Takashi Yamano

Lecture Notes on Advanced Econometrics

Lecture 8: Instrumental Variables Estimation

Endogenous Variables

Consider a population model:

$$y_{i1} = \alpha_1 y_{i2} + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i$$

We call y_2 an **endogenous** variable when y_2 is correlated with u. As we have studied earlier, y_2 would be correlated with u if (*a*) there are omitted variables that are correlated with y_2 and y_1, (*b*) y_2 is measured with errors, and (*c*) y_1 and y_2 are simultaneously determined (we will cover this issue later). All of these problems, we can identify the source of the problems as the correlation between the error term and one or some of the independent variables.

For all of these problems, we can apply **instrumental variables (IV)** estimations because instrumental variables are used to cut correlations between the error term and independent variables. To conduct IV estimations, we need to have instrumental variables (or instruments in short) that are **(R1) uncorrelated with *u*** but **(R2) partially and sufficiently strongly correlated with y_2 once the other independent variables are controlled for**.

It turns out that it is very difficult to find proper instruments!

In practice, we can test the second requirement (*b*), but we can not test the first requirement (*a*) because u is unobservable. To test the second requirement (*b*), we need to express **a reduced form** equation of y_2 with all of **exogenous** variables. Exogenous variables include all of independent variables that are not correlated with the error term and the instrumental variable, z. The reduced form equation for y_2 is

$$y_2 = \delta_Z z + \delta_1 + \delta_2 x_2 + \dots + \delta_{k-1} x_{k-1} + u$$

For the instrumental variable to satisfy the second requirement (R2), the estimated coefficient of z must be significant.

In this case, we have one endogenous variable and one instrumental variable. When we have the same number of endogenous and instrumental variables, we say the endogenous variables are **just identified**. When we have more instrumental variables than endogenous variables, we say the endogenous variables are **over-identified**. In this case, we need to use "two stage least squares" (**2SLS**) estimation. We will come back to 2SLS later.

Define **x** = (y_2, 1, x_2, …, x_{k-1}) as a 1-by-k vector, **z** = (z, 1, x_2, …, x_{k-1}) a 1-by-k vector of all exogenous variables, ***X*** as a n-by-k matrix that includes one endogenous variable and k-1 independent variables, and ***Z*** as a n-by-k matrix that include one instrumental variable and (k-1) independent variables:

$$X = \begin{bmatrix} y_{12} & 1 & x_{12} & x_{13} & & x_{1k-1} \\ y_{22} & 1 & x_{22} & x_{23} & & x_{2k-1} \\ & & & & & \\ y_{n2} & 1 & x_{n2} & x_{n3} & & x_{nk-1} \end{bmatrix}, \quad Z = \begin{bmatrix} z_1 & 1 & x_{12} & x_{13} & & x_{1k-1} \\ z_2 & 1 & x_{22} & x_{23} & & x_{2k-1} \\ & & & & & \\ z_n & 1 & x_{n2} & x_{n3} & & x_{nk-1} \end{bmatrix}.$$

The instrumental variables (IV) estimator is

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'Y$$

Notice that we can take the inverse of $Z'X$ because both Z and X are n-by-k matrices and $Z'X$ is a k-by-k matrix which has full rank, k. This indicates that there is no perfect co linearity in Z. The condition that $Z'X$ has full rank of k is called the **rank condition**.

Problems with Small property.

The consistency of the IV estimators can be shown by using the two requirements for IVs:

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'(X\beta + u)$$

$$= \beta + (Z'X)^{-1} Z'u$$
$$= \beta + (Z'X/n)^{-1} Z'u/n$$

From the first requirement (R1), $p\lim\ Z'u/n \to 0$.
From the second requirement (R2), $p\lim\ Z'X/n \to A, where\ A \equiv E(z'x)$.

Therefore, the IV estimator is consistent when IVs satisfy the two requirements.

A Bivariate IV model

Let's consider a simple bivariate model:

$$y_1 = \beta_0 + \beta_1 y_2 + u$$

We suspect that y_2 is an endogenous variable, cov(y_2, u) $\neq 0$. Now, consider a variable, z, which is correlated y_2 but not correlated with u: cov(z, y_2) $\neq 0$ but cov(z, u) = 0.
Consider cov(z, y_1):

$$\begin{aligned} \mathrm{cov}(z, y_1) &= \mathrm{cov}(z, \beta_0 + \beta_1 y_2 + u) \\ &= \beta_0 \mathrm{cov}(z,1) + \beta_1 \mathrm{cov}(z, y_2) + \mathrm{cov}(z,u) \end{aligned}$$

Because $\mathrm{cov}(z,1) = 0\ and\ \mathrm{cov}(z,u) = 0$, we find that

$$\beta_1 = \frac{\mathrm{cov}(z, y_1)}{\mathrm{cov}(z, y_2)}$$

$$= \frac{\sum_{i=1}^{n} (z_i - \bar{z})(y_{i1} - \bar{y}_1)}{\sum_{i=1}^{n} (z_i - \bar{z})(y_{i2} - \bar{y}_2)}, \text{ which is}$$

$$= (Z'X)^{-1} Z'Y.$$

Thus, we find the same conclusion as using the matrix form.

The problem in practice is the first requirement, cov(z, u) = 0. We can not empirically confirm this requirement because u cannot be observed. Thus, the validity of this assumption is left to economic theory or economists' common sense.

Recent studies show that even the first requirement can be problematic when the correlation between the endogenous and instrumental variables is weak. We will discuss this later.

Example 1: Card (1995), CARD.dta.

A dummy variable grew up near a 4 year collage as an IV on *educ*.

