```
------------------------------------------------------------------------------
. /*** To demonstrate use of 2SLS ***/
.
. * Case:  In the early 1990's Tanzania implemented a FP program to reduce fertility, which was among
the highest in the world
. * The FP program had two main components: provision of FP methods through health facilities, and
IEC (information, education, and communication)
. * An evaluation was carried out in early 1990's
. * We are evaluating the impact of the IEC component on Family Size Preferences
. * A DHS was conducted in 1996
. * Outcome of interest: Ideal number of children (idealnum), it is a continuous variable
. * Program variable of interest: Number of FP messages seen  (numfpmess)
.
. /* File 1996 Tanzania DHS */
. clear

. use newtanz96

. count
  6,866
.
. rename cluster clusterid
.
. * Review structure of the data
.
. * Basic information
. tab idealnum
    Ideal number of |
            children |      Freq.     Percent        Cum.
---------------------+-----------------------------------
                   0 |          6        0.09        0.09
                   1 |         39        0.57        0.66
                   2 |        407        5.93        6.58
                   3 |        736       10.72       17.30
                   4 |      1,783       25.97       43.27
                   5 |      1,173       17.08       60.36
                   6 |      1,267       18.45       78.81
                   7 |        314        4.57       83.38
                   8 |        458        6.67       90.05
                   9 |        103        1.50       91.55
                  10 |        468        6.82       98.37
                  11 |         12        0.17       98.54
                  12 |         58        0.84       99.39
                  13 |          5        0.07       99.46
                  14 |          1        0.01       99.48
                  15 |         16        0.23       99.71
                  16 |          4        0.06       99.77
                  20 |         15        0.22       99.99
                  30 |          1        0.01      100.00
---------------------+-----------------------------------
               Total |      6,866      100.00

. tab numfpmess
  numfpmess |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      2,377       34.62       34.62
          1 |      1,338       19.49       54.11
          2 |        930       13.55       67.65
          3 |        639        9.31       76.96
          4 |        491        7.15       84.11
          5 |        373        5.43       89.54
          6 |        324        4.72       94.26
          7 |        257        3.74       98.00
          8 |        137        2.00      100.00
------------+-----------------------------------
      Total |      6,866      100.00

.
```

```
. ****************************************
. * Case I: Two continuous dependent variables
. ****************************************
. /*
> We are interested in the impact of the IEC program (numfpmess) on Ideal Number of Children
> To estimate the impact of the IEC program you need to specify an equation for idealnum. So,
> Equation 1: idealnum = f(age, education, area of residence, being poor, availability of FP methods,
and number of FP messages seen)
>
> But, we suspect that numfpmess is endogenous, so you need to specify a second equation for
numfpmess:
> Equation 2:  numfpmess=f(age, education, area of residence, being poor, availability of FP methods,
and IV variables)
> */
.
. /* This is the main equation of interest without correction for endogeneity */
.
. regress idealnum age20_24 age25_29 age30_34 age35_39 age40_44 age45_49 edu16 edu7 edu8p poor rural
hofp5 hcfp5 difp5 numfpmess, vce(cluster clusterid)

Linear regression                               Number of obs     =      6,866
                                                F(15, 326)        =      96.65
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2204
                                                Root MSE          =     2.0828
                            (Std. Err. adjusted for 327 clusters in clusterid)
------------------------------------------------------------------------------
             |               Robust
    idealnum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    age20_24 |     .54018   .0709194     7.62   0.000     .4006625    .6796975
    age25_29 |   .9008975   .0770944    11.69   0.000     .7492322    1.052563
    age30_34 |   1.234149   .0867003    14.23   0.000     1.063586    1.404712
    age35_39 |   1.490688    .094516    15.77   0.000     1.304749    1.676626
    age40_44 |   1.671066    .128447    13.01   0.000     1.418376    1.923756
    age45_49 |   1.852991   .1569793    11.80   0.000     1.544171    2.161811
       edu16 |  -.7927907    .097326    -8.15   0.000    -.9842569   -.6013244
        edu7 |  -1.107775   .1069401   -10.36   0.000    -1.318155   -.8973956
       edu8p |  -1.916019   .1397935   -13.71   0.000     -2.19103   -1.641008
        poor |   .2857614   .1206765     2.37   0.018     .0483584    .5231643
       rural |   .5607607   .1175471     4.77   0.000     .3295142    .7920073
       hofp5 |  -.0781175    .135658    -0.58   0.565     -.344993     .188758
       hcfp5 |   .0159511   .1196267     0.13   0.894    -.2193867    .2512889
       difp5 |  -.0796897   .1059243    -0.75   0.452    -.2880713    .1286918
   numfpmess |   -.094716   .0141647    -6.69   0.000    -.1225817   -.0668502
       _cons |   5.049447   .1742879    28.97   0.000     4.706576    5.392317
------------------------------------------------------------------------------

.
