Introduction to Stata – Lecture 4: Instrumental variables

Hayley Fisher

3 March 2010

Key references: Cameron and Trivedi (2009) chapter 6, Angrist and Pischke (2009) chapter 4, Wooldridge (2009) chapter 15, Greene (2008) chapter 13.

1 Introduction

For OLS to give consistent estimators the error term must be unrelated to the regressors – that is $E(u|\mathbf{x}) = 0$. This is often not an assumption that can be made. One common approach is to use an instrumental variable (IV) estimator. The instrumental variable, z, needs to be correlated with the endogenous variable x, and uncorrelated with the error term so $E(u|z) = 0$. Finding a valid and strong instrument is often very hard – see Angrist and Pischke (2009) for a full discussion.

This lecture focuses on the implementation of IV estimation in Stata and the related tests available. I use the dataset from Angrist and Krueger's (1991) article examining the economic return to schooling using quarter of birth and its interaction with compulsory school attendance laws as an instrument. The paper is available at http://www.jstor.org/stable/2937954 and the dataset is available on my website.

The problem with estimating the economic returns to education is that of omitted variable bias. We can estimate the following equation:

$$lwage_i = \alpha + \beta schooling_i + \gamma X_i + \varepsilon_i \tag{1}$$

However, there is a key unmeasured determinant of wages – ability. This means that ability is subsumed into the error term ε_i. Since ability is also likely linked to schooling, this means that $E(\varepsilon|schooling) \neq 0$. So an OLS estimate of β is inconsistent.

The key to Angrist and Krueger's approach is to recognise that there is a relationship between quarter of birth and the amount of schooling an individual receives:

> "...most states require students to enter school in the calendar year in which they turn 6. School start age is therefore a function of date of birth. Specifically, those born late in the year are young for their grade. In states with a December 31 birthday cutoff, children born in the fourth quarter enter school shortly before they turn 6, while those born in the first quarter enter school at around $6\frac{1}{2}$. Furthermore, because compulsory schooling laws typically require students to remain in school only until their 16th birthday, these groups of students will be in different grades, or through a given grade to a different degree, when they reach the legal dropout age. The combination of school start-age policies and compulsory schooling laws creates a natural experiment in which children are compelled to attend school for different lengths of time, depending on their birthdays."
>
> Angrist and Pischke (2009) p.117

This means that quarter of birth can be used as an instrumental variable for schooling – it is correlated with schooling but there is no reason to believe that it also affects wages.

2 Data

We can load `angristkrueger.dta` into Stata and look at the data. `summarize` shows:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         age |    247199    44.72566    2.898535         40         50
        ageq |    247199    45.10029    2.877965      40.25         50
        educ |    247199    11.49334    3.360663          0         18
       lwage |    247199    5.155175    .6512804  -.0198026   8.947976
     married |    247199    .8928151    .3093488          0          1
-------------+--------------------------------------------------------
      census |    247199          70           0         70         70
         qob |    247199    2.488068    1.112981          1          4
        race |    247199    .0820675    .2744681          0          1
        smsa |    247199    .3020036    .4591278          0          1
         yob |    247199    1924.528    2.861746       1920       1929
-------------+--------------------------------------------------------
      region |    247199    3.433169    2.611136          0          8
```

So we have a large dataset of 247,199 observations with no missing observations. The minimum and maximum values look reasonable – the dataset is already cleaned. We need to create a variable for age squared – here using the `ageq` variable which has quarter years of age included.

```
. generate ageq2=ageq^2
```

3 Implementing instrumental variables estimation

We start by estimating the impact of years of education on log wages using OLS to give a baseline.

