This turns out to be exactly the coefficient on `mothereduc` in the first-stage regression. This is no coincidence since regression coefficients are the effect of one variable on another, holding the remaining regressors constant.[2]

```
First Stage Regression:  OLS, using observations 1-428
Dependent variable: educ

                coefficient   std. error   t-ratio    p-value
  ------------------------------------------------------------
  const          9.77510      0.423889     23.06     7.57e-077 ***
  exper          0.0488615    0.0416693     1.173    0.2416
  sq_exper      -0.00128106   0.00124491   -1.029    0.3040
  mothereduc     0.267691     0.0311298     8.599    1.57e-016 ***
```

The correlation between the two sets of residuals yields what is called a partial correlation. This is a correlation where the common effects of `const`, `exper`, and `sq_exper` have been removed. The partial correlation between `e1` and `e2` is 0.3854. Partial correlations play a key role in testing for weak instruments.

## 10.3 Specification Tests

There are three specification tests you will find useful with instrumental variables estimation. By default, **Gretl** computes each of these whenever you estimate a model using two-stage least squares. Below I'll walk you through doing it manually and we'll compare the manual results to the automatically generated ones.

### 10.3.1 Hausman Test

The first test is to determine whether the independent variable(s) in your model is (are) in fact uncorrelated with the model's errors. If so, then least squares is more efficient than the IV estimator. If not, least squares is inconsistent and you should use the less efficient, but consistent, instrumental variable estimator. The null and alternative hypotheses are $H_o: \ Cov(x_i, e_i) = 0$ against $H_a: \ Cov(x_i, e_i) \neq 0$. The first step is to use least squares to estimate the first stage of TSLS

$$x_i = \gamma_1 + \theta_1 z_{i1} + \theta_2 z_{i2} + \nu_i \tag{10.3}$$

and to save the residuals, $\hat{\nu}_i$. Then, add the residuals to the original model

$$y_i = \beta_1 + \beta_2 x_i + \delta\hat{\nu}_i + e_i \tag{10.4}$$

Estimate this equation using least squares and use the $t$-ratio on the coefficient $\delta$ to test the hypothesis. If it is significantly different from zero then the regressor, $x_i$ is not exogenous or

[2]This demonstrates an important outcome of the Frisch-Waugh-Lovell Theorem.

predetermined with respect to $e_i$ and you should use the IV estimator (TSLS) to estimate $\beta_1$ and $\beta_2$. If it is not significant, then use the more efficient estimator, OLS.

The **gretl** script for the Hausman test applied to the wage equation is:

```
open "c:\Program Files\gretl\data\poe\mroz.gdt"
logs wage
list x = const educ exper sq_exper
list z2 = const exper sq_exper mothereduc fathereduc
ols educ z2 --quiet
series ehat2 = $uhat
ols l_wage x ehat2
```

Notice that the equation is **overidentified**. There are two additional instruments, `mothereduc` and `fathereduc`, that are being used for a lone endogenous regressor, `educ`. Overidentification basically means that you have more instruments than necessary to estimate the model. Lines 5 and 6 of the script are used to get the residuals from least squares estimation of the first stage regression, and the last line adds these to the wage model, which is estimated by least squares. The $t$-ratio on `ehat2` =1.671, which is not significant at the 5% level. We would conclude that the instruments are exogenous.

You may have noticed that whenever you use two-stage least squares in **gretl** that the program automatically produces the test statistic for the Hausman test. There are several different ways of computing this statistic so don't be surprised if it differs from the one you compute manually using the above script.

