.699646     4.52   0.000       4.3448    11.01785
------------------------------------------------------------------------------
```

i. A unit increase of **educ** would bring the median of **bmi** up by .053927, holding all others fixed.

ii.

```
. nlcom _b[mombmi] + _b[dadbmi], l(90)

       _nl_1: _b[mombmi] + _b[dadbmi]

------------------------------------------------------------------------------
         bmi | Coefficient  Std. err.      z    P>|z|     [90% conf. interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .3092614   .0458747     6.74   0.000     .2338042    .3847185
------------------------------------------------------------------------------
```

The median of **bmi** is expected to increase by .3092614, holding all others fixed. The 90% confidence interval for this effect is given in the Stata result.

iii.

```
. sqreg bmi educ age mombmi dadbmi, reps(500)
(fitting base model)

Bootstrap replications (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................   50
..................................................  100
..................................................  150
..................................................  200
..................................................  250
..................................................  300
..................................................  350
..................................................  400
..................................................  450
..................................................  500
Simultaneous quantile regression                    Number of obs =        770
  bootstrap(500) SEs                                .50 Pseudo R2 =     0.0796

------------------------------------------------------------------------------
             |              Bootstrap
         bmi | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
q50          |
        educ |    .053927   .0957065     0.56   0.573    -.1339516    .2418056
         age |   .3358445    .069755     4.81   0.000     .1989106    .4727784
      mombmi |   .1315307   .0303472     4.33   0.000      .071957    .1911045
      dadbmi |   .1777306   .0452244     3.93   0.000      .088952    .2665092
       _cons |   7.681324   1.726007     4.45   0.000     4.293052     11.0696
------------------------------------------------------------------------------
```

The standard errors and z-statistics are slightly different because it is from different bootstrap samples than before.

(e)

```
. sqreg bmi educ age mombmi dadbmi, q(.1 .25 .5 .75 .90) reps(500)
(fitting base model)

Bootstrap replications (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................   50
..................................................  100
..................................................  150
..................................................  200
..................................................  250
..................................................  300
```

```
............................................   350
............................................   400
............................................   450
............................................   500

Simultaneous quantile regression                Number of obs =        770
  bootstrap(500) SEs                            .10 Pseudo R2 =     0.0478
                                                .25 Pseudo R2 =     0.0771
                                                .50 Pseudo R2 =     0.0796
                                                .75 Pseudo R2 =     0.0755
                                                .90 Pseudo R2 =     0.1118
```

bmi	Coefficient	Bootstrap std. err.	t	P>|t|	[95% conf.	interval]
q10						
educ	.0772834	.1201185	0.64	0.520	-.1585175	.3130843
age	.1946482	.0705716	2.76	0.006	.0561113	.3331851
mombmi	.0804205	.0357001	2.25	0.025	.0103386	.1505023
dadbmi	.1088267	.0340647	3.19	0.001	.0419554	.175698
_cons	10.43543	1.834004	5.69	0.000	6.835156	14.03571
q25						
educ	.0720729	.1145663	0.63	0.529	-.1528287	.2969745
age	.2089615	.0737352	2.83	0.005	.0642142	.3537088
mombmi	.1358798	.0308052	4.41	0.000	.075407	.1963527
dadbmi	.1055299	.0388365	2.72	0.007	.0292911	.1817687
_cons	10.02518	1.713144	5.85	0.000	6.662158	13.3882
q50						
educ	.053927	.0969542	0.56	0.578	-.136401	.2442549
age	.3358445	.0710474	4.73	0.000	.1963735	.4753155
mombmi	.1315307	.0314753	4.18	0.000	.0697426	.1933189
dadbmi	.1777306	.0457923	3.88	0.000	.0878371	.2676242
_cons	7.681324	1.743933	4.40	0.000	4.257861	11.10479
q75						
educ	.042771	.1969077	0.22	0.828	-.3437726	.4293146
age	.3557238	.108574	3.28	0.001	.1425855	.5688622
mombmi	.2982242	.072172	4.13	0.000	.1565455	.4399029
dadbmi	.216229	.1000768	2.16	0.031	.0197713	.4126867
_cons	5.228177	2.998677	1.74	0.082	-.6584344	11.11479
q90						
educ	-.1577382	.2437446	-0.65	0.518	-.6362258	.3207494
age	.3642897	.1732562	2.10	0.036	.0241757	.7044037
mombmi	.3438461	.0715972	4.80	0.000	.2032958	.4843964
dadbmi	.2529488	.0793759	3.19	0.001	.0971284	.4087691
_cons	8.846463	5.218547	1.70	0.090	-1.397909	19.09083

i. The slope estimate of **educ** becomes negative for q90.

ii. Yes, it looks the original linear regression had heteroskedastic errors. It is obvious because if errors were homoskedastic, the slope estimates of any covariate in the quantile regression should be consistent over different τ values.

iii.