. /* To control for the endogeneity of numfpmess we need to apply 2SLS: Two Stage Least Squares */
. /* First Stage:
> Run the regression for numfpmess */
. /* The identifying variables are lisradio and soapop  */
.
. regress numfpmess age20_24 age25_29 age30_34 age35_39 age40_44 age45_49 edu16 edu7 edu8p poor rural
hofp5 hcfp5 difp5 lisradio soapop, vce(cluster clusterid)

Linear regression                               Number of obs     =      6,866
                                                F(16, 326)        =     172.13
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4261
                                                Root MSE          =     1.6716

                            (Std. Err. adjusted for 327 clusters in clusterid)
------------------------------------------------------------------------------
             |               Robust
   numfpmess |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    age20_24 |   .5066266   .0657321     7.71   0.000     .3773141    .6359391
    age25_29 |   .6029298   .0703205     8.57   0.000     .4645906    .7412689
    age30_34 |   .7152608   .0719955     9.93   0.000     .5736265    .8568952
    age35_39 |   .7908484    .078767    10.04   0.000     .6358926    .9458041
    age40_44 |   .7314801   .0808561     9.05   0.000     .5724146    .8905456
```

```
    age45_49 |   .3812395   .0817277     4.66   0.000     .2204592    .5420198
       edu16 |    .466958    .060911     7.67   0.000     .3471297    .5867864
        edu7 |   .8745185   .0578633    15.11   0.000      .760686    .9883511
       edu8p |   1.737677   .1410064    12.32   0.000     1.460279    2.015074
        poor |  -.2333752   .0551657    -4.23   0.000    -.3419008   -.1248496
       rural |  -.5734665   .1102345    -5.20   0.000    -.7903272   -.3566058
       hofp5 |    .235844   .0906072     2.60   0.010     .0575954    .4140926
       hcfp5 |  -.0052702   .0709928    -0.07   0.941    -.1449321    .1343917
       difp5 |   .1620166   .0650989     2.49   0.013     .0339496    .2900835
    lisradio |   .5614473   .0614692     9.13   0.000     .4405209    .6823737
      soapop |   1.744989   .0736808    23.68   0.000     1.600039    1.889939
       _cons |   .4568662   .1337546     3.42   0.001     .1937351    .7199972
------------------------------------------------------------------------------

.
. /* Test that the identifying variables (lisaradio, soapop) are good predictors */.
. test lisradio soapop

 ( 1)  lisradio = 0
 ( 2)  soapop = 0

       F(  2,   326) =  384.77
            Prob > F =    0.0000

.
. /* Obtain the estimated predicted values of numfpmess */
. predict pnumfpmess
(option xb assumed; fitted values)

.