### 10.3.2 Testing for Weak Instruments

To test for weak instruments, regress each independent variable suspected of being contemporaneously correlated with the error ($x_k$) onto all of the instruments (internal and external). Suppose $x_K$ is the endogenous regressor. The first stage regression is:

$$x_K = \gamma_1 + \gamma_2 x_2 + \cdots + \gamma_{K-1} x_{K-1} + \theta_1 z_1 + \cdots + \theta_L z_L + \nu_K \tag{10.5}$$

In this notation, the $z_1, \ldots, z_L$ are the **external** instruments. The others, $x_2, \ldots, z_{K-1}$ are exogenous and are used as instruments for themselves (i.e., internal to the model). If the $F$-statistic associated with the hypothesis that the coefficients on the external instruments, $\theta_1, \ldots, \theta_L$ are jointly zero is less than 10, then you conclude that the instruments are weak. If it is greater than 10, you conclude that the instruments are strong enough. The following script uses least squares to perform three such tests. The first regression assumes there is only one instrument, $z1$; the second that the single instrument is $z2$; the third assumes both are instruments.

```
open "@gretldir\data\poe\mroz.gdt"
smpl wage>0 --restrict
```

```
logs wage
square exper
list x = const educ exper sq_exper
list z2 = const exper sq_exper mothereduc fathereduc
ols educ z2
omit mothereduc fathereduc
```

When `omit` follows an OLS regression, **gretl** estimates a restricted model where the variables listed after it are omitted from the model above. It then performs a joint hypothesis test that the coefficients of the omitted variables are zero against the alternative that one or more are not zero. The `--quiet` option reduces the amount of output you have to wade through by suppressing the regressions; only the test results are printed. The output from **gretl** appears in Figure 10.3 below: Since the $F$ value $= 55.4$, which is well beyond 10. We reject the hypothesis that the (external) instruments `mothereduc` and `fathereduc` are weak in favor of the alternative that they are strong.

**Gretl** proves its worth here. Whenever you estimate a model using two stage least squares, **gretl** will compute the test statistic for the weak instruments test.

### 10.3.3 Sargan Test

The final test is the **Sargan test** of the overidentifying restrictions implied by an overidentified model. Recall that to be overidentified just means that you have more instruments than you have endogenous regressors. In our example we have a single endogenous regressor (`educ`) and two instruments, (`mothereduc` and `fatehreduc`). The first step is to estimate the model using TSLS using all the instruments. Save the residuals and then regress these on the instruments alone. $TR^2$ from this regression is approximately $\chi^2$ with the number of surplus instruments as your degrees of freedom. **Gretl** does this easily since it saves $TR^2$ as a part of the usual regression output, where $T$ is the sample size (which we are calling $N$ in cross-sectional examples). The script for the Sargan test follows:

```
1 open "@gretldir\data\poe\mroz.gdt"
2 smpl wage>0 --restrict
3 logs wage
4 square exper
5 list x = const educ exper sq_exper
6 list z2 = const exper sq_exper mothereduc fathereduc
7 tsls l_wage x; z2
8 series ehat2 = $uhat
9 ols ehat2 z2
10 scalar test = $trsq
11 pvalue X 2 test
```

The first 6 lines open the data, restricts the sample, generates logs and squares, and creates the lists of regressors and instruments. In line 7 the model is estimated using TSLS with the variables in list x as regressors and those in z2 as instruments. In line 8 the residuals are saved as ehat2. Then in line 9 a regression is estimated by ordinary least squares using the residuals and instruments as regressors. $TR^2$ is collected and the $p$-value computed in the last line.

The result is:

```
Generated scalar test = 0.378071

Chi-square(2): area to the right of 0.378071 = 0.827757
(to the left: 0.172243)
```

The $p$-value is large and the null hypothesis that the overidentifying restrictions are valid cannot be rejected. The instruments are determined to be ok. Rejection of the null hypothesis can mean that the instruments are either correlated with the errors or that they are omitted variables in the model. In either case, the model as estimated is misspecified.

