Lecture Notes on Advanced Econometrics

Lecture 8: Instrumental Variables Estimation

Endogenous Variables

Consider a population model:

\[ y_1 = \beta_2 y_2 + \beta_1 + \beta_2 x_2 + \ldots + \beta_{k-1} x_{k-1} + u \]  

(7-1)

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 in the next lecture note). 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 finding proper instruments is very difficult!

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

\[ y_2 = \delta_z z + \delta_1 + \delta_2 x_2 + \ldots + \delta_{k-1} x_{k-1} + u \]

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 \( x = (y_2, 1, x_2, \ldots, x_{k-1}) \) as a 1-by-k vector, \( z = (z, 1, x_2, \ldots, x_{k-1}) \) 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:

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

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 collinearity in \( Z \). The condition that \( Z'X \) has full rank of k is called the rank condition.

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 \( A = E(z'x) \).

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

**A Bivariate IV model**

Let’s consider a simple bivariate model:

\[
y_1 = \beta_{y2} y_2 + \beta_1 + u
\]

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

\[
\text{cov}(z, y_1) = \text{cov}(z, \beta_{y2} y_2 + \beta_1 + u) \\
= \beta_{y2} \text{cov}(z, y_2) + \text{cov}(z, u)
\]
Because $\text{cov}(z, u) = 0$,

$$
\beta_2 = \frac{\text{cov}(z, y_1)}{\text{cov}(z, y_2)} = \frac{\sum_{i=1}^{n} (z_i -  \bar{z}) (y_{1i} -  \bar{y}_1)}{\sum_{i=1}^{n} (z_i -  \bar{z}) (y_{2i} -  \bar{y}_2)}
$$

The problem in practice is the first requirement, $\text{cov}(z, u) = 0$. We cannot 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. Here is a bivariate case.

**Weak Correlation between the IVs and the Endogenous Variables**

In a bivariate model, we write

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

because $\text{Corr}(z, u) = \frac{\text{cov}(z, u)}{\text{sd}(z)\text{sd}(u)}$ (see Wooldridge pp714)

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

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

Thus if $z$ is only weakly correlated with the endogenous variable, $y_2$, $-\text{corr}(z, y_2)$ is very small, the IV estimator could be severely biased.

*Example 7-1: Card (1995), CARD.dta.*

A dummy variable grew up near a 4 year college as an IV on $educ$.

**OLS**

```
. reg lwage educ
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 3010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>58.5153536</td>
<td>1</td>
<td>58.5153536</td>
<td>F(  1, 3008) = 329.54</td>
</tr>
<tr>
<td>Residual</td>
<td>534.126258</td>
<td>3008</td>
<td>.17756857</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
</tbody>
</table>

R-squared = 0.0987
Correlation between nearc4 (an IV) and educ
.reg educ nearc4
Source | SS df MS
---------+------------------------------
Model | 448.604204 1 448.604204
Residual | 21113.4759 3008 7.01910767
Total | 21562.0801 3009 7.16586243
------------------------------------------------------------------------------
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
---------+------------------------------
Model | -340.111443 1 -340.111443
Residual | 932.753054 3008 .310090776
Total | 592.641611 3009 .196956335
------------------------------------------------------------------------------
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
------------------------------------------------------------------------------

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 7-1*
The Two-Stage Least Squares Estimation

Again, let’s consider a population model:

\[ y_1 = \beta_2 y_2 + \beta_1 + \beta_2 x_2 + \ldots + \beta_{k-1} x_{k-1} + u \]  

(7-3)

where \( y_2 \) is an endogenous variable. Suppose that there are \( m \) instrumental variables. Instruments, \( z = (1, x_2, \ldots, x_{k-1}, z_1, \ldots, z_m) \), are correlated with \( y_2 \). From the reduced form equation of \( y_2 \) with all exogenous variables (exogenous independent variables plus instruments), we have

\[ y_2 = \delta_1 + \delta_2 x_2 + \ldots + \delta_{k-1} x_{k-1} + \delta_{k} z_1 + \ldots + \delta_{k+m-1} z_m + r_{y2} \]  

(7-4)

\( \hat{y}_2 \) 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 (7-3), \( \hat{y}_2 \) is not correlated with \( u \), while \( r_{y2} \) is correlated with \( u \). 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:

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

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

\[ \hat{\beta}_{2SLS} = (\hat{X}'X)^{-1}\hat{X}'Y = (X'P_Z X)^{-1}X'P_Z Y = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \]  

(7-5)
This can be also written as
\[ \hat{\beta}_{2SLS} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{Y} \]  

(7-6)

This is the 2SLS estimator. It is called as two-stage because it looks like we take two steps by creating projected \( \mathbf{X} \) to estimate 2SLS estimator in (10-6). As matter of a fact, we do not need to take two steps, as you can see in (10-6), we can just estimate 2SLS estimators in one step by using \( \mathbf{X} \) and \( \mathbf{Z} \). (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 from (10-5) or (10-6). 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, it is always safe not to use the two-step procedure.

Avoid using predicted variables as much as you can!

Econometric packages will provide you 2SLS results based on (10-5) or (10-6). 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 some 1-by-(k+m-1) vector \( \mathbf{z} \), \( \text{E}(\mathbf{z}'u) = 0 \),

where \( \mathbf{z} = (1, x_2, \ldots, x_{k-1}, z_1, \ldots, z_m) \).

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

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

The order condition is \( k-1+m \geq k-1+h \), 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,

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

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: \( q \) against \( z \)'s and \( x \)'s.
(ii) Obtain the residual, \( \hat{v} \)
(iii) Estimate
\[ y = \beta_0 + \beta_1 q + \beta_2 x_1 + \ldots + \beta_{k-1} x_{k-1} + \hat{v} + u \]
(iv) If \( \beta_1 \) is significant, then \( q \) is endogenous.

**Testing the Over-Identification**

(i) Estimate \( \hat{\beta}_V \) and obtain \( \hat{v} \).
(ii) Regress \( \hat{v} \) on \( z \)'s and \( x \)'s.
(iii) Get \( R^2 \) and get \( NR^2 \), which is chi-squared.

**Example 7-2: Card (1995), card.dta again.**

```
. 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 =  0.2905               Root MSE      =  .37419
        |               ---------+------------------------------               Adj R-squared =  0.2891               Root MSE      =  .37419
Total |  592.641611  3009  .196956335
        |               ---------+------------------------------               Adj R-squared =  0.2891               Root MSE      =  .37419
        | lwage |      Coef.    Std. Err.      t    P>|t|     [95% Conf. Interval]
--------+-----------------------------------------------
        | lwage |      Coef.    Std. Err.      t    P>|t|     [95% Conf. Interval]
--------+-----------------------------------------------
```
| Variable | 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 |

IV Estimation: nearc2 nearc4 as IVs

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

Instrumental variables (2SLS) regression

```
Source | SS       | df       | MS
---------|----------|----------|----------
Model    | 86.2368644 | 6     | 14.3728107 |
Residual | 506.404747   | 3003 | .168632949 |
Total    | 592.641611   | 3009 | .196956335 |
```

```
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

```
Instrumented: educ
Instruments: nearc2 nearc4 + exper expersq ... south
```

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

Instrumental variables (2SLS) regression

```
Source | SS       | df       | MS
---------|----------|----------|----------
Model    | 108.419483 | 6     | 18.0699139 |
Residual | 320.580001  | 2213 | .144862178 |
Total    | 428.999484  | 2219 | .193330097 |
```

```
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  | -.0002449 .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

```
Instrumented: educ
Instruments: nearc2 nearc4 fatheduc motheduc + exper expersq ... south
```

8
**IV Estimation: fatheduc motheduc as IVs**

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

Instrumental variables (2SLS) regression

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 2220</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>108.477154</td>
<td>6</td>
<td>18.0795257</td>
<td>F(  6,  2213) = 83.44</td>
</tr>
<tr>
<td>Residual</td>
<td>320.52233</td>
<td>2213</td>
<td>0.144836118</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>428.999484</td>
<td>2219</td>
<td>0.193330097</td>
<td>R-squared = 0.2529</td>
</tr>
</tbody>
</table>

```
| 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 | -.00501 | 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 7-2*
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 \( u_j \):

\[
E(u'u') = \begin{bmatrix}
E(u_1u_1) & E(u_1u_2) & \cdots & E(u_1u_n) \\
E(u_2u_1) & E(u_2u_2) & \cdots & E(u_2u_n) \\
\vdots & \vdots & \ddots & \vdots \\
E(u_nu_1) & E(u_nu_2) & \cdots & E(u_nu_n)
\end{bmatrix}
= \begin{bmatrix}
\sigma_1^2 & 0 & \cdots & 0 \\
0 & \sigma_2^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_n^2
\end{bmatrix}
= \Sigma
\]

or

\[
E(u'u') = \sigma^2 \begin{bmatrix}
\frac{\sigma_1^2}{\sigma^2} & 0 & \cdots & 0 \\
0 & \frac{\sigma_2^2}{\sigma^2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{\sigma_n^2}{\sigma^2}
\end{bmatrix} = \sigma^2 \Omega
\]

Notice that under homoskedasticity, \( \Omega = I \).

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

\[
\text{Var}(\hat{B}) = \text{Var} [B + (X'X)^{-1} X' u]
= E [ (X'X)^{-1} X' u u' X (X'X)^{-1} ]
= (X'X)^{-1} X' E [u u'] X (X'X)^{-1}
= (X'X)^{-1} X' \sigma^2 \Omega X (X'X)^{-1}
= (X'X)^{-1} X' \Sigma X (X'X)^{-1}
\]

(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:

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

Thus, in the presence of heteroskedasticity, the statistical inference based on \( s^2(X'X)^{-1} \) would be biased, and t-statistics and F-statistics are inappropriate. Instead, we should use (7-1) 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:

\[ \hat{\beta} | X ~ N[B, \sigma^2 (X'X)^{-1} (X\Omega X)(X'X)^{-1}] \]

**Asymptotic Properties of OLS Estimators**
If \( Q = \text{plim}(X'X) \) and \( \text{plim}(X'\Omega X/n) \) are both finite positive definite matrices, then \( \hat{\beta} \) is consistent for \( B \). Under the assumed conditions,

\[ \text{plim}\hat{\beta} = B. \]

(See Greene (2003), page 193-194, for details.)

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

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

(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 \( \hat{\Sigma} \), we have
\[
X' \hat{\Sigma} X = X' \begin{bmatrix}
\hat{u}_1^2 & 0 & 0 \\
0 & \hat{u}_2^2 & 0 \\
0 & 0 & \hat{u}_n^2
\end{bmatrix} X = \sum_{i=1}^{n} \hat{u}_i^2 \begin{bmatrix}
x_{i1}^2 & x_{i1}x_{i2} & \ldots & x_{i1}x_{ki} \\
x_{i2}x_{i1} & x_{i2}^2 & \ldots & x_{i2}x_{ki} \\
\vdots & \vdots & \ddots & \vdots \\
x_{in}x_{i1} & x_{in}x_{i2} & \ldots & x_{in}^2
\end{bmatrix}.
\]

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

\[
\text{Var}(\hat{B}) = (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1}
\]

(7-2)

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 \texttt{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 OLS 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

\[
H_0: \ RB = r
\]

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

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

\[ H_0: \ E(u| X) = \sigma^2 \]

To test this null hypothesis, we estimate

\[ \hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + e. \quad (7-3) \]

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

\[ LM = n R^2 \sim \chi^2_k \]

**The White Test for Heteroskedasticity**
White proposed to add the squares and cross products of all independent variables to (9-2):

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

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 \( \delta_2 = 0 \).

Example 7-1: Step-by-Step Estimation for Robust Standard Errors

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

\[ \log(wage) = \beta_0 + \beta_1 \text{female} + \beta_2 \text{educ} + \beta_3 \text{exper} + \beta_4 \text{expersq} + u \]

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., \texttt{reg y x1 x2 x3 x4, robust}). But at least you know how robust standard errors are calculated by STATA.

\[ \text{Number of obs} = 526 \]

\[ \text{F( 4, 521) = 86.69} \]

\[ \text{Prob > F} = 0.0000 \]

\[ \text{R-squared} = 0.3996 \]

\[ \text{Adj R-squared} = 0.3950 \]

\[ \text{Root MSE} = 0.41345 \]
predict e, residual
.
. * Create variables with squared residuals
  gen esq=e*esq
  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

. * 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.

  ------------------------------------------------------------------------------
  | female | -3371868 | 0.0363214 | 9.283 | 0.000 | -4085411 | -2658324 |
  | educ | 0.0841361 | 0.0069568 | 12.094 | 0.000 | 0.0704692 | 0.0978029 |
  | exper | .03891 | 0.0048235 | 8.067 | 0.000 | 0.029434 | 0.0483859 |
  | expsq | -0.000686 | 0.0001074 | -6.389 | 0.000 | -0.000897 | -0.0004751 |
  | _cons | .390483 | 0.1022096 | 3.820 | 0.000 | 0.1896894 | 0.5912767 |
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

| Coef.       | Std. Err. | t     | P>|t|       | 95% Conf. Interval |
|-------------|-----------|-------|-----------|-------------------|
| logwage     | -.3371868 | .0361838 | -9.319    | 0.000             |- .4082709, -.2661026 |
| female      | 0.0841361 | .00769 | 10.941    | 0.000             | .069029, .0992432   |
| educ        | 0.03891   | .0046752 | 8.323    | 0.000             | .0297253, .0480946  |
| exper       | -.000686  | .0001005 | -6.829   | 0.000             | -.0008834, -.0004887 |
| expsq       | 3.596     | .1085985 | 3.596    | 0.000             | .1771383, .6038278  |

end of do-file

Example 7-2: 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

| 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

End of Example 7-1
Example 7-2

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 7-2

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
          +-------------------------------------------
Total |   40.608506   525  .077349535               Adj R-squared =  0.0111
          +-------------------------------------------

         |      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 7-3
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
\[ \Omega^{-1} = PP' \]
\(P\) is a “n x n” matrix

Pre-multiply \(P\) on a regression model
\[ Py = PXB + Pu \]
or
\[ y* = X* B + u* \quad (9-1) \]

In this model, the variance of \(u*\) is
\[ E(u* u'^* ) = E(uu')P = P \sigma^2 \Omega P' = \sigma^2 P \Omega P' = \sigma^2 I. \]

Note that \(PSPP' = I\), because define \(PSP\) is \(A\), then \(P\Omega P' = P' A\), by the definition of \(P \Omega^{-1} P' = P' A\), thus \(P' = P' A\). Therefore, \(A\) must be \(I\).

