Yang Huan A0224968N

Question 1

(a)

We should pay attention in angrist90 dataset that which factors also influenced, we may consider adding something related to both mother and father's own sides: father's and mother's race; father's and mother's education; mother's age

(b)

```
. reg p kid3 if yearschm!=. //model1

      Source |       SS           df       MS      Number of obs   =   614,389
-------------+----------------------------------   F(1, 614387)    =   8755.88
       Model |  1831.06681         1  1831.06681   Prob > F        =    0.0000
    Residual |  128483.283   614,387  .209124351   R-squared       =    0.0141
-------------+----------------------------------   Adj R-squared   =    0.0140
       Total |  130314.349   614,388  .212104321   Root MSE        =     .4573

------------------------------------------------------------------------------
           p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kid3 |   -.115594   .0012353   -93.57   0.000    -.1180152   -.1131728
       _cons |   .7334874   .0007159  1024.60   0.000     .7320843    .7348905
------------------------------------------------------------------------------

est store model1

reg p kid3 yearschm //model2

      Source |       SS           df       MS      Number of obs   =   614,389
-------------+----------------------------------   F(2, 614386)    =  12845.74
       Model |  5230.56842         2  2615.28421   Prob > F        =    0.0000
    Residual |  125083.781   614,386  .203591522   R-squared       =    0.0401
-------------+----------------------------------   Adj R-squared   =    0.0401
       Total |  130314.349   614,388  .212104321   Root MSE        =    .45121

------------------------------------------------------------------------------
           p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kid3 |  -.0965608   .0012278   -78.65   0.000    -.0989671   -.0941544
    yearschm |   .0303201   .0002346   129.22   0.000     .0298602      .03078
       _cons |   .3984259   .0026874   148.25   0.000     .3931586    .4036932
------------------------------------------------------------------------------

est store model2
```

| | (1) | (2) |
|---|---|---|
| | model1 | model2 |
| VARIABLES | p | p |
| | | |
| kid3 | -0.116*** | -0.0966*** |
| | (0.00124) | (0.00123) |
| yearschm | | 0.0303*** |
| | | (0.000235) |
| Constant | 0.733*** | 0.398*** |
| | (0.000716) | (0.00269) |
| | | |
| Observations | 614,389 | 614,389 |
| R-squared | 0.014 | 0.040 |

Standard errors in parentheses

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

We can see that the effect of kid3 is decreasing(down) when add/control the mother's education. Maybe it is mainly because there is a positive relationship ($\gamma_1 > 0$) between mother labor supply and mother's education.

Due to $\beta_1 = \beta_2 + \gamma_1 \frac{cov(kid3,edum)}{var(kid3)}$ <0 and since $\beta_2$ also < 0; $\gamma_1 > 0$ and **absolute magnitude** $\beta_1 > \beta_2$, we can predict that $\boldsymbol{cov(kid3, edum) < 0}$. So, **the sign of the correlation between mother's education and the kid3 is negative**.

**(c)**

```
. reg yearschm kid3

      Source |       SS           df       MS      Number of obs   =   614,389
-------------+----------------------------------   F(1, 614387)    =   8971.93
       Model |  54000.4999         1  54000.4999   Prob > F        =    0.0000
    Residual |  3697887.94   614,387  6.01882517   R-squared       =    0.0144
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |  3751888.44   614,388  6.10670853   Root MSE        =    2.4533

------------------------------------------------------------------------------
    yearschm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kid3 |  -.6277433   .0066273   -94.72   0.000    -.6407327    -.614754
       _cons |    11.0508   .0038405  2877.42   0.000     11.04327    11.05833
------------------------------------------------------------------------------
```

**It shows that our prediction from (b) is correct.**

(d)

```
. reg kid3 samesex, robust //see if it's OK to proceed as IV

Linear regression                               Number of obs     =    614,389
                                                F(1, 614387)      =    2509.06
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0041
                                                Root MSE          =     .47131

------------------------------------------------------------------------------
             |               Robust
        kid3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     samesex |   .0602308   .0012024    50.09   0.000     .0578741    .0625876
       _cons |   .3056107   .0008324   367.14   0.000     .3039792    .3072422
------------------------------------------------------------------------------

. ivregress 2sls p (kid3=samesex), robust //IV method
```