Finally, **gretl** produces these tests whenever you estimate a model using tsls. If the model is exactly identified, then the Sargan test results are omitted. Here is what the output looks like in the wage example:

Hausman test –
  Null hypothesis: OLS estimates are consistent
  Asymptotic test statistic: $\chi^2(1) = 2.8256$
  with p-value = 0.0927721

Sargan over-identification test –
  Null hypothesis: all instruments are valid
  Test statistic: LM = 0.378071
  with p-value = $P(\chi^2(1) > 0.378071) = 0.538637$

Weak instrument test –
  First-stage $F(2, 423) = 55.4003$

```
Critical values for desired TSLS maximal size, when running
  tests at a nominal 5% significance level:

     size     10%     15%     20%     25%
    value   19.93   11.59    8.75    7.25

  Maximal size is probably less than 10%
```

You can see that the Hausman test statistic differs from the one we computed manually using the script. However, the $p$-value associated with this version and ours above are virtually the same. The results from the instrument strength test and from the Sargan test for overdentification are the same. In conclusion, there is no need to compute any of these tests manually, unless you want to.

Finally, you will also see that some additional information is being printed at the bottom of the test for weak instruments. The rule-of-thumb we have suggested is that if the $F > 10$ then instruments are relatively strong. This begs the question, why not use the usual 5% critical value from the $F$-distribution to conduct the test? The answer is that instrumental variables estimators (though consistent) are biased in small samples. The weaker the instruments, the greater the bias. In fact, the bias is inversely related to the value of the $F$-statistic. An $F = 10$ is roughly equivalent to $1/F = 10\%$ bias in many cases. The other problem caused by weak instruments is that they affect the asymptotic distribution of the usual $t$- and $F$-statistics. This table is generated to give you a more specific idea of what the actual size of the weak instruments test is. For instance, if you are willing to reject weak instruments 10% of the time, then use a critical value of 19.93. The rule-of-thumb value of 10 would lead to actual rejection of weak instruments somewhere between 15% and 20% of the time. Since our $F = 55.4 > 19.93$ we conclude that our test has a size less than 10%. If so, you would expect the resulting TSLS estimator based on these very strong instruments to exhibit relatively small bias.

### 10.3.4 Multiple Endogenous Regressors and the Cragg-Donald F-test

[3]Cragg and Donald (1993) have proposed a test statistic that can be used to test for weak identification (i.e., weak instruments). In order to compute it manually, you have to obtain a set of canonical correlations. These are not computed in **gretl** so we will use another free software, **R**, to do part of the computations. On the other hand, **gretl** prints the value of the Cragg-Donald statistic by default so you won't have to go to all of this trouble. Still, to illustrate a very powerful feature of **gretl** we will use **R** to compute part of this statistic.

One solution to identifying weak instruments in models with more than one endogenous regressor is based on the use of canonical correlations. Canonical correlations are a generalization of the usual concept of a correlation between two variables and attempt to describe the association between two **sets** of variables.

Let $N$ denote the sample size, $B$ the number of righthand side endogenous variables, $G$ the number of exogenous variables included in the equation (including the intercept), $L$ the number of external instruments–i.e., ones not included in the regression. If we have two variables in the first set of variables and two variables in the second set then there are two canonical correlations, $r_1$ and $r_2$.

A test for weak identification–which means that the instruments are correlated with endogenous regressors, but not very highly–is based on the Cragg-Donald $F$-test statistic

$$\text{Cragg-Donald} - F = [(N - G - B)/L] \times [r_B^2/(1 - r_B^2)] \tag{10.6}$$

The Cragg-Donald statistic reduces to the usual weak instruments $F$-test when the number of endogenous variables is $B = 1$. Critical values for this test statistic have been tabulated by Stock and Yogo (2005), so that we can test the null hypothesis that the instruments are weak, against the alternative that they are not, for two particular consequences of weak instruments.

[3]The computations in this section use **R**. You should refer to D for some hints about using **R**.