Because \(E(u* u'^* ) = \sigma^2 I\), the model (9-3) satisfies the assumption of homoskedasticity. Thus, we can estimate the model (9-3) by the conventional OLS estimation.

Hence,
\[ \hat{B} = (X'X_*)_^{-1} X'_* y_* \]
\[ = (X'PX)^{-1} X'Py \]
\[ = (X'\Omega^{-1} X)^{-1} X'\Omega^{-1} y \]
is the efficient estimator of \(B\). This is called the Generalized Least Square (GLS) estimator. Note that the GLS estimators are unbiased when \(E(u*| X_*) = 0\). The variance of GLS estimator is
\[ \text{var}(\hat{B}) = \sigma^2 (X'X_*)_^{-1} = \sigma^2 (X'\Omega^{-1} X)^{-1}. \]

Note that, under homoskedasticity, i.e., \(\Omega^{-1} = I\), GLS becomes OLS.
The problem is, as usual, that we don’t know $\sigma^2 \Omega$ or $\Sigma$. Thus we have to either assume $\Sigma$ or estimate $\Sigma$ 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,

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

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

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

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 $\omega_i$?

**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 $\Omega$.

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

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

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

19
By taking log of the both sides and using $\hat{u}^2$ instead of $u^2$, we can estimate

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

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

**Example 9-1**

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

. * Generate weighted variables
. 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
```

```
<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 524</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F( 5, 519) = 1660.16</td>
</tr>
<tr>
<td>Model</td>
<td>113.916451</td>
<td>5 22.7832901</td>
<td>Prob &gt; F = 0.0000</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>7.12253755</td>
<td>519 .013723579</td>
<td>R-squared = 0.9412</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.9406</td>
</tr>
<tr>
<td>Total</td>
<td>121.038988</td>
<td>524 .230990435</td>
<td>Root MSE = .11715</td>
<td></td>
</tr>
</tbody>
</table>

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

End of Example 9-1
```

**Example 9-2**

```plaintext
* 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 9-2
Lecture 11: Unobserved Effects and Panel Analysis

Panel Data

There are two types of panel data sets: a pooled cross section data set and a longitudinal data set. A pooled cross section data set is a set of cross-sectional data across time from the same population but independently sampled observations each time. A longitudinal data set follows the same individuals, households, firms, cities, regions, or countries over time.

Many governments conduct nationwide cross sectional surveys every year or every once in a while. Census is an example of such surveys. We can create a pooled cross section data set if we combine these cross sectional surveys over time. Because these surveys are often readily available from governments (well this is not true for most of the times because of many government officials do not see any benefits of making publicly funded surveys public!), it is relatively easier to obtain pooled cross section data.

Longitudinal data, however, can provide much more detailed information as we see in this lecture. Because longitudinal data follow the same samples over time, we can analyze behavioral changes over time of the samples.

Nonetheless, pooled cross section data can provide information that a single cross section data cannot.

Pooled Cross Section Data

With pooled cross section data, we can examine changes in coefficients over time. For instance,

\[ y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 x_{it1} + \beta_3 x_{it2} + \beta_4 x_{it3} + u_{it} \]  

(1)

where \( T_{it} = 1 \) if \( t = 2 \) and \( 0 \) if \( t = 1 \).

The coefficient of the time dummy \( T_{it} \) measures a change in the constant term over time. If we are interested in a change in a potential effect of one of the variables, then we can use an interaction term between the time dummy and one of the variables:

\[ y_{it} = \beta_0 + \beta_1 d_{it} + \beta_2 x_{it1} + \delta_2 (T_{it} \times x_{it1}) + \beta_3 x_{it2} + \beta_4 x_{it3} + u_{it} \]

(2)

\( \delta_2 \) measures a change in the coefficient of \( x_1 \) over time.

How about if there are changes in all of the coefficients over time? To examine if there is a structural change, we can use the Chow test. To conduct the Chow test, consider the following model:
\[ y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{it3} + \beta_4 x_{it4} + u_{it} \]  
for \( t = 1, 2 \).

We consider this model as a **restricted model** because we are imposing restrictions that all the coefficients remain the same over time. There are \( k+1 \) restrictions (in this case 5 restrictions).

**Unrestricted models** are

\[ y_{it} = \delta_0 + \delta_1 x_{it1} + \delta_2 x_{it2} + \delta_3 x_{it3} + \delta_4 x_{it4} + v_{it} \quad \text{for } t = 1 \]  
\[ y_{it} = \gamma_0 + \gamma_1 x_{it1} + \gamma_2 x_{it2} + \gamma_3 x_{it3} + \gamma_4 x_{it4} + e_{it} \quad \text{for } t = 2 \]  

The coefficients of the first model (\( t = 1 \)) are not restricted to be the same as in the second model (\( t = 2 \)). If all of the coefficients remain the same over time, i.e., \( \beta_j = \delta_j = \gamma_j \), then the sum of squared residuals from the restricted model (SSR\(_r\)) should be equal to the sum of the sums of squared residuals from the two unrestricted models (SSR\(_{ur1}\) + SSR\(_{ur2}\)).

On the other hand, if there is a structural change, i.e., changes in the coefficients over time, then the sum of SSR\(_{ur1}\) and SSR\(_{ur2}\) should be smaller than SSR\(_r\), because unrestricted coefficients in unrestricted models should match the data more precisely than the restricted model. Then we take the difference between SSR\(_r\) and (SSR\(_{ur1}\) + SSR\(_{ur2}\)) and examine if there is statistically significant difference between the two:

\[ F = \frac{\text{SSR}_r - (\text{SSR}_{ur1} + \text{SSR}_{ur2})}{(k + 1)} / \frac{(\text{SSR}_{ur1} + \text{SSR}_{ur2})/([N_1 + N_2 - 2(k + 1)])}{[k(N_1 + N_2 - 2(k + 1))]} \]

This is called the **Chow test**.

Alternatively, we can create interaction terms on all of independent variables (including the constant term) and conduct a F-test on the coefficients of the \( k+1 \) interaction terms. This is just the same the Chow test.

**Difference-in-Differences Estimator**

In many economic analyses, we are interested in some of policy variables or politically interesting variables and how these variables affect people’s lives. However, evaluating
the impacts of various policies is difficult because most of policies are not done under experimental designs.

For instance, suppose a government of a low-income country decided to invest in health facilities to improve child health. Suppose this particular government decided to start with the most-needy communities. After some years, the government wanted to evaluate the impacts of the investment in health facilities. The government conducts a cross-sectional survey. However, the government finds negative correlation between newly-build health facilities and child health. What happened?

The problem is that the government built health facilities in communities with poor child health.

In the figure above, the child health in poor community \( i \) with the government-investments \( (z) \) has improved over time, but its absolute level is still not as good as the child health in rich communities without the government investments. Thus, an OLS model with a dummy variable for the government investments in health facilities will find a coefficient of \( z \):

\[
H_{it} = \beta_0 + \beta_1 z_{it} + u_{it} \tag{6}
\]

for \( i = 1, \ldots, N \) communities.

When we find a negative coefficient or an opposite effect of what expected, we call it the reverse causality.

From the figure, it is obvious that we need to measure a difference between the two groups for each time period and measure a net change in the differences over time:

\[
\delta = [E(H_{i2}; z = 1) - E(H_{i2}; z = 0)] - [E(H_{i1}; z = 1) - E(H_{i2}; z = 0)]. \tag{7}
\]
Although both differences are negative, the difference between the two groups in the second period is much smaller than the difference in the first period. Thus, the net change is positive, which measures the net impact of $z$ on $H$. We call the $\delta$ in (10-7) the difference-in-differences (DID) estimator.

The difference-in-differences estimator can be estimated by estimating the following model:

$$H_{it} = \beta_0 + \beta_1 T + \beta_2 z_i + \delta (T \times z_i) + u_{it}$$  \hspace{1cm} (8)

We can think this example as a kind of the omitted variables problem. We can rewrite (9-6) as

$$H_i = \beta_0 + \beta_1 z_i + \alpha_i + u_i$$  \hspace{1cm} (9)

where $\alpha_i$ is an important unobserved variable (or an unobserved fixed effect) which is correlated with both the government investments and the child health. Let’s say that $\alpha_i$ measures the lack of basic infrastructure in community $i$: the larger the $\alpha_i$, the poorer the basic infrastructure. Because the government targets the poor communities for the investments, $\alpha_i$ and $z_i$ are correlated positively. But $\alpha_i$ and $H_i$ are correlated negatively because $\alpha_i$ measures the lack of basic infrastructure. Therefore, the estimated coefficient of $z_i$ will be biased downward, which produces a reverse causality.

**The First Differenced Estimation**

Let’s go back to the DID estimator and rearrange it so that the first term measure a difference in $H_{it}$ of community $i$ over time:

$$\delta = [E(H_{T2}; z = 1) – E(H_{T1}; z = 1)] – [E(H_{C2}; z = 0) – E(H_{C1}; z = 0)].$$  \hspace{1cm} (10)

Here the first term measures a change over time for the treatment group ($T$) and the second term measures for the comparison group ($C$).

In a regression form, we can also rearrange (8). Let’s write the equation (8) with an unobserved fixed effect:

$$H_{it} = \beta_0 + \beta_1 T + \beta_2 z_i + \delta (T \times z_i) + \alpha_i + u_{it}$$  \hspace{1cm} (11)

Now, the problem is that $z$ could be correlated with $\alpha_i$, which may be also correlated with $H_{it}$. For the first period (thus $T = 0$), the equation (11) is

$$H_{it} = \beta_0 + \beta_2 z_i + \alpha_i + u_{it},$$

and for the second period ($T = 1$):
\[ H_{it+1} = (\beta_0 + \beta_1) + (\beta_2 + \delta) z_i + \alpha_i + u_{it+1}. \]

Then by taking the first-difference, we have

\[ H_{it+1} - H_{it} = \beta_0 + \delta z_i + v_{it+1} \tag{12} \]

Notice that the unobserved fixed effect, \( \alpha_i \), has been excluded from this model because the unobserved fixed effect is fixed over time. In the first-differenced equation (12), \( z \) will not be correlated with the error term.

**Quasi-experimental and Experimental Designs**

From this point of view, it is obvious that under a nonrandom assignment of \( z \) (or a quasi-experimental design), \( \delta \) in (8) could be biased because \( z \) (a program indicator) could be correlated with unobserved factors which may be also correlated with \( H \) (a dependent variable).

In contrast, under a *random* assignment of \( z \) (or an experimental design), \( z \) will not be correlated with any of unobserved factors. Thus the difference-in-differences estimator will provide reliable estimators of the impacts of programs on outcomes.

Under an experimental design, a group of individuals or observations that receive benefits from a give policy is called a **treatment group** or an experimental group. And a group of non-beneficiaries is called a **control group**. Under a quasi-experimental design, a group of non-beneficiaries is called a comparison group, and reserve the name “the control group” for experimental designs.

In social science, it is difficult to conduct experimental designed programs because of ethics and political difficulties. But experimental designed programs can provide very useful information about the effectiveness of public (or private) policies.

Recent Example of an experimental designed project: PROGRESA in Mexico, designed and research by IFPRI. See www.ifpri.org.


Thus, we have dealt with an omitted variable problem by taking a difference over time. Next, we study the omitted fixed effect problem in general.

**The Omitted Variables Problem Revisited**

Suppose that a correctly specified regression model would be
\[ y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u \]

\(X_1\) and \(X_2\) have \(k_1\) and \(k_2\) columns, respectively. But, suppose we regress \(y\) on \(X_1\) without including \(X_2\) (\(X_2\) represent omitted variables). The OLS estimator is

\[
\hat{\beta}_1 = (X'_1X_1)^{-1}X'_1Y = (X'_1X_1)^{-1}X'_1(X_1\beta_1 + X_2\beta_2 + u) \\
= \beta_1 + (X'_1X_1)^{-1}X'_1X_2\beta_2 + (X'_1X_1)^{-1}X'_1u
\]

By taking the expectation on both sides, we have

\[
E(\hat{\beta}_1) = \beta_1 + (X'_1X_1)^{-1}X'_1X_2\beta_2 = \beta_1 + \hat{\delta}_{12}\beta_2
\]

Note, however, that the second term indicates the column of slopes (\(\hat{\delta}_{12}\)) in least squares regression of the corresponding column of \(X_2\) on the columns of \(X_1\).

Thus, unless either \(\hat{\delta}_{12} = 0\) or \(\beta_2 = 0\), \(\hat{\beta}_1\) is biased.

To overcome the omitted variables problem, we can take two different methods. First method is to use panel data. As you see later, by using panel (longitudinal) data, we can eliminate unobserved variables that are specific to each sample and fixed (or time-invariant or time-constant) over time. Second method is to use instrumental variables that are correlated with independent variables that are considered to be correlated with unobserved variables but uncorrelated with the dependent variable. We will discuss instrumental variables in the next lecture note.

Before we discuss about estimation methods that use panel data, let us start with types of panel data.

**Linear Unobserved Effects**

What are unobserved variables? It is impossible to collect all variables in surveys that affect people’s economic activities. Thus, it is inevitable to have unobserved variables in our estimation models. What, then, we should do? First, we should start with characterizing possible unobserved variables.

The most common type of unobserved variables is a fixed effect. A fixed effect is a time invariant characteristic of an individual or a group (or cluster). For instance, \(a_i\) may represent a fixed characteristic of group \(i\). This could be a regional fixed effect or a cluster fixed effect. Another example is \(a_j\) which represents a fixed characteristic of a group (cluster) \(j\).