```
. xi: regress lwage educ race married smsa i.yob i.region ageq ageq2, vce(robust)
```

output omitted

So we are also controlling for race, marital status, size of town lived in, age and including a set of dummies for year of birth and region. The output suggests that an additional year of education is associated with a 7.0% increase in wages.

To get a better estimate using quarter of birth as an instrument for education, we use the `ivregress` command. The syntax is as follows:

`ivregress` *estimator depvar* [*varlist1*] (*varlist2*=*varlistiv*) [*if*] [*in*] [*weight*] [, *options*]

Here *estimator* is one of `2sls`, `gmm` or `liml`. These are different methods of estimating the model – you must include one. *varlist1* contains the exogenous regressors, *varlist2* the endogenous regressors and *varlistiv* the instruments.

So, to estimate the effect of education on wages, using a set of quarter of birth dummies as instruments, we type:

```
. xi: ivregress 2sls lwage race married smsa ageq ageq2 i.region i.yob (educ=i.qob),
vce(robust)
i.region          _Iregion_0-8        (naturally coded; _Iregion_0 omitted)
i.yob             _Iyob_1920-1929     (naturally coded; _Iyob_1920 omitted)
i.qob             _Iqob_1-4           (naturally coded; _Iqob_1 omitted)

Instrumental variables (2SLS) regression                Number of obs =  247199
                                                        Wald chi2(23) =29586.85
```

```
                                                  Prob > chi2   =  0.0000
                                                  R-squared     =  0.2242
                                                  Root MSE      =  .57364

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0849189   .0653187     1.30   0.194    -.0431034    .2129411
        race |  -.2636648    .151434    -1.74   0.082    -.5604699    .0331403
     married |   .2867815   .0268903    10.66   0.000     .2340775    .3394855
        smsa |  -.1256149   .0385288    -3.26   0.001    -.2011301   -.0500998
        ageq |   .1166074   .0654006     1.78   0.075    -.0115755    .2447902
       ageq2 |  -.0012151   .0007409    -1.64   0.101    -.0026672    .0002371
  _Iregion_1 |    .030666   .0484413     0.63   0.527    -.0642772    .1256093
  _Iregion_2 |  -.1833031    .108244    -1.69   0.090    -.3954574    .0288513
  _Iregion_3 |  -.0063643    .030069    -0.21   0.832    -.0652984    .0525699
  _Iregion_4 |  -.1245374   .0121238   -10.27   0.000    -.1482996   -.1007752
  _Iregion_5 |  -.0266789   .0274541    -0.97   0.331    -.0804879    .0271301
  _Iregion_6 |  -.0874231   .0723285    -1.21   0.227    -.2291844    .0543382
  _Iregion_7 |  -.1328462   .0384853    -3.45   0.001    -.2082759   -.0574164
  _Iregion_8 |  -.1345381   .0745719    -1.80   0.071    -.2806962    .0116201
  _Iyob_1921 |  -.0009953   .0122851    -0.08   0.935    -.0250737    .0230831
  _Iyob_1922 |  -.0037629   .0279409    -0.13   0.893     -.058526    .0510003
  _Iyob_1923 |  -.0003771   .0393341    -0.01   0.992    -.0774706    .0767164
  _Iyob_1924 |   .0040516   .0524957     0.08   0.938    -.0988382    .1069414
  _Iyob_1925 |   .0186145   .0678595     0.27   0.784    -.1143877    .1516167
  _Iyob_1926 |   .0255744   .0786054     0.33   0.745    -.1284894    .1796382
  _Iyob_1927 |   .0366671   .0956002     0.38   0.701    -.1507059    .2240401
  _Iyob_1928 |   .0461339   .1062542     0.43   0.664    -.1621204    .2543883
  _Iyob_1929 |   .0421031    .113661     0.37   0.711    -.1806683    .2648746
       _cons |   1.233629   1.930212     0.64   0.523    -2.549517    5.016774
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   race married smsa ageq ageq2 _Iregion_1 _Iregion_2 _Iregion_3
               _Iregion_4 _Iregion_5 _Iregion_6 _Iregion_7 _Iregion_8
               _Iyob_1921 _Iyob_1922 _Iyob_1923 _Iyob_1924 _Iyob_1925
               _Iyob_1926 _Iyob_1927 _Iyob_1928 _Iyob_1929 _Iqob_2 _Iqob_3
               _Iqob_4
```

Instrumenting for education here has led to a coefficient of 0.085 on education, but this is insignificantly different from zero (though note that with unrobust standard errors the coefficient retains its significance).

Angrist and Krueger actually use the interactions of the quarter of birth dummies with year of birth dummies, so have a larger set of instruments. A useful option for `ivregress` is `first` which reports the first stage regression as well as the full output. This is shown below:

```
. xi: ivregress 2sls lwage race married smsa ageq ageq2 i.region i.yob (educ=i.yob*i.qob),
vce(robust) first
i.region          _Iregion_0-8        (naturally coded; _Iregion_0 omitted)
i.yob             _Iyob_1920-1929     (naturally coded; _Iyob_1920 omitted)
i.qob             _Iqob_1-4           (naturally coded; _Iqob_1 omitted)
i.yob*i.qob       _IyobXqob_#_#       (coded as above)
note: _IyobXqob_1929_4 dropped due to collinearity

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

```
                                                  Number of obs =   247199
                                                  F( 50, 247148) =  318.76
                                                  Prob > F      =   0.0000
                                                  R-squared     =   0.0699
                                                  Adj R-squared =   0.0697
                                                  Root MSE      =   3.2414

------------------------------------------------------------------------------
             |               Robust
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        race |  -2.317499   .0263069   -88.09   0.000     -2.36906   -2.265938
     married |     .40606   .0225021    18.05   0.000     .3619564    .4501636
        smsa |  -.5882687   .0144997   -40.57   0.000    -.6166877   -.5598497
        ageq |    .442292   .5560267     0.80   0.426    -.6475056     1.53209
       ageq2 |  -.0056273   .0061598    -0.91   0.361    -.0177004    .0064458
  _Iregion_1 |  -.7393628   .0219049   -33.75   0.000    -.7822959   -.6964297
  _Iregion_2 |  -1.652889   .0355642   -46.48   0.000    -1.722594   -1.583185
  _Iregion_3 |   -.455914   .0223444   -20.40   0.000    -.4997083   -.4121196
  _Iregion_4 |   -.155019   .0370815    -4.18   0.000    -.2276977   -.0823403
  _Iregion_5 |  -.4105518   .0323985   -12.67   0.000    -.4740521   -.3470515
  _Iregion_6 |  -1.104921   .0258032   -42.82   0.000    -1.155495   -1.054348
  _Iregion_7 |  -.5818435   .0288952   -20.14   0.000    -.6384772   -.5252097
  _Iregion_8 |  -1.139735   .0309536   -36.82   0.000    -1.200404   -1.079067
  _Iyob_1921 |   .1376454   .0630084     2.18   0.029     .0141505    .2611402
  _Iyob_1922 |   .0031841   .0922253     0.03   0.972     -.177575    .1839433
  _Iyob_1923 |   .0358809   .1167224     0.31   0.759    -.1928919    .2646537
  _Iyob_1924 |   .0087671   .1319694     0.07   0.947    -.2498895    .2674237
  _Iyob_1925 |  -.0293336   .1361694    -0.22   0.829    -.2962222    .2375549
  _Iyob_1926 |  -.0539437   .1294869    -0.42   0.677    -.3077346    .1998472
  _Iyob_1927 |  -.1149449   .1115026    -1.03   0.303    -.3334871    .1035973
  _Iyob_1928 |  -.0977601   .0852882    -1.15   0.252    -.2649227    .0694025
  _Iyob_1929 |  (dropped)
     _Iqob_2 |   .0904968   .0546323     1.66   0.098     -.016581    .1975746
     _Iqob_3 |   .2115748   .0544615     3.88   0.000     .1048317    .3183179
     _Iqob_4 |   .2005132   .0419578     4.78   0.000      .118277    .2827494
_IyobXqo~1_2 |  -.1003874   .0814898    -1.23   0.218    -.2601052    .0593304
_IyobXqo~1_3 |  -.2249199   .0801459    -2.81   0.005    -.3820038   -.0678361
_IyobXqo~1_4 |  -.1309525   .0782446    -1.67   0.094    -.2843099    .0224049
_IyobXqo~2_2 |  -.0854015   .0820744    -1.04   0.298    -.2462652    .0754623
_IyobXqo~2_3 |  -.0504703   .0798497    -0.63   0.527    -.2069737    .1060331
_IyobXqo~2_4 |  -.1099786   .0765667    -1.44   0.151    -.2600474    .0400901
_IyobXqo~3_2 |  -.0911235    .080527    -1.13   0.258    -.2489544    .0667073
_IyobXqo~3_3 |  -.1085343   .0794706    -1.37   0.172    -.2642946     .047226
_IyobXqo~3_4 |  -.1499147   .0726453    -2.06   0.039    -.2922975   -.0075319
_IyobXqo~4_2 |  -.0964844   .0798402    -1.21   0.227    -.2529692    .0600003
_IyobXqo~4_3 |  -.1395907   .0790225    -1.77   0.077    -.2944728    .0152914
_IyobXqo~4_4 |   -.100248   .0713929    -1.40   0.160    -.2401763    .0396803
_IyobXqo~5_2 |   -.121421   .0802058    -1.51   0.130    -.2786222    .0357803
_IyobXqo~5_3 |  -.1286412   .0800122    -1.61   0.108    -.2854629    .0281806
_IyobXqo~5_4 |  -.1701297   .0722058    -2.36   0.018    -.3116511   -.0286083