The problem with weak instruments is summarized by Hill et al. (2011, p. 435):

**Relative Bias:** In the presence of weak instruments the amount of bias in the IV estimator can become large. Stock and Yogo consider the bias when estimating the coefficients of the endogenous variables. They examine the maximum IV estimator bias relative to the bias of the least squares estimator. Stock and Yogo give the illustration of estimating the return to education. If a researcher believes that the least squares estimator suffers a maximum bias of 10%, and if the relative bias is 0.1, then the maximum bias of the IV estimator is 1%.

**Rejection Rate (Test Size):** When estimating a model with endogenous regressors, testing hypotheses about the coefficients of the endogenous variables is frequently of interest. If we choose the $\alpha = 0.05$ level of significance we expect that a true null hypothesis is rejected 5% of the time in repeated samples. If instruments are weak, then the actual rejection rate of the null hypothesis, also known as the test size, may be larger. Stock and Yogo's second criterion is the maximum rejection rate of a true null hypothesis if we choose $\alpha = 0.05$. For example, we may be willing to accept a maximum rejection rate of 10% for a test at the 5% level, but we may not be willing to accept a rejection rate of 20% for a 5% level test.

The script to compute the statistic manually is given below:

```
open "@gretldir\data\poe\mroz.gdt"
smpl wage>0 --restrict
logs wage
square exper
series nwifeinc = (faminc-wage*hours)/1000
list x = mtr educ kidsl6 nwifeinc const
list z = kidsl6 nwifeinc mothereduc fathereduc const
tsls hours x ; z
scalar df = $df
```

This first section loads includes much that we've seen before. The data are loaded, the sample restricted to the wage earners, the log of wage is taken, the square is experience is added to the data. Then a new variable is computed to measure family income from all other members of the household. The next part estimates a model of `hours` as a function of `mtr`, `educ`, `kidsl6`, `nwifeinc`, and a constant. Two of the regressors are endogenous: `mtr` and `educ`. The external instruments are `mothereduc` and `fathereduc`; these join the internal ones (`const, kidsl6, nwifeinc`) in the instrument list. The degrees of freedom from this regression is saved to compute $(N - G - B)/L$.

The next set of lines partial's out the influence of the internal instruments on each of the endogenous regressors and on the external instruments.

```
list w = const kidsl6 nwifeinc
ols mtr w --quiet
```

```
12 series e1 = $uhat
13 ols educ w --quiet
14 series e2 = $uhat
15 ols mothereduc w --quiet
16 series e3 = $uhat
17 ols fathereduc w --quiet
18 series e4 = $uhat
```

Now this is where it gets interesting. From here we are going to call a separate piece of software called **R** to do the computation of the canonical correlations. Lines 19-25 hold what **gretl** refers to as a **foreign block**.

```
19 foreign language=R --send-data --quiet
20     set1 <- gretldata[,29:30]
21     set2 <- gretldata[,31:32]
22     cc1 <- cancor(set1,set2)
23     cc <- as.matrix(cc1$cor)
24     gretl.export(cc)
25 end foreign
26
27 vars = mread("@dotdir/cc.mat")
28 print vars
29 scalar mincc = minc(vars)
30 scalar cd = df*(mincc^2)/(2*(1-mincc^2))
31 printf "\nThe Cragg-Donald Statistic is %10.4f.\n",cd
```

A foreign block takes the form

Foreign Block syntax

```
foreign language=R [--send-data] [--quiet]
... R commands ...
end foreign
```

and achieves the same effect as submitting the enclosed **R** commands via the GUI in the noninteractive mode (see section 30.3 of the **Gretl** Users Guide). In other words, it allows you to use **R** commands from within **gretl** . Of course, you have to have installed **R** separately, but this greatly expands what can be done using **gretl**. The `--send-data` option arranges for auto-loading of the data from the current **gretl** session. The `--quiet` option prevents the output from **R** from being echoed in the **gretl** output. The block is closed with an `end foreign` command.