Suppose, we want to estimate the following model with a group fixed effect,
\[ y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + a_j + u_i \]

(13)

In this case, as long as unobserved variables (that are correlated with individual variables and the dependent variable) are fixed characteristics of groups, then we can eliminate the omitted variables problem by explicitly including group dummies:

\[ y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \alpha' d_j + u_i \]

(14)

where \( \alpha' \) is a 1-by-\( j \) vector, and \( d_j \) is a \( j \)-by-1 vector. For instance, it is common practice to include district or village dummies in cross-sectional data. However, in a cross-sectional data set, it is impossible to include individual dummies for all samples because we have only one observation per sample. We will have \( n-1 \) dummies for \( n \) observations.

If we have multiple observations for each sample (thus we need longitudinal data not a pooled cross-sectional data over time), then it is possible to have \( n-1 \) dummies for \( s \times n \) observations. \( (s \) is the number of observations per sample.) Thus, we estimate

\[ y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + \alpha' d_i + u_{it} \]

(15)

This is called the **Dummy Variable Regression** model. In this model, we have eliminated the unobserved fixed effects by explicitly including individual dummy variables.

A different way of eliminating the fixed effects is to use the **first difference model**, as we have seen earlier. Here let us reconsider the first difference model in a general treatment. Suppose, again, that we have the following model for time \( t=1 \) and \( t=2 \):

\[ y_{i1} = \beta_0 + \beta_1 x_{i11} + \cdots + \beta_k x_{i1k} + a_j + u_{i1} \]

and

\[ y_{i2} = \beta_0 + \beta_1 x_{i21} + \cdots + \beta_k x_{i2k} + a_j + u_{i2} \]

By subtracting the model for \( t=1 \) from the model for \( t=2 \), we have

\[ y_{i2} - y_{i1} = \beta_1 (x_{i21} - x_{i11}) + \cdots + \beta_k (x_{i2k} - x_{i1k}) + u_{i2} - u_{i1} \]

or

\[ \Delta y = \beta_1 \Delta x_{i1} + \cdots + \beta_k \Delta x_{ik} + \Delta u_i \]

This is called the first difference model. Notice that the individual fixed effect, \( a_j \), has been eliminated. Thus as long as the new error term is uncorrelated with the new independent variables, then the estimators should be unbiased.

Some notes: First, a first differenced independent variable, \( \Delta x_{ik} \), must have some variation across \( i \). For instance, a gender dummy variable does not change over time, the first-differenced gender dummy is zero for all \( i \). Thus, you can not estimate coefficient on time-invariant independent variables in first difference models. Second, differenced independent variables loose variation. Thus, estimators often have large standard errors. Large sample size helps to estimate parameters precisely.
Fixed Effect Estimation

In the previous lecture, we studied the first differenced model, concerning the correlation between a policy variable and an unobserved fixed effect. In this lecture, we generalize the model. Consider the following model with T period and k variables

\[ y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_k x_{itk} + \alpha_i + u_{it} \]  

(16)

The omitted unobserved fixed effect could be correlated with any of k independent variables.

To take the fixed effect away, one can subtract the mean of each variable:

\[ y_{it} - \bar{y}_i = \beta_1 (x_{it1} - \bar{x}_{it1}) + \ldots + \beta_k (x_{itk} - \bar{x}_{itk}) + v_{it} \]  

(17)

As you can see, the unobserved fixed effect has been excluded from the model. This model is called the fixed effect estimation. To estimate the fixed effect model, you need to transform each variable by taking the mean out and estimate the OLS with the transformed data (the time-demeaned data). In STATA, you don’t need to transfer the data yourself. Instead you just need to use a command “xtreg y x_1 x_2 \ldots x_k, fe i(id).” See the manuals under “xtreg.”

One drawback of the fixed effect estimation is that some of time-invariant variables will be also excluded from the model. For instance, consider a typical wage model, where the dependent variable is \( \log(wage) \). Some of individual characteristics, such as education and gender are time-invariant (or fixed over time). Thus if you are interested in the effects of time-invariant variables you cannot estimate the coefficients of such variables. However, what you can do is to estimate the changes in the effects of such time-invariant variables.

Instead of taking the fixed effects out, we can explicitly include them as dummies:

\[ y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_k x_{itk} + \delta_1 \alpha_1 + \ldots + \delta_n \alpha_n + u_{it} \]  

(18)

This is called the least squares dummy variable (LSDV) model. The LSDV model provides the exactly the same results as Fixed Effect model.

When T=2, the first differenced (FD) model, the Fixed Effect (FE) model, and LSDV model all provide the same results.
### Example 11-1: OLS, Fixed Effect, First-Differenced, and LSDV models

```
. use c:\docs\fasid\econometrics\homework\JTRAIN.dta;
. keep if year==1988|year==1989;
(157 observations deleted)
. replace sales=sales/10000;
(254 real changes made)

** OLS;
. reg hrsemp grant employ sales union d89;

Source |       SS       df       MS                  Number of obs =     220
---------+------------------------------               F(  5,   214) =    2.07
Model |  64302349.8     5  12860470.0               Prob > F      =  0.0703
Residual |  1.3291e+09   214  6210560.99               R-squared     =  0.0461
---------+------------------------------               Adj R-squared =  0.0239
Total |  1.3934e+09   219  6362385.39               Root MSE      =  2492.1

------------------------------------------------------------------------------
hrsemp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
grant |   824.4124   419.8616      1.964   0.051      -3.181518    1652.006
employ |  -3.411366   4.232316     -0.806   0.421      -11.75373    4.931001
sales |   .0431946   .3313198      0.130   0.896      -.6098735    .6962627
union |   942.0082   442.0818      2.131   0.034       70.61581    1813.401
d89 |  -287.6238   338.4361     -0.850   0.396      -954.7191    379.4715
_cons |   156.7909   300.7423      0.521   0.603      -436.0056    749.5874
------------------------------------------------------------------------------

** Fixed Effect Model;
. xtreg hrsemp grant employ sales union d89, fe i(fcode);

Fixed-effects (within) regression               Number of obs      =       220
Group variable (i) : fcode                      Number of groups   =       114
R-sq:  within  = 0.0322                         Obs per group: min =         1
between = 0.0106                                        avg =       1.9
overall = 0.0212                                        max =         2
corr(u_i, Xb)  = -0.0144                        F(4,102)           =      0.85
obs per group: min =         1                      Prob > F           =    0.4972
avg =       1.9                                     F(4,102) =     0.87
max =         2                                     Prob > F =  0.7716

------------------------------------------------------------------------------
hrsemp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
grant |   846.6938   552.6861      1.532   0.129      -249.5565    1942.944
employ |  -3.079166   5.235437     -0.600   0.551      -13.43076    7.272438
sales |  -0.0847216   .9174962     -0.092   0.927      -1.904871    1.735428
union |  (dropped)
d89 |  -292.2915   368.5441     -0.793   0.430      -1023.297     438.714
_cons |   137.1313   917.5104      0.149   0.881      -1682.746    1957.009
------------------------------------------------------------------------------

sigma_u |  1742.3611
sigma_e |   2578.396
rho |  .31349012   (fraction of variance due to u_i)

F test that all u_i=0:     F(113,102) =     0.87             Prob > F = 0.7716
```

---

30
** LSDV model;**
.xi: reg hrsemp grant employ sales union d89 i.fcode;
i.fcode               Ifcod1-157   (Ifcod1 for fcode==410032 omitted)

Source |       SS       df       MS                  Number of obs =     220
---------+------------------------------               F(117,   102) =    0.92
Model |   71523529    117   6113278.03               Prob > F      =  0.6706
Residual |   678108872   102   6648126.19               R-squared     =  0.5133
---------+------------------------------               Adj R-squared = -0.0449
Total |  1.3934e+09   219  6362385.39               Root MSE      =  2578.4

------------------------------------------------------------------------------
hrsemp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
grant |   846.6938   552.6861      1.532   0.129      -249.5565    1942.944
employ |   1.504937   20.69735      0.073   0.942      -39.54816    42.55803
sales |  -.0847216   .9174962     -0.092   0.927      -1.904571    1.735128
union |  -838.5586   5723.478     -0.147   0.884      -12191.05    10513.93
d89 |  -292.2915   368.5441     -0.793   0.430      -1023.297     438.714
Ifcod2 |  -192.7619   3889.521     -0.050   0.961      -7907.609    7712.085
Ifcod3 |  -203.2925   4021.587     -0.051   0.960      -8180.091    7773.506

Output omitted...

** First Differenced Model;**
.reg dhrsemp dgrant demploy dsales;

Source |       SS       df       MS                  Number of obs =     106
---------+------------------------------               F(  3,   102) =    0.81
Model |  32189365.8     3  10729788.6               Prob > F      =  0.4928
Residual |  1.3562e+09   102  13296252.1               R-squared     =  0.0232
---------+------------------------------               Adj R-squared = -0.0055
Total |  1.3884e+09   105  13222924.6               Root MSE      =  3646.4

------------------------------------------------------------------------------
dhrsemp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
dgrant |   846.6938   552.6861      1.532   0.129      -249.5565    1942.944
demploy |   1.504937   20.69735      0.073   0.942      -39.54816    42.55803
dsales |  -.0847216   .9174962     -0.092   0.927      -1.904571    1.735128
_cons |  -292.2915   368.5441     -0.793   0.430      -1023.297     438.714

End of Example 11-1

** Example 11-2: Transferring The Panel Data**

In general, the panel data are stacked vertically. For instance, in JTRAIN.dta, two observations (actually there are three years of observations, but I dropped one year) for each firm is stacked vertically:

.list fcode year d89 employ

<table>
<thead>
<tr>
<th>firm code</th>
<th>year</th>
<th>d89</th>
<th># of employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 410032</td>
<td>1988</td>
<td>0</td>
<td>131</td>
</tr>
<tr>
<td>2. 410032</td>
<td>1989</td>
<td>1</td>
<td>123</td>
</tr>
<tr>
<td>3. 410440</td>
<td>1988</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>4. 410440</td>
<td>1989</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>5. 410495</td>
<td>1988</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>6. 410495</td>
<td>1989</td>
<td>1</td>
<td>24</td>
</tr>
</tbody>
</table>
To construct differenced variables, you need to “linearize” the vertical data. Here is an example:

. do "C:\WINDOWS\TEMP\STD0c0000.tmp"
. #delimit;
delimiter now ;
. clear;
. set more off;
. set matsize 800;
. set memory 100m;
(102400k)

. *** Obtain data from 1988;
. clear;
. use c:\docs\fasid\econometrics\homework\JTRAIN.dta;
. keep year fcode employ;
. keep if year==1988;
(314 observations deleted)
. rename employ employ88;
. drop year;
. sort fcode;
. save c:\docs\tmp\jtrain88.dta, replace;
file c:\docs\tmp\jtrain88.dta saved

. *** Obtain data from 1989;
. clear;
. use c:\docs\fasid\econometrics\homework\JTRAIN.dta;
. keep fcode year employ;
. keep if year==1989;
(314 observations deleted)
. rename employ employ89;
. drop year;
. sort fcode;
. save c:\docs\tmp\jtrain89.dta, replace;
file c:\docs\tmp\jtrain89.dta saved

. ** Combine the two years;
. clear;
. use c:\docs\tmp\jtrain89.dta;
. sort fcode;
. merge fcode using c:\docs\tmp\jtrain88.dta;
. gen demploy=employ89-employ88;
(11 missing values generated)

. list fcode employ89 employ88 demploy in 1/3;
     fcode | employ89 | employ88 | demploy
----------|----------|----------|----------
     1.  410032 |  123     |  131     |     -8
     2.  410440 |   14     |   13     |      1
     3.  410495 |   24     |   25     |     -1

End of Example 11-2
Lecture 12: Random Effect Models and the Hausman Test

Random Effect Estimation

Let’s go back to a general longitudinal model with $T$ periods:

$$ y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_k x_{itk} + \alpha_i + u_{it} \quad (1) $$

The purpose of the fixed effect estimation and the first differenced estimation is to eliminate $\alpha_i$ because we suspect $\alpha_i$ to be correlated with some of independent variables. However, what if we are wrong (that there is no correlation between $\alpha_i$ and any of independent variables)? If this is case, then the FE and FD estimations will be inefficient because we lose $n-1$ degree of freedom (think about having $n-1$ dummies).

Even when there is no correlation between $\alpha_i$ and independent variables, the OLS estimation on (11-4) will have a problem because of heteroskedasticity.

In (11-4) the new error term is

$$ v_{it} = \alpha_i + u_{it} \quad (2) $$

Let’s make some assumptions on $\alpha_i$ and $u_i$:

- $E(\alpha_i) = 0$, $E(u_i) = 0$
- $\text{Var}(\alpha_i) = \sigma^2_{\alpha}$, $\text{Var}(u_i) = \sigma^2_u$
- $E(\alpha_i \alpha_j) = 0$, $E(u_i u_j) = 0$

Then for the same observation unit $i$,

$$ \text{Var}(v_{it}) = \sigma^2_{\alpha} + \sigma^2_u \quad \text{and} \quad \text{cov}(v_{it} v_{is}) = \sigma^2_{\alpha} $$

So for the same observation unit $i$ for $T$ period, the variance-covariance matrix is

$$ \Sigma_{y_{i,T}} = E(vv') = \begin{bmatrix} \sigma^2_{\alpha} & \sigma^2_{\alpha} & \sigma^2_u \\ \sigma^2_{\alpha} & \sigma^2_u & \sigma^2_u \\ \sigma^2_u & \sigma^2_u & \sigma^2_u + \sigma^2_u \end{bmatrix} $$

Because there are $N$ observation units for $T$ periods, the variance-covariance matrix for all observation $NT$:
The feasible GLS (FGLS) estimation based on the above variance-covariance matrix is called the **Random Effect estimation**.

**The Hausman Test**

So, the Fixed Effect model and Random Effect model, which one should we use? For this question, we can use the Hausman test. The Hausman test (a kind of Wald test) is

\[
W = [\hat{\beta}_{FE} - \hat{\beta}_{RE}][Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})]^{-1}[\hat{\beta}_{FE} - \hat{\beta}_{RE}].
\]

This test is base on an observation that under the assumption of zero conditional mean, FE estimators are consistent \((\hat{\beta}_{FE} - \hat{\beta}_{RE} = 0)\) but inefficient \([Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE}) > 0]\). Thus, \(W\) will be small.