_IyobXqo~6_2 |  -.1053616   .0809432    -1.30   0.193    -.2640082     .053285
_IyobXqo~6_3 |  -.0492929   .0815955    -0.60   0.546    -.2092179    .1106321
_IyobXqo~6_4 |  -.0674566    .073929    -0.91   0.362    -.2123555    .0774423
_IyobXqo~7_2 |  -.0674867   .0800262    -0.84   0.399    -.2243359    .0893624
```

```
_IyobXqo~7_3 |  -.1247802   .0823747    -1.51   0.130    -.2862324    .0366721
_IyobXqo~7_4 |  -.0731652   .0747712    -0.98   0.328    -.2197147    .0733843
_IyobXqo~8_2 |  -.0369743   .0810478    -0.46   0.648    -.1958259    .1218773
_IyobXqo~8_3 |  -.1049247   .0851365    -1.23   0.218      -.27179    .0619405
_IyobXqo~8_4 |  -.0316991   .0792725    -0.40   0.689    -.1870712     .123673
_IyobXqo~9_2 |  -.0442215   .0738958    -0.60   0.550    -.1890554    .1006124
_IyobXqo~9_3 |  -.1141512   .0723279    -1.58   0.115    -.2559121    .0276097
       _cons |   3.673014   12.42174     0.30   0.767    -20.67326    28.01929
------------------------------------------------------------------------------


Instrumental variables (2SLS) regression               Number of obs =  247199
                                                       Wald chi2(23) =28822.61
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.2065
                                                       Root MSE      =  .58017

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1007152   .0335512     3.00   0.003     .0349561    .1664742
        race |  -.2270555   .0779375    -2.91   0.004    -.3798102   -.0743008
     married |   .2803622   .0143412    19.55   0.000     .2522539    .3084705
        smsa |  -.1163201   .0199067    -5.84   0.000    -.1553365   -.0773037
        ageq |   .1170352   .0661708     1.77   0.077    -.0126573    .2467276
       ageq2 |  -.0011772   .0007366    -1.60   0.110    -.0026209    .0002664
  _Iregion_1 |   .0423372   .0251233     1.69   0.092    -.0069035    .0915779
  _Iregion_2 |  -.1571906   .0558419    -2.81   0.005    -.2666387   -.0477424
  _Iregion_3 |   .0008335   .0158882     0.05   0.958    -.0303069    .0319738
  _Iregion_4 |  -.1220909   .0085285   -14.32   0.000    -.1388065   -.1053752
  _Iregion_5 |  -.0201888   .0149419    -1.35   0.177    -.0494742    .0090967
  _Iregion_6 |   -.069971   .0373572    -1.87   0.061    -.1431898    .0032478
  _Iregion_7 |  -.1236594   .0204492    -6.05   0.000    -.1637391   -.0835796
  _Iregion_8 |  -.1165475   .0386157    -3.02   0.003    -.1922328   -.0408622
  _Iyob_1921 |   .0010063    .010104     0.10   0.921    -.0187972    .0208099
  _Iyob_1922 |   .0020042   .0190431     0.11   0.916    -.0353196     .039328
  _Iyob_1923 |   .0079007   .0261617     0.30   0.763    -.0433753    .0591767
  _Iyob_1924 |   .0154163   .0335622     0.46   0.646    -.0503645     .081197
  _Iyob_1925 |   .0336961   .0416298     0.81   0.418    -.0478969     .115289
  _Iyob_1926 |    .043175   .0475898     0.91   0.364    -.0500992    .1364492
  _Iyob_1927 |   .0584277   .0562273     1.04   0.299    -.0517758    .1686313
  _Iyob_1928 |   .0703787   .0622065     1.13   0.258    -.0515438    .1923013
  _Iyob_1929 |   .0679547   .0670612     1.01   0.311    -.0634828    .1993922
       _cons |   .9314844   1.624969     0.57   0.566    -2.253397    4.116366
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   race married smsa ageq ageq2 _Iregion_1 _Iregion_2 _Iregion_3
               _Iregion_4 _Iregion_5 _Iregion_6 _Iregion_7 _Iregion_8
               _Iyob_1921 _Iyob_1922 _Iyob_1923 _Iyob_1924 _Iyob_1925
               _Iyob_1926 _Iyob_1927 _Iyob_1928 _Iyob_1929 _Iqob_2 _Iqob_3
               _Iqob_4 _IyobXqob_1921_2 _IyobXqob_1921_3 _IyobXqob_1921_4
               _IyobXqob_1922_2 _IyobXqob_1922_3 _IyobXqob_1922_4
               _IyobXqob_1923_2 _IyobXqob_1923_3 _IyobXqob_1923_4
               _IyobXqob_1924_2 _IyobXqob_1924_3 _IyobXqob_1924_4
               _IyobXqob_1925_2 _IyobXqob_1925_3 _IyobXqob_1925_4
```

```
                  _IyobXqob_1926_2 _IyobXqob_1926_3 _IyobXqob_1926_4
                  _IyobXqob_1927_2 _IyobXqob_1927_3 _IyobXqob_1927_4
                  _IyobXqob_1928_2 _IyobXqob_1928_3 _IyobXqob_1928_4
                  _IyobXqob_1929_2 _IyobXqob_1929_3
```