OLS

```
. reg lwage educ

  Source |       SS       df       MS              Number of obs =    3010
---------+------------------------------           F(  1,  3008) =  329.54
   Model |  58.5153536     1  58.5153536           Prob > F      =  0.0000
Residual |  534.126258  3008   .17756857           R-squared     =  0.0987
---------+------------------------------           Adj R-squared =  0.0984
   Total |  592.641611  3009  .196956335           Root MSE      =  .42139
------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .0520942   .0028697     18.153   0.000       .0464674     .057721
   _cons |   5.570883   .0388295    143.470   0.000       5.494748    5.647017
------------------------------------------------------------------------------

Correlation between nearc4 (an IV) and educ
. reg educ nearc4

  Source |       SS       df       MS              Number of obs =    3010
---------+------------------------------           F(  1,  3008) =   63.91
   Model |  448.604204     1  448.604204           Prob > F      =  0.0000
Residual |  21113.4759  3008  7.01910767           R-squared     =  0.0208
---------+------------------------------           Adj R-squared =  0.0205
   Total |  21562.0801  3009  7.16586243           Root MSE      =  2.6494
------------------------------------------------------------------------------
    educ |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  nearc4 |    .829019   .1036988      7.994   0.000       .6256913    1.032347
```

```
   _cons |   12.69801   .0856416    148.269   0.000      12.53009    12.86594
------------------------------------------------------------------------------
```

Thus, nearc4 satisfies the one of the two requirements to be a good candidate as an IV.

IV Estimation:

```
. ivreg lwage (educ=nearc4)

Instrumental variables (2SLS) regression

  Source |       SS       df       MS               Number of obs =    3010
---------+------------------------------            F(  1,  3008) =   51.17
   Model | -340.111443     1 -340.111443            Prob > F      =  0.0000
Residual |  932.753054  3008  .310090776            R-squared     =       .
---------+------------------------------            Adj R-squared =       .
   Total |  592.641611  3009  .196956335            Root MSE      =  .55686

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .1880626   .0262913      7.153   0.000       .1365118    .2396134
   _cons |   3.767472   .3488617     10.799   0.000        3.08344    4.451504
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   nearc4
------------------------------------------------------------------------------
```

Note that you can obtain the same coefficient by estimating OLS on *lwage* with the predicted *educ* (predicted by *nearc4*). However, the standard error would be incorrect. In the above IV Estimation, the standard error is already corrected.

End of Example 1

The Two-Stage Least Squares Estimation

Again, let's consider a population model:

$$y_1 = \alpha_1 y_2 + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u \qquad (1)$$

where y_2 is an endogenous variable. Suppose that there are m instrumental variables. Instruments, $\mathbf{z} = (1, x_1, \ldots, x_k, z_1, \ldots, z_m)$, are correlated with y_2. From the reduced form equation of y_2 with all exogenous variables (exogenous independent variables plus instruments), we have

$$y_2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \delta_{k+1} z_1 + ... + \delta_{k+m} z_m + \varepsilon$$
$$y_2 = \hat{y}_2 + \varepsilon$$

$\hat{y}_2$ is a linear projection of y_2 with all exogenous variables. Because $\hat{y}_2$ is projected with all exogenous variables that are not correlated with the error term, u, in (1), $\hat{y}_2$ <u>is not correlated with u</u>, while ε <u>is correlated with u</u>. Thus, we can say that by estimating y_2 with all exogenous variables, we have divided into two parts: one is correlated with u and the other is not.

The projection of y_2 with Z can be written as

$$\hat{y}_2 = Z\hat{\delta} = Z(Z'Z)^{-1} Z'y_2$$

When we use the two-step procedure (as we discuss later), we use this $\hat{y}_2$ in the place of y_2. But now, we treat y_2 as a variable in X and project X itself with Z:

$$\hat{X} = Z\hat{\Pi} = Z(Z'Z)^{-1} Z'X = P_Z X$$

$\hat{\Pi}$ is a (k+m-1)-by-k matrix with coefficients, which should look like:

$$\hat{\Pi} = \begin{bmatrix} \delta_1 & 1\,0 & 0 \\ \delta_2 & 0\,1 & 0 \\ & & \\ \delta_{k+m-1} & 0\,0 & 1 \end{bmatrix}.$$

Thus, y_2 in X should be expressed as a linear projection, and other independent variables in X should be expressed by itself. $P_Z = Z(Z'Z)^{-1}Z'$ is a n-by-n symmetric matrix and idempotent (i.e., $P_Z' P_Z = P_Z$). We use $\hat{X}$ as instruments for X and apply the IV estimation as in

$$\begin{aligned} \hat{\beta}_{2SLS} &= (\hat{X}'X)^{-1}\hat{X}'Y \\ &= (X'P_Z X)^{-1} X'P_Z Y \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'Y \qquad (2) \end{aligned}$$

This can be also written as

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y$$

This is the 2SLS estimator. It is called as two-stage because it looks like we take two steps by creating projected X to estimate the 2SLS estimators. We do not need to take two steps as we show in (2). We can just estimate 2SLS estimators in one step by using X and Z. (This is what econometrics packages do.)

The Two-Step procedure

It is still a good idea to know how to estimate the 2SLS estimators by a two-step procedure:

Step 1: Obtain $\hat{y}_2$ by estimating an OLS against all of exogenous variables, including all of instruments (**the first-stage regression**)

Step 2: Use $\hat{y}_2$ in the place of y_2 to estimate y_1 against $\hat{y}_2$ and all of exogenous independent variables, not instruments (**the second stage regression**)

The estimated coefficients from the two-step procedure should exactly the same as 2SLS. **However, you must be aware that the standard errors from the two-step procedure are incorrect, usually smaller than the correct ones.** Thus, in practice,

<u>avoid using predicted variables as much as you can</u>!

Econometric packages will provide you 2SLS results based on (2). So you do not need to use the two-step procedure.

We use the first step procedure to test the second requirement for IVs. In the first stage regression, we should conduct a F-test on all instruments to see if instruments are jointly significant in the endogenous variable, y_2. As we discuss later, instruments should be strongly correlated with y_2 to have reliable 2SLS estimators.

Consistency of 2SLS

Assumption 2SLS.1: For vector **z**, E($\mathbf{z}'u$)=0,

where $\mathbf{z} = (1, x_1, \ldots, x_{k,} z_1, \ldots, z_m)$.

Assumption 2SLS.2: (a) rank E($\mathbf{z}'\mathbf{z}$) = k+m+1; (b) rank E($\mathbf{z}'\mathbf{x}$) = k.