If we are only interested in one important variable, then we can do the Hausman test on one variable. This requires a simple calculation by using usual standard errors:

\[
W = [\hat{\beta}_{FE,k} - \hat{\beta}_{RE,k}]/[se(\hat{\beta}_{FE,k})^2 - se(\hat{\beta}_{RE,k})^2]^{1/2}
\]

Thus, we just need to take a difference between the FE and RE coefficients of one variable, and divide it by a squared root of a difference in standard errors of the FE and RE estimators.

**Example 12-1: Random Effect Model**

Here we estimate the model in Example 11-1 with the random effect model.

```stata
. ** Random Effect Model;
. xtreg hrsemp grant employ sales union d89, i(fcode);

Random-effects GLS regression                   Number of obs      =       220
Group variable (i) : fcode                      Number of groups   =       114
R-sq:  within  = 0.0316                         Obs per group: min =         1
between = 0.0589                                 avg =       1.9
overall = 0.0461                                 max =         2
Random effects u_i ~ Gaussian                   Wald chi2(5)       =     10.35
corr(u_i, X) = 0 (assumed)                      Prob > chi2        =    0.0658
```

34
\[ \text{hrsemp} \]  
\[ \text{grant} \quad 824.4124 \quad 419.8616 \quad 1.964 \quad 0.050 \quad 1.498773 \quad 1647.326 \]  
\[ \text{employ} \quad -3.411366 \quad 4.232316 \quad -0.806 \quad 0.420 \quad -11.70655 \quad 4.883822 \]  
\[ \text{sales} \quad 0.0431946 \quad 0.3313198 \quad 0.130 \quad 0.896 \quad -0.6061802 \quad 0.6925694 \]  
\[ \text{union} \quad 942.0082 \quad 442.0818 \quad 2.131 \quad 0.033 \quad 75.5438 \quad 1808.473 \]  
\[ \text{d89} \quad -287.6238 \quad 338.4361 \quad -0.850 \quad 0.395 \quad -950.9464 \quad 375.6988 \]  
\[ \text{cons} \quad 156.7909 \quad 300.7423 \quad 0.521 \quad 0.602 \quad -432.6531 \quad 746.235 \]  
\[ \text{sigma_u} \quad 0 \]  
\[ \text{sigma_e} \quad 2578.396 \]  
\[ \text{rho} \quad 0 \quad \text{(fraction of variance due to u_i)} \]  

End of Example 12-1

Example 12-2 The Hausmann Test

\[ \text{** Fixed Effect Model;} \]  
\[ \text{. xtreg tothrs grant employ d88 d89, fe i(fcode);} \]  
\[ \text{Fixed-effects (within) regression Number of obs} = 390 \]  
\[ \text{Group variable (i) : fcode Number of groups} = 135 \]  
\[ \text{R-sq: within} = 0.3778 \]  
\[ \text{between} = 0.0348 \]  
\[ \text{overall} = 0.0821 \]  
\[ \text{F(4,251) = 38.10} \]  
\[ \text{corr(u_i, Xb) = -0.0268} \]  
\[ \text{F test that all u_i=0: F(134, 251) = 19.17} \]  
\[ \text{prob > F = 0.0000} \]  

\[ \text{** Random Effect Model;} \]  
\[ \text{. xtreg tothrs grant employ d88 d89, i(fcode);} \]  
\[ \text{Random-effects GLS regression Number of obs} = 390 \]  
\[ \text{Group variable (i) : fcode Number of groups} = 135 \]  
\[ \text{R-sq: within} = 0.3776 \]  
\[ \text{between} = 0.0348 \]  
\[ \text{overall} = 0.0830 \]  
\[ \text{Wald chi2(4) = 157.26} \]
corr(u_i, X) = 0 (assumed)  Prob > chi2 = 0.0000

|     | Coef.   | Std. Err. |    z  |     P>|z| |     [95% Conf. Interval] |
|-----|---------|-----------|-------|--------|--------------------------|
|grant| 30.54896| 2.906927  | 10.51 | 0.000  | 24.85149  36.24644 |
|employ| -1.152185| 0.406756 | -2.83 | 0.005  | -.194913  -.0354957 |
|d88| 1.185915| 2.28278 | 0.52  | 0.603  | -3.288252  5.660082 |
|d89| 8.415649| 2.265225 | 3.72  | 0.000  | 3.975891  12.85541 |
|_cons| 29.33547| 4.651854 | 6.31  | 0.000  | 20.21801  38.45294 |

sigma_u | 43.622037    
sigma_e | 17.236702    
rho | .86495191 (fraction of variance due to u_i)

. hausman:

<table>
<thead>
<tr>
<th></th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>grant</td>
<td>30.49303</td>
<td>30.54896</td>
<td>-.0559286</td>
<td>.4174025</td>
</tr>
<tr>
<td>employ</td>
<td>-.1312958</td>
<td>-.1152185</td>
<td>-.0160773</td>
<td>.0478801</td>
</tr>
<tr>
<td>d88</td>
<td>1.359526</td>
<td>1.185915</td>
<td>.173611</td>
<td>.204822</td>
</tr>
<tr>
<td>d89</td>
<td>8.754239</td>
<td>8.415649</td>
<td>.3385899</td>
<td>.4686923</td>
</tr>
</tbody>
</table>

Test: Ho: difference in coefficients not systematic

Test: Ho: difference in coefficients not systematic

chi2( 4) = (b-B)’[(V_b-V_B)^(-1)](b-B) = 2.86
Prob>chi2 = 0.5810

The Hausman test indicates that the RE estimators are consistent and efficient. Thus, the test suggests to use the Random Effect model. If we want to apply the Hausman test on grant, then S

W = (30.49 - 30.55) / (2.937^2 - 2.907^2)^(1/2) = -0.06/0.623 = -0.096

End of Example 12-2
**Lecture 13: Attrition in Panel Data**

**Figure 1.** Attrition

**Attrition in Panel Data**

*Motivation*

In panel data, it is often the case where observations (countries, states, households, individuals, etc.) drop out of sample after \( t = 2 \) (Attrition). If the attrition happens randomly, then attrition does not create any serious problems. We would simply have a smaller number of observations. However, if the attrition takes place systematically, then the attrition may create a sample selection bias.

Suppose that we are interested in estimating the first difference model to eliminate the fixed effects:

\[
\Delta y_t = \Delta x_t \beta + \Delta u_t \tag{1}
\]

But suppose that samples identified in the circle in Figure 1 are not available at \( t = 2 \). Because the samples lost due to attrition have low values in both \( \Delta y_t \) and \( \Delta x_t \), the estimated coefficient of \( \Delta x_t \) would be biased downward.

The attrition selection model is

\[
s_t = I[w_t \delta_t + v_t > 0] \tag{2}
\]
\( s_t \) is a dummy variable which is one if a sample is dropped out at time \( t \). Since we only consider absorbing states, \( s_t \) remains one if a sample drops out. \( w_t \) is a vector of variables that affect the attrition. \( w_t \) must be available for all observations.

Because the attrition problem can be considered as a sample selection problem, we can correct the sample selection by using the inverse Mills ratio, provided we can identify the selection with some identifying variables. From the selection model, we can estimate the probability of each sample to be dropped out in each time period and use the predicted probability to create the inverse Mills ratio as

\[
\hat{V}(w_t \delta_t) = \phi(w_t \delta_t) / \Phi(w_t \delta_t)
\]

Thus we can estimate

\[
\Delta y_{it} = \Delta x_{it} \beta + \rho_2 d_2 \hat{\lambda}_{i2} + \cdots + \rho_T d_T \hat{\lambda}_{iT} + \text{error}_t
\]

where \( d_2, \ldots, d_T \) are time dummies.

**Inverse Probability Weighting (IPW) Approach**

Another method is based on inverse probability weighting (IPW). This method relies on an assumption that the observed information in the first period, \( z_{i1} \), is a strong predictor of selection in each time period (selection on observables). Notice that the information in the first period is observable for all samples. So, we have

\[
P(s_{it} = 1 | y_{it}, x_{it}, z_{it}) = P(s_{it} = 1 | z_{it}) \quad t = 2, \ldots, T
\]

Inverse Probability Weighting Approach

Step 1  Estimate the selection model (2) and obtain predicted probability,

Step 2  Use \( 1 / \hat{p}_{it} \) as the weight in the regression models.

For the observations after \( t = 3 \), the weight should be calculated by multiplying the probabilities: \( 1 / \hat{p}_{i2} (1 / \hat{p}_{i3}) \cdots 1 / \hat{p}_{iT} \).

Table 1 shows an example of attrition in panel data. Of the 1,500 households in the original data, 1,422 households (94.8 percent) were re-interviewed in 2000, and of those, 1,266 households (84.4 percent) were re-interviewed in 2002.

In Table 1, households that were interviewed in the first round, not in the second round, but again re-interviewed in the third round are not included. All the samples that were not interviewed in the second round are not included in Table 1. This is called “attrition is in an absorbing state.” In this lecture note, we only discuss attrition in an absorbing state.

Table 1. Sample Households and Working-Age Adult Mortality in Rural Kenya

<table>
<thead>
<tr>
<th>Province</th>
<th>Original sample households in 1997</th>
<th>Percentage of households re-interviewed among the original (1997) households</th>
<th>Percentage of households incurring working-age mortality</th>
<th>HIV prevalence at urban sentinel sites in 1990-94a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
<td>(C)</td>
<td>(D)</td>
</tr>
<tr>
<td>Coastal</td>
<td>91</td>
<td>(96.7%)</td>
<td>(78.0%)</td>
<td>6.8</td>
</tr>
<tr>
<td>Eastern</td>
<td>242</td>
<td>(96.3%)</td>
<td>(88.8%)</td>
<td>4.3</td>
</tr>
<tr>
<td>Nyunza</td>
<td>280</td>
<td>(93.6%)</td>
<td>(87.5%)</td>
<td>12.2</td>
</tr>
<tr>
<td>Western</td>
<td>303</td>
<td>(95.7%)</td>
<td>(89.8%)</td>
<td>4.8</td>
</tr>
<tr>
<td>Central</td>
<td>181</td>
<td>(96.1%)</td>
<td>(90.6%)</td>
<td>2.9</td>
</tr>
<tr>
<td>Rift Valley</td>
<td>403</td>
<td>(93.1%)</td>
<td>(74.2%)</td>
<td>0.4</td>
</tr>
<tr>
<td>Total</td>
<td>1,500</td>
<td>1,422</td>
<td>1,266</td>
<td>5.8</td>
</tr>
</tbody>
</table>

Districts in Nyunza Province

<table>
<thead>
<tr>
<th>Districts in Nyunza Province</th>
<th>Percentage of households in 1997</th>
<th>Percentage of households incurring working-age mortality</th>
<th>HIV prevalence at urban sentinel sites in 1990-94a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kisumu/Siaya</td>
<td>188</td>
<td>16.9</td>
<td>21.6</td>
</tr>
<tr>
<td>Kisii</td>
<td>92</td>
<td>2.4</td>
<td>3.6</td>
</tr>
</tbody>
</table>


Note: (a) Working-age is defined as 15-49 for women and 15-54 for men. (b) The average percentage of pregnant women who visited the urban sentinel-surveillance sites and tested HIV positive in 1990-1994. Data are taken from 11 urban sentinel-surveillance sites (NASCOP, 2001).
Table 4. Household-level re-interview model (Probit\textsuperscript{a})

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
<td>(C)</td>
</tr>
<tr>
<td><strong>Lagged HIV Prevalence Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio of HIV+ Pregnant Women in 1990-94 / 1992-96</td>
<td>-0.071</td>
<td>0.002</td>
<td>-0.529</td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
<td>(0.02)</td>
<td>(2.75)**</td>
</tr>
<tr>
<td><strong>Household Characteristics in 1997</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male: Some Primary School (=1)</td>
<td>0.008</td>
<td>-0.020</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.62)</td>
<td>(0.84)</td>
</tr>
<tr>
<td>Male: Primary Finished (=1)</td>
<td>-0.019</td>
<td>-0.024</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(0.93)</td>
<td>(1.03)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>Female: Some Primary School (=1)</td>
<td>0.007</td>
<td>-0.009</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
<td>(0.52)</td>
<td>(1.16)</td>
</tr>
<tr>
<td>Female: Primary Finished (=1)</td>
<td>0.002</td>
<td>-0.012</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.79)</td>
<td>(1.03)</td>
</tr>
<tr>
<td>Female Headed in 1997 (=1)</td>
<td>-0.035</td>
<td>-0.038</td>
<td>-0.032</td>
</tr>
<tr>
<td></td>
<td>(1.15)</td>
<td>(1.32)</td>
<td>(0.68)</td>
</tr>
<tr>
<td>Polygamous Household in 1997 (=1)</td>
<td>0.060</td>
<td>0.021</td>
<td>N.A.\textsuperscript{b}</td>
</tr>
<tr>
<td></td>
<td>(3.73)**</td>
<td>(1.42)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of Male Adults in 1997</strong></td>
<td>0.007</td>
<td>0.002</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
<td>(0.39)</td>
<td>(1.84)</td>
</tr>
<tr>
<td><strong>Number of Female Adults in 1997</strong></td>
<td>-0.005</td>
<td>-0.002</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(1.15)</td>
<td>(0.39)</td>
<td>(0.69)</td>
</tr>
<tr>
<td>Land Tenure in 1997 (=1)</td>
<td>-0.001</td>
<td>-0.005</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.80)</td>
<td>(0.75)</td>
</tr>
<tr>
<td><strong>Ln (Landholding Size in Acres in 1997)</strong></td>
<td>0.005</td>
<td>-0.007</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(1.15)</td>
<td>(2.26)*</td>
</tr>
<tr>
<td><strong>Ln (Asset Value in 1997)</strong></td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(1.15)</td>
<td>(1.63)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>Distance to Bus Stop in 1997 (km)</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(0.78)</td>
<td>(1.33)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Distance to Piped Water in 1997 (km)</td>
<td>0.000</td>
<td>-0.000</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.77)</td>
<td>(1.30)</td>
</tr>
<tr>
<td>Year 2002 (=1)</td>
<td>-0.012</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>(0.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Enumeration Team Dummies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team 2 in 2000</td>
<td>0.028</td>
<td>0.233</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
<td>(10.09)**</td>
<td></td>
</tr>
<tr>
<td>Team 3 in 2000</td>
<td>-0.043</td>
<td>-0.036</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>(1.76)</td>
<td>(1.86)</td>
<td></td>
</tr>
<tr>
<td>Team 4 in 2000</td>
<td>-0.062</td>
<td>-0.039</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>(2.20)*</td>
<td>(1.35)</td>
<td></td>
</tr>
<tr>
<td>Team 2 in 2002</td>
<td>-0.133</td>
<td>N.A.</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>(4.47)**</td>
<td></td>
<td>(3.34)**</td>
</tr>
<tr>
<td>Team 3 in 2002</td>
<td>-0.028</td>
<td>N.A.</td>
<td>-0.117</td>
</tr>
<tr>
<td></td>
<td>(0.83)</td>
<td></td>
<td>(2.79)**</td>
</tr>
<tr>
<td>Province Dummies Included\textsuperscript{c}</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td><strong>Joint Test for Team Effects (\chi^2)</strong></td>
<td>39.1 [0.00]**</td>
<td>111.5 [0.00]**</td>
<td>24.6 [0.00]**</td>
</tr>
<tr>
<td><strong>Joint Test for HH Characteristics (\chi^2)</strong></td>
<td>17.2 [0.14]</td>
<td>14.5 [0.34]</td>
<td>21.2 [0.04]**</td>
</tr>
<tr>
<td>E[y]</td>
<td>0.920</td>
<td>0.948</td>
<td>0.890</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are absolute z-scores, calculated using heteroskedasticity robust standard errors clustered for households. ** indicates 1 percent significance level; * indicates 5 percent significance level. (a) Estimated coefficients are marginal changes in probability. (b) Because all the polygamous households were re-interviewed in 2002, the dummy is excluded in column C. (c) Five province dummies are included but not reported in this table.
### Table 5. Factors Associated with Households Afflicted by Working-age Adult Mortality (Probit*)

