Measurement Error and IV estimation

The data set `ivdat.dta` contains information on the number of GCSE passes of a sample of 16 year olds and the total income of the household in which they live. Income tends to be measured with error. Individuals tend to mis-report incomes, particularly third-party incomes and non-labour income. The following regression may therefore be subject to measurement error in one of the right hand side variables, (the gender dummy variable is less subject to error).

```plaintext
.reg nqfede inc1 female
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 252</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>274.029395</td>
<td>2</td>
<td>137.014698</td>
<td>F( 2, 249) = 14.55</td>
</tr>
<tr>
<td>Residual</td>
<td>2344.9706</td>
<td>249</td>
<td>9.41755263</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2619.00</td>
<td>251</td>
<td>10.4342629</td>
<td>Adj R-squared = 0.0974</td>
</tr>
</tbody>
</table>

| nqfede   | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-----------|-------|-----------|-------|------|----------------------|
| inc1  | .0396859 | 0.0087786 | 4.52  | 0.000 | .022396 - .0569758  |
| female | 1.172351 | .387686  | 3.02  | 0.003 | .4087896 - 1.935913 |
| _cons  | 4.929297 | .4028493 | 12.24 | 0.000 | 4.13587 - 5.722723  |

Econometric theory tells us that in the presence of classical measurement error, (the error is not correlated with the unobserved true value of income), then the OLS coefficient on income is biased toward zero – and the coefficient on gender is biased in some unknown way.

The income variables is therefore replaced with an instrument – in this case the rank of the household in the income distribution (ie 1st, 252nd etc) – assuming that the rank is unaffected by measurement error so that the error is not so large as to change the ranking of households in the income distribution, but correlated with the level of income (by construction).

```plaintext
.ivreg nqfede (inc1=ranki) female, first
```

**First-stage regressions**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 252</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>81379.4112</td>
<td>2</td>
<td>40689.7056</td>
<td>F( 2, 249) = 247.94</td>
</tr>
<tr>
<td>Residual</td>
<td>40863.626</td>
<td>249</td>
<td>164.110948</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>122243.037</td>
<td>251</td>
<td>487.024053</td>
<td>Adj R-squared = 0.6630</td>
</tr>
</tbody>
</table>

| nqfede   | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-----------|-------|-----------|-------|------|----------------------|
| incl     | .2342779 | 1.618777 | 0.14  | 0.885 | -2.953962 - 3.422518 |
| female   | .2470712 | .0110979 | 22.26 | 0.000 | .2252136 - .2689289 |
| ranki    | .7722511 | 1.855748 | 0.42  | 0.678 | -2.882712 - 4.427214 |

Can see that instrument satisfies 1st assumption since it is highly correlated with income, net of the other exogenous variables.
The 2nd stage of the estimation takes the predicted values from this regression and regresses number of GCSE passes on these predicted values (Note can show predicted value of an exogenous variable is equal to the actual value so the variable “female” enters the second stage regression as its own instrument).

Instrumental variables (2SLS) regression

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 252</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>270.466601</td>
<td>2</td>
<td>135.2333</td>
<td>F( 2, 249) = 13.09</td>
</tr>
<tr>
<td>Residual</td>
<td>2348.5334</td>
<td>249</td>
<td>9.43186104</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2619.00</td>
<td>251</td>
<td>10.4342629</td>
<td>R-squared = 0.1033</td>
</tr>
</tbody>
</table>

Instrumented: inc1
Instruments: female ranki

When this is done the size of the coefficient on income rises (as expected) – though the standard error around this estimate is larger (as expected).

To get estimates robust to heteroskedasticity of unknown form type

`ivreg nqfede (inc1=ranki) female, first robust`

First-stage regressions - omitted

IV (2SLS) regression with robust standard errors

| nqfede | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|--------|-------|-----------|------|------|-----------------------|
| inc1   | 0.0450854 | 0.0107683  | 4.19 | 0.000 | 0.0238768 .066294    |
| female | 1.176652  | 0.3880121  | 3.03 | 0.003 | 0.4124481 1.940856   |
| _cons  | 4.753386  | 0.4513194  | 10.53| 0.000 | 3.864496 5.642277    |

Instrumented: inc1
Instruments: female ranki

In this example, allowance for heteroskedasticity makes little difference to the estimated standard errors.
Note in the simple 2 variable model we can get an idea of the likely size of the attenuation bias, since the reciprocal on the coefficient on GCSE passes from a reverse regression of income on exam passes gives an upper bound on the true unobserved effect of income on GCSE passes.

The simple regression gives

