15. Instrumental Variable Method

Another way of measuring the impact of the program when treatment has not been randomly assigned is by using the instrumental variable (IV) method. The IV estimation regards the treatment variable (in this case, participation in microfinance programs) as endogenous. The idea is to find an observable exogenous variable or variables (instruments) that influence the participation variable but do not influence the outcome of the program if participating. Thus, one would want at least one instrument that is not in the covariates and that satisfies the preceding requirements. IV estimation is a two-step process. First, the treatment variable is run against all covariates, including the instruments. Then, the predicted value of the treatment—instead of the actual value—is used in the second stage.

IV Implementation Using the "ivreg" Command

The first step in IV implementation is to find an instrument. In the example, a household's choice to participate in the microcredit program is used as the instrument variable. The household's choice depends on two factors: availability of microcredit programs in the village and the household's eligibility to participate (which is determined by its landholding). Even though program placement in the village may be endogenous, a household's eligibility is not, and the combination of these two factors is therefore exogenous.

Using the 1998 data (hh_98.dta), create a village program variable for females and then a female program choice variable at the household level.[1] As mentioned in earlier exercises, a household is eligible to participate in microcredit programs if it has fewer than 50 decimals of land.

```
egen villfmf=max(dmmfd), by(vill);
gen fchoice=villfmf==1 & hhland<50;
```

Next, create additional instruments by interacting the choice variable with all covariates. The Stata "for" command is used to do so in one command:

```
for var agehead-educhead lnland vaccess pcirr rice-oil: gen
fchX=fchoice*X;
```

The next step is the IV implementation, which uses the Stata “ivreg” command. The first-stage equation appears within parentheses in the syntax, and the first option displays the first-stage results:

```
ivreg lexptot agehead-educhead lnland vaccess pcirr rice-oil
(dfmfd= agehead-educhead lnland vaccess pcirr rice-oil fch*),
first;
```

The output shows the first-stage results first and then the second-stage results. According to the first-stage output, household head’s education and household’s land asset negatively influence microcredit program participation; so do the instruments. The second-stage results show that after controlling for the endogeneity of program participation, female participation in microcredit programs has a significant impact (32.6 percent) on household’s per capita expenditure ($t = 2.28$).

```
First-stage regressions
-----------------------

     Source |    SS         df        MS             Number of obs =   1129
------------+------------------------------          F( 23,  1105) =   6.15
      Model | 31.9544747    23   1.38932499          Prob > F      = 0.0000
   Residual | 249.471566  1105   .225766123          R-squared     = 0.1135
------------+------------------------------          Adj R-squared =  .0951
      Total | 281.426041  1128   .249491171          Root MSE      = .47515

------------------------------------------------------------------------------
      dfmfd |     Coef.    Std. Err.     t     P>|t|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
    agehead | -.0017996     .001853   -0.97   0.332   -.0054354    .0018362
    sexhead |  -.090353    .0949407   -0.95   0.341   -.2766374    .0959314
   educhead | -.0111658     .006549   -1.70   0.088   -.0240157    .0016841
     lnland | -.0743253    .0463394   -1.60   0.109   -.1652485    .0165979
    vaccess | -.1696796     0699002   -2.43   0.015   -.3068316   -.0325275
      pcirr | -.0459691     0831373   -0.55   0.580   -.2090939    .1171558
       rice |  .0085986    .0155203    0.55   0.580   -.0218539    .0390511
      wheat |  .0102826    .0292563    0.35   0.725   -.0471216    .0676869
       milk | -.0211565    .0104327   -2.03   0.043   -.0416267   -.0006864
     potato | (dropped)
        egg |  .0043442    .0934236    0.05   0.963   -.1789635    .1876519
        oil |  .0017818    .0065519    0.27   0.786   -.0110737    .0146373
    fchoice |   -.97571    .4857339   -2.01   0.045   -1.928775    -.022645
 fchagehead |  .0062515    .0023876    2.62   0.009    .0015669    .0109362
 fchsexhead |  .1562665    .1116846    1.40   0.162   -.0628713    .3754043
fcheduchead | -.0083186    .0088998   -0.93   0.350   -.0257811    .0091439
  fchlnland | -.0028382    .1781701   -0.02   0.987   -.3524282    .3467517
 fchvaccess |  .1823573     .084952    2.15   0.032    .0156719    .3490427
   fchpcirr |  .1830853    .1025273    1.79   0.074   -.0180849    .3842554
    fchrice | -.0253889     .019694   -1.29   0.198   -.0640307    .0132529
   fchwheat |  -.019292    .0365608   -0.53   0.598   -.0910284    .0524444
    fchmilk |  .0319648    .0126207    2.53   0.011    .0072016     .056728
     fchegg |  .0802827    .1110378    0.72   0.470    -.137586    .2981513
     fchoil |  .0097549     .007933    1.23   0.219   -.0058106    .0253203
      _cons |  .7880826    .3962508    1.99   0.047    .0105937    1.565571
------------------------------------------------------------------------------
```

```
Instrumental variables (2SLS) regression

     Source |       SS       df       MS              Number of obs =    1129
-------------+------------------------------           F( 12,  1116) =   22.94
       Model | 48.1621199     12   4.01350999           Prob > F      = 0.0000
    Residual |  249.69781   1116   .223743557           R-squared     = 0.1617
-------------+------------------------------           Adj R-squared = 0.1527
       Total |  297.85993   1128   .264060221           Root MSE      = .47302

------------------------------------------------------------------------------
     lexptot |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dfmfd |   .3255436   .1426528     2.28   0.023     .0456457    .6054415
     agehead |   .0030299   .0011679     2.59   0.010     .0007383    .0053214
     sexhead |  -.0566001   .0494292    -1.15   0.252    -.1535847    .0403844
    educhead |   .0533665   .0048684    10.96   0.000     .0438142    .0629188
      lnland |   .2210422   .0408664     5.41   0.000     .1408586    .3012258
     vaccess |  -.0030504   .0403496    -0.08   0.940      -.08222    .0761193
       pcirr |   .1389462   .0496316     2.80   0.005     .0415644    .2363281
        rice |   .0054628    .009462     0.58   0.564    -.0131025    .0240281
       wheat |  -.0401031   .0173472    -2.31   0.021    -.0741399   -.0060664
        milk |   .0207911   .0058035     3.58   0.000     .0094042     .032178
      potato |  (dropped)
         egg |   .1005972   .0508165     1.98   0.048     .0008905    .2003039
         oil |   .0081386   .0038401     2.12   0.034     .0006041    .0156732
       _cons |   7.407985   .2280463    32.48   0.000     6.960537    7.855433
------------------------------------------------------------------------------
Instrumented:  dfmfd
Instruments:   agehead sexhead educhead lnland vaccess pcirr rice
               wheat milk
               potato egg oil fchoice fchagehead fchsexhead
               fcheduchead
               fchlnland fchvaccess fchpcirr fchrice fchwheat
               fchmilk fchegg
               fchoil
------------------------------------------------------------------------------
```