<table>
<thead>
<tr>
<th></th>
<th>Pooled Model (A)</th>
<th>Adult mortality in 1997-2000 Corrected for attrition (B)</th>
<th>Adult mortality in 2000-2002 Corrected for attrition (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Logged HIV Prevalence rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio of HIV+ Pregnant Women</td>
<td>0.260</td>
<td>0.261</td>
<td>0.316</td>
</tr>
<tr>
<td></td>
<td>(3.61)**</td>
<td>(3.74)**</td>
<td>(4.07)**</td>
</tr>
<tr>
<td><strong>Household Level Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male: Some Primary (=1)</td>
<td>-0.010</td>
<td>-0.010</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.78)</td>
<td>(0.71)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>Male: Primary Finished (=1)</td>
<td>-0.011</td>
<td>-0.011</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(0.74)</td>
<td>(0.81)</td>
</tr>
<tr>
<td>Female: Some Primary (=1)</td>
<td>0.004</td>
<td>0.005</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.37)</td>
<td>(0.44)</td>
<td>(0.53)</td>
</tr>
<tr>
<td>Female: Primary Finished (=1)</td>
<td>-0.005</td>
<td>-0.004</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td>(0.41)</td>
<td>(0.35)</td>
</tr>
<tr>
<td>Female Headed in 1997 (=1)</td>
<td>0.002</td>
<td>0.001</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.08)</td>
<td>(0.61)</td>
</tr>
<tr>
<td>Polygamous HH in 1997 (=1)</td>
<td>-0.014</td>
<td>-0.015</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>(1.53)</td>
<td>(1.60)</td>
<td>(1.83)</td>
</tr>
<tr>
<td># of Male Adults in 1997</td>
<td>0.005</td>
<td>0.005</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(1.95)</td>
<td>(1.99)**</td>
<td>(1.92)</td>
</tr>
<tr>
<td># of Female Adults in 1997</td>
<td>0.002</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.73)</td>
<td>(0.73)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>Land Tenure in 1997 (=1)</td>
<td>-0.010</td>
<td>-0.010</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(2.12)**</td>
<td>(2.05)**</td>
<td>(1.17)</td>
</tr>
<tr>
<td>ln (Landholding in 1997)</td>
<td>0.005</td>
<td>0.006</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td>(1.29)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>ln (Asset Value in 1997)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.52)</td>
<td>(2.36)**</td>
</tr>
<tr>
<td>Distance to Bus Stop in 1997</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(0.75)</td>
<td>(1.28)</td>
</tr>
<tr>
<td>Distance to Piped Water in 1997</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(1.07)</td>
<td>(0.97)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Year 2002 (=1)</td>
<td>-0.033</td>
<td>-0.034</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>(4.38)**</td>
<td>(4.50)**</td>
<td>(1.37)</td>
</tr>
<tr>
<td><strong>Province Dummies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nyasa province</td>
<td>0.026</td>
<td>0.029</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(1.37)</td>
<td>(1.27)</td>
<td>(0.93)</td>
</tr>
<tr>
<td>Coastal province</td>
<td>0.018</td>
<td>0.019</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(1.30)</td>
<td>(1.35)</td>
<td>(1.72)</td>
</tr>
<tr>
<td>Western province</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Central province</td>
<td>0.004</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.27)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Rift Valley province</td>
<td>-0.008</td>
<td>-0.009</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(0.84)</td>
<td>(0.94)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are absolute z-scores, calculated using heteroskedasticity robust standard errors clustered for households. ** indicates 1 percent significance level; * indicates 5 percent significance level.

(a) Estimated coefficients are marginal changes in probability.
Lecture 14: Propensity Score Matching

Propensity Score Matching Method

When we studied program evaluations previously, we define the probability of being in the treatment group as:

\[ p(x) = P(w = 1 | x) . \]

This is called the propensity score. For simple analyses, the propensity score can be used to obtain the average treatment effect by calculating

\[ ATE = N^{-1} \sum_{i=1}^{N} (w_i - \hat{p}(x_i))y_i / [\hat{p}(x_i)(1 - \hat{p}(x_i))] \]

\[ = N^{-1} \left[ \sum_{Treatment} y_i / \hat{p}(x_i) - \sum_{Control} y_i / (1 - \hat{p}(x_i)) \right] \]

Or we can simply estimate a regression model:

\[ y_i = \delta w_i + \beta_0 + \beta \hat{p}(x_i) + u_i \]

But it is not clear if this is any better than estimating a model

\[ y_i = \delta w_i + \beta_0 + x_i \beta + u_i . \]

where \( x_i \) is a 1-by-k vector.

A more popular approach is to use the propensity score to obtain matching estimators. The idea is to match two observations that have the same or a very similar propensity score and measure the difference in the outcome:

\[ E[y_1 | w = 1, p(x)] - E[y_0 | w = 0, p(x)] = E[y_i - y_0 | p(x)] \]

as defined before, \( y_1 \) is the outcome when the sample has actually received benefits from a program and \( y_0 \) is the outcome when the sample has not received any benefits from the program. \( w \) indicates who are in the treatment group. For the sake of counterfactual discussions, the treatment group may not receive benefits from the program, i.e., \( E[y_0 | w = 1, p(x)] \).

The main assumption that one indicator, the propensity score, can control for everything except the participation in the program. For instance, suppose that a very critical component is missing in the propensity score estimation, then the propensity score with missing component would match two observations that should not be matched.
See examples in the class.
Lecture 15: Simultaneous Equations Models

Simultaneous Bias

Consider a two-equation structural model

\[ y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1 \quad (1) \]
\[ y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2 \quad (2) \]

These two-equations are examples of structural equations because both equations contain an endogenous variable. Or we call a set of these equations as a simultaneous equations model (SEM) because these two equations simultaneously determine both \( y_1 \) and \( y_2 \). \( y_1 \) and \( y_2 \) are called endogenous variables, and \( z_1 \) and \( z_2 \) are called exogenous variables. Finally \( u_1 \) and \( u_2 \) are called structural errors.

Because we have two endogenous variables and two equations, we can solve for each endogenous variable in terms of exogenous variables and structural errors:

\[ y_2 = \alpha_2 (\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2 \]
\[ y_2 = \alpha_2 \alpha_1 y_2 + \alpha_2 \beta_1 z_1 + \alpha_2 u_1 + \beta_2 z_2 + u_2 \]
\[ (1-\alpha_2 \alpha_1)y_2 = \alpha_2 \beta_1 z_1 + \beta_2 z_2 + \alpha_2 u_1 + u_2 \]

Assuming that \( \alpha_2 \alpha_1 \neq 1 \), we have

\[ y_2 = (\alpha_2 \beta_1 / (1-\alpha_2 \alpha_1)) z_1 + (\beta_2 / (1-\alpha_2 \alpha_1)) z_2 + (\alpha_2 u_1 + u_2) / (1-\alpha_2 \alpha_1) \]
\[ y_2 = \delta_1 z_1 + \delta_2 z_2 + v \quad (3) \]

Thus, in (3), \( y_2 \) is expressed in terms of exogenous variables. This is called a reduced form equation, which has only exogenous variables and error terms, for \( y_2 \). From (13-3), we can show

\[ \text{Cov} (y_2, u_1) = \text{Cov} (\delta_1 z_1 + \delta_2 z_2 + (\alpha_2 u_1 + u_2) / (1-\alpha_2 \alpha_1), u_1) \]
\[ = E(\frac{\alpha_2 u_1^2}{1-\alpha_2 \alpha_1}) \]
\[ = \frac{\alpha_2}{1-\alpha_2 \alpha_1} \sigma_1^2 \]

This is not zero if \( \alpha_2 \alpha_1 \neq 1 \) and \( \alpha_2 \neq 0 \). Thus, in equation (13-1), \( y_2 \) is correlated with the error term, \( u_1 \), and the estimated coefficient of \( y_2 \) will be biased. This is called a simultaneous bias. In equation (13-2), the estimated coefficient of \( y_1 \) will be biased because of the simultaneous bias.
Identification

Consider a two-equation model:

Supply curve: \[ q = \alpha_1 p + \beta_1 z + u_1 \]

Demand curve: \[ q = \alpha_2 p + u_2 \]

In these simple equations, we have one exogenous variable in the supply curve, \( z \). The supply curve shifts up and down as \( z \) changes. Thus, we may call it as a supply-curve shifter.

The demand curve does not have any demand-curve shifters, thus the demand curve does not move. Therefore, when we find changes in price and quantity, we can assume that the changes are caused by \( z \) and that the observed points (A, B, C) are on the demand curve. Thus, by connecting these points, we can identify the demand curve.

In the previous lecture note, we have studied that we need at least one instrument to identify one endogenous variable. In the demand equation, the endogenous variable is the price, \( p \), and the instrument is the supply-curve shifter, \( z \). Thus, the price in the demand curve equation is identified.

In contrast, we can not identify the supply curve because we do not have any instrument to identify price in the supply equation.

Suppose now that we have these demand and supply curves. The only difference is that \( z \) is included in the both equations:

Supply curve: \[ q = \alpha_1 p + \beta_1 z + u_1 \]

Demand curve: \[ q = \alpha_2 p + \beta_2 z + u_2 \]
In this case, we can not identify either curve, because $z$ is included in the both equations. Thus, $z$ does not satisfy the exclusion restrictions. This is illustrated in Panel 2. We observe changes in price and quantity from A to C. But, because both curves can shift from A to C, we can not trace curves. Both curves in Panel 2 could be steeper or flatter.

**Reduced Form Equations in General Cases**

In general, we can have more than two simultaneous equations (more than two endogenous variables) at once. Suppose there are 3-endogenous variables:

\[
y_1 = f_1(y_2, y_3, z_1, z_2, z_3, z_4)
\]
\[
y_2 = f_2(y_1, y_3, z_1, z_2, z_5, z_6, z_7)
\]
\[
y_3 = f_3(y_1, y_2, z_1, z_2, z_8, z_9, z_{10})
\]

The reduced form equation for each endogenous variable will contain all of the exogenous variables:

\[
y_1 = f_1(z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10})
\]
\[
y_2 = f_2(z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10})
\]
\[
y_3 = f_3(z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10})
\]

From these reduced form equations, we can identify endogenous variables in the structural from equations.

**Example 15-1: Inflation and Openness by using Openness.dta**

See Example 16.4 and 16.6 in Wooldridge:

**The Reduced Form Equation for “Open”**

```
. reg open lpcinc lland
```

```
Source |       SS       df       MS              Number of obs =     114
-------------+------------------------------           F(  2,   111) =   45.17
Model |  28606.1936     2  14303.0968           Prob > F      =  0.0000
Residual |  35151.7966   111  316.682852           R-squared     =  0.4487
-------------+------------------------------           Adj R-squared =  0.4387
Total |  63757.9902   113  564.230002           Root MSE      =  17.796

------------------------------------------------------------------------------
open |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
lpcinc |   .5464812    1.49324     0.37   0.715    -2.412473    3.505435
lland |  -7.567103   .8142162    -9.29   0.000    -9.180527   -5.953679
_cons |   117.0845    15.8483     7.39   0.000     85.68006     148.489
------------------------------------------------------------------------------
```

**The OLS for inflation**

```
. reg inf open lpcinc
```

```
Source |       SS       df       MS              Number of obs =     114
-------------+------------------------------           F(  2,   111) =    2.63
Model |  2945.92812     2  1472.96406           Prob > F      =  0.0764
Residual |  29912.0621   111  270.378745           R-squared     =  0.4889
-------------+------------------------------           Adj R-squared =  0.4784
Total |  32857.9902   113  289.853471           Root MSE      =  16.447

------------------------------------------------------------------------------
inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
open |   3.567103   .6142162    5.85   0.000     2.350679    4.783527
lpcinc |  -2.546812   1.49324    -1.72   0.089    -5.487320    .393696
_cons |  -97.0845   15.84832   -6.18   0.000    -128.4890   -65.67998
------------------------------------------------------------------------------
```

46
The estimated coefficient changes from -0.215 to -0.337. There is no over-identification problem because the number of endogenous variable and the instrument is both one.
Lecture 16: Systems of Equations (SUR, GMM, and 3SLS)

Seemingly Unrelated Regression (SUR) Model

Consider a set of $G$ linear equations:

\[
\begin{align*}
    y_1 &= X_1 \beta_1 + \varepsilon_1 \\
    y_2 &= X_2 \beta_2 + \varepsilon_2 \\
    &\vdots \\
    y_G &= X_G \beta_G + \varepsilon_G
\end{align*}
\]

where $X_g$ is a $N \times k_g$ matrix ($k_g$ may differ across equations but often not in applications). $(g = 1, 2, \ldots, G)$ $y_g$, $\beta_g$, $\varepsilon_2$ are all $N \times 1$ vectors. $X_g$ contains only exogenous variables.