```
reg nqfede incl

Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =   19.32
Model |  187.911492     1  187.911492           Prob > F      =  0.0000
Residual | 2431.08851   250  9.72435403           R-squared     =  0.0717
-------------+------------------------------           Adj R-squared =  0.0680
Total |  2619.00   251 10.4342629           Root MSE      =  3.1184

------------------------------------------------------------------------------
nqfede |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
incl |   .0392071    .008919     4.40   0.000      .021641    .0567731
_cons |   5.572737   .3475979    16.03   0.000     4.888143     6.25733
```

The reverse regression gives

```
reg incl nqfede

Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =   19.32
Model |  8770.85587     1  8770.85587           Prob > F      =  0.0000
Residual | 113472.181   250  453.888725           R-squared     =  0.0717
-------------+------------------------------           Adj R-squared =  0.0680
Total | 122243.037   251  487.024053           Root MSE      =  21.305

------------------------------------------------------------------------------
nqfede |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
incl |   1.830009   .4163005     4.40   0.000     1.010106    2.649912
_cons |   19.64721   3.145406     6.25   0.000     13.45233    25.84208
```

So the reciprocal of the estimated coefficient on number of gcses is

```
. display 1/_b[nqfede]
  0.54644538
```

Hence we know the true value of the income effect lies in the range

```
0.039 < b < .545
```

which is a very large bound.

Another way of looking at the issue is to recognise that since the R^2 from the two regressions is the same we have

```
R^2 = b_{y,x}/1/b_{x,y} = .039/(.546) = 0.0717
```

So that low R^2 typical of cross section regressions means that the potential for measurement error to affect the results significantly is quite high, (but less so in time series regressions where R^2 are typically in the region of 0.9)
Weak Instruments
Contrast the above IV results with the use of an alternative instrument for household income – the number of children in the household. Number of children is weakly negatively correlated with wealth and so is likely to be a weak instrument – as the following output shows.

ivreg2 nqfede (incl=nkids) female, first small

First-stage regression of incl:

|                     | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|---------------------|-------|-----------|-------|-----|----------------------|
| female              | -0.6897 | 2.7896 | -0.25 | 0.805 | -6.1840 to 4.7049   |
| nkids               | -1.7851 | 1.349 | -1.32 | 0.187 | -4.4514 to 0.9809   |
| _cons               | 34.0590 | 2.327 | 14.64 | 0.000 | 29.4755 to 38.6423  |

Partial R-squared of excluded instruments: 0.0070
Test of excluded instruments:
F( 1, 249) = 1.75
Prob > F = 0.1869

Instrumental variables (2SLS) regression

|                     | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|---------------------|-------|-----------|-------|-----|----------------------|
| incl                | 0.0065 | 0.1080 | 0.06  | 0.952 | -0.2062 to 0.2192    |
| female              | 1.1459 | 0.4078 | 2.81  | 0.005 | 0.3428 to 1.9490     |
| _cons               | 6.0103 | 3.5306 | 1.70  | 0.090 | -0.9434 to 12.9639   |

Sargan statistic (overidentification test of all instruments): 0.000
(equation exactly identified)

The result is that when used as an instrument the IV estimate is far away from the original OLS estimate, the original IV estimate and has a large standard error so is also statistically insignificant.

Note: When there is only a single endogenous regressor, can use a rule of thumb that the F value in the 1st stage of the regression of a test that the coefficients on the instruments is less than 10 indicates that the instrument(s) are weak (or a low partial R² ) and that 2SLS will be biased in this case.
Can see that this is the case above, but not when income rank is used as the instrument.

ivreg2 nqfede (incl=ranki) female, first small

First-stage regressions

|                | Coef. | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|----------------|-------|-----------|------|------|---------------------|
| female         | .2342779 | 1.618777   | 0.14 | 0.885 | -2.953963 - 3.422518 |
| ranki          | .2470712 | .0110979   | 22.26| 0.000 | .2252136 - .2689289 |
| _cons          | .7722511 | 1.855748   | 0.42 | 0.678 | -2.882712 - 4.427215 |

Partial R-squared of excluded instruments:  0.6656

Test of excluded instruments:
  F( 1, 249) = 495.64
  Prob > F = 0.0000

Note: too high a correlation between X and Z and we may suspect that the instrument is unlikely to satisfy the second requirement that Z be uncorrelated with the residual.
Tests of Endogeneity

Since the assumption $\text{Cov}(Z,u)$ can never be observed, have to use alternative ways of testing this assumption. The most commonly used test is the Hausman test, based on a comparison of the OLS and IV estimates. Under the null of no endogeneity, OLS is consistent and efficient while IV is consistent but inefficient. If endogeneity exists then only IV is consistent.

The test can be done automatically in Stata. Simply type the following commands:

```stata
. quietly ivreg nqfede (inc1=ranki) female
. est store iv
. quietly reg nqfede inc1 female
. hausman iv
```

<table>
<thead>
<tr>
<th></th>
<th>Prior</th>
<th>Current</th>
<th>Difference</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>incl</td>
<td>0.0450854</td>
<td>0.0396859</td>
<td>0.0053995</td>
<td>0.0062363</td>
</tr>
<tr>
<td>female</td>
<td>1.176652</td>
<td>1.172351</td>
<td>0.0043008</td>
<td>0.0159046</td>
</tr>
</tbody>
</table>

Test:  Ho: difference in coefficients not systematic

$$\text{chi2}(2) = (b-B)'[(V_{b-V_B})^{-1}](b-B)$$

$$= 0.75$$

Prob>chi2 = 0.6874

In this case can’t reject the null: there is no significant difference between the IV and OLS estimates, and so can **not reject** the null of no endogeneity.

Note 1: one of the reasons that can’t reject the null is that there is a larger standard error associated with the IV estimates. Even a poor instrument can pass this test - as can be seen using number of children rather than income rank.

```stata
. quietly ivreg nqfede (inc1=nkids) female
. est store iv
. quietly reg nqfede inc1 female
. hausman iv
```

<table>
<thead>
<tr>
<th></th>
<th>Prior</th>
<th>Current</th>
<th>Difference</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>incl</td>
<td>0.0065052</td>
<td>0.0396859</td>
<td>-0.0331807</td>
<td>0.1076429</td>
</tr>
<tr>
<td>female</td>
<td>1.145922</td>
<td>1.172351</td>
<td>-0.0264292</td>
<td>0.12638</td>
</tr>
</tbody>
</table>

Test:  Ho: difference in coefficients not systematic

$$\text{chi2}(2) = (b-B)'[(V_{b-V_B})^{-1}](b-B)$$

$$= 0.10$$

Prob>chi2 = 0.9536

Again can’t reject the null.
Moral: Test is only as good as the instruments used.

Note 2: can also use heteroskedastic “robust” option here if heteroskedasticity is suspected.

Note 3: An asymptotic equivalent version of the test, (Wu-Hausman), is to take the residuals from the 1st stage of the regression (the regression of the endogenous variable on its instruments), and include them as additional regressor(s) in the original OLS equation (include an estimated residual for each endogenous rhs variable).

```
reg incl ranki female
Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  2,   249) =  247.94
Model |  81379.4112     2  40689.7056           Prob > F      =  0.0000
Residual |  40863.626   249  164.110948           R-squared     =  0.6657
-------------+------------------------------           Adj R-squared =  0.6630
Total |  122243.037   251  487.024053           Root MSE      =  12.811

------------------------------------------------------------------------------
       incl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    ranki |   .2470712   .0110979    22.26   0.000     .2252136    .2689289
  female |   .2342779   1.618777     0.14   0.885    -2.953962    3.422518
    _cons |   .7722511   1.855748     0.42   0.678    -2.882712    4.427214
------------------------------------------------------------------------------