(b) is **the rank condition** for identification that z is sufficiently linearly related to x so that rank E($\mathbf{z}'\mathbf{x}$) has ful column rank.

The order condition is k-1+m≥ k-1+h, where h is the number of endogenous variable. Thus, the order condition indicates that we must have at least as many instruments as endogenous variables.

Under assumption 2SLS1 and 2SLS2, the 2SLS estimators in (10-5) are consistent.

Under homoskedasticity,

$$\hat{\sigma}^2(\hat{X}'\hat{X})^{-1} = (n-k)^{-1}\sum_{i=1}^{n}\hat{u}_i(\hat{X}'\hat{X})^{-1}$$

is a valid estimator of the asymptotic variance of $\hat{\beta}_{2SLS}$.

Under heteroskedasticity, the heteroskedasticity-robust (White) standard errors is

$$(\hat{X}'\hat{X})^{-1}\hat{X}\hat{\Sigma}\hat{X}(\hat{X}'\hat{X})^{-1}$$

Here are some tests for IV estimation. See Wooldridge (Introductory Econometrics) for details.

Testing for Endogeneity

(i) Estimate the reduced form model using the endogenous variable as the dependent variable:
$y_2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \delta_{k+1} z_1 + ... + \delta_{k+m} z_m + \varepsilon$

(ii) Obtain the residual, $\hat{\varepsilon}$.

(iii) Estimate
$y_1 = \alpha_1 y_2 + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \delta\hat{\varepsilon} + u$

(iv) If $\hat{\delta}$ is significant, then y_2 is endogenous.

Testing the Over-Identification

(i) Estimate $\hat{\beta}_{2SLS}$ and obtain $\hat{u}$.

(ii) Regress $\hat{u}$ on *z* and *x*.

(iii) Get R^2 and get nR^2, which is chi-squared. If this is significantly different from zero, then at least some of the IVs are not exogenous.

Example 2: Card (1995), card.dta again.

OLS

```
. reg lwage educ  exper expersq black smsa south
```

```
  Source |       SS       df       MS              Number of obs =    3010
---------+------------------------------           F(  6,  3003) =  204.93
   Model |  172.165615     6  28.6942691           Prob > F      =  0.0000
Residual |  420.475997  3003  .140018647           R-squared     =  0.2905
---------+------------------------------           Adj R-squared =  0.2891
   Total |  592.641611  3009  .196956335           Root MSE      =  .37419

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |    .074009   .0035054     21.113   0.000       .0671357    .0808823
   exper |   .0835958   .0066478     12.575   0.000       .0705612    .0966305
 expersq |  -.0022409   .0003178     -7.050   0.000      -.0028641   -.0016177
   black |  -.1896316   .0176266    -10.758   0.000      -.2241929   -.1550702
    smsa |    .161423   .0155733     10.365   0.000       .1308876    .1919583
   south |  -.1248615   .0151182     -8.259   0.000      -.1545046   -.0952184
   _cons |   4.733664   .0676026     70.022   0.000       4.601112    4.866217
------------------------------------------------------------------------------
```

2SLS Estimation: nearc2 nearc4 as IVs

```
. ivreg lwage (educ= nearc2 nearc4)  exper expersq black smsa south

Instrumental variables (2SLS) regression

  Source |       SS       df       MS              Number of obs =    3010
---------+------------------------------           F(  6,  3003) =  110.30
   Model |  86.2368644     6  14.3728107           Prob > F      =  0.0000
Residual |  506.404747  3003  .168632949           R-squared     =  0.1455
---------+------------------------------           Adj R-squared =  0.1438
   Total |  592.641611  3009  .196956335           Root MSE      =  .41065

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .1608487   .0486291      3.308   0.001        .065499    .2561983
```

```
   exper |   .1192111   .0211779      5.629   0.000       .0776865    .1607358
 expersq |  -.0023052   .0003507     -6.574   0.000      -.0029928   -.0016177
   black |  -.1019727   .0526187     -1.938   0.053       -.205145    .0011996
    smsa |   .1165736   .0303135      3.846   0.000       .0571363    .1760109
   south |  -.0951187   .0234721     -4.052   0.000      -.1411418   -.0490956
   _cons |   3.272103   .8192562      3.994   0.000       1.665743    4.878463
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   nearc2 nearc4 + exper expersq ... south
------------------------------------------------------------------------------

. ivreg lwage (educ= nearc2 nearc4  fatheduc motheduc)  exper expersq black sms
> a south

Instrumental variables (2SLS) regression

  Source |       SS       df       MS              Number of obs =    2220
---------+------------------------------           F(  6,  2213) =   83.66
   Model |  108.419483     6  18.0699139           Prob > F      =  0.0000
Residual |  320.580001  2213  .144862178           R-squared     =  0.2527
---------+------------------------------           Adj R-squared =  0.2507
   Total |  428.999484  2219  .193330097           Root MSE      =  .38061

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .1000713     .01263      7.923   0.000       .0753034    .1248392
   exper |   .0989441    .009482     10.435   0.000       .0803496    .1175385
 expersq |   -.002449   .0004013     -6.103   0.000      -.0032359   -.0016621
   black |  -.1504635   .0259113     -5.807   0.000      -.2012765   -.0996505
    smsa |    .150854   .0195975      7.698   0.000       .1124226    .1892854
   south |  -.1072406   .0180661     -5.936   0.000      -.1426688   -.0718123
   _cons |    4.26178    .216812     19.657   0.000       3.836604    4.686956
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   nearc2 nearc4 fatheduc motheduc + exper expersq ... south
```

2SLS Estimation: fatheduc motheduc as IVs

```
. ivreg lwage (educ= fatheduc motheduc)  exper expersq black smsa south

Instrumental variables (2SLS) regression

  Source |       SS       df       MS              Number of obs =    2220
---------+------------------------------           F(  6,  2213) =   83.44
   Model |  108.477154     6  18.0795257           Prob > F      =  0.0000
Residual |   320.52233  2213  .144836118           R-squared     =  0.2529
---------+------------------------------           Adj R-squared =  0.2508
   Total |  428.999484  2219  .193330097           Root MSE      =  .38057

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |    .099931    .012756      7.834   0.000       .0749161     .124946
   exper |    .098884   .0095123     10.395   0.000       .0802299     .117538
 expersq |  -.0024487   .0004012     -6.103   0.000      -.0032356   -.0016619
   black |  -.1505902   .0259598     -5.801   0.000      -.2014983   -.0996822
    smsa |   .1509271   .0196181      7.693   0.000       .1124553    .1893988
   south |  -.1072797   .0180714     -5.936   0.000      -.1427183    -.071841
   _cons |    4.26415   .2189075     19.479   0.000       3.834865    4.693436
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   fatheduc motheduc + exper expersq ... south
------------------------------------------------------------------------------
```

Which ones are better?