The system (1) is called “Seemingly Unrelated” because each equation seems unrelated with other equations. However, it is often assumed that error terms are correlated across equations. For example, many applications estimated a set of demand functions for families. In this case, error terms are most likely to be correlated across equations.

The system (1) can be written as:

\[
\begin{bmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_G
\end{bmatrix} = 
\begin{bmatrix}
    X_1 & 0 & \ldots & 0 \\
    0 & X_2 & \ldots & 0 \\
    0 & 0 & \ldots & X_G
\end{bmatrix} 
\begin{bmatrix}
    \beta_1 \\
    \beta_2 \\
    \vdots \\
    \beta_G
\end{bmatrix} + 
\begin{bmatrix}
    \varepsilon_1 \\
    \varepsilon_2 \\
    \vdots \\
    \varepsilon_G
\end{bmatrix}
\]

Where $K$ is a sum of $k_g$ for $g = 1, \ldots, G$.

The error term has the following properties:

\[
E[\varepsilon | X_1, X_2, \ldots, X_G] = 0 \\
E[\varepsilon \varepsilon'] = \Omega
\]

$\varepsilon$ is a $NG \times 1$ vector, and $E[\varepsilon \varepsilon'] = \Omega$ is a $NG \times NG$ matrix. We further assume that

\[
E[\varepsilon_t \varepsilon_s | X_1, X_2, \ldots, X_G] = \sigma_{ij} \\
\quad \quad \quad \text{if } t = s \text{ and } 0 \text{ otherwise.}
\]

For instance, this indicates that the error term for the demand for pork (i-th equation) at time t, $\varepsilon_t$, is correlated with the error term for the demand for chicken (j-th equation) at time t, $\varepsilon_t$. But the error terms are uncorrelated across time.
For one equation, the variance-covariance matrix of $\varepsilon_i$ is

$$E[\varepsilon_i | X_1, X_2, \ldots, X_G] = \sigma_{ii} I_N$$

The variance-covariance matrix of $\varepsilon_i$ and $\varepsilon_j$ is

$$E[\varepsilon_i \varepsilon_j' | X_1, X_2, \ldots, X_G] = \sigma_{ij} I_N$$

Thus, for all the observations

$$E[\varepsilon \varepsilon' | X_1, X_2, \ldots, X_G] = \Omega = \begin{bmatrix} \sigma_{11} I & \sigma_{12} I & \ldots & \sigma_{1G} I \\ \sigma_{21} I & \sigma_{22} I & \ldots & \sigma_{2G} I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1} I & \sigma_{G2} I & \ldots & \sigma_{GG} I \end{bmatrix}$$  \hspace{1cm} (3)$$

(What is the order of this square matrix?)

From (2), we have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ 0 & 0 & \ldots & X_G \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_G \end{bmatrix} = X\beta + \varepsilon$$ \hspace{1cm} (4)$$

Thus, the error term of the SUR model in (4) has a variance-covariance of (3). Thus, a GLS estimation can be applied to the SUR model. Define

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1G} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1} & \sigma_{G2} & \ldots & \sigma_{GG} \end{bmatrix}$$

so, in (3),

$$\Omega_{GN} = \Sigma_G \otimes I_N \hspace{0.5cm} \text{and} \hspace{0.5cm} \Omega_{GN}^{-1} = \Sigma_G^{-1} \otimes I_N$$

Denoting the $ij$th element of $\Sigma_G^{-1}$ by $\sigma_{ij}^0$, thus
\[ \Sigma_G^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1G} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1} & \sigma_{G2} & \ldots & \sigma_{GG} \end{bmatrix}. \]

The GLS estimator is
\[ \hat{\beta}_{GLS-SUR} = [X\Omega^{-1}X']^{-1}X\Omega^{-1}y = [X'(\Sigma_G^{-1} \otimes I_N)X]^{-1}X'(\Sigma_G^{-1} \otimes I_N)y. \]

See Greene pp342 for more details.

Some important results on the SUR Model:

**SUR1** If the equations are actually unrelated, then there is no payoff to GLS estimation.

**SUR2** If the same set of independent variables are used for each equation, \( X_1 = X_2 = \ldots = X_G \), then GLS and OLS are identical.

**SUR3** If the independent variables on one block of equations are a subset of those in another, then GLS brings on efficiency gain over OLS in estimation of the smaller set of equations.

We will examine these results in examples.

**Testing SUR2**

Define \( y = (y'_1, y'_2, \ldots, y'_G), \beta = (\beta'_1, \beta'_2, \ldots, \beta'_G), u = (u'_1, u'_2, \ldots, u'_G), \text{and } X = \text{diag}(X_1, X_2, \ldots, X_G). \) Then, we can write (12-4) as
\[ y = X\beta + u \]

Then the GLS estimator is
\[ \hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X\Omega^{-1}y = (X'(\Sigma_G^{-1} \otimes I_N)X)^{-1}X'(\Sigma_G^{-1} \otimes I_N)y \] (5)

Suppose all equations have the same set of independent variables, \( X_1 = X_2 = \ldots = X_G = X \). Then, we can write \( X = I_G \otimes X \) and \( X' = I_G \otimes X' \). Thus,
\[ \hat{\beta}_{GLS} = [(I_G \otimes X')(\Sigma_G^{-1} \otimes I_N)(I_G \otimes X)]^{-1}(I_G \otimes X')(\Sigma_G^{-1} \otimes I_N)y \]
\[ = [(\Sigma_G^{-1} \otimes X')I_G \otimes X')^{-1}(\Sigma_G^{-1} \otimes X')y \]
\[ = [(\Sigma_G^{-1} \otimes X')^{-1}(\Sigma_G^{-1} \otimes X')y \]
\[ = [(\Sigma_G^{-1} \otimes (XX'))^{-1}(\Sigma_G^{-1} \otimes X')y \]
\[
\hat{\beta}_G = \left[(I_G \otimes (XX)^{-1}X')y\right]
\]
\[
= (XX')^{-1}X'y
\]
\[
\begin{bmatrix}
(XX)^{-1}X' & 0 & 0 \\
0 & (XX)^{-1}X' & 0 \\
0 & 0 & (XX)^{-1}X'
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2 \\
y_G
\end{bmatrix}
\]

(We have used a rule on Kronecker Product that \( (A \otimes B)(C \otimes D) = (AC \otimes BD) \) when AC and BD can be defined. See Greene A.5.5. in Appendix A.) Thus, the GLS estimators are identical to the OLS estimators.

**Example 14-1: Wage and Fringe Benefits (FRINGE.dta)**

In this example, we estimate a two-equation system for hourly wage and hourly benefits. There are 616 workers in the data set. The FGLS results for a SUR model are presented below.

**FGL Estimation for a SUR Model**

```
. sureg (hrearn educ exper exper2 union) (hrbens educ exper exper2 union), small dfk
```

```
Seemingly unrelated regression

<table>
<thead>
<tr>
<th>Equation</th>
<th>Obs</th>
<th>Parms</th>
<th>RMSE</th>
<th>&quot;R-sq&quot;</th>
<th>F-Stat</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrearn</td>
<td>616</td>
<td>4</td>
<td>4.46</td>
<td>0.155</td>
<td>28.06</td>
<td>0.000</td>
</tr>
<tr>
<td>hrbens</td>
<td>616</td>
<td>4</td>
<td>0.56</td>
<td>0.297</td>
<td>64.56</td>
<td>0.000</td>
</tr>
</tbody>
</table>

|           | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|-----------|-------|-----------|------|-----|---------------------|
| hrearn    |       |           |      |     |                     |
| educ      | .5069603 | .0685878 | 7.39 | 0.000 | .3723975 - .6415231 |
| exper     | -.0131478 | .05253 | -0.25 | 0.802 | -.1162067 .0899112 |
| exper2    | .0030274 | .0011343 | 2.67 | 0.008 | .0008021 .0052527 |
| union     | 1.098166 | .3965588 | 2.77 | 0.006 | .3925394 1.876177 |
| _cons     | -1.728478 | 1.009515 | -1.71 | 0.087 | -.3709053 .2520974 |

| hrbens    |       |           |      |     |                     |
| educ      | .0790536 | .0068578 | 7.39 | 0.000 | .0621783 .0959288 |
| exper     | .0485195 | .0065877 | 7.37 | 0.000 | .0355951 .061444 |
| exper2    | -.000781 | .0001422 | -5.49 | 0.000 | -.0010601 -.0005019 |
| union     | .4901085 | .0497317 | 9.86 | 0.000 | .3925394 .5876775 |
| _cons     | -.7495522 | .1266015 | -5.92 | 0.000 | -.9979326 -.5011719 |
```

51
Here, we examine a correlation between residuals from two equations. The test indicates that the residuals from the two equations are correlated with a correlation of 0.3377. The Breush-Pagan test suggests that the two are correlated significantly. It is not surprising to find a strong correlation because the same unobserved variables, such as ability, across two equations should be correlated.

Correlation matrix of residuals:

<table>
<thead>
<tr>
<th></th>
<th>hrearn</th>
<th>hrbens</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrearn</td>
<td>1.0000</td>
<td>0.3377</td>
</tr>
<tr>
<td>hrbens</td>
<td>0.3377</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Breusch-Pagan test of independence: chi2(1) = 70.259, Pr = 0.0000

Now, we estimate the two equations separately by OLS. The results below are exactly the same as in the FGLS-SUR model above. This confirms SUR2, GLS and OLS are identical if the same set of independent variables is used in a SUR model. The model above is not useless because we could at least test if the residuals from equations are correlated.

**OLS on Hourly Wage:**

```
. reg hrearn educ exper exper2 union
```

```
Source |       SS       df       MS              Number of obs =     616
--------+------------------------------           F(  4,   611) =   28.06
Model |  2232.63395     4  558.158487           Prob > F      =  0.0000
Residual |  12155.2624   611  19.8940465           R-squared     =  0.1552
--------+------------------------------           Adj R-squared =  0.1496
Total |  14387.8963   615  23.3949534           Root MSE      =  4.4603

------------------------------------------------------------------------------
|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+---------------------------------------------------------------------
hrearn | .5069603   .0685878     7.39   0.000      .372264    .6416567
educ | .0790536   .0086015     9.19   0.000     .0621615    .0959456
exper | .0485195   .0065877     7.37   0.000     .0355823    .0614568
exper2 | -.000781   .0001422    -5.49   0.000    -.0010603   -.0005016
union |  1.098166   .3965588     2.77   0.006     .3193819    1.876949
_cons | -.749552   1.009515    -0.78   0.436    -.3520159    .2529093
------------------------------------------------------------------------------
```

**OLS on Hourly Benefits:**

```
. reg hrbens educ exper exper2 union
```

```
Source |       SS       df       MS              Number of obs =     616
--------+------------------------------           F(  4,   611) =   64.56
Model |  80.8036641     4  20.200916           Prob > F      =  0.0000
Residual |   191.16841   611  .312877922           R-squared     =  0.2971
--------+------------------------------           Adj R-squared =  0.2925
Total |  271.972074   615  .442231015           Root MSE      =  .55935

------------------------------------------------------------------------------
|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+---------------------------------------------------------------------
hrbens | .0790536   .0086015     9.19   0.000     .0621615    .0959456
educ | .0485195   .0065877     7.37   0.000     .0355823    .0614568
exper | -.000781   .0001422    -5.49   0.000    -.0010603   -.0005016
exper2 |  .4901085   .0497317     9.86   0.000     .3924426    .5877743
union |  .4901085   .0497317     9.86   0.000     .3924426    .5877743
_cons | -.749552   1.009515    -0.78   0.436    -.3520159    .2529093
------------------------------------------------------------------------------
```

To test SUR 3, we add two independent variables, married and male, to the first equation but not in the second. Thus, the independent variables in the second equation are a subset
of the set of independent variables in the first equation. The results below confirms
SUR3, the results below on the second equation are identical to the OLS results.

```
.sureg (hrearn educ exper exper2 union married male) ( hrbens educ exper exper2 union),
small dfk corr
```

Seemingly unrelated regression

<table>
<thead>
<tr>
<th>Equation</th>
<th>Obs</th>
<th>Parms</th>
<th>RMSE</th>
<th>&quot;R-sq&quot;</th>
<th>F-Stat</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrearn</td>
<td>616</td>
<td>6</td>
<td>4.378737</td>
<td>0.1884</td>
<td>21.87</td>
<td>0.0000</td>
</tr>
<tr>
<td>hrbens</td>
<td>616</td>
<td>4</td>
<td>0.5593549</td>
<td>0.2971</td>
<td>64.56</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

|                 | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-----------------|-------|-----------|-------|------|----------------------|
| hrearn          | educ  |.4866347   |.0674135 |7.22 |0.000     |.3543755    |.6188939|
|                 | exper |-.0385265  |.0525284 |-.73 |0.463     |-.1415826    |.0645295|
|                 | exper2| .0032235   |.0011205 |2.88 |0.004     |.0010252    |.0054218|
|                 | union | .8654272   |.3933037 |2.20 |0.028     |.5938007    |1.137054|
|                 | married| .4591233 | .3970173 |1.16 |0.248     |-.3197891    |1.238036|
|                 | male  |1.212324 | .3788968 |3.20 |0.001     |.5214057    |1.895685|
|                 | _cons |-.117965 | .9962684 |-2.13|0.034     |-.6075577    |-.7056757|
| hrbens          | educ  |.0790536   |.0086015 |9.19 |0.000     |.0621782    |.0959289|
|                 | exper |.0485195   |.0065877 |7.37 |0.000     |.0355951    |.061444|
|                 | exper2| -.000781   |.0001422 |-.54 |0.593     |-.0010601    |-.0005019|
|                 | union | .4901085   |.0497317 |9.86 |0.000     |.3925393    |.5876776|
|                 | _cons |-.7495522  |.1266015 |5.92 |0.000     |-.997933    |-.5011715|

End of Example 14-1

Generalized Method of Moments (GMM) Estimation

Suppose we have a set of equations and for observation $i$, we have

$$y_i = X_i \beta + \epsilon_i$$

where $y_i$ is a G-by-1 vector (2-by-1 if there are two equations), $X_i$ is a G-by-K matrix, and

$\epsilon_i$ is a G-by-1 error vector.