```
. testnl [q10]_b[age]=[q25]_b[age]=[q50]_b[age]=[q75]_b[age]=[q90]_b[age]

  (1)  [q10]_b[age] = [q25]_b[age]
  (2)  [q10]_b[age] = [q50]_b[age]
  (3)  [q10]_b[age] = [q75]_b[age]
  (4)  [q10]_b[age] = [q90]_b[age]

               chi2(4) =        4.16
           Prob > chi2 =        0.3848
```

It is 0.3848.

iv.

```
. testnl ([q50]_b[mombmi] = [q90]_b[mombmi]) ([q50]_b[dadbmi] = [q90]_b[dadbmi])

  (1)  [q50]_b[mombmi] = [q90]_b[mombmi]
  (2)  [q50]_b[dadbmi] = [q90]_b[dadbmi]

               chi2(2) =       10.89
           Prob > chi2 =        0.0043
```

P-value is 0.0043.

V.

```
. preserve

. collapse (mean) educ age mombmi dadbmi

. tempfile means

. save `means´
file /var/folders/sl/twzsfyy90bq0k_r245410_100000gn/T//S_27975.000004 saved as
    .dta format

. restore

. append using `means´
(variable educ was byte, now float to accommodate using data´s values)
(variable age was byte, now float to accommodate using data´s values)

. sqreg bmi educ age mombmi dadbmi, q(.1 .25 .5 .75 .90) reps(500)
(fitting base model)

Bootstrap replications (500)
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500

Simultaneous quantile regression                    Number of obs =        770
  bootstrap(500) SEs                                .10 Pseudo R2 =     0.0478
                                                    .25 Pseudo R2 =     0.0771
                                                    .50 Pseudo R2 =     0.0796
                                                    .75 Pseudo R2 =     0.0755
                                                    .90 Pseudo R2 =     0.1118
```

bmi	Coefficient	Bootstrap std. err.	t	P>|t|	[95% conf.	interval]
q10						
educ	.0772834	.1156081	0.67	0.504	-.1496634	.3042303
age	.1946482	.0685983	2.84	0.005	.059985	.3293114
mombmi	.0804205	.036183	2.22	0.027	.0093906	.1514503
dadbmi	.1088267	.0331613	3.28	0.001	.0437288	.1739246
_cons	10.43543	1.652902	6.31	0.000	7.190673	13.6802
q25						
educ	.0720729	.1169336	0.62	0.538	-.1574759	.3016218
age	.2089615	.0706963	2.96	0.003	.0701797	.3477433
mombmi	.1358798	.0303327	4.48	0.000	.0763346	.1954251
dadbmi	.1055299	.0386207	2.73	0.006	.0297147	.1813451
_cons	10.02518	1.765514	5.68	0.000	6.559353	13.49101
q50						
educ	.053927	.0962937	0.56	0.576	-.1351042	.2429582
age	.3358445	.0735369	4.57	0.000	.1914864	.4802026
mombmi	.1315307	.0319572	4.12	0.000	.0687965	.194265
dadbmi	.1777306	.0441134	4.03	0.000	.0911329	.2643283
_cons	7.681324	1.728408	4.44	0.000	4.288337	11.07431
q75						
educ	.042771	.1878694	0.23	0.820	-.3260297	.4115717
age	.3557238	.1167391	3.05	0.002	.1265568	.5848909
mombmi	.2982242	.0701453	4.25	0.000	.1605241	.4359243
dadbmi	.216229	.0994086	2.18	0.030	.021083	.411375
_cons	5.228177	2.928817	1.79	0.075	-.5212944	10.97765
q90						
educ	-.1577382	.2577391	-0.61	0.541	-.663698	.3482216
age	.3642897	.1776034	2.05	0.041	.0156419	.7129376
mombmi	.3438461	.0691785	4.97	0.000	.208044	.4796482
dadbmi	.2529488	.0769468	3.29	0.001	.1018968	.4040007
_cons	8.846463	5.13658	1.72	0.085	-1.237002	18.92993