Testing for Endogeneity: OLS versus IV

A few tests can be used to determine whether an ordinary least squares (OLS) or IV is more appropriate. Stata has a command "ivendog" that performs an *F*-test and chi-square test following methodologies called the Wu-Hausman test and the Durbin-Wu-Hausman test, respectively. The null hypothesis is that OLS is consistent (in this case, it implies treatment is exogenous). If the null hypothesis is not rejected, an OLS should suffice; otherwise an IV method should be used. The "ivendog" command is used after the "ivreg" command:

```
ivendog;
```

The results show that the null hypothesis is rejected at the 10 percent level, implying that IV is a better model than OLS:

```
Tests of endogeneity of: dfmfd
H0: Regressor is exogenous
    Wu-Hausman F test:                3.01281   F(1,1115)    P-value = 0.08289
    Durbin-Wu-Hausman chi-sq test:    3.04242   Chi-sq(1)    P-value = 0.08111
```

IV Method for Binary Treatment: "treatreg" Command

The preceding IV estimation methods apply when the endogenous regressor is continuous. When the endogenous regressor is binary (participant/nonparticipant), using a linear model in the first stage of the IV procedure may or may not be appropriate. Another method that fits a treatment-effects model when the endogenous regressor is binary is the "treatreg" command in Stata. The "treatreg" command fits a treatment-effects model using either the full maximum likelihood or the two-step consistent estimator. The "treatreg" command takes into account the effect of the binary endogenous variable on the outcome of interests conditional on the two sets of exogenous variables. The command estimates two regressions simultaneously. The first equation is estimated using probit regression to predict the probability of treatment. The second is either a linear or probit regression for the outcome variables. The two error terms are assumed to be jointly normally distributed.

Following is an example of how the "treatreg" command is used with the Bangladesh 1998 data. Its syntax is very similar to that of the "ivreg" command:

```
treatreg lexptot agehead-educhead lnland vaccess pcirr
rice-oil, treat (dfmfd= agehead-educhead lnland vaccess pcirr
rice-oil fch*);
```