End of Example 2

Lecture 8A: Weak Instruments

Remember the two requirements for instrumental variables to be valid: **(R1) uncorrelated with *u*** but **(R2) partially and sufficiently strongly correlated with y_2 once the other independent variables are controlled for**.

The first requirement is difficult to be confirmed because we cannot observe u. So, we rely on economic theory or natural experiments to find instrumental variables that satisfy the first requirement. The second requirement can be checked by conducting some analyses.

One way of checking the second requirement is to estimate a regression model of endogenous variables on exogenous variables which include instrumental variables and other exogenous variables. Suppose that we have one endogenous variable, y_2, and m instrumental variables, z_1, …, z_m. Then estimate the following model:

$$y_{i2} = \beta_0 + \delta_1 z_{i1} + ... + \delta_m z_{im} + \beta_1 x_{i1} + ... + \beta\delta_k x_{ik} + u_i . \qquad (1)$$

Then obtain a F-statistics on the estimators of the instrumental variables: $H_0 : \delta_1 = ... = \delta_m = 0$. If the F-statistics is small, then we call the instrumental variables *weak*. When the instrumental variables are weak, the IV or 2SLS estimators could be inconsistent or have large standard errors.

Inconsistency

To examine how weak instruments can make IV and 2SLS estimators inconsistent, let us consider a simple bivariate model:

$$y_1 = \beta_0 + \beta_1 y_2 + u$$

In a bivariate model, we write

$$\text{plim } \hat{\beta}_{IV} = \beta + \frac{\text{cov}(z,u)}{\text{cov}(z, y_2)}$$

because Corr(z,u)=cov(z,u)/[sd(z)sd(u)] (see Wooldridge pp714)

$$\text{plim } \hat{\beta}_{IV} = \beta + \frac{corr(z,u)(sd(z)sd(u))}{corr(z,y_2)(sd(z)sd(y_2))}$$

$$\text{plim } \hat{\beta}_{IV} = \beta + \frac{corr(z,u)}{corr(z,y_2)}\frac{sd(u)}{sd(y_2)}$$

Thus if z is only weakly correlated with the endogenous variable, y_2, i.e., $corr(z, y_2)$ is very small, the IV estimator could be severely biased even when the correlation between z and u is very small.

Consider now the OLS estimator of the bivariate model ignoring the endogeneity of y_2:

$$\begin{aligned}\text{plim } \hat{\beta}_{OLS} &= \beta + \frac{\text{cov}(y_2,u)}{\text{var}(y_2)} \\ &= \beta + \frac{corr(y_2,u)\,sd(y_2)\,sd(u)}{sd(y_2)^2} \\ &= \beta + \frac{corr(y_2,u)\,sd(u)}{sd(y_2)}\end{aligned}$$

Here we have the endogenous bias. But the size of the endogenous bias could be smaller than the bias created by the weak instrument. To compare these two, let us take the ratio of these two:

$$\frac{p\lim\hat{\beta}_{IV} - \beta}{p\lim\hat{\beta}_{OLS} - \beta} = \frac{corr(z,u)}{corr(z,y_2)}\frac{1}{corr(y_2,u)}.$$

When this ratio is larger than 1, the bias in the IV estimator is larger than the OLS estimator. When the instrumental variable is completely uncorrelated with u, the bias in the IV estimator is zero. When it has even a very small correlation with u, the size of the ratio depends on the size of the denominator, which could be very small when the correlation between the instrumental variable and the endogenous variable, $corr(z, y_2)$, is small. The implication from this simple model could be also applied on a more complicated model where there are more than one endogenous variable and one instrumental variable.

Low Precision

Weak instrumental variables can lead to large standard errors of the IV/2SLS estimators. The variance of the IV estimator is

$$V(\hat{\beta}_{IV}) = \sigma^2 (Z'X)^{-1} Z'Z(Z'X)^{-1}.$$

This could be rearranged as:

$$V(\hat{\beta}_{IV}) = \sigma^2 (X'X)^{-1} X'X\,(Z'X)^{-1} Z'Z(Z'X)^{-1}$$

$$= V(\hat{\beta}_{OLS}) X'X\,(Z'X)^{-1} Z'Z(Z'X)^{-1}$$

$$= V(\hat{\beta}_{OLS})[(X'X)^{-1} Z'X]^{-1}\ [(Z'Z)^{-1} Z'X]^{-1}$$

$$= V(\hat{\beta}_{OLS})[\hat{\Pi}_{Z,X}]^{-1}\ [\hat{\Pi}_{X,Z}]^{-1}$$

where $\hat{\Pi}_{Z,X}$ and $\hat{\Pi}_{X,Z}$ are projections of *Z* on *X* and *X* on *Z*, respectively. If the correlations between *Z* and *X* are low, then $\hat{\Pi}_{Z,X}$ and $\hat{\Pi}_{X,Z}$ have low values, which would make the variance of the IV estimators large. This could be applied on the 2SLS estimators.

In empirical estimation, we often find large standard errors in the IV/2SLS estimators. This could be caused by weak instruments.

Thus, weak instrumental variables can cause inconsistency and imprecision in the IV/2SLS estimators. But how weak is weak?