```
. predict q10bmihat, equation(#1)
(option xb assumed; fitted values)

. predict q90bmihat, equation(#5)
```

```
(option xb assumed; fitted values)
. list q10bmihat q90bmihat in -1

     +---------------------+
     | q10bmi~t   q90bmi~t |
     |---------------------|
771. | 20.73985   30.82972 |
     +---------------------+
```

(20.73985, 30.82972). Unconditional: (20.2, 30.95)

Question 6

(a)

```
. use "voucher.dta", clear
. count if select == 0
  468
. count if selectyrs == 4
  108
. count if choiceyrs == 4
  56
```

468 students were never awarded a voucher. 108 had a voucher available for four years. 56 attended a choice school for four years.

(b)

```
. regr choiceyrs selectyrs, r
Linear regression                               Number of obs     =        990
                                                F(1, 988)         =    1665.27
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7898
                                                Root MSE          =       .576

------------------------------------------------------------------------------
             |               Robust
   choiceyrs | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   selectyrs |   .7668317   .0187913    40.81   0.000     .7299562    .8037073
       _cons |   .0199189   .0105037     1.90   0.058    -.0006931     .040531
------------------------------------------------------------------------------
```

They are highly positively related, as expected. The p-value is 0.000 and the relationship is very strong. *selectyrs* is a sensible IV candidate for *choiceyrs* as it is highly correlated with *selectyrs* and not correlated with u_1 because it was randomly selected.

(c)

```
. regr mnce choiceyrs, r
Linear regression                               Number of obs     =        990
                                                F(1, 988)         =      13.58
                                                Prob > F          =     0.0002
                                                R-squared         =     0.0122
                                                Root MSE          =     20.754

------------------------------------------------------------------------------
             |               Robust
        mnce | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   choiceyrs |  -1.837014   .4985704    -3.68   0.000    -2.815393    -.858636
       _cons |    46.2344   .8973782    51.52   0.000     44.47342    47.99539
------------------------------------------------------------------------------

. regr mnce choiceyrs black hispanic female, r
Linear regression                               Number of obs     =        990
                                                F(4, 985)         =      20.28
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0868
                                                Root MSE          =     19.986

------------------------------------------------------------------------------
             |               Robust
        mnce | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   choiceyrs |  -.5652475   .4940605    -1.14   0.253     -1.53478    .4042845
       black |  -16.01743   1.926572    -8.31   0.000    -19.79808   -12.23677
    hispanic |  -13.40287    2.41094    -5.56   0.000    -18.13404   -8.671704
      female |   1.352745   1.279764     1.06   0.291    -1.158633    3.864123
       _cons |   57.12192   1.879984    30.38   0.000     53.43268    60.81115
------------------------------------------------------------------------------
```

The slope estimator of *choiceyrs* is -1.837014. It does not makes sense as the more a student attends a choice school, the less they do well on the math exam. It becomes -.5652475 when I add *black*, *hispanic*, and *female*, making it less negatively related.

(d) *choiceyrs* might be endogenous in this equation because student have to apply for the voucher in order to be considered awarded, which is not a random process even though the vouchers were chosen by lottery among those who applied.

(e)