Following is the treatment-effect method using maximum likelihood estimation. It shows that women's participation does have a positive significant impact on household's expenditure ($t = 3.49$):

```
Treatment effects model -- MLE                     Number of obs   =      1129
                                                   Wald chi2(12)   =    271.45
Log likelihood = -1427.6651                        Prob > chi2     =    0.0000

------------------------------------------------------------------------------
             |    Coef.    Std. Err.     z      P>|z|    [95% Conf. Interval]
------------+-----------------------------------------------------------------
  lexptot    |
    agehead |  .0028983   .0011858    2.44    0.015    .0005742    .0052225
    sexhead | -.0558392   .0504364   -1.11    0.268   -.1546927    .0430142
   educhead |  .0547403   .0048088   11.38    0.000    .0453152    .0641654
     lnland |  .2386945   .0384969    6.20    0.000     .163242    .3141469
    vaccess |  .0026497   .0408488    0.06    0.948   -.0774125    .0827118
      pcirr |  .1305888   .0500755    2.61    0.009    .0324427     .228735
       rice |  .0060323   .0096418    0.63    0.532    .0128654      .02493
      wheat | -.0404817    .017699   -2.29    0.022   -.0751711   -.0057923
       milk |  .0208849   .0059217    3.53    0.000    .0092787    .0324912
        egg |  .0944399   .0515543    1.83    0.067   -.0066047    .1954846
        oil |  .0074181   .0038636    1.92    0.055   -.0001545    .0149906
      dfmfd |  .4168906   .1196073    3.49    0.000    .1824647    .6513166
      _cons |  7.391633   .2322404   31.83    0.000     6.93645    7.846816
------------+-----------------------------------------------------------------

dfmfd        |
    agehead |  -.004252   .0050252   -0.85    0.397   -.0141012    .0055973
    sexhead | -.1799594   .2534342   -0.71    0.478   -.6766813    .3167625
```

```
    educhead | -.0453168   .0184985   -2.45    0.014   -.0815733   -.0090604
      lnland | -.1791062   .1315339   -1.36    0.173   -.4369079    .0786956
     vaccess | -.5458849   .1822059   -3.00    0.003   -.9030019   -.1887679
       pcirr |  -.121319   .2202852   -0.55    0.582   -.5530702    .3104321
        rice |  .0093552   .0406836    0.23    0.818   -.0703831    .0890935
       wheat |  .0082867   .0782386    0.11    0.916   -.1450581    .1616316
        milk | -.0605588   .0294469   -2.06    0.040   -.1182737   -.002844
         egg |  .0366651   .2578851    0.14    0.887   -.4687804    .5421107
         oil | -.0017389   .0177263   -0.10    0.922   -.0364818     .033004
     fchoice | -3.391314   1.291503   -2.63    0.009   -5.922613   -.8600159
   fchagehead |  .0156243   .0063892    2.45    0.014    .0031018    .0281468
   fchsexhead |  .3432873   .2937005    1.17    0.242   -.2323551    .9189296
  fcheduchead |  .0056506   .0247551    0.23    0.819   -.0428685    .0541698
    fchlnland | -.2419577   .4632756   -0.52    0.601   -1.149961    .6660458
   fchvaccess |  .6105495   .2173745    2.81    0.005    .1845032    1.036596
     fchpcirr |  .4829752   .2662667    1.81    0.070    -.038898    1.004848
     fchrice | -.0446986    .050703   -0.88    0.378   -.1440747    .0546775
    fchwheat | -.0191072   .0959983   -0.20    0.842   -.2072604     .169046
     fchmilk |  .0866831   .0345121    2.51    0.012    .0190407    .1543255
      fchegg |  .1975426    .297008    0.67    0.506   -.3845824    .7796676
      fchoil |  .0345253   .0207377    1.66    0.096   -.0061198    .0751704
       _cons |  1.309823   1.095342    1.20    0.232   -.8370069    3.456653
------------+----------------------------------------------------------------
     /athrho | -.4622307   .1677019   -2.76    0.006   -.7909205    -.133541
    /lnsigma | -.7283617   .0440104  -16.55    0.000   -.8146205    -.642103
------------+----------------------------------------------------------------
         rho | -.4319006   .1364191                     -.6589302   -.1327528
       sigma |  .4826991   .0212438                      .4428074    .5261847
      lambda |  -.208478   .0740375                     -.3535888   -.0633673
-----------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0): chi2(1) = 5.14 Prob > chi2 = 0.0234
-----------------------------------------------------------------------------
```

IV with Fixed Effects: Cross-Sectional Estimates

Instrumental variable regression can be combined with fixed effects. Here a demonstration using cross-sectional data is shown. The command to use is “xtivreg” with the “fe” option. A village-level fixed-effects regression is run using the same hh_98.dta. Here is the command for women’s participation in microcredit:

```
xtivreg lexptot year agehead-educhead lnland vaccess pcirr
rice-oil (dfmfd= agehead-educhead lnland vaccess pcirr rice-oil
mch*), fe i(vill);
```