Looking at the first stage we see that few of the instruments have coefficients significantly different from zero – we might be concerned that the instrument is weak. We will see how to test for this later. The coefficient on education is 0.10 and is significantly different from zero.

We can estimate this model using the `gmm` and `liml` options:

```
. xi: ivregress liml lwage race married smsa ageq ageq2 i.region i.yob (educ=i.yob*i.qob),
vce(robust)
```

(output omitted)

```
. xi: ivregress gmm lwage race married smsa ageq ageq2 i.region i.yob (educ=i.yob*i.qob),
vce(robust)
```

(output omitted)

Storing the estimates, we can then create a table to compare them:

```
. esttab ols twosls liml gmm, b se keep(educ ageq ageq2 race married smsa) mtitles
```

	(1)	(2)	(3)	(4)
	ols	twosls	liml	gmm
educ	0.0701***	0.101**	0.282	0.100**
	(0.000388)	(0.0336)	(0.358)	(0.0335)
race	-0.298***	-0.227**	0.193	-0.228**
	(0.00470)	(0.0779)	(0.829)	(0.0779)
married	0.293***	0.280***	0.207	0.280***
	(0.00440)	(0.0143)	(0.145)	(0.0143)
smsa	-0.134***	-0.116***	-0.00960	-0.117***
	(0.00259)	(0.0199)	(0.211)	(0.0199)
ageq	0.116	0.117	0.122	0.119
	(0.0651)	(0.0662)	(0.103)	(0.0663)
ageq2	-0.00125	-0.00118	-0.000743	-0.00120
	(0.000722)	(0.000737)	(0.00141)	(0.000738)
N	247199	247199	247199	247199

```
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

We see that the LIML coefficients are less precisely estimated than the other approaches. Instrumenting appears to increase the effect of schooling on wages.

4 Postestimation tests

Stata has a series of commands that can be used after `ivregress` to test various hypotheses. We first rerun the 2SLS regression.