```
. ivregress 2sls mnce (choiceyrs = selectyrs) black hispanic female, r

Instrumental variables 2SLS regression           Number of obs   =        990
                                                 Wald chi2(4)    =      80.28
                                                 Prob > chi2     =     0.0000
                                                 R-squared       =     0.0864
                                                 Root MSE        =     19.939

------------------------------------------------------------------------------
             |               Robust
        mnce | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   choiceyrs |  -.2413189   .5917809    -0.41   0.683    -1.401188    .9185504
       black |  -16.31692   1.947611    -8.38   0.000    -20.13417   -12.49967
    hispanic |   -13.7754    2.41278    -5.71   0.000    -18.50436    -9.04644
      female |   1.319709   1.277918     1.03   0.302    -1.184964    3.824383
       _cons |   57.06804    1.87722    30.40   0.000     53.38875    60.74732
------------------------------------------------------------------------------
Instrumented: choiceyrs
 Instruments: black hispanic female selectyrs
```

Using IV does not produce a positive effect of attending a choice school as the coefficient of *choiceyrs* is -.2413189, which is still negative. However, the coefficient did increase from -.5652475. As the other explanatory variables are not correlated with the instrument variable *selectyrs* and therefore their coefficients do not differ from the OLS ones.

(f)

```
. regr mnce choiceyrs black hispanic female mnce90, r

Linear regression                               Number of obs     =        328
                                                F(5, 322)         =      49.77
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4237
                                                Root MSE          =     16.029

------------------------------------------------------------------------------
             |               Robust
        mnce | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   choiceyrs |   .4105823   .7544828     0.54   0.587    -1.073756    1.894921
       black |  -8.305183   2.552212    -3.25   0.001     -13.3263   -3.284067
    hispanic |   -4.10498   3.465353    -1.18   0.237    -10.92257    2.712612
      female |   -.882847   1.784385    -0.49   0.621    -4.393373    2.627679
      mnce90 |   .6203655   .0473119    13.11   0.000     .5272861    .7134449
       _cons |    22.1529   3.599737     6.15   0.000     15.07092    29.23487
------------------------------------------------------------------------------

. ivregress 2sls mnce (choiceyrs = selectyrs) black hispanic female mnce90, r

Instrumental variables 2SLS regression           Number of obs   =        328
                                                 Wald chi2(5)    =     258.25
                                                 Prob > chi2     =     0.0000
                                                 R-squared       =     0.4173
                                                 Root MSE        =     15.969

------------------------------------------------------------------------------
             |               Robust
        mnce | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   choiceyrs |   1.799385   .9378768     1.92   0.055    -.0388202    3.637589
       black |  -9.067109   2.556081    -3.55   0.000    -14.07694   -4.057283
    hispanic |   -5.00373    3.43925    -1.45   0.146    -11.74453    1.737076
      female |  -1.020484   1.773235    -0.58   0.565    -4.495961    2.454992
      mnce90 |   .6288128   .0468642    13.42   0.000     .5369606    .7206649
       _cons |   21.53886   3.585963     6.01   0.000      14.5105    28.56722
------------------------------------------------------------------------------
Instrumented: choiceyrs
 Instruments: black hispanic female mnce90 selectyrs
```

β_1 in OLS is .4105823 and β_1 in IV is 1.799385. For the IV estimate, each year in a choice school is worth 1.80 on the math percentile score. Not even 2 percentile change can't be seen as a practically large effect.

(g) Compared to part (d) where there are 990 observations, there are only 328 observations for the analysis from part (f). This makes the analysis from part (f) not entirely convincing.

(h)