Inside our foreign block we create two sets of variables. The first set includes the residuals, `e1` and `e2` computed above. There are obtained from a matrix called `gretldata`. This is the name that **gretl** gives to data matrices that are passed into **R**. You have to pull the desired variables out of `gretldata`. In this case I used a rather inartful but effective means of doing so. These two variables are located in the 29th and 30th columns of `gretldata`. These also happen to be their ID numbers in **gretl**. Line 20 puts these two variables into `set1`.

The second set of residuals is put into `set2`. Then, **R**'s `cancor` function is used to find the canonical correlations between the two sets of residuals. The entire set of results is stored in **R** as `cc`. This object contains many results, but we only need the actual canonical correlations between the two sets. The canonical correlations are stored within `cc` as `cor`. They are retrieved as `cc$cor` and put into a matrix with **R**'s as.matrix command. These are exported to **gretl** as `cc.mat`. **R** adds the `.mat` suffix. `cc.mat` is placed in your working directory.

The next step is to read the `cc.mat` into **gretl**. Then in line we take the smallest canonical correlation and use it in line to compute the Cragg-Donald statistic. The result printed to the screen is:

```
? printf "\nThe Cragg-Donald Statistic is %6.4f.\n",cd
The Cragg-Donald Statistic is     0.1006.
```

It matches the automatic one produced by `tsls`, which is shown below, perfectly! It also shows that these instruments are **very** weak.

```
Weak instrument test -
  Cragg-Donald minimum eigenvalue = 0.100568
  Critical values for desired TSLS maximal size, when running
  tests at a nominal 5% significance level:

     size      10%      15%      20%      25%
    value     7.03     4.58     3.95     3.63

  Maximal size may exceed 25%
```

Of course, you can do this exercise without using **R** as well. **Gretl**'s matrix language is very powerful and you can easily get the canonical correlations from two sets of regressors. The following funcrion[4] does just that.

```
function matrix cc(list Y, list X)
   matrix mY = cdemean({Y})
   matrix mX = cdemean({X})

   matrix YX = mY'mX
   matrix XX = mX'mX
   matrix YY = mY'mY

   matrix ret = eigsolve(qform(YX, invpd(XX)), YY)
```

[4]Function supplied by **gretl** guru Riccardo Lucchetti.

```
    return sqrt(ret)
end function
```

The function is called `cc` and takes two arguments, just as the one in **R**. Feed the function two lists, each containing the variable names to be included in each set for which the canonical correlations are needed. Then, the variables in each set are demeaned using the very handy `cdemean` function. This function centers the columns of the matrix argument around the column means. Then the various cross-products are taken (YX, XX, YY) and the eigenvalues for $|Q - \lambda YY| = 0$, where $Q = (YX)(XX)^{-1}(YX)^T$, are returned.

Then, to get the value of the Cragg-Donald $F$, assemble the two sets of residuals and use the `cc` function to get the canonical correlations.

```
list E1 = e1 e2
list E2 = e3 e4

l = cc(E1, E2)
scalar mincc = minc(l)
scalar cd = df*(mincc^2)/(2*(1-mincc^2))
printf "\nThe Cragg-Donald Statistic is %10.4f.\n",cd
```

## 10.4 Simulation

In appendix 10F of *POE4*, the authors conduct a Monte Carlo experiment comparing the performance of OLS and TSLS. The basic simulation is based on the model

$$y = x + e \tag{10.7}$$

$$x = \pi z_1 + \pi z_2 + \pi z_3 + v \tag{10.8}$$

The $z_i$ are exogenous instruments that are each N(0,1). The errors, $e$ and $v$, are

$$\begin{pmatrix} e \\ v \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right] \tag{10.9}$$

The parameter $\pi$ controls the strength of the instruments and is set to either 0.1 or 0.5. The parameter $\rho$ controls the endogeneity of $x$. When $\rho = 0$, $x$ is exogenous. When $\rho = 0.8$ it is seriously endogenous. Sample size is set to 100 and 10,000 simulated samples are drawn.

The **gretl** script to perform the simulation appears below:

```
scalar N = 100
nulldata N
```