```
. xi: ivregress 2sls lwage race married smsa ageq ageq2 i.region i.yob (educ=i.yob*i.qob),
vce(robust)
```

(output omitted)

We can test for the endogeneity of education. The IV approach assumes that education is endogenous – if it is in fact exogenous then OLS would be more efficient. The command `estat endogenous` performs a test for this:

```
. estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Robust score chi2(1)            =  .857777  (p = 0.3544)
  Robust regression F(1,247174)   =  .857705  (p = 0.3544)
```

This implements the Durbin-Wu-Hausman test (see p.183 of Cameron and Trivedi (2009) for details). The null hypothesis is that education is exogenous. Here we see that the hypothesis is not rejected so in fact we cannot reject exogeneity of education in this model.

A second test is a test of overidentifying restrictions. This is possible when there are more instruments than endogenous variables. The test assumes that one instrument is valid and then tests for the validity of all other instruments (ie. whether the instruments are uncorrelated with the error term in the second stage).

```
. estat overid

  Test of overidentifying restrictions:

  Score chi2(28)          =  29.0241  (p = 0.4113)
```

We do not rejected the overidentifying restrictions. One note of caution here is that the test assumes that at least one instrument is valid. All of our instruments here are drawn from the same concept of quarter of birth affecting the amount of schooling – so, we would expect either all instruments, or no instruments, to be valid.

We can also test the strength of the instruments using `estat firststage`:

```
. estat firststage
note: _Iyob_1929 dropped because of collinearity
```

(first stage regression output omitted)

```
 ( 1)  _Iqob_2 = 0
 ( 2)  _Iqob_3 = 0
 ( 3)  _Iqob_4 = 0
 ( 4)  _IyobXqob_1921_2 = 0
 ( 5)  _IyobXqob_1921_3 = 0
 ( 6)  _IyobXqob_1921_4 = 0
 ( 7)  _IyobXqob_1922_2 = 0
 ( 8)  _IyobXqob_1922_3 = 0
 ( 9)  _IyobXqob_1922_4 = 0
 (10)  _IyobXqob_1923_2 = 0
 (11)  _IyobXqob_1923_3 = 0
 (12)  _IyobXqob_1923_4 = 0
 (13)  _IyobXqob_1924_2 = 0
 (14)  _IyobXqob_1924_3 = 0
 (15)  _IyobXqob_1924_4 = 0
 (16)  _IyobXqob_1925_2 = 0
```

```
(17)  _IyobXqob_1925_3 = 0
(18)  _IyobXqob_1925_4 = 0
(19)  _IyobXqob_1926_2 = 0
(20)  _IyobXqob_1926_3 = 0
(21)  _IyobXqob_1926_4 = 0
(22)  _IyobXqob_1927_2 = 0
(23)  _IyobXqob_1927_3 = 0
(24)  _IyobXqob_1927_4 = 0
(25)  _IyobXqob_1928_2 = 0
(26)  _IyobXqob_1928_3 = 0
(27)  _IyobXqob_1928_4 = 0
(28)  _IyobXqob_1929_2 = 0
(29)  _IyobXqob_1929_3 = 0