A Rule of Thumb to find Weak Instruments

Staiger and Stock (1997) suggest that the F-statistics of instrumental variables in (1), of this lecture note, should be larger than 10 to ensure that the maximum bias in IV estimators to be less than 10 %. If you are willing to accept the maximum bias in IV estimators to be less than 20 %, the threshold is F-stat being larger than 5. If the number of instrumental variables is one, the F-statistics should be replaced by the t-statistics.

In practice, it is quite difficult to find valid instrumental variables that are not weak. I personally find searching for instrumental variables time-consuming and not so rewarding. One can never be sure about validness of IVs. You should look for natural experiments or randomaized experiments that could be used as instrumental variables.

Lecture 9: Heteroskedasticity and Robust Estimators

In this lecture, we study heteroskedasticity and how to deal with it. Remember that we did not need the assumption of Homoskedasticity to show that OLS estimators are unbiased under the finite sample properties and consistency under the asymptotic properties. What matters is how to correct OLS standard errors.

Heteroskedasticity

In this section, we consider heteroskedasticity, while maintaining the assumption of no-autocorrelation. The variance of disturbance *i*, u_i, is not constant across observations but not correlated with u_j:

$$E(uu') = \begin{bmatrix} E(u_1u_1) & E(u_1u_2) & E(u_1u_n) \\ E(u_2u_1) & E(u_2u_2) & E(u_2u_n) \\ \\ E(u_nu_1) & E(u_nu_2) & E(u_nu_n) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ \\ 0 & 0 & \sigma_n^2 \end{bmatrix} = \Sigma$$

or

$$E(uu') = \sigma^2 \begin{bmatrix} \sigma_1^2/\sigma^2 & 0 & 0 \\ 0 & \sigma_2^2/\sigma^2 & 0 \\ \\ 0 & 0 & \sigma_n^2/\sigma^2 \end{bmatrix} = \sigma^2 \Omega$$

Notice that under homoskedasticity, $\Omega = I$.

Under heteroskedasticity, the sample variance of OLS estimator (under finite sample properties) is

$$Var(\hat{\beta}) = Var[\beta + (X'X)^{-1}X'u]$$

$$= E[(X'X)^{-1}X'uu'X(X'X)^{-1}]$$

$$= (X'X)^{-1} X'E(uu')X(X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1} X'\Omega X(X'X)^{-1} \qquad (1)$$

(See Theorem 10.1 in Greene (2003))

Unless you specify, however, econometric packages automatically assume homoskedasticity and will calculate the sample variance of OLS estimator based on the homoskedasticity assumption:

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

Thus, in the presence of heteroskedasticity, the statistical inference based on $\sigma^2 (X'X)^{-1}$ would be biased, and *t*-statistics and *F*-statistics are inappropriate. Instead, we should use (1) to calculate standard errors and other statistics.

Finite Sample Properties of OLS Estimators

The OLS estimators are unbiased and have the sampling variance specified in (6-1). If u is normally distributed, then the OLS estimators are also normally distributed:

$$\hat{\mathrm{B}} \mid X \sim N[B, \sigma^2 (X'X)^{-1} (X'\Omega X)(X'X)^{-1}]$$

Asymptotic Properties of OLS Estimators

If $p\lim(X'X/n) = Q$ and $p\lim(X'\Omega X/n)$ are both finite positive definite matrices, then $Var(\hat{\beta})$ is consistent for $Var(\beta)$.

Robust Standard Errors

If Σ is known, we can obtain efficient least square estimators and appropriate statistics by using formulas identified above. However, as in many other problems, Σ is unknown. One common way to solve this problem is to estimate Σ empirically: First, estimate an OLS model, second, obtain residuals, and third, estimate Σ:

$$\hat{\Sigma} = \begin{bmatrix} \hat{u}_1^2 & 0 & 0 \\ 0 & \hat{u}_2^2 & 0 \\ & & \\ 0 & 0 & \hat{u}_n^2 \end{bmatrix}$$

(We may multiply this by (n/(n-k-1)) as a degree-of-freedom correction. But when the number of observations, n, is large, this adjustment does not make any difference.)

Thus by using the estimated Σ, we have

$$X'\hat{\Sigma}X = X' \begin{bmatrix} \hat{u}_1^2 & 0 & 0 \\ 0 & \hat{u}_2^2 & 0 \\ & & \\ 0 & 0 & \hat{u}_n^2 \end{bmatrix} X .$$

Therefore, we can estimate the variances of OLS estimators (and standard errors) by using $\hat{\Sigma}$:

$$Var(\hat{\beta}) = (X'X)^{-1} X'\Sigma X (X'X)^{-1}$$

Standard errors based on this procedure are called (**heteroskedasticity) robust standard errors** or **White-Huber standard errors**. Or it is also known as the **sandwich estimator of variance** (because of how the calculation formula looks like). This procedure is reliable but entirely empirical. We do not impose any assumptions on the structure of heteroskedasticity.

Sometimes, we may impose assumptions on the structure of the heteroskedasticity. For instance, if we suspect that the variance is homoskedastic within a group but not across groups, then we obtain residuals for all observations and calculate average residuals for each group. Then, we have $\hat{\Sigma}$ which has a constant $\hat{u}_j^2$ for group j. (In STATA, you can specify groups by using ***cluster***.)

In practice, we usually do not know the structure of heteroskedasticity. Thus, it is safe to use the robust standard errors (especially when you have a large sample size.) Even if there is no heteroskedasticity, the robust standard errors will become just conventional OLS standard errors. Thus, the robust standard errors are appropriate even under homoskedasticity.