. /* Second Stage */
. * Replace the original numfpmess variable by its predicted value (pnumfpmess) in the first
equation, and run the modified equation

regress idealnum  age20_24 age25_29 age30_34 age35_39 age40_44 age45_49 edu16 edu7 edu8p poor rural
hofp5 hcfp5 difp5 pnumfpmess, vce(cluster clusterid)

Linear regression                               Number of obs     =      6,866
                                                F(15, 326)        =     100.07
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2228
                                                Root MSE          =     2.0797
                            (Std. Err. adjusted for 327 clusters in clusterid)
------------------------------------------------------------------------------
             |               Robust
    idealnum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    age20_24 |   .6324292   .0725749     8.71   0.000     .4896548    .7752035
    age25_29 |   .9995317    .079045    12.65   0.000     .8440291    1.155034
    age30_34 |   1.354261   .0901298    15.03   0.000     1.176951     1.53157
    age35_39 |   1.626333   .0938612    17.33   0.000     1.441683    1.810983
    age40_44 |   1.792447   .1298266    13.81   0.000     1.537044    2.047851
    age45_49 |   1.898088   .1545886    12.28   0.000     1.593971    2.202205
       edu16 |  -.6976763   .0988575    -7.06   0.000    -.8921554   -.5031973
        edu7 |  -.9038628   .1080823    -8.36   0.000     -1.11649    -.691236
       edu8p |   -1.51835   .1492792   -10.17   0.000    -1.812022   -1.224678
        poor |   .1889899   .1217295     1.55   0.122    -.0504845    .4284643
       rural |    .385538     .12273     3.14   0.002     .1440953    .6269808
       hofp5 |  -.0119034   .1354915    -0.09   0.930    -.2784515    .2546446
       hcfp5 |   .0164262   .1190232     0.14   0.890    -.2177244    .2505767
       difp5 |  -.0439007   .1059769    -0.41   0.679    -.2523856    .1645842
  pnumfpmess |  -.2520981   .0291238    -8.66   0.000    -.3093924   -.1948038
       _cons |   5.248691    .180579    29.07   0.000     4.893444    5.603939
------------------------------------------------------------------------------

.
. /* Compare the results */
.
```

```
. /* Exogeneity Test */

./* For the exogeneity test you need to obtain the estimated residuals from the numfpmess equation */

. regress numfpmess age20_24 age25_29 age30_34 age35_39 age40_44 age45_49 edu16 edu7 edu8p poor rural
hofp5 hcfp5 difp5 lisradio soapop, vce(cluster clusterid)

Linear regression                               Number of obs     =      6,866
                                                F(16, 326)        =     172.13
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4261
                                                Root MSE          =     1.6716
                                (Std. Err. adjusted for 327 clusters in clusterid)
------------------------------------------------------------------------------
             |               Robust
   numfpmess |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    age20_24 |   .5066266   .0657321     7.71   0.000     .3773141    .6359391
    age25_29 |   .6029298   .0703205     8.57   0.000     .4645906    .7412689
    age30_34 |   .7152608   .0719955     9.93   0.000     .5736265    .8568952
    age35_39 |   .7908484    .078767    10.04   0.000     .6358926    .9458041
    age40_44 |   .7314801   .0808561     9.05   0.000     .5724146    .8905456
    age45_49 |   .3812395   .0817277     4.66   0.000     .2204592    .5420198
       edu16 |    .466958    .060911     7.67   0.000     .3471297    .5867864
        edu7 |   .8745185   .0578633    15.11   0.000      .760686    .9883511
       edu8p |   1.737677   .1410064    12.32   0.000     1.460279    2.015074
        poor |  -.2333752   .0551657    -4.23   0.000    -.3419008   -.1248496
       rural |  -.5734665   .1102345    -5.20   0.000    -.7903272   -.3566058
       hofp5 |    .235844   .0906072     2.60   0.010     .0575954    .4140926
       hcfp5 |  -.0052702   .0709928    -0.07   0.941    -.1449321    .1343917
       difp5 |   .1620166   .0650989     2.49   0.013     .0339496    .2900835
    lisradio |   .5614473   .0614692     9.13   0.000     .4405209    .6823737
      soapop |   1.744989   .0736808    23.68   0.000     1.600039    1.889939
       _cons |   .4568662   .1337546     3.42   0.001     .1937351    .7199972
------------------------------------------------------------------------------

. predict resnumfp,res
.