**Assumption SIV 1:** $E(Z_i' u_i) = 0$, where $Z_i$ is a G-by-L matrix of observable
instrumental variables.

**Assumption SIV 2:** $\text{rank } (Z_i' X_i) = K$.

SIV stands for “system instrumental variables.” We also assume $E(u_i)=0$.
The order condition, $L \geq K$, should be satisfied when the rank condition (SIV 2) is met.

We have a set of equations with potentially endogenous variables:

$$y_1 = X_1 \beta_1 + \epsilon_1$$
$$y_2 = X_2 \beta_2 + \epsilon_2$$
$$\cdots$$
$$y_G = X_G \beta_G + \epsilon_G$$

where we can also write
\[ Y = X \beta + \epsilon \]

Where \( Y \) is a NG-by-1 vector, \( X \) is a NG-by-K matrix, and \( \epsilon \) is a NG-by-1 vector. Under Assumptions SIV 1 and SIV 2, \( \hat{\beta}_{SIV} \) is the unique K-by-1 vector solving the linear set population moment conditions:

\[
E[Z' (Y - X \hat{\beta}_{SIV})] = 0
\]  

(7)

Because sample averages are consistent estimators of population moments, the analogy applied to condition (7) suggests choosing \( \hat{\beta}_{SIV} \) to solve

\[
N^{-1} \sum_{i=1}^{N} Z_i'(Y_i - X_i \hat{\beta}_{SIV}) = 0
\]  

(8)

\( Z \) is a NG-by-L matrix. When \( L = K \),

\[
\hat{\beta}_{SIV} = (Z'X)^{-1}Z'Y
\]

This is called the **system IV (SIV) estimator**.

When \( L > K \) (so we have more columns in the IV matrix \( Z \) than we need for identification), choosing \( \hat{\beta}_{SIV} \) is more complicated. The equation (13-8) does not have a solution. Instead, we choose \( \hat{\beta}_{SIV} \) to make the vector in equation (13-8) as “small” as possible in the sample. One idea is to minimize this:

\[
\begin{bmatrix}
\sum_{i=1}^{N} Z_i'(Y_i - X_i \hat{\beta}_{SIV}) \\
\sum_{i=1}^{N} Z_i'(Y_i - X_i \hat{\beta}_{SIV})
\end{bmatrix}
\begin{bmatrix}
\sum_{i=1}^{N} Z_i'(Y_i - X_i \hat{\beta}_{SIV}) \\
\sum_{i=1}^{N} Z_i'(Y_i - X_i \hat{\beta}_{SIV})
\end{bmatrix}'
\]

But, a more general class of estimators is obtained by using a weighting matrix in the quadratic form. Let \( \hat{W} \) be an L-by-L symmetric, positive semidefinite matrix. The generalized method of moments (GMM) estimator that solves the problem

\[
\min_{\beta} \left[ \sum_{i=1}^{N} Z_i'(Y_i - X_i \hat{\beta}_{SIV}) \right]' \hat{W} \left[ \sum_{i=1}^{N} Z_i'(Y_i - X_i \hat{\beta}_{SIV}) \right]
\]

or

\[
\min_{\beta} \left[ Z'(Y - X\hat{\beta}_{SIV}) \right]' \hat{W} \left[ Z'(Y - X\hat{\beta}_{SIV}) \right]
\]

\[
\min_{\beta} \left[ Y'Z - \hat{\beta}'_{SIV} X'Z \right] \hat{W} \left[ Z'(Y - X\hat{\beta}_{SIV}) \right]
\]

\[
\min_{\beta} Y'Z \hat{W} Z'Y - \hat{\beta}'_{SIV} X'Z \hat{W} Z'Y - Y'Z \hat{W} Z'X \hat{\beta}_{SIV} + \hat{\beta}'_{SIV} X'Z \hat{W} Z'X \hat{\beta}_{SIV}
\]

F.O.C. is

\[
-X'Z \hat{W} Z' Y - Y'Z \hat{W} Z' X + 2 X'Z \hat{W} Z' X \hat{\beta}_{SIV} = 0
\]
\[ X'Z\hat{W}Z'X\hat{\beta}_{SIV} = X'Z\hat{W}Z'Y \]
\[ \hat{\beta}_{GMM} = (X'Z\hat{W}Z'X)^{-1}X'Z\hat{W}Z'Y \]  

(9)

**Assumption SIV 3:** \( \hat{W} \rightarrow W \) as \( N \rightarrow \infty \), where \( W \) is nonrandom, symmetric, \( L \)-by-\( L \) positive definite matrix.

Theorem (Consistency of GMM): Under assumptions SIV 1-3, the GMM estimators are consistent and asymptotically normally distributed.

### The System 2SLS Estimator

Suppose we choose
\[ \hat{W} = (Z'Z/N)^{-1} \]

which is a consistent estimator of \( [E(Z'_iZ_i)]^{-1} \). Assumption SIV 3 simply requires that \( E(Z'_iZ_i) \) exist and be nonsingular. When we plug this into equation (9), we get

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

(10)

This is called the **system 2SLS estimator**.

### The Optimal Weighting Matrix

The Procedure
1. Let \( \tilde{\beta}_{iv} \) be an initial consistent estimator of \( \beta \). In most cases this is the system 2SLS estimator.
2. Obtain the G-by-1 residual vectors
   \[ \tilde{e}_i = y_i - X_i\tilde{\beta}_{iv} \quad \text{for } i = 1, 2, \ldots, N \]
3. A generally consistent estimator of \( \Lambda \) is
   \[ \hat{\Lambda} = N^{-1}\sum_{i=1}^{N}Z'_i\tilde{e}_i\tilde{e}_i'Z_i \]
4. Choose
   \[ \hat{W} = \hat{\Lambda}^{-1} = \left( N^{-1}\sum_{i=1}^{N}Z'_i\tilde{e}_i\tilde{e}_i'Z_i \right)^{-1} \]
   and use this matrix to obtain the asymptotically optimal GMM estimator.

### The Three-Stage Least Squares (3SLS) Estimator

The 3SLS estimator is a GMM estimation that uses a particular weighting matrix. Let \( \tilde{u}_i = y_i - X_i\hat{\beta}_{3SLS} \) be the residual from an initial estimation (usually a system 2SLS), then define

\[ \Omega = N^{-1}\sum_{i=1}^{N}\tilde{u}_i\tilde{u}_i' \]
Using the same arguments as in the FGLS case, \( \hat{\Omega} \xrightarrow{p} \Omega = E(\tilde{u}_i\tilde{u}_j) \). The weighting matrix used by 3SLS is

\[
\hat{W} = \left( N^{-1} \sum_{i=1}^{N} Z_i^\prime \hat{\Omega} Z_i \right)^{-1} = \left[ Z^\prime (I_N \otimes \hat{\Omega}) Z / N \right]^{-1}
\]

Plugging this into (9), we get

\[
\hat{\beta}_{3SLS} = \left( X' Z (I_N \otimes \hat{\Omega}) Z X' \right)^{-1} X' Z (I_N \otimes \hat{\Omega}) Z Y
\]

(11)

This is different from the traditional 3SLS estimator in most textbooks, such as Greene (pp. 406 in 5e). The traditional 3SLS estimator is

\[
\hat{\beta}_{3SLS} = \left[ \hat{X}' (I_N \otimes \hat{\Omega}^{-1}) \hat{X} \right]^{-1} \hat{X}' (I_N \otimes \hat{\Omega}^{-1}) Y
\]

where \( \hat{X}_i \equiv Z_i \hat{\Gamma} \) and \( \hat{\Gamma} = (Z'Z)^{-1} Z'X \). (Note that in Greene notations are different.)

In short, the GMM definition of 3SLS is more generally valid, and it reduces to the standard definition in the traditional simulation equations setting.
Lecture 17: Single Equation GMM and STATA Programs

In Lecture Note 14, we have studied GMM estimations in terms of systems of equations. In this lecture note, we take a step back and study a single equation GMM.

Generalized Method of Moments (GMM) Estimation

Suppose we have an equation and for observation $i$, we have

$$ y_i = X_i \beta + \varepsilon_i \quad i = 1, \ldots, N. $$

where $X_i$, which is a 1-by-$K$ vector, may contain some endogenous variables that potentially correlated with the error term. Further, suppose that we have instrumental variables that are correlated with the endogenous variables but not with the error term. We denote a matrix with instrumental variables and exogenous independent variables as $Z_i$, which is 1-by-$L$ vector.

Thus, we can also write

$$ Y = X \beta + \varepsilon, \quad E(Z \varepsilon) = 0 $$

Where $Y$ is a $N$-by-1 vector, $X$ is a $N$-by-$K$ matrix, $Z$ is a $N$-by-$L$ matrix, and $\varepsilon$ is a $N$-by-1 vector. Under usual assumptions on instrumental variables estimation, $\hat{\beta}_W$ is the unique $K$-by-1 vector solving the linear set population moment conditions:

$$ E[Z' (Y - X \hat{\beta}_W)] = 0 \quad (1) $$

Because sample averages are consistent estimators of population moments, the analogy applied to condition (15-1) suggests choosing $\hat{\beta}_W$ to solve

$$ N^{-1} \sum_{i=1}^{N} Z_i' (Y_i - X_i \hat{\beta}_W) = 0 \quad (2) $$

When $L = K$,

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

As we know already, this is called the **IV estimator**, but this can be thought as a special case of GMM estimators. The basic principle of the method of moments is to choose the parameter estimate so that the corresponding sample moments, Equation (1) in this case, are also equal to zero.

When $L > K$ (so we have more columns in the IV matrix $Z$ than we need for identification), choosing $\hat{\beta}_{GMM}$ is more complicated. Instead, we choose $\hat{\beta}_{GMM}$ to make
the distance between the two vectors below as “small” as possible in the sample:
\[
\sum_{i=1}^{N} (Z_i'Y_i - Z_i'X_i \hat{\beta}_{GMM})
\]

To measure the distance between the two vectors, we construct a quadratic form with a weighting matrix. Let \( \hat{W} \) be an L-by-L symmetric, positive semidefinite matrix. Thus, the generalized method of moments (GMM) estimator solves the minimizing problem:

\[
\min_{\beta} \left[ \sum_{i=1}^{N} (Y_i - X_i \hat{\beta}_{SV})' \hat{W} \left( \sum_{i=1}^{N} (Y_i - X_i \hat{\beta}_{SV}) \right) \right]
\]

or

\[
\min_{\beta} \left[ Z'(Y - X\hat{\beta}_{SV})' \hat{W} Z'(Y - X\hat{\beta}_{SV}) \right]
\]

\[
\min_{\beta} \left[ Y'Z - \hat{\beta}_{SV}' XZ \hat{W} Z'Y - \hat{W} Z'X \hat{\beta}_{SV} \right]
\]

\[
\min_{\beta} Y'Z \hat{W} Z'Y - \hat{\beta}_{SV}' XZ \hat{W} Z'X \hat{\beta}_{SV} + \hat{\beta}_{SV}' XZ \hat{W} Z'X \hat{\beta}_{SV}
\]

F.O.C. is

\[
- X'Z \hat{W} Z' Y - Y'Z \hat{W} Z' X + 2 X'Z \hat{W} Z' X \hat{\beta}_{SV} = 0
\]

\[
X'Z \hat{W} Z' X \hat{\beta}_{SV} = X'Z \hat{W} Z' Y
\]

\[
\hat{\beta}_{GMM} = (X'Z \hat{W} Z' X)^{-1} X'Z \hat{W} Z' Y
\]

(3)

The asymptotically consistent estimator of the variance of the GMM estimator is

\[
V(\hat{\beta}_{GMM}) = (X'Z \hat{W} Z' X)^{-1} X'Z \hat{W} \Lambda \hat{W} Z' X (X'Z \hat{W} Z' X)^{-1}
\]

(4)

where \( \Lambda \equiv E(\varepsilon_i^2 Z_i' Z_i) = Var(Z \varepsilon_i) \).

As we will see below, GMM estimator becomes to various least square estimators depending on how we choose \( \Lambda \) and \( \hat{W} \). First, we start with the optimal (i.e., the most efficient) GMM estimator.

**Efficient GMM Estimator**

In (4), we recognize that the variance of the GMM estimator will have a least asymptotic variance if we choose \( \hat{W} = \Lambda^{-1} \). Then, the least asymptotic variance of the GMM estimator is

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

\[
= (X'Z \Lambda^{-1} Z' X)^{-1} X'Z \Lambda^{-1} Z' X (X'Z \Lambda^{-1} Z' X)^{-1}
\]

\[
= (X'Z \Lambda^{-1} Z' X)^{-1}
\]

(5)
We can obtain a consistent estimator of \( \Lambda = E(e_i^2 Z_i'Z_i) = Var(Z\varepsilon_i) \) from \( E(\hat{e}_i^2 Z_i'Z_i) \), where the residual is obtained from consistent estimators (but could be inefficient estimators such as 2SLS estimators). Therefore, we can construct a procedure to obtain the optimal weighting matrix as:

**The Optimal Weighting Matrix**

**The Procedure**

5. Let \( \tilde{\beta}_{ini} \) be an initial consistent estimator of \( \beta \). In most cases this is the system 2SLS estimator.