      F( 29,247148) =    3.02
           Prob > F =    0.0000

 First-stage regression summary statistics
 --------------------------------------------------------------------------
             |            Adjusted      Partial       Robust
    Variable |   R-sq.       R-sq.        R-sq.   F(29,247148)  Prob > F
 -------------+------------------------------------------------------------
        educ |  0.0699      0.0697       0.0001       3.01972    0.0000
 --------------------------------------------------------------------------
```

This tests the joint significance of all of the instruments. We see that they are jointly significantly different from zero, with a p=value of 0.0000. Their partial R squared is just 0.0001. This casts concern over the strength of the instruments. Generally an F statistic over 10 is required to suggest instruments are sufficiently strong. If the instruments are weak, we may find that 2SLS gives standard errors which are too small. LIML is thought to be a better approach if instruments are weak. See Murray (2006) for a good discussion on approaches to take to avoid problems with weak instruments.

5 An alternative command

Instead of using `ivregress` we can use the user written command `ivreg2`. This automatically calculates many additional statistics. I show it here with the `first` option to give all first stage statistics. Note that to get robust standard errors here the option is `robust` and not `vce(robust)`. Otherwise the syntax is similar to that for `ivregress`.

```
. xi: ivreg2 lwage race married smsa ageq ageq2 i.region i.yob (educ=i.yob*i.qob),
robust first
```

(first stage regression output omitted)

```
Partial R-squared of excluded instruments:   0.0001
Test of excluded instruments:
  F( 28,247148) =     1.03
  Prob > F      =   0.4217



Summary results for first-stage regressions
-------------------------------------------


Variable     | Shea Partial R2 |   Partial R2    |  F( 28,247148)    P-value
educ         |     0.0001      |     0.0001      |        1.03       0.4217
```

```
NB: first-stage F-stat heteroskedasticity-robust

Underidentification tests
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Kleibergen-Paap rk LM statistic            Chi-sq(28)=28.83   P-val=0.4211
Kleibergen-Paap rk Wald statistic          Chi-sq(28)=28.84   P-val=0.4206

Weak identification test
Ho: equation is weakly identified
Kleibergen-Paap Wald rk F statistic                1.03
See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and overidentifying restrictions are valid
Anderson-Rubin Wald test     F(28,247148)=1.21   P-val=0.2051
Anderson-Rubin Wald test     Chi-sq(28)=33.88    P-val=0.2048
Stock-Wright LM S statistic  Chi-sq(28)=33.94    P-val=0.2028

NB: Underidentification, weak identification and weak-identification-robust
    test statistics heteroskedasticity-robust

Number of observations              N  =     247199
Number of regressors                K  =         24
Number of instruments               L  =         51
Number of excluded instruments      L1 =         28

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                      Number of obs =   247199
                                                      F( 23,247175) =  1253.02
                                                      Prob > F      =   0.0000
Total (centered) SS     =  104853.0198                Centered R2   =   0.2064
Total (uncentered) SS   =  6674371.774                Uncentered R2 =   0.9875
Residual SS             =  83215.58688                Root MSE      =    .5802

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1007806   .0335533     3.00   0.003     .0350173    .1665438
        race |  -.2269039   .0779425    -2.91   0.004    -.3796683   -.0741394
     married |   .2803356   .0143421    19.55   0.000     .2522255    .3084456
        smsa |  -.1162816    .019908    -5.84   0.000    -.1553005   -.0772627
        ageq |   .1178817   .0661751     1.78   0.075    -.0118191    .2475826
       ageq2 |  -.0011864   .0007366    -1.61   0.107    -.0026301    .0002574
   _Iregion_1 |   .0423854   .0251249     1.69   0.092    -.0068584    .0916292
   _Iregion_2 |  -.1570824   .0558455    -2.81   0.005    -.2665375   -.0476272
   _Iregion_3 |   .0008633   .0158892     0.05   0.957    -.0302791    .0320056
   _Iregion_4 |  -.1220805   .0085292   -14.31   0.000    -.1387974   -.1053637
```