. /* Include the estimated residuals in the original first equation */
. regress idealnum age20_24 age25_29 age30_34 age35_39 age40_44 age45_49 edu16 edu7 edu8p poor rural
hofp5 hcfp5 difp5 numfpmess resnumfp, vce(cluster clusterid)

Linear regression                               Number of obs     =      6,866
                                                F(16, 326)        =      94.73
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2243
                                                Root MSE          =     2.0778
                                (Std. Err. adjusted for 327 clusters in clusterid)
------------------------------------------------------------------------------
             |               Robust
    idealnum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    age20_24 |   .6324292   .0724912     8.72   0.000     .4898195    .7750388
    age25_29 |   .9995317   .0788049    12.68   0.000     .8445014    1.154562
    age30_34 |   1.354261   .0901098    15.03   0.000     1.176991    1.531531
    age35_39 |   1.626333   .0936483    17.37   0.000     1.442101    1.810564
    age40_44 |   1.792447   .1298866    13.80   0.000     1.536926    2.047969
    age45_49 |   1.898088   .1548276    12.26   0.000     1.593501    2.202675
       edu16 |  -.6976763    .098489    -7.08   0.000    -.8914306   -.5039221
        edu7 |  -.9038628   .1076775    -8.39   0.000    -1.115693   -.6920324
       edu8p |   -1.51835   .1498853   -10.13   0.000    -1.813214   -1.223485
        poor |   .1889899   .1209297     1.56   0.119    -.0489113     .426891
       rural |    .385538   .1216905     3.17   0.002     .1461402    .6249359
       hofp5 |  -.0119034   .1341035    -0.09   0.929    -.2757209     .251914
       hcfp5 |   .0164262   .1180333     0.14   0.889    -.2157768    .2486292
       difp5 |  -.0439007   .1051377    -0.42   0.677    -.2507347    .1629334
   numfpmess |  -.2520981   .0289951    -8.69   0.000    -.3091393   -.1950569
    resnumfp |   .1967283   .0313777     6.27   0.000     .1349999    .2584567
       _cons |   5.248691   .1801139    29.14   0.000     4.894359    5.603023
------------------------------------------------------------------------------
.
. /* Examine the significance of the residuals: variable resnumfp */
```

```
. /* in this case resnumfp is significant,therefore variable numfpmess is endogenous, you did well by
using 2SLS */
. /* what if you would've found that resnumfp was not significant? */
.
.
. /* Alternative procedure: ivregress  */
. /* it generates better standard errors */

.
. ivregress 2sls idealnum  age20_24 age25_29 age30_34 age35_39 age40_44 age45_49 edu16 edu7 edu8p
poor rural hofp5 hcfp5 difp5 (numfpmess=lisradio soapop), vce(cluster clusterid)

Instrumental variables (2SLS) regression          Number of obs   =      6,866
                                                  Wald chi2(15)   =    1486.42
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2049
                                                  Root MSE        =      2.101

                            (Std. Err. adjusted for 327 clusters in clusterid)
------------------------------------------------------------------------------
             |               Robust
    idealnum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   numfpmess |  -.2520981   .0295549    -8.53   0.000    -.3100246   -.1941716
    age20_24 |   .6324292   .0736917     8.58   0.000     .4879962    .7768622
    age25_29 |   .9995317   .0795878    12.56   0.000     .8435424    1.155521
    age30_34 |   1.354261    .091386    14.82   0.000     1.175148    1.533374
    age35_39 |   1.626333    .094583    17.19   0.000     1.440953    1.811712
    age40_44 |   1.792447   .1311723    13.66   0.000     1.535354     2.04954
    age45_49 |   1.898088   .1563132    12.14   0.000      1.59172    2.204456
       edu16 |  -.6976763    .097931    -7.12   0.000    -.8896175   -.5057352
        edu7 |  -.9038628   .1070122    -8.45   0.000    -1.113603   -.6941228
       edu8p |   -1.51835   .1557308    -9.75   0.000    -1.823577   -1.213123
        poor |   .1889899   .1183616     1.60   0.110    -.0429945    .4209743
       rural |    .385538   .1203968     3.20   0.001     .1495646    .6215115
       hofp5 |  -.0119034   .1303021    -0.09   0.927    -.2672908    .2434839
       hcfp5 |   .0164262   .1151891     0.14   0.887    -.2093403    .2421926
       difp5 |  -.0439007   .1028442    -0.43   0.669    -.2454715    .1576702
       _cons |   5.248691   .1804954    29.08   0.000     4.894927    5.602456
------------------------------------------------------------------------------
Instrumented:  numfpmess
Instruments:   age20_24 age25_29 age30_34 age35_39 age40_44 age45_49 edu16
               edu7 edu8p poor rural hofp5 hcfp5 difp5 lisradio soapop

.
. log close
      name:  <unnamed>
       log:
--------------------------------------------------------------------------------
```