6. Obtain residuals: \( \tilde{e}_i = y_i - X_i\tilde{\beta}_{ini} \) for \( i = 1, 2, \ldots, N \)

7. A generally consistent estimator of \( \Lambda \) is \( \hat{\Lambda} = N^{-1} \sum_{i=1}^{N} \tilde{e}_i^2 Z_i'Z_i \)

8. Choose \( \hat{W} = \hat{\Lambda}^{-1} = \left( N^{-1} \sum_{i=1}^{N} \tilde{e}_i^2 Z_i'Z_i \right)^{-1} \)

and use this weighting matrix to obtain the asymptotically optimal GMM estimator. (See Examples below)

**The GMM=2SLS Estimator**

Up to this point, we have not made a homoskedasticity assumption. But, what happens if we do? Under a conditional homoskedasticity, we have

\[
E(e_i^2 | X_i) = \sigma^2.
\]

Then,

\[
\Lambda = E(e_i^2 Z_i'Z_i) = \sigma^2 E(Z_i'Z_i).
\]

A consistent estimator of \( \Lambda \) is \( \hat{\Lambda} = \hat{\sigma}^2 Z'Z \). By following the general rule to obtain the least variance, we choose the weighting matrix as

\[
\hat{W} = \hat{\Lambda}^{-1} = (\hat{\sigma}^2)^{-1}(Z'Z)^{-1}.
\]

When we plug this into equations (4) and (5), we get

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

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

This is the **2SLS estimator and its variance**.

**STATA Programs on OLS, robust-OLS, IV, and GMM.**

**Example 15-1: A STATA Program of OLS on CARD.dta**

. #delimit;

59
clear
use c:\docs\fasid\classes\econometrics\wooldridge_data\card.dta

** A matrix with everything;
mat accum x=lwage educ exper expersq black smsa south;
(obs=3010)

** YY;
matrix yy=x[1,1];

** XY;
matrix xy=x[2...,1];

** XX;
matrix xx=x[2...,2...];

** Obtain b;
matrix b=(syminv(xx)*xy)';

** Obtain SSR;
matrix s=(yy-2*b*xy+b*xx*b')/(3010-7);

mat list s;
symmetric s[1,1]
lwage
lwage .14001865

** Obtain variances;
matrix v=s*syminv(xx);

ereturn post b v;
ereturn display;

---------------------------------------------+-----------------------------
        |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
educ |   .0740097   .0035054    21.11   0.000     .0671385    .0808795
exper |   .0835958   .0066478    12.57   0.000     .0705664    .0966253
expersq |  -.0022409   .0003178    -7.05   0.000    -.0028638   -.0016179
black |  -.1896315   .0176266   -10.76   0.000    -.2241790   -.1550841
sma |   .1614230   .0155733    10.37   0.000     .1308999     .1919460
south |  -.1248615   .0151182    -8.26   0.000    -.1544927   -.0952303
_cons |   4.733664   .0676026    70.02   0.000     4.601166    4.866163
-------------+----------------------------------------------------------------
.reg lwage educ exper expersq black smsa south;

Source |       SS       df       MS              Number of obs =    3010
-------------+------------------------------           F(  6,  3003) =  204.93
Model |  172.165628     6  28.6942714           Prob > F      =  0.0000
Residual |  420.476016  3003  .140018653           R-squared     =  0.2905
-------------+------------------------------           Adj R-squared =  0.2891
Total |  592.641645  3009  .196956346           Root MSE      =  .37419
-------------+------------------------------
lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
**Example 15-2: A STATA Program of robust-OLS on CARD.dta**

```
.* #delimit; delimiter now;
.* set more off;
.* clear;
.* use c:\docs\fasid\classes\econometrics\wooldridge_data\card.dta;
.* reg lwage educ exper expersq black smsa south;
.* predict rhat, residuals;
.* ** creating rx variables;
.* gen reduc=rhat*educ;
.* gen rexper=rhat*exper;
.* gen rexpersq=rhat*expersq;
.* gen rblack=rhat*black;
.* gen rsmsa=rhat*smsa;
.* gen rsouth=rhat*south;
.* ** A matrix with everything;
.* gen one=1;
.* mat accum x=lwage educ exper expersq black smsa south one, noc;
```
Example 15-3: A STATA Program of IV on CARD.dta

```
.*delimit*
.delimiter now *
.set more off*

(mat accum xrrx=reduc rexpersq rblack rsmasa rsouth rhat, noc; * 
* YY;  
.matrix yy=x[1,1];  
* XY;  
.matrix xy=x[2...,1];  
* XX;  
.matrix xx=x[2...,2...];  
* Obtain b;  
.matrix b=(syminv(xx)*xy)';  
* Obtain variances;  
.matrix v=syminv(xx)*xrrx*syminv(xx);  
ereturn post b v;  
ereturn display;  
)*
regexp l wage educ exper expersq black smsa south, robust;  
Regression with robust standard errors Number of obs = 3010  
F( 6, 3003) = 217.74  
Prob > F = 0.0000  
R-squared = 0.2905  
Root MSE = .37419  
------------------------------------------------------------------------------  
|       Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]  
-------------+----------------------------------------------------------------  
educ |   0.074009   0.003637   20.34   0.000     0.066889    0.081139  
exper |   0.083596   0.006725   12.43   0.000     0.070416    0.096776  
expersq |  -0.002241   0.000318  -7.04    0.000    -0.002865   -0.001618  
black |  -0.189632   0.017412  -10.88   0.000    -0.223759   -0.155504  
smsa |   0.161423   0.015157   10.65   0.000     0.131749     0.191131  
south |  -0.124861   0.015333  -8.14    0.000    -0.154914   -0.094763  
one |   4.733664   0.070076   67.55   0.000     4.596318    4.871011  
------------------------------------------------------------------------------
```
. set matsize 800;
. clear;

. use c:\docs\fasid\classes\econometrics\wooldridge_data\card.dta;

** A matrix with everything;
. mat accum x=lwage educ nearc4 nearc2 exper expersq black smsa south; (obs=3010)

. mat accum y=lwage educ exper expersq black smsa south; (obs=3010)

** YY;
. matrix yy=x[1,1];

** ZZ;
. matrix zz=x[3...,3...];

** XZ;
. matrix xz1=x[2,3...];
. matrix xz2=x[5...,3...];

** ZZ;
. matrix zz=x[3...,3...];

** XX;
. matrix xx=y[2...,2...];

** XY;
. matrix xy=y[2...,1];

** Obtain b;
. matrix b=(syminv(xz*syminv(zz)*xz')*xz*syminv(zz)*zy)';

** Obtain SSR;
. matrix s=(yy-2*b*xy+b*xx*b')/(3010-7);

** Obtain variances;
. matrix v=s*syminv(xz*syminv(zz)*xz');

** Obtain b v;
. ereturn post b v;

** Obtain variances;
. matrix v=s*syminv(xz*syminv(zz)*xz');

. ereturn display;

|            | Coef.    | Std. Err. | z       | P>|z|    | [95% Conf. Interval]        |
|------------|----------|-----------|---------|-------|-----------------------------|
| educ       |       .1608487 | .0486291  | 3.31    | 0.001 | .0655375 .25616            |
| exper      |       .1192112 | .0211779  | 5.63    | 0.000 | .077033 .1607191           |
| expersq    |      -.0023052 | .0003507  | -6.57   | 0.000 | -.0029925 -.001618         |
| black      |      -.1019726 | .0526187  | -1.94   | 0.053 | -.2051033 .0011582         |
| smsa       |       .1165736 | .0303135  | 3.85    | 0.000 | .0571602 .175987           |
| south      |      -.0951187 | .0234721  | -4.05   | 0.000 | -.1411233 -.0491141        |
| _cons      |       3.272102 | .8192563  | 3.99    | 0.000 | 1.666389 4.877815          |
. ivreg lwage (educ=nearc4 nearc2) exper expersq black smsa south;

Instrumental variables (2SLS) regression

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 3010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>86.2367703</td>
<td>6</td>
<td>14.3727951</td>
<td>F(6, 3003) = 110.30</td>
</tr>
<tr>
<td>Residual</td>
<td>506.404874</td>
<td>3003</td>
<td>.168632992</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>592.641645</td>
<td>3009</td>
<td>.196956346</td>
<td>R-squared = 0.1455</td>
</tr>
</tbody>
</table>

------------------------------------------------------------------------------
| lwage | Coef.   | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-------|---------|-----------|-------|------|----------------------|
| educ  | .1608487| .0486291  | 3.31  | 0.001| .065499 .2561984     |
| exper | .1192112| .0211779  | 5.63  | 0.000| .0776866 .1607358    |
| expersq | -.0023052 | .0003507 | -6.57 | 0.000| -.0029928 -.0016177  |
| black | -.1019726| .0526187  | -1.94 | 0.053| -.2051449 .0011997   |
| smsa  | .1165736 | .0303135  | 3.85  | 0.000| .0571363 .1760109    |
| south | -.0951187| .0234721  | -4.05 | 0.000| -.1411418 -.0490956  |
| _cons | 3.272102 | .8192563  | 3.99  | 0.000| 1.665742 4.878462    |

------------------------------------------------------------------------------

Instrumented: educ
Instruments: exper expersq black smsa south nearc4 nearc2

Example 15-4: A STATA Program of GMM on CARD.dta

. #delimit;
. delimiter now;
. set more off;

. set matsize 800;
. clear;
. use c:\docs\fasid\classes\econometrics\wooldridge_data\card.dta;
. ivreg lwage (educ=nearc4 nearc2) exper expersq black smsa south;

Instrumental variables (2SLS) regression

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 3010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>86.2367703</td>
<td>6</td>
<td>14.3727951</td>
<td>F(6, 3003) = 110.30</td>
</tr>
<tr>
<td>Residual</td>
<td>506.404874</td>
<td>3003</td>
<td>.168632992</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>592.641645</td>
<td>3009</td>
<td>.196956346</td>
<td>R-squared = 0.1455</td>
</tr>
</tbody>
</table>

------------------------------------------------------------------------------
| lwage | Coef.   | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-------|---------|-----------|-------|------|----------------------|
| educ  | .1608487| .0486291  | 3.31  | 0.001| .065499 .2561984     |
| exper | .1192112| .0211779  | 5.63  | 0.000| .0776866 .1607358    |
| expersq | -.0023052 | .0003507 | -6.57 | 0.000| -.0029928 -.0016177  |
| black | -.1019726| .0526187  | -1.94 | 0.053| -.2051449 .0011997   |
| smsa  | .1165736 | .0303135  | 3.85  | 0.000| .0571363 .1760109    |
| south | -.0951187| .0234721  | -4.05 | 0.000| -.1411418 -.0490956  |
| _cons | 3.272102 | .8192563  | 3.99  | 0.000| 1.665742 4.878462    |

------------------------------------------------------------------------------

Instrumented: educ

64
Instruments: exper expersq black smsa south nearc4 nearc2

. predict rhat, residual;

. ** Create Variables multiplied by rhat;
. gen Znearc4=nearc4*rhat;
. gen Znearc2=nearc2*rhat;
. gen Zexper=exper*rhat;
. gen Zexpersq=expersq*rhat;
. gen Zblack=black*rhat;
. gen Zsmsa=smsa*rhat;
. gen Zsouth=south*rhat;

. ** Create S-hat matrix;
. mat accum ZrrZ=Znearc4 Znearc2 Zexper Zexpersq Zblack Zsmsa Zsouth rhat, noc;
(obs=3010)

. ** A matrix with y;
. gen one=1;

. mat accum x=lwage educ nearc4 nearc2 exper expersq black smsa south one, noc;
(obs=3010)

. mat accum y=lwage educ exper expersq black smsa south one, noc;
(obs=3010)

. ** YY;
. matrix yy=x[1,1];

. ** ZZ;
. matrix zz=x[3...,3...];

. ** XZ;
. matrix xz1=x[2,3...];

. ** XX;
. matrix xx=y[2...,2...];

. ** XY;
. matrix xy=y[2...,1];

. ** Obtain GMMb;
. matrix b=(syminv(xz*syminv(ZrrZ)*xz')*xz*syminv(ZrrZ)*zy)';

. ** Obtain variances;
. matrix v=syminv(xz*syminv(ZrrZ)*xz');

. ereturn post b v;
. ereturn display;

------------------------------------------------------------------------------
|      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval] 
------------------------------------------------------------------------------
educ |   .1588387   .0484983     3.28   0.001     .0637838    .2538935
exper |   .1182042   .0212941     5.55   0.000     .0764684    .1599399
expersq |  -.0022962   .0003686    -6.23   0.000    -.0030186   -.0015738
black |  -.1056934   .0519689    -2.03   0.042    -.2075506   -.0038361
smsa |   .1170294   .0302564     3.87   0.000     .0577281    .1763308
south |  -.096691   .0233983    -4.11   0.000    -.1419508   -.0502311
one |   3.307021   .8165958     4.05   0.000     1.706523    4.907519
------------------------------------------------------------------------------

.ivgmm0 lwage (educ=nearc4 nearc2) exper expersq black smsa south;

Instrumental Variables Estimation via GMM
Number of obs = 3010
Root MSE = 0.4086
Hansen J = 2.6532
Chi-sq(1) P-val = 0.1034

------------------------------------------------------------------------------
|                GMM
lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval] 
------------------------------------------------------------------------------
educ |   .1588387   .0484983     3.28   0.001     .0637838    .2538935
exper |   .1182042   .0212941     5.55   0.000     .0764684    .1599399
expersq |  -.0022962   .0003686    -6.23   0.000    -.0030186   -.0015738
black |  -.1056934   .0519689    -2.03   0.042    -.2075506   -.0038361
smsa |   .1170294   .0302564     3.87   0.000     .0577281    .1763308
south |  -.096691   .0233983    -4.11   0.000    -.1419508   -.0502311
_cons |   3.307021   .8165958     4.05   0.000     1.706523    4.907519
------------------------------------------------------------------------------

Instrumented: educ
Instruments: exper expersq black smsa south nearc4 nearc2

------------------------------------------------------------------------------