Next, run a village-level fixed-effects regression with the same hh_98.dta. Using village-level fixed effects causes the participation impacts to disappear:

```
Fixed-effects (within) IV regression          Number of obs      =      1129
Group variable: vill                          Number of groups   =       104

R-sq: within  = 0.1550                        Obs per group: min =         4
      between = 0.2246                                       avg =      10.9
      overall = 0.1618                                       max =        19

                                              Wald chi2(5)       = 453021.37
corr(u_i, Xb) = 0.0511                        Prob > chi2        =    0.0000
```

```
------------------------------------------------------------------------------
   lexptot |     Coef.    Std. Err.     z     P>|z|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
     dfmfd |   .1901029   .1956837    0.97    0.331     -.19343    .5736359
   agehead |   .0020665   .0011244    1.84    0.066   -.0001373    .0042703
   sexhead |  -.0352392   .0472055   -0.75    0.455   -.1277602    .0572818
  educhead |   .0433888   .0056147    7.73    0.000    .0323842    .0543934
    lnland |   .2283189   .0470498    4.85    0.000    .1361029    .3205349
   vaccess |  (dropped)
     pcirr |  (dropped)
      rice |  (dropped)
     wheat |  (dropped)
      milk |  (dropped)
       egg |  (dropped)
       oil |  (dropped)
     _cons |    8.10043   .1268782   63.84    0.000    7.851754    8.349107
-----------+------------------------------------------------------------------
   sigma_u |  .24105185
   sigma_e |  .42196914
       rho |  .24604092   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:    F(103,1020) =      3.04        Prob > F  = 0.0000
------------------------------------------------------------------------------
Instrumented:  dfmfd
Instruments:   agehead sexhead educhead lnland vaccess pcirr rice wheat
milk egg oil mchoice mchagehead mchsexhead mcheduchead mchlnland mchvaccess
mchpcirr mchrice mchwheat mchmilk mchegg mchoil
```

IV with Fixed Effects: Panel Estimates

An implementation of "xtivreg" using panel data is now shown with the panel data hh_9198.dta. After creating necessary variables as before, issue the "xtivreg" command.

```
xtivreg lexptot year agehead-educhead lnland vaccess pcirr
rice-oil (dfmfd= agehead-educhead lnland vaccess pcirr rice-oil
mch*), fe i(nh);
```

The results do not show any participation impact on expenditure.

```
Fixed-effects (within) IV regression       Number of obs      =      1652
Group variable: nh                         Number of groups   =       826

R-sq: within  = 0.1667                     Obs per group: min =         2
      between = 0.1924                                    avg =       2.0
      overall = 0.1733                                    max =         2

                                           Wald chi2(14)      = 866855.47
corr(u_i, Xb) = 0.1215                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
   lexptot |    Coef.     Std. Err.     z     P>|z|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
     dfmfd |  .0430727    .124483     0.35   0.729    -.2009096     .287055
      year |  .2360629   .0707606     3.34   0.001     .0973747    .3747511
   agehead |   .000021   .0017636     0.01   0.990    -.0034355    .0034775
   sexhead | -.0536457   .0727231    -0.74   0.461    -.1961803     .088889
  educhead |  .0136537    .008419     1.62   0.105    -.0028472    .0301546
```

```
    lnland |   .1362576   .0629346    2.17   0.030     .0129079    .2596072
   vaccess |  -.0991489     .05371   -1.85   0.065    -.2044186    .0061207
     pcirr |   .0954609   .0642934    1.48   0.138    -.0305519    .2214737
      rice |   .0199218   .0131231    1.52   0.129     -.005799    .0456426
     wheat |  -.0244967   .0128117   -1.91   0.056    -.0496072    .0006138
      milk |  -.0028403   .0065394   -0.43   0.664    -.0156572    .0099766
    potato |     -.0199   .0165334   -1.20   0.229    -.0523049    .0125048
       egg |   .1703499   .0483323    3.52   0.000     .0756203    .2650795
       oil |   .0045626   .0031518    1.45   0.148    -.0016148      .01074
     _cons |   7.833876   .2515847   31.14   0.000     7.340779    8.326973
-----------+----------------------------------------------------------------
   sigma_u |  .34559734
   sigma_e |  .36468826
       rho |  .47314159   (fraction of variance due to u_i)
----------------------------------------------------------------------------
F test that all u_i=0:    F(825,812) =     1.57        Prob > F    = 0.0000
----------------------------------------------------------------------------
Instrumented: dfmfd
Instruments:  year agehead sexhead educhead lnland vaccess pcirr rice wheat
milk potato egg oil fchoice fchagehead fchsexhead fcheduchead fchlnland
fchvaccess fchpcirr fchrice fchwheat fchmilk fchpotato fchegg fchoil
```

Note

1. By 1998, all sample villages had microcredit programs, but for purposes of demonstrating the process of creating the variable, this exercise creates the village program variable.