. predict uhat, resid
.
. reg nqfede incl female uhat
Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  3,   248) =    9.94
Model |  281.121189     3  93.7070629           Prob > F      =  0.0000
Residual |  2337.87881   248  9.42693069           R-squared     =  0.1073
-------------+------------------------------           Adj R-squared =  0.0965
Total |     2619.00   251  10.4342629           Root MSE      =  3.0703

------------------------------------------------------------------------------
nqfede |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   incl |   .0450854   .0107655     4.19   0.000     .0238819    .0662888
  female |   1.176652   .3879107     3.03   0.003     .4126329    1.940672
   uhat |  -.0161473   .0186169    -0.87   0.387    -.0528147    .0205201
    _cons |   4.753386   .4512015    10.53   0.000     3.864711    5.642062
------------------------------------------------------------------------------

In this case the t value on the residual is insignificant which again suggests that endogeneity may not be an issue here.

Note you can also get this result by typing the following command:

ivendog
```
Tests of endogeneity of: incl
H0: Regressor is exogenous
Wu-Hausman F test: 0.75229 F(1,248) P-value = 0.38659
Durbin-Wu-Hausman chi-sq test: 0.76211 Chi-sq(1) P-value = 0.38267
the first test is simply the square of the t value in the last regression – since $t^2 = F$)

the second test is the Hausman test based solely on a comparison of the OLS and IV estimates of income (and not female)

Note that the estimated coefficients on income and gender in this “augmented” regression are identical to those in the IV estimation on page 2. Why? – Essentially, the OLS residuals “$uhat$” are orthogonal to female by construction – see notes on algebra of least squares and problem set 6 – and the residuals net out the effect of the rank of income on actual income effectively creating the predicted value of income so the regression is equivalent to the second regression in 2SLS.

The standard errors in this augmented equation are however invalid in the presence of significant predicted variables: Intuitively, OLS ignores the fact that these predictions are based on sample not population estimates.
Test of Over-identifying Restrictions

If have more instruments then strictly necessary, can test whether the additional instruments are “valid” ie uncorrelated with the residual in the original model. Do this by comparing the IV estimates from a just identified 2SLS model with the IV estimates when all possible instruments are used. Under the null that all instruments are valid both estimates should be consistent (and therefore close to each other, allowing for statistical variation). If the null is not satisfied (some of the extra instruments are invalid), then only the just identified estimates are consistent and there should be a significant difference between the just identified and over-identified IV estimates.

Can do this in 2 ways. First using the automatic version of the Hausman test provided in stata.

. quietly ivreg nqfede (inc1=ranki) female /* just identified */
. hausman, save
. quietly ivreg nqfede (inc1=ranki nkids) female
/* more instruments (2) than strictly necessary so model is over-identified */
. hausman

```
---- Coefficients ----
|      (b)          (B)            (b-B)   sqrt(diag(V_b-V_B))
|     Prior       Current       Difference        S.E.
-------------+-------------------------------------------------------------
inc1 |   .0450854     .0450498         .0000356     .0001079
female |   1.176652     1.176624         .0000284     .0017349
---------------------------------------------------------------------------

b = less efficient estimates obtained previously from ivreg
B = fully efficient estimates obtained from ivreg

Test:  Ho:  difference in coefficients not systematic

chi2(  2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
=     0.11
Prob>chi2 =     0.9470

Estimated coefficients from both IV estimates not significantly different, so can’t reject null that additional instruments are valid (orthogonal to endogenous variable)

Does it matter which instruments are used in the just identified model? – No. If all instruments are valid then the estimates should differ only as a result of sampling variation.
An asymptotic equivalent version of this test is to take the residuals from the 2SLS regression of the model based on all the available instruments (rank, no. children) and regress these residuals on all the instruments. In this case NR² from this auxiliary regression is $X^2_{(l-k)}$ where $l-k$ is the no. of overidentifying restrictions – total no. instruments – total no. endogenous rhs variables

Intuitively these residuals should not be correlated with any exogenous variable if these variables are to be used as instruments

```
. ivreg nqfede (incl=ranki nkids) female
```

