IV example: Return to Education

The main equation for estimating return to education is from Mincerian wage equation:

```
reg lwage educ exper expersq
```

```
      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  3,   424) =   26.29
       Model |  35.0222967     3  11.6740989            Prob > F      =  0.0000
    Residual |  188.305144   424  .444115906           R-squared     =  0.1568
-------------+------------------------------           Adj R-squared =  0.1509
       Total |  223.327441   427  .523015084           Root MSE      =  .66642

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1074896   .0141465     7.60   0.000     .0796837    .1352956
       exper |   .0415665   .0131752     3.15   0.002     .0156697    .0674633
     expersq |  -.0008112   .0003932    -2.06   0.040    -.0015841   -.0000382
       _cons |  -.5220406   .1986321    -2.63   0.009    -.9124667   -.1316144
------------------------------------------------------------------------------
```

```
est sto reg1
```

The error term *u* from this main equation is thought to be correlated with *educ* because of omitted ability and other factors, such as quality of education and family background. Here, we can collect some information on family background (mother's and father's education). We will use the variables *motheduc* and *fatheduc* as an instrument for *educ*.

```
reg educ exper expersq motheduc fatheduc
```

```
      Source |       SS       df       MS              Number of obs =     753
-------------+------------------------------           F(  4,   748) =   66.52
       Model |  1025.94324     4   256.48581            Prob > F      =  0.0000
    Residual |   2884.0966   748  3.85574412           R-squared     =  0.2624
-------------+------------------------------           Adj R-squared =  0.2584
       Total |  3910.03984   752  5.19952106           Root MSE      =  1.9636

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |    .085378   .0255485     3.34   0.001     .0352228    .1355333
     expersq |  -.0018564   .0008276    -2.24   0.025    -.0034812   -.0002317
    motheduc |   .1856173   .0259869     7.14   0.000     .1346014    .2366331
    fatheduc |   .1845745   .0244979     7.53   0.000     .1364817    .2326674
       _cons |   8.366716   .2667111    31.37   0.000     7.843125    8.890307
------------------------------------------------------------------------------
```

```
predict educ_hat2
predict v_hat, resid
```

As we can see, mother's and father's education variables are correlated with education (of their child). However, we cannot guarantee whether these two variables are uncorrelated with the omitted factor *u*. Mother's education might be correlated with child's ability from early childhood development. For now, assume that they do not have this relationship.

Using 2SLS to estimate the wage equation, we get

```
ivregress 2sls lwage (educ = motheduc fatheduc) exper expersq
```

```
Instrumental variables (2SLS) regression              Number of obs =     428
                                                      Wald chi2(3)  =   24.65
                                                      Prob > chi2   =  0.0000
                                                      R-squared     =  0.1357
                                                      Root MSE      =  .67155
```

```
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0613966   .0312895     1.96   0.050     .0000704    .1227228
       exper |   .0441704   .0133696     3.30   0.001     .0179665    .0703742
     expersq |   -.000899   .0003998    -2.25   0.025    -.0016826   -.0001154
       _cons |   .0481003    .398453     0.12   0.904    -.7328532    .8290538
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   exper expersq motheduc fatheduc

est sto iv1
```

What if we estimate 2SLS manually?

```
reg lwage educ_hat2 exper expersq

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  3,   424) =    7.36
       Model |  11.0582283     3  3.68607609           Prob > F      =  0.0001
    Residual |  212.269213   424  .500634935           R-squared     =  0.0495
-------------+------------------------------           Adj R-squared =  0.0428
       Total |  223.327441   427  .523015084           Root MSE      =  .70756

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   educ_hat2 |   .0568605   .0310692     1.83   0.068    -.0042083    .1179292
       exper |   .0421082    .014286     2.95   0.003     .0140281    .0701884
     expersq |  -.0008565   .0004255    -2.01   0.045    -.0016929   -.0000201
       _cons |   .1332093   .3817364     0.35   0.727    -.6171221    .8835407
------------------------------------------------------------------------------

est sto reg2


estout reg1 iv1 reg2, cells(b(star fmt(3)) se(par fmt(4))) stats(r2 N, fmt(3 0))
starlevels(* 0.10 ** 0.05 *** 0.01)

------------------------------------------------------------
                      reg1            iv1            reg2
                      b/se            b/se            b/se
------------------------------------------------------------
educ                 0.107***        0.061**
                   (0.0141)        (0.0313)
exper                0.042***        0.044***        0.042***
                   (0.0132)        (0.0134)        (0.0143)
expersq             -0.001**        -0.001**        -0.001**
                   (0.0004)        (0.0004)        (0.0004)
educ_hat2                                           0.057*
                                                   (0.0311)
_cons               -0.522***        0.048           0.133
                   (0.1986)        (0.3985)        (0.3817)
------------------------------------------------------------
r2                   0.157           0.136           0.050
N                      428             428             428
------------------------------------------------------------
```

Test for endogeneity

```
hausman iv1 reg1, constant sigmamore

                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      iv1          reg1        Difference          S.E.
-------------+----------------------------------------------------------------
        educ |    .0613966     .1074896        -.046093        .0276406
       exper |    .0441704     .0415665        .0026039        .0015615
     expersq |    -.000899    -.0008112       -.0000878        .0000526
       _cons |    .0481003    -.5220406        .5701409        .3418964
------------------------------------------------------------------------------
```

```
                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)
                       =       2.78
             Prob>chi2 =      0.0954
             (V_b-V_B is not positive definite)
```

Another way to test for endogeneity: using the residuals from the first stage regression and adding into the structural equation. Then, do the t-test.

```
reg lwage educ exper expersq v_hat
```

```
      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  4,   423) =   20.53
       Model |   36.306365     4  9.07659124           Prob > F      =  0.0000
    Residual |  187.021076   423  .442130203           R-squared     =  0.1626
-------------+------------------------------           Adj R-squared =  0.1547
       Total |  223.327441   427  .523015084           Root MSE      =  .66493

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0639033   .0292123     2.19   0.029     .0064841    .1213226
       exper |   .0463071   .0134368     3.45   0.001     .0198959    .0727183
     expersq |  -.0009444   .0004001    -2.36   0.019    -.0017308    -.000158
       v_hat |   .0558771    .032788     1.70   0.089    -.0085706    .1203248
       _cons |   -.011404   .3592486    -0.03   0.975    -.7175388    .6947307
------------------------------------------------------------------------------
```

Test for overidentification restriction

Key: If all instruments are exogenous, the 2SLS residuals should be uncorrelated with the instruments.

Generate residuals from 2SLS regression, then regress on all exogenous variables.

```
reg uiv exper expersq motheduc fatheduc
```

```
      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  4,   423) =    0.09
       Model |  .170503136     4  .042625784           Prob > F      =  0.9845
    Residual |   192.84951   423  .455909007           R-squared     =  0.0009
-------------+------------------------------           Adj R-squared = -0.0086
       Total |  193.020013   427  .452037502           Root MSE      =  .67521

------------------------------------------------------------------------------
         uiv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |  -.0000183   .0133291    -0.00   0.999    -.0262179    .0261813
     expersq |   7.34e-07   .0003985     0.00   0.999    -.0007825     .000784
    motheduc |  -.0066065   .0118864    -0.56   0.579    -.0299704    .0167573
    fatheduc |   .0057823   .0111786     0.52   0.605    -.0161902    .0277547
       _cons |   .0109641   .1412571     0.08   0.938    -.2666892    .2886173
------------------------------------------------------------------------------
```

Calculate $nR^2 \sim$ chi-sq(q) under H_0: all IVs are uncorrelated with 2SLS error term.