We can see that t=50, so it's okay to let samesex be IV

```
. ivregress 2sls p (kid3=samesex), robust //IV method

Instrumental variables (2SLS) regression          Number of obs   =    614,389
                                                  Wald chi2(1)    =      16.67
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0127
                                                  Root MSE        =     .45762

------------------------------------------------------------------------------
             |               Robust
           p |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kid3 |  -.0791607   .0193863    -4.08   0.000    -.1171571   -.0411643
       _cons |   .7212525   .0065336   110.39   0.000     .7084468    .7340582
------------------------------------------------------------------------------
Instrumented:  kid3
Instruments:   samesex
```

| | (1) | (2) |
|---|---|---|
| | model1 | model1_2 |
| VARIABLES | p | p |
| | | |
| kid3 | -0.116*** | -0.0792*** |
| | (0.00124) | (0.0194) |
| Constant | 0.733*** | 0.721*** |
| | (0.000716) | (0.00653) |

| | | |
|---|---|---|
| Observations | 614,389 | 614,389 |
| R-squared | 0.014 | 0.013 |

Standard errors in parentheses

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

**With IV equation model 1_2 , it is meaning that mother have children that first two are same gender can reduce 0.0792 working hours.**

(e)

$$plim(\hat{\beta}_{1,OLS}) = \beta_1 + \frac{\sigma_u}{\sigma_x} \cdot Corr(x,u)$$
$$plim(\hat{\beta}_{1,IV}) = \beta_1 + \frac{\sigma_u}{\sigma_x} \cdot \frac{Corr(z,u)}{Corr(z,x)}$$

```
. corr kid3 yearschm
(obs=614,389)

             |     kid3 yearschm
-------------+------------------
        kid3 |   1.0000
    yearschm |  -0.1200   1.0000
```

```
. corr samesex kid3
(obs=614,389)

             |  samesex     kid3
-------------+------------------
     samesex |   1.0000
        kid3 |   0.0638   1.0000
```

```
. corr samesex yearschm
(obs=614,389)

             |  samesex yearschm
-------------+------------------
     samesex |   1.0000
    yearschm |  -0.0005   1.0000
```

```
. di -0.0005/0.0638
-.00783699
```

So, the bias from IV (-0.078) is smaller than the bias from OLS(-0.12). But we also should pay attention corr(z,x), the bias of IV could be blown up by it.

(f)

We use Hausman method to test whether is endogenous. We test model 1 and model 2.

```
. hausman IV model1, constant sigmamore

Note: the rank of the differenced variance matrix (1) does not equal the number
        of coefficients being tested (2); be sure this is what you expect, or
        there may be problems computing the test.  Examine the output of your
        estimators for anything unexpected and possibly consider scaling your
        variables so that the coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       IV          model1       Difference          S.E.
-------------+----------------------------------------------------------------
        kid3 |   -.0791607     -.115594        .0364333        .0193334
       _cons |    .7212525     .7334874       -.0122349        .0064925
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from ivregress
            B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        3.55
                Prob>chi2 =      0.0595
```

Model 1 p=0.0595>0.05, cannot reject H0 at 5%, so seems like no endogenous problem. But it is on the edge of level. And we know we need at least one outside exogenous variable,so it may be not trusted. **We need to use more precise model, model2 to test.**

```
. hausman IV2 model2, constant sigmamore //hausman 检验是否内生性, 原假设无内生
> , 拒绝则有内生

Note: the rank of the differenced variance matrix (1) does not equal the number
        of coefficients being tested (3); be sure this is what you expect, or
        there may be problems computing the test.  Examine the output of your
        estimators for anything unexpected and possibly consider scaling your
        variables so that the coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      IV2          model2       Difference          S.E.
-------------+----------------------------------------------------------------
        kid3 |   -.0778864    -.0965608        .0186744        .0190936
    yearschm |    .0307483     .0303201        .0004282        .0004378
       _cons |    .3875135     .3984259       -.0109125        .0111575
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from ivregress
            B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.96
                Prob>chi2 =      0.3281
```