Instrumental variables (2SLS) regression

```
<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 252</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>270.513461</td>
<td>2</td>
<td>135.256731</td>
<td>F( 2, 249) = 13.08</td>
</tr>
<tr>
<td>Residual</td>
<td>2348.48654</td>
<td>249</td>
<td>9.43167285</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2619.00</td>
<td>251</td>
<td>10.4342629</td>
<td>R-squared = 0.1033</td>
</tr>
</tbody>
</table>

| nqfede | Coef.   | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------|---------|-----------|-------|-------|---------------------|
| incl   | 0.0450498 | 0.0107678 | 4.18  | 0.000 | 0.0238422 - 0.0662573 |
| female | 1.176624  | 0.3880082 | 3.03  | 0.003 | 0.4124274 - 1.94082 |
| _cons  | 4.754547  | 0.451304  | 10.54 | 0.000 | 3.865687 - 5.643407 |
```

Instrumented: incl
Instruments: female ranki nkids

```
. predict ivres, resid /* save IV residuals */
. reg ivres ranki nkids female /* regress on all exogenous variables */
```

```
<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 252</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1.28181828</td>
<td>3</td>
<td>.427272761</td>
<td>F( 3, 248) = 0.05</td>
</tr>
<tr>
<td>Residual</td>
<td>2347.20471</td>
<td>248</td>
<td>9.4645351</td>
<td>Prob &gt; F = 0.9872</td>
</tr>
<tr>
<td>Total</td>
<td>2348.48652</td>
<td>251</td>
<td>9.35652002</td>
<td>R-squared = 0.0005</td>
</tr>
</tbody>
</table>

| ivres  | Coef.   | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------|---------|-----------|-------|-------|---------------------|
| ranki  | 0.0001009 | 0.0026769 | 0.04  | 0.970 | -.0051714 - 0.0053732 |
| nkids  | 0.0694712  | 0.1887812 | 0.37  | 0.713 | -.3023477 - 0.4412901 |
| female | -0.037353  | 0.3888828 | -0.01 | 0.992 | -.7696694 - 0.7621987 |
| _cons  | -0.0705879  | 0.4839739 | -0.15 | 0.884 | -1.023811 - 0.8826353 |
```

So NR² = 252* 0.0005 = 0.126

From tables critical value of Chi-squared distribution with 1 degree of freedom at 5% level is 3.84

So can’t reject null that additional instruments are valid.

Can obtain these results automatically using the command:
Tests of overidentifying restrictions:
Sargan N*R-sq test         0.138  Chi-sq(1)    P-value = 0.7107
Basmann test              0.135  Chi-sq(1)    P-value = 0.7129

If do same test for just 1 instrument

.ivreg  nqfede (incl=ranki) sex

Instrumental variables (2SLS) regression

Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  2,   249) =   13.09
Model |  270.466601     2    135.2333           Prob > F      =  0.0000
Residual |  2348.5334   249  9.43186104           R-squared     =  0.1033
-------------+------------------------------           Adj R-squared =  0.0961
Total |        2619   251  10.4342629           Root MSE      =  3.0711

------------------------------------------------------------------------------
nqfede |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
incl |   .0450854   .0107683     4.19   0.000     .0238768     .066294
sex |   1.176652   .3880121     3.03   0.003     .4124481    1.940856
_cons |   3.576734   .7221421     4.95   0.000     2.154449    4.999019
------------------------------------------------------------------------------

Instrumented:  incl
Instruments:   sex ranki

predict uhat, resid
.reg uhat sex ranki

Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  2,   249) =    0.00
Model |  1.3642e-12     2  6.8212e-13           Prob > F      =  1.0000
Residual |  2348.53339   249  9.43186102           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared = -0.0080
Total |  2348.53339   251  9.35670675           Root MSE      =  3.0711

------------------------------------------------------------------------------
 uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
sex |  -1.77e-08   .3880761    -0.00   1.000    -.7643301    .7643301
ranki |   2.50e-11   .0026605     0.00   1.000    -.00524     .00524
_cons |   2.19e-08   .7192765     0.00   1.000    -1.416642    1.416642
------------------------------------------------------------------------------

Then method fails (since rhs covariates are uncorrelated with residuals by construction)