A **heteroskedasticity-robust t statistic** can be obtained by dividing an OSL estimator by its robust standard error (for zero null hypotheses). The usual F-statistic, however, is invalid. Instead, we need to use the heteroskedasticity-robust Wald statistic. Suppose the hypotheses can be written as

$$H_0 : R\beta = r$$

Where R is a *q* x *(k+1)* matrix (q < (k+1)) and ***r*** is a *q* x 1 vector with zeros for this case. Thus,

$$R = \begin{bmatrix} 1\;0 & 0\;0\;0 \\ 0\;1 & 0\;0\;0 \\ & \\ 0 & 1\;0\;0 \end{bmatrix} = \left[I_q : 0_{q\times(k+1-q)} \right],\; r = \begin{bmatrix} 0 \\ 0 \\ \\ 0 \end{bmatrix}.$$

The **heteroskedasticity-robust Wald statistics** for testing the null hypothesis is

$$W = (R\hat{\beta} - r)'(R\hat{V}R')^{-1}(R\hat{\beta} - r)$$

where $\hat{V}$ is given in (7-2). The heteroskedasticity-robust Wald statistics is asymptotically distributed chi-squared with *q* degree of freedom. The Wald statistics can be turned into an appropriate F-statistics (*q*, *q-k-1*) by dividing it by *q*.

Tests for Heteroskedasticity

When should we use robust standard errors? My personal answer to this question is "almost always." As you will see in Example 7-1, it is very easy to estimate robust standard errors with STATA or other packages. Thus, at least I suggest that you estimate robust standard errors and see if there are any significant differences between conventional standard errors and robust standard errors. If results are robust, i.e., when you do not find any significant differences between two sets of standard errors, then you could be confident in your results based on homoskedasticity.

Statistically, you can use following two heteroskedasticity tests to decide if you have to use robust standard errors or not.

The Breusch-Pagan Test for Heteroskedasticity

If the homoskedasticity assumption is true, then the variance of error terms should be constant. We can make this assumption as a null hypothesis:

$$H_0\text{:}\ \ E(u|\ X) = \Phi^2$$

To test this null hypothesis, we estimate

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + e$$

Under the null hypothesis, independent variables should not be jointly significant. The F-statistics that test a joint significance of all independent variables is

$$F_{k,n-k-1} = \frac{R^2 / k}{(1 - R^2)/(n - k - 1)}$$

The LM test statistics is

$$LM = n\,R^2 \sim \Pi^2_k$$

*The **White** Test for Heteroskedasticity*

White proposed to add the squares and cross products of all independent variables:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \lambda_1 x_1^2 + \lambda_2 x_2^2 + ... + \lambda_k x_k^2 + \phi_1 x_1 x_2 + \phi_2 x_1 x_3 + ... + \phi \ x_{k-1} x_k + v$$

Because $\hat{y}$ includes all independent variables, this test is equivalent of conducting the following test:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + v$$

We can use F-test or LM-test on H: $\delta_1 = 0 \, and \ \delta_2 = 0$.

Example1: Step-by-Step Estimation for Robust Standard Errors

In the following do-file, I first estimate a wage model:

$$\log Wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 \exp er + \beta_4 \exp ersq + u$$

by using WAGE1.dta. Then, by using residuals from this conventional OLS, I estimate $\hat{\Sigma}$ and obtain robust standard errors by step-by-step with matrix. Finally, I verify what I get with robust standard errors provided by STATA. Of course, you do not need to use matrix to obtain robust standard errors. You just need to use STATA command, "robust," to get robust standard errors (e.g., **`reg y x1 x2 x3 x4, robust`**). But at least you know how robust standard errors are calculated by STATA.

```
. *** on WAGE1.dta
. *** This do-file estimates White-Huber robust standard errors

. set matsize 800
. clear
. use c:\docs\fasid\econometrics\homework\wage1.dta
.
. * Variable construction
```

```
. gen logwage=ln(wage)

. gen expsq=exper*exper

. gen x0=1

.

. * Obtain conventional OLS residuals
```

```
. reg logwage female educ exper expsq

  Source |       SS       df       MS                  Number of obs =     526
---------+------------------------------               F(  4,   521) =   86.69
   Model |  59.2711314     4  14.8177829               Prob > F      =  0.0000
Residual |    89.05862   521   .17093785               R-squared     =  0.3996
---------+------------------------------               Adj R-squared =  0.3950
   Total |  148.329751   525   .28253286               Root MSE      =  .41345
------------------------------------------------------------------------------
 logwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  female |  -.3371868   .0363214     -9.283   0.000      -.4085411   -.2658324
    educ |   .0841361   .0069568     12.094   0.000       .0704692    .0978029
   exper |     .03891   .0048235      8.067   0.000        .029434    .0483859
   expsq |   -.000686   .0001074     -6.389   0.000       -.000897   -.0004751
   _cons |    .390483   .1022096      3.820   0.000       .1896894    .5912767
------------------------------------------------------------------------------
. predict e, residual

. gen esq=e*e

.

. * Create varaibles with squared residuals

. gen efemale=esq*female

. gen eeduc=esq*educ

. gen eexper=esq*exper

. gen eexpsq=esq*expsq

. gen ex0=esq*x0

.

. * Matrix construction

. mkmat logwage, matrix(y)

. mkmat efemale eeduc eexper eexpsq ex0, matrix(ex)

.
```

```
. * White-Huber robust standard errors
. matrix ixx=syminv(x'*x)
. matrix sigma=ex'*x
. matrix b_robust=(526/(526-5))*ixx*sigma*ixx

. * Here is the White-Huber robust var(b^)

. mat list b_robust
b_robust[5,5]
            female        educ       exper       expsq          x0
female   .00130927   .00003556  -1.277e-07   1.647e-09  -.00109773
  educ   .00003556   .00005914  -3.028e-06   1.570e-07  -.00077218
 exper  -1.277e-07  -3.028e-06   .00002186  -4.500e-07  -.00010401
 expsq   1.647e-09   1.570e-07  -4.500e-07   1.009e-08   5.805e-07
    x0  -.00109773  -.00077218  -.00010401   5.805e-07   .01179363

. * Take square root of the diagonal elements>> sd.error
. * Thus sd.er. for female is (0.0013)^.5=0.036, educ is
(0.000059)^.5=0.00768, . * and so on.
```