But model2 with added edum variable, hausman test p>0.05, **that prove that we cannot reject kid3 is exogenous(无法拒绝外生原假设)**. **We use model 2 and believe model2 better because this model is more fitted with higher R-squared. [Kid3 is exogenous]**

| | (1) |
|---|---|
| | Hausman_model2 |
| VARIABLES | p |
| | |
| kid3 | -0.0779*** |
| | (0.0191) |
| yearschm | 0.0307*** |
| | (0.000497) |
| Constant | 0.388*** |
| | (0.0115) |
| | |
| Observations | 614,389 |
| R-squared | 0.040 |

Standard errors in parentheses

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

| | (1) | (2) |
|---|---|---|
| | model1 | model2 |
| VARIABLES | p | p |
| | | |
| kid3 | -0.116*** | -0.0966*** |
| | (0.00124) | (0.00123) |
| yearschm | | 0.0303*** |
| | | (0.000235) |
| Constant | 0.733*** | 0.398*** |
| | (0.000716) | (0.00269) |

| | | |
|---|---|---|
| Observations | 614,389 | 614,389 |
| R-squared | 0.014 | 0.040 |

Standard errors in parentheses

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

```
reg kid3 yearschm samesex
predict v, resid
reg p kid3 yearschm v, r
outreg2 using  "/Users/eric/Desktop/学业/1研一上/5103/Assignments/3/Yang Huan A0224968N
Assignment3/Output/edogenoustest_model2.doc" , word excel replace
//using method learnt from class to test model2, the result is same
```

```
Linear regression                               Number of obs     =    614,389
                                                F(3, 614385)      =    8445.31
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0401
                                                Root MSE          =     .45121
```

| p | Coef. | Robust Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| kid3 | -.0778864 | .0191328 | -4.07 | 0.000 | -.1153861 | -.0403867 |
| yearschm | .0307483 | .0004982 | 61.72 | 0.000 | .0297718 | .0317248 |
| v | -.0187515 | .019173 | -0.98 | 0.328 | -.05633 | .0188269 |
| _cons | .3875135 | .0114994 | 33.70 | 0.000 | .3649751 | .4100519 |

**using method learnt from class to test model2, the result is same, i.e. cannot reject kid3 is exogenous. [kid3 is exogenous]**

Q2

(a)

$$Pr(\text{accurate}=1 \mid \text{high\_temp}=1, \text{duration}=180.9)$$

$$= \frac{e^{-0.396-0.386-0.001\times 180.9}}{1+e^{-0.396-0.386-0.001\times 180.9}} = 0.2762979424 \approx 0.277$$

(excel)

$$Pr(\text{Accurate}=1 \mid \text{hightemp}=0, \text{duration}=180.9)$$

$$= \frac{e^{-0.396-0.001\times 180.9}}{1+e^{-0.396-0.001\times 180.9}} = 0.359646 \approx 0.360$$

$\therefore$ $0.277 - 0.360 = -0.083$

Hence, we can see when temperature is higher, probability of the accuracy of umpire reduces by 0.083

(b)

from part (a) the Partial Effect is $-0.083$,

in this linear model, the effect of high temp is $-0.0843$,

$\therefore$ the difference is very small, not surprisingly.

(c)

$$Pr(\text{accurate}=1 \mid \text{high\_temp}=1, \text{duration}=180.9, \text{year}=2017)$$

$$= \frac{e^{-1.003-0.425-0.003\times180.9+1.181}}{1+e^{-1.003-0.425-0.003\times180.9+1.181}} = 0.312233 \approx 0.312$$

$$Pr(\text{accurate}=1 \mid \text{high\_temp}=1, \text{duration}=180.9, \text{year}=2008)$$

$$= \frac{e^{-1.003-0.425-0.003\times180.9}}{1+e^{-1.003-0.425-0.003\times180.9}} = 0.122313 \approx 0.122$$

Difference $\Rightarrow$ $0.312 - 0.122 = 0.190$

So the probability of the accuracy for the umpire is 0.190 higher in 2017 than in 2008