```
. ivregress 2sls mnce (choiceyrs1 choiceyrs2 choiceyrs3 choiceyrs4 = /*
> */ selectyrs1 selectyrs2 selectyrs3 selectyrs4) black hispanic female, r

Instrumental variables 2SLS regression           Number of obs   =        990
```

```
                                                  Wald chi2(7)    =      83.85
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0850
                                                  Root MSE        =     19.955

------------------------------------------------------------------------------
             |               Robust
        mnce | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
 choiceyrs1 |    .3899757   2.461086     0.16   0.874    -4.433664    5.213616
 choiceyrs2 |    .7736516   3.919113     0.20   0.844    -6.907669    8.454972
 choiceyrs3 |   -4.284797   3.559596    -1.20   0.229    -11.26148    2.691882
 choiceyrs4 |    2.407061   4.071456     0.59   0.554    -5.572846    10.38697
       black |   -16.29717   2.016116    -8.08   0.000    -20.24869   -12.34566
    hispanic |   -13.36599   2.568347    -5.20   0.000    -18.39985   -8.332119
      female |     1.36639   1.279023     1.07   0.285    -1.140448    3.873228
       _cons |    56.88582   1.895133    30.02   0.000     53.17143    60.60021
------------------------------------------------------------------------------
Instrumented: choiceyrs1 choiceyrs2 choiceyrs3 choiceyrs4
 Instruments: black hispanic female selectyrs1 selectyrs2 selectyrs3
              selectyrs4
```

(i)

```
. ivregress 2sls mnce (choiceyrs = /*
> */ selectyrs1 selectyrs2 selectyrs3 selectyrs4) black hispanic female, r

Instrumental variables 2SLS regression            Number of obs   =        990
                                                  Wald chi2(4)    =      80.28
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0865
                                                  Root MSE        =     19.939

------------------------------------------------------------------------------
             |               Robust
        mnce | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   choiceyrs |    -.252745   .5906766    -0.43   0.669     -1.41045    .9049599
       black |   -16.30635   1.946441    -8.38   0.000    -20.12131    -12.4914
    hispanic |   -13.76226     2.4129    -5.70   0.000    -18.49146   -9.033064
      female |    1.320875   1.278125     1.03   0.301    -1.184203    3.825953
       _cons |    57.06994   1.876938    30.41   0.000     53.39121    60.74867
------------------------------------------------------------------------------
Instrumented: choiceyrs
 Instruments: black hispanic female selectyrs1 selectyrs2 selectyrs3
              selectyrs4
```

Here, β_1 estimate is -.252745 whereas it was -.2413189 in part (e). They are very similar to each other.

```
. ivregress gmm mnce (choiceyrs = /*
> */ selectyrs1 selectyrs2 selectyrs3 selectyrs4) black hispanic female, r

Instrumental variables GMM regression             Number of obs   =        990
                                                  Wald chi2(4)    =      82.53
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0864
GMM weight matrix: Robust                         Root MSE        =     19.939

------------------------------------------------------------------------------
             |               Robust
        mnce | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   choiceyrs |   -.2437544   .5907315    -0.41   0.680    -1.401567    .9140581
       black |   -16.46039   1.938796    -8.49   0.000    -20.26036   -12.66042
    hispanic |   -13.83923   2.409442    -5.74   0.000    -18.56164   -9.116806
      female |     1.34696   1.275053     1.06   0.291    -1.152097    3.846017
       _cons |    57.17299   1.873959    30.51   0.000     53.50009    60.84588
------------------------------------------------------------------------------
Instrumented: choiceyrs
 Instruments: black hispanic female selectyrs1 selectyrs2 selectyrs3
              selectyrs4

. estat overid

  Test of overidentifying restriction:

  Hansen´s J chi2(3) = 1.71092 (p = 0.6345)
```

The p-value for the overidentification test is 0.6345.

(j) We could include *selectyrs* as a covariate. This would enable us to see the effect of each of other covariates holding *selectyrs* constant.