But, you do not need to go through this calculation yourself. STATA has a command called "robust."

```
. * Verify with STATA version of robust standard errors

. reg logwage female educ exper expsq, robust
Regression with robust standard errors                 Number of obs =     526
                                                       F(  4,   521) =   81.97
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3996
                                                       Root MSE      =  .41345
------------------------------------------------------------------------------
         |               Robust
 logwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  female |  -.3371868   .0361838     -9.319   0.000      -.4082709   -.2661026
```

```
    educ |   .0841361      .00769     10.941   0.000        .069029    .0992432
   exper |     .03891    .0046752      8.323   0.000       .0297253    .0480946
   expsq |   -.000686    .0001005     -6.829   0.000      -.0008834   -.0004887
   _cons |    .390483    .1085985      3.596   0.000       .1771383    .6038278
------------------------------------------------------------------------------
end of do-file
```

End of Example 1

Example 2: the Breusch-Pagan *test*

```
. *** on WAGE1.dta
. *** This do-file conducts the Breusch-Pagan heteroskedasticity test
. set matsize 800
. clear
. use c:\docs\fasid\econometrics\homework\wage1.dta
. * Variable construction
. gen logwage=ln(wage)
. gen expsq=exper*exper

. * Obtain residuals from the level model
. reg wage female educ exper expsq
(Output is omitted)

. predict u, residual
. gen uu=u*u

. * the BP test
. reg uu female educ exper expsq
  Source |       SS       df       MS                  Number of obs =     526
---------+------------------------------               F(  4,   521) =   12.79
   Model |  24632.1712     4   6158.0428               Prob > F      =  0.0000
Residual |  250826.691   521  481.433189               R-squared     =  0.0894
---------+------------------------------               Adj R-squared =  0.0824
   Total |  275458.863   525  524.683548               Root MSE      =  21.942
------------------------------------------------------------------------------
      uu |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
```

```
---------+--------------------------------------------------------------------
  female |  -5.159055   1.927575    -2.676   0.008     -8.945829   -1.372281
    educ |   1.615551   .3691974     4.376   0.000      .8902524    2.340849
   exper |   1.067343   .2559852     4.170   0.000      .5644535    1.570233
   expsq |  -.0189783   .0056986    -3.330   0.001     -.0301733   -.0077833
   _cons |  -18.15532   5.424263    -3.347   0.001     -28.81144   -7.499208
------------------------------------------------------------------------------
. test female educ exper expsq
 ( 1)  female = 0.0
 ( 2)  educ = 0.0
 ( 3)  exper = 0.0
 ( 4)  expsq = 0.0
       F(  4,   521) =   12.79
```

Prob > F = 0.0000

The LM test statistics is 526 x 0.0894 = 47.02. This is significant at 1 percent level because the critical level is 13.28 for a chi-square distribution of four degree of freedom.

End of Example 2

Example 7-3: the White test

```
. * Obtain residuals from the log model
. reg logwage female educ exper expsq
(Output omitted)

. predict yhat
(option xb assumed; fitted values)
. predict v, residual
. gen yhatsq=yhat*yhat
. gen vsq=v*v

. * the White test
. reg vsq yhat yhatsq

  Source |       SS       df       MS                  Number of obs =     526
```

```
---------+------------------------------           F(  2,   523) =    3.96
   Model |  .605241058     2  .302620529           Prob > F      =  0.0197
Residual |   40.003265   523  .076488078           R-squared     =  0.0149
---------+------------------------------           Adj R-squared =  0.0111
   Total |   40.608506   525  .077349535           Root MSE      =  .27656
------------------------------------------------------------------------------
     vsq |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    yhat |   -.187119   .2637874     -0.709   0.478      -.7053321     .331094
  yhatsq |   .0871914   .0812258      1.073   0.284      -.0723774    .2467603
   _cons |   .2334829   .2096599      1.114   0.266      -.1783961     .645362
------------------------------------------------------------------------------
. test yhat yhatsq

 ( 1)  yhat = 0.0
 ( 2)  yhatsq = 0.0

       F(  2,   523) =    3.96
            Prob > F =    0.0197
```

LM stat is 526 x 0.0149 = 7.84, which is significant at 5 percent level but not at 1 percent level.

End of Example 3

Lecture 10: GLS, WLS, and FGLS

Generalized Least Square (GLS)

So far, we have been dealing with heteroskedasticity under OLS framework. But if we knew the variance-covariance matrix of the error term, then we can make a heteroskedastic model into a homoskedastic model.

As we defined before

$$E(uu') = \sigma^2 \Omega = \Sigma .$$

Define further that

$$\Omega^{-1} = P'P$$

P is a "n x n" matrix

Pre-multiply P on a regression model

$$Py = PX\beta + Pu$$

or

$$\tilde{y} = \tilde{X}\beta + \tilde{u}$$

In this model, the variance of $\tilde{u}$ is

$$E(\tilde{u}\tilde{u}') = E(Puu'P') = PE(uu')P' = P\sigma^2\Omega P' = \sigma^2 P\Omega P' = \sigma^2 I$$

Note that $P\Omega P' = I$, because define $P\Omega P' = A$, then $P'P\Omega P' = P'A$. By the definition of P, $\Omega^{-1}\Omega P' = P'A$**,** thus $P' = P'A$. Therefore, $\boldsymbol{A}$ must be $\boldsymbol{I}$.

Because $E(\tilde{u}\tilde{u}') = \sigma^2 I$, the model satisfies the assumption of homoskedasticity. Thus, we can estimate the model by the conventional OLS estimation.

Hence,

$$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$

$$= (X'P'PX)^{-1} X'P'Py$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

is the efficient estimator of β. This is called the **Generalized Least Square (GLS)** estimator. Note that the GLS estimators are unbiased when $E(\widetilde{u} \mid \widetilde{X}) = 0$. The variance of GLS estimator is

$$\text{var}(\hat{\text{B}}) = \sigma^2(\widetilde{X}'\widetilde{X})^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}.$$

Note that, under homoskedasticity, i.e., Ω^{-1}=***I***, GLS becomes OLS.

The problem is, as usual, that we don't know $\sigma^2\,\Omega$ *or* Σ. Thus we have to either assume Σ or estimate Σ empirically. An example of the former is Weighted Least Squares Estimation and an example of the later is Feasible GLS (FGLS).