```
  _Iregion_5 |   -.020162   .0149428    -1.35   0.177    -.0494494    .0091253
  _Iregion_6 |  -.0698987   .0373596    -1.87   0.061    -.1431221    .0033248
  _Iregion_7 |  -.1236214   .0204505    -6.04   0.000    -.1637036   -.0835392
  _Iregion_8 |  -.1164728   .0386181    -3.02   0.003    -.1921629   -.0407827
  _Iyob_1921 |   .0009451   .0101047     0.09   0.925    -.0188597    .0207499
  _Iyob_1922 |   .0019085   .0190443     0.10   0.920    -.0354176    .0392346
  _Iyob_1923 |    .007784   .0261633     0.30   0.766    -.0434951    .0590632
  _Iyob_1924 |   .0152995   .0335643     0.46   0.649    -.0504853    .0810843
  _Iyob_1925 |   .0336004   .0416324     0.81   0.420    -.0479976    .1151984
  _Iyob_1926 |   .0431142   .0475927     0.91   0.365    -.0501658    .1363942
  _Iyob_1927 |   .0584271   .0562308     1.04   0.299    -.0517833    .1686376
  _Iyob_1928 |   .0704497   .0622104     1.13   0.257    -.0514805    .1923798
  _Iyob_1929 |   .0681133   .0670654     1.02   0.310    -.0633324     .199559
       _cons |   .9112057   1.625074     0.56   0.575    -2.273881    4.096292
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):             28.832
                                                   Chi-sq(28) P-val =   0.4211
------------------------------------------------------------------------------
Weak identification test (Kleibergen-Paap rk Wald F statistic):          1.030
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias    21.42
                                         10% maximal IV relative bias    11.34
                                         20% maximal IV relative bias     6.13
                                         30% maximal IV relative bias     4.32
                                         10% maximal IV size             81.40
                                         15% maximal IV size             42.37
                                         20% maximal IV size             29.12
                                         25% maximal IV size             22.43
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):        29.020
                                                   Chi-sq(27) P-val =   0.3599
------------------------------------------------------------------------------
Instrumented:         educ
Included instruments: race married smsa ageq ageq2 _Iregion_1 _Iregion_2
                      _Iregion_3 _Iregion_4 _Iregion_5 _Iregion_6 _Iregion_7
                      _Iregion_8 _Iyob_1921 _Iyob_1922 _Iyob_1923 _Iyob_1924
                      _Iyob_1925 _Iyob_1926 _Iyob_1927 _Iyob_1928 _Iyob_1929
Excluded instruments: _Iqob_2 _Iqob_3 _IyobXqob_1921_2 _IyobXqob_1921_3
                      _IyobXqob_1921_4 _IyobXqob_1922_2 _IyobXqob_1922_3
                      _IyobXqob_1922_4 _IyobXqob_1923_2 _IyobXqob_1923_3
                      _IyobXqob_1923_4 _IyobXqob_1924_2 _IyobXqob_1924_3
                      _IyobXqob_1924_4 _IyobXqob_1925_2 _IyobXqob_1925_3
                      _IyobXqob_1925_4 _IyobXqob_1926_2 _IyobXqob_1926_3
                      _IyobXqob_1926_4 _IyobXqob_1927_2 _IyobXqob_1927_3
                      _IyobXqob_1927_4 _IyobXqob_1928_2 _IyobXqob_1928_3
                      _IyobXqob_1928_4 _IyobXqob_1929_2 _IyobXqob_1929_3
Duplicates:           _Iyob_1921 _Iyob_1922 _Iyob_1923 _Iyob_1924 _Iyob_1925
                      _Iyob_1926 _Iyob_1927 _Iyob_1928 _Iyob_1929
Dropped collinear:    _Iqob_4 _IyobXqob_1929_4
------------------------------------------------------------------------------
```

References

Angrist, Joshua D. and Aan B. Krueger, "Does compulsory school attendance affect schooling and earnings?," *The Quarterly Journal of Economics*, November 1991, *106* (4), 979–1014.

___ **and Jörn-Steffen Pischke**, *Mostly Harmless Econometrics*, Princeton University Press, 2009.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics Using Stata*, Texas: Stata Press, 2009.

Greene, William H., *Econometric Analysis*, 6th ed., Pearson/Prentice Hall, 2008.

Murray, Michael P., "Avoiding invalid instruments and coping with weak instruments," *The Journal of Economic Perspectives*, 2006, *20* (4), 111–132.

Wooldridge, Jeffrey M., *Introductory Econometrics: A Modern Approach*, 4th ed., South Western / Cengage Learning, 2009.