Weighted Least Squares Estimation (WLS)

Consider a general case of heteroskedasticity.

$$\text{Var}(u_i) = \sigma_i^2 = \sigma^2\omega_i.$$

Then,

$$E(uu') = \sigma^2\begin{bmatrix}\omega_1 & 0 & 0\\ 0 & \omega_2 & 0\\ & & \\ 0 & 0 & \omega_n\end{bmatrix} = \sigma^2\,\Omega,\ \ thus\ \Omega^{-1} = \begin{bmatrix}\omega_1^{-1} & 0 & 0\\ 0 & \omega_2^{-1} & 0\\ & & \\ 0 & 0 & \omega_n^{-1}\end{bmatrix}.$$

Because of $\Omega^{-1} = P'P$, ***P*** is a n x n matrix whose *i*-th diagonal element is $1/\sqrt{\omega_i}$. By pre-multiplying ***P*** on ***y*** and ***X***, we get

$$y_* = Py = \begin{bmatrix} y_1 / \sqrt{\omega_1} \\ y_2 / \sqrt{\omega_2} \\ \\ y_n / \sqrt{\omega_n} \end{bmatrix} \qquad and \qquad X_* = PX = \begin{bmatrix} 1/\sqrt{\omega_1} & x_{11}/\sqrt{\omega_1} \; ... \; x_{1k}/\sqrt{\omega_1} \\ 1/\sqrt{\omega_2} & x_{21}/\sqrt{\omega_2} \; ... \; x_{2k}/\sqrt{\omega_2} \\ & \\ 1/\sqrt{\omega_n} & x_{n1}/\sqrt{\omega_n} \; ... \; x_{nk}/\sqrt{\omega_n} \end{bmatrix}.$$

The OLS on y_* *and* X_* is called the Weighted Least Squares (WLS) because each variable is weighted by $\sqrt{\omega_i}$. The question is: where can we find ω_i?

Feasible GLS (FGLS)

Instead of assuming the structure of heteroskedasticity, we may estimate the structure of heteroskedasticity from OLS. This method is called Feasible GLS (FGLS). First, we estimate $\hat{\Omega}$ from OLS, and, second, we use $\hat{\Omega}$ *instead of* Ω.

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$$

There are many ways to estimate FGLS. But one flexible approach (discussed in Wooldridge page 277) is to assume that

$$\text{var}(u \mid X) = u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k)$$

By taking log of the both sides and using $\hat{u}^2$ instead of u^2, we can estimate

$$\log(\hat{u}^2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + e.$$

The predicted value from this model is $\hat{g}_i = \widehat{\log(\hat{u}^2)}$. We then convert it by taking the exponential into $\hat{\omega}_i = \exp(\hat{g}_i) = \exp(\widehat{\log(\hat{u}^2)}) = \hat{\hat{u}}^2$. We now use WLS with weights $1/\hat{\omega}_i$ *or* $1/\hat{\hat{u}}^2$.

Example 1

```
. * Estimate the log-wage model by using WAGE1.dta with WLS
. * Weight is educ

. * Generate weighted varaibles
. gen w=1/(educ)^0.5
```

```
. gen wlogwage=logwage*w
. gen wfemale=female*w
. gen weduc=educ*w
. gen wexper=exper*w
. gen wexpsq=expsq*w

. * Estimate weighted least squares (WLS) model

. reg wlogwage weduc wfemale wexper wexpsq w, noc

  Source |       SS       df       MS              Number of obs =     524
---------+------------------------------           F(  5,   519) = 1660.16
   Model |  113.916451     5  22.7832901           Prob > F      =  0.0000
Residual |  7.12253755   519  .013723579           R-squared     =  0.9412
---------+------------------------------           Adj R-squared =  0.9406
   Total |  121.038988   524  .230990435           Root MSE      =  .11715

------------------------------------------------------------------------------
wlogwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   weduc |    .080147    .006435     12.455   0.000        .0675051    .0927889
 wfemale |  -.3503307   .0354369     -9.886   0.000       -.4199482   -.2807133
  wexper |   .0367367   .0045745      8.031   0.000        .0277498    .0457236
  wexpsq |  -.0006319    .000099     -6.385   0.000       -.0008264   -.0004375
       w |   .4557085   .0912787      4.992   0.000        .2763872    .6350297
------------------------------------------------------------------------------
```

End of Example 1

Example 2

```
. * Estimate reg
. reg logwage educ female exper expsq
(Output omitted)
. predict e, residual
. gen logesq=ln(e*e)
```

```
. reg logesq educ female exper expsq
(output omitted)
. predict esqhat
(option xb assumed; fitted values)
. gen omega=exp(esqhat)

. * Generate weighted varaibles
. gen w=1/(omega)^0.5
. gen wlogwage=logwage*w
. gen wfemale=female*w
. gen weduc=educ*w
. gen wexper=exper*w
. gen wexpsq=expsq*w

. * Estimate Feasible GLS (FGLS) model
. reg wlogwage weduc wfemale wexper wexpsq w, noc

  Source |       SS       df       MS                  Number of obs =     524
---------+------------------------------               F(  5,   519) = 1569.72
   Model |  31164.1981     5  6232.83962               Prob > F      =  0.0000
Residual |  2060.77223   519  3.97065941               R-squared     =  0.9380
---------+------------------------------               Adj R-squared =  0.9374
   Total |  33224.9703   524   63.406432               Root MSE      =  1.9927

------------------------------------------------------------------------------
wlogwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   weduc |   .0828952   .0069779     11.880   0.000       .0691868    .0966035
 wfemale |  -.2914609   .0349884     -8.330   0.000      -.3601971   -.2227246
  wexper |   .0376525    .004497      8.373   0.000       .0288179    .0464872
  wexpsq |  -.0006592   .0001008     -6.540   0.000      -.0008573   -.0004612
       w |   .3848487   .0950576      4.049   0.000       .1981038    .5715936
------------------------------------------------------------------------------